

Progress in digitizing US herbaria, 2011–2012

Mary E. Barkworth¹, Ellen Dean², Ben Legler³, Mark Mayfield⁴, Zack Murrell⁵, Eric Ribbens⁶

¹Intermountain Herbarium, Department of Biology, Utah State University, Logan, Utah 84322-5305, U.S.A.;

²UC Davis Center for Plant Diversity, Plant Sciences MS#7, University of California, 1 Shields Avenue, Davis, California 95616-8780, U.S.A.;

³Burke Museum, Box 355325, University of Washington, Seattle, WA 98195-5325;

⁴Herbarium, Division of Biology, Kansas State University, Ackert Hall, Manhattan, Kansas 66506-U.S.A.;

⁵Herbarium, Biology Department, Appalachian State University, Rankin Science Building, Boone, North Carolina 28608, U.S.A.;

⁶R. M. Myers Herbarium, Biological Sciences Department, Western Illinois University, Macomb, Illinois 61455, U.S.A.

Corresponding Author: Mary E. Barkworth (mary.barkworth@usu.edu)

Abstract

US Biological Collections are engaged in a massive enterprise: making their specimen information freely available via the web. In 2011, we surveyed US herbaria to determine how much progress herbaria have made in three key areas: databasing, specimen imaging, and data sharing. In 2012, we added a fourth area, provision of latitude/longitude information. The surveys were sent, by regular mail, to the herbaria listed in Index herbariorum (the recognized listing of the world's research herbaria). We received 176 responses to the survey in 2011, 315 in 2012; the latter number was 52% of the listed herbaria in the US in October, 2012 that were not known to be inactive. The percentage of specimens in each of the three categories surveyed in both years was higher in 2012 than in 2011. In 2012, the responding herbaria held over 70 million specimens, of which they had databased over 22 million, imaged over 1 million, and georeferenced over 7 million. Of these herbaria, 71 made some or all of their records available through a regional web site, 59 only through a local web site, and 118 had databases that were not web accessible. Of the 2012 responses, 81 were from herbaria that had not initiated activity in any of the four areas surveyed. Every state had one herbarium that was databasing and providing latitude/longitude information, most had more; eleven states had no herbarium that had begun imaging. Greatest progress in digitization has been made in areas where there is a regional network. Enabling browser-based data entry has encouraged participation by herbaria with fewer than 200,000 specimens. The regional portals offer visitors several options for searching and easy access to images of living representatives of the included organism in addition to specimen images; most also provide access to published descriptions of the taxa. Progress could be accelerated by acquisition of imaging equipment in states where there is none at present and establishment of national networks for vascular plants, algae, and microfungi because these would permit better use of tools for collaboration.

Keywords

Herbaria, Digitization, Databasing, Imaging, Georeferencing, Biological Collections, U.S.A.

Introduction

In 2008, the US Virtual Herbarium project (USVH) was initiated by representatives of multiple US Herbaria. Those present represented a wide range of backgrounds. Some of those present were interested in finding out how to get their collections online, others in how specimen information from herbaria could be made available. In 2009, the US Virtual Herbarium Project (USVH) was formally established with the goal of enabling creation of a single portal that would provide access to the specimen records of all specimens in all US herbaria. The portal itself was of less concern than helping herbaria, all herbaria, make their specimen information in a manner that would permit tapping into the resulting information with a single search. In 2010, the National Science Foundation announced a program, Advancing Digitization of Biological Collections, that has a similar goal but for all biological collections, not just herbaria.

The USVH project works towards attainment of its goal by working with regional coalitions of herbaria and individual herbaria and by assisting in disseminating information. As with any project, it is important that progress towards its goals be measured on a regular basis because, as Kelvin (1883, p. 73) stated, "... when you cannot measure [what you are speaking about], when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind...". In this paper, we compare the results of two surveys of digitization in US herbaria, one conducted in 2011 and the other in 2012. We conclude with recommendations, based on the survey results and conversations with many of those involved, on how progress being could be accelerated.

The surveys focused on four core aspects of digitization: databasing, imaging, provision of latitude/longitude information [henceforth "latlonging"], and data sharing. The information gained will help those involved with US herbaria identify how best to move towards the ultimate goal, enabling creation of a single portal to records of all specimens in US herbaria, and provide valuable information to those interested in digitization of other natural history collections, whether in the US or elsewhere.

Methods

Survey Design

In both 2011 and 2012, the surveys (App. 1, 2a, 2b) were designed to elicit information concerning two areas of herbarium digitization: specimen processing and making the resulting information available. Both surveys asked respondents to estimate the total number of specimens in their collection, the number databased, the number imaged, and the url(s) of sites where the information was made available. The 2012 survey also asked for an estimate of the number of specimens that had latitude/longitude information and for separate estimates of the number of specimens imaged and the number of labels imaged.

The surveys were deliberately kept short and factual in the hope that this would persuade more recipients to respond than a longer survey asking more for more detailed information or subjective judgments. Questions about funding and support were omitted because a meaningful survey of these aspects would require significantly more time of the respondents in addition to a greater commitment to data analysis on the part of the project. For instance, the range of support sources includes grants and contracts from a wide range of sources, students earning course credits, volunteers, and donations. The absence of questions relating to the taxonomic and/or geographic scope of digitizing activities reflected the project's interest in seeing *all* specimens digitized. Within a herbarium, priorities are determined first by funding after which priority is probably given to specimens from the home state and/or the research interests of the person in charge.

To compensate for the brevity and narrow focus of the surveys, recipients were invited to submit comments on any aspect of the survey or the project; many did so.

Survey Delivery

In 2011, all the surveys were sent out by the US postal service using the addresses in *Index herbariorum* (IH; Thiers continuously updated). In July, because of the low return rate, Barkworth sent emails and/or called several non-responding herbaria, often visiting the web site of the owning institution to determine whom to contact. These efforts substantially increased the number of surveys returned but were very time-consuming. They sometimes revealed that the herbarium was no longer in existence, information that was shared with Dr. Barbara Thiers who maintains IH. On a happier note, they sometimes resulted in an expression of pleasure at the interest being shown.

In 2012, the survey was initially distributed in April as an online survey using Survey Monkey (<http://www.surveymonkey.com>). Requests to complete it were sent through two listservs, TAXACOM (<http://mailman.nhm.ku.edu/mailman/listinfo/taxacom>) and HERBARIA (<http://www.nacse.org/mailman/listinfo/herbaria>). After four weeks, because the number of returns received was low, the online site was closed and a paper copy developed (App. 2). This was mailed out to the 592 US herbaria in IH (Thiers continuously updated) that were not described as transferred, closed, or discarded. In addition, the survey was posted, as a downloadable pdf file, to <http://herbarium.usu.edu>. By mid-June, the number of answers received was still lower than in 2011 so reminders were posted to the two listservs. In addition, non-responding herbaria with over a million specimens were sent a letter asking for a response. On July 8 2012, Barkworth presented a preliminary report to the USVH annual meeting. In response to comments made by the attendees, a revised version of the survey was posted to <http://herbarium.usu.edu> and those attending were asked to encourage their colleagues to submit responses. In addition, another request for completion of the survey was posted to the two listservs, several regional representatives to the US Virtual Herbarium project were asked to encourage responses, and emails were sent to herbaria that had responded in 2011 but not 2012. Each effort resulted in additional surveys being returned.

Herbarium status

In analyzing the data for this paper, it became evident that it would be helpful to develop a list of US herbaria known to have been transferred, closed, or discarded so a field for “Status” was added to the database. Herbaria that responded to the survey were classified as “Responding” unless the respondent stated otherwise. Other herbaria were classified as transferred (T), closed (C), or discarded (D) or, there was no information in IH, as unknown (U). The “transferred” category (T) includes herbaria that are maintained as separate entities but administered jointly (e.g., JEPS and UC; herbarium codes from Thiers (updated continuously) as well as those that have been integrated into another herbarium. “Closed” (C) refers to herbaria that are in storage; it was not used for herbaria that do not make loans, that are open only by appointment, or are used solely to assist in teaching. “Discarded” (D) herbaria are those stated to be discarded and those about which IH states something to the effect that “no one seems to know what has become of it”. Some of the non-responding herbaria were obviously active because their IH entry had been updated in 2012. Others were probably closed or discarded but, without evidence to the contrary, “unknown” is the most accurate description of their status. The survey’s database does not in-

clude all the active herbaria in US because minimal effort was made to identify US herbaria that were not listed in IH.

We also decided to summarize the ownership of herbaria in the US. For this purpose, we used Wikipedia to determine whether owning institution, as indicated in IH, was a public academic institution, private academic institution, federal agency, state agency, or some other private institution such as an independent garden, arboretum, museum, or preserve.

Data analysis

In 2011, the survey data were entered into a spreadsheet (Microsoft 2010). This was used to construct histograms and plot the geographic distribution of responding herbaria. The results were cited in Barkworth and Murrell (2012) but have not been published. In 2012, both the 2011 and 2012 data were entered into a database created using Alpha 5 (Alpha Software 2012). This database was updated as surveys were received. In mid-June, a synopsis of the data received was posted to the Web at <http://herbarium.usu.edu/surveyresults.html>, together with a list of the responding herbaria. These pages were updated daily until July 1, after which they were updated more irregularly. The purpose in publishing the results as they were received was to encourage additional returns. This paper reflects surveys returned before October 19, 2012. Seven surveys were returned after that date, six of them in 2013.

The data were analyzed using histograms and scatterplots generated in Excel (Microsoft Office 2010). In addition, maps were prepared using ArcGIS10 (ESRI 2011) to portray the geographic distribution of digitizing activities and the development of regional and taxonomic networks in 2012. Latitude and longitude data for the herbaria were provided by Dr. Barbara Thiers from IH.

Results

Herbarium Status

Herbaria are scattered throughout the US, with the lowest concentration being in the areas that are least densely populated (Fig. 1) such as Alaska, Nevada, and the Central Plains. Table 1 summarizes the status of US herbaria in 2011 and 2012. The most striking difference between the two years was in the number of herbaria responding to the survey. The number of responses in 2012, 331, represents about half the herbaria in the US, a proportion that add to the credibility of the data although the difference between the two years makes some comparisons more difficult.

One herbarium was accidentally discarded between the two surveys. The institution is seeking donation of duplicates in order to re-establish its collection. One herbarium was closed between the two years. According to the respondent, the curator had retired and the institution had not yet decided whether to keep or transfer the collection. Seven herbaria were added to those

Table 1. Status of US herbaria in 2011 and 2012. See text for explanation of terms used.

	Discarded	Closed	Transferred	Responding	Not responding	Total
2011	19	13	113	190	415	750
2012	20	14	120	331	266	751

known to have been transferred to another institution. In at one case, the respondent noted that the decision was made in order to provide better curation for the collection.

Ownership: Most herbaria in the US are owned by an academic institution, the majority by a public academic institution (Table 2). There are about as many herbaria associated with independent gardens, museums, etc. as with a government agency. There were three herbaria whose ownership was did not fit any of the five categories. Most of the large non-academic herbaria have well-established connections with one or more universities in their area. Many of the small academic herbaria do not offer graduate degrees. Most of the federal herbaria are used primarily for documenting the plants in a particular unit of federal land and/or training field personnel. There is, however, considerable variability within each category.

Table 2. Ownership of US herbaria. Responding herbaria replied to one of the two surveys; non-responding herbaria did not reply to either survey. Transferred, closed, and discarded herbaria are not included.

	Academic - Public	Academic - Private	Museums, Gardens, etc.	Federal Agencies	State Agencies	Other
Responding	207	58	58	30	10	2
Non-responding	113	68	20	25	8	1

Table 2 indicates that the US Virtual Herbarium project needs to reach out more to herbaria associated with private academic institutions. Many of these are primarily teaching institutions but it is important that all students engaged in biodiversity studies, no matter what the nature of their institution, learn how record and contribute specimen data in ways that maximize the value of their efforts.

Distribution. On September 9, 2012, *Index herbariorum* (IH) listed 738 herbaria for the US. This included herbaria in the Virgin Islands, Puerto Rico, and the U.S. National Historic Park in Guam. Two more, BNFH and CALVIN were added in October 2012, bringing the total to 740. They were not included in the data for 2012. The project’s data base included, in addition to the 738 herbaria listed in IH, 11 herbaria that were not listed there but responded to one or both of the surveys bringing the total number of herbaria in the data base at the end of 2012 to 749.

The geographic distribution of the responding herbaria was similar to that of the herbaria of unknown status (Fig. 1) but the size of herbaria in the two groups was quite different. In 2012, the 318 responding herbaria varied in size from 100 to 7,300,000 specimens, with the average size being 220,724. The 273 herbaria of unknown status with data in IH ranged in size from

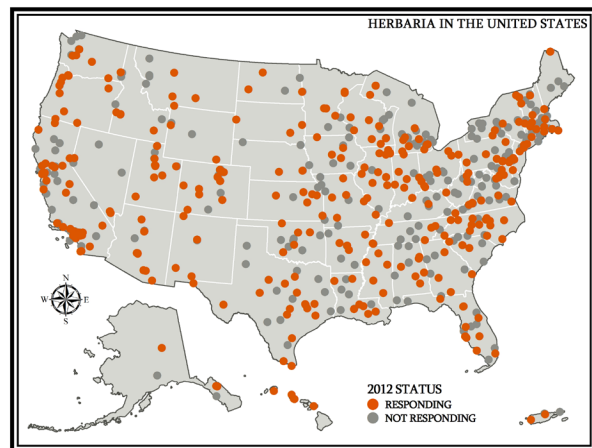


Fig. 1. Distribution of herbaria in the US (excluding transferred, closed, and discarded herbaria).

250 to 600,000 with the average being 25,492. There no data available for two herbaria. Only 13 had 100,000 or more specimens.

Digitization – Specimen data

Total number of specimens

In 2011, the 194 responding herbaria held a total of 58,417,336 specimens, the mean being 310,120 specimens and the median 41,000. In 2012, the 331 responding herbaria held 70,912,154 specimens for a mean of 214,236 specimens and median of 28,000. The difference between the mean and median in each year shows that there are many small herbaria and a few very large herbaria. For this reason, the horizontal axes in figures 2 and 3 are logarithmic, the upper limit of each size class being twice that of the one before. To facilitate comparison between the years, all charts have the same scale.

The drop in both the mean and median from 2011 to 2012 reflects the larger number of small herbaria responding in 2012. Of the 171 herbaria that responded for the first time in 2012, 140 had fewer than 100,000 specimens; none had more than 900,000. The two smallest herbaria to respond in 2012, neither of which is listed in IH, had 100 specimens; the smallest listed herbarium had 500 specimens.

Of the 160 herbaria that responded in both years, 65 grew by a total of 3,706,140 specimens, but 42 apparently shrank by a total of 674,967 specimens. The remaining 53 herbaria did not change in size. Some of the respondents reporting a decrease in size explained that specimens outside the herbarium's focus had been transferred to another institution. In most cases, however, the lower number probably reflected a more accurate estimate based on closer familiarity, a change in how the estimate was made, or the difficulty of making the estimate.

Databasing

In 2011, 152 (78.3%) of the responding herbaria had begun databasing their collections [for convenience and brevity, "herbaria" is used to mean the people associated with a herbarium]. The total number of specimens databased was 16,995,004 or 29.1% of the specimens housed in the responding herbaria. In 2012, the number of databasing herbaria was 240 (72.3% of the responding herbaria) and the number of their databased specimens was 22,788,976 or 32.1% of their total. The reduction in the proportion of herbaria that had begun databasing in 2012 resulted primarily from the many herbaria that responded for the first time in 2012, many of which had not started databasing their collections. Comparison of databasing data for the two years (Fig. 2) shows that there were more herbaria in each range for number databased, particularly in the number of herbaria with <1,000 databased and those with 8,000-256,000 databased.

Imaging

In 2011, 67 or 34.5% of the responding herbaria had started imaging their specimens. In 2012, 121 or 36.8% had begun to do so. The number of specimens imaged in 2012 was more than twice that in 2011: 1,436,191 (2.5% of the total) in 2011, 3,383,947 (5.5% of the total) in 2012. The number of herbaria in each range shown is greater in 2012 than in 2011, with the greatest increase being the number of herbaria with 32,000 – 256,000 imaged (Fig. 2). A few herbaria had many more specimens imaged than databased or "latlonged". This reflects work flows in which priority is given to specimen imaging. The increase in the proportion of herbaria engaged in imag-

ing reflects the establishment of regional networks, all of which have equipment for imaging and routinely database from images rather than specimens.

Latitude/Longitude data

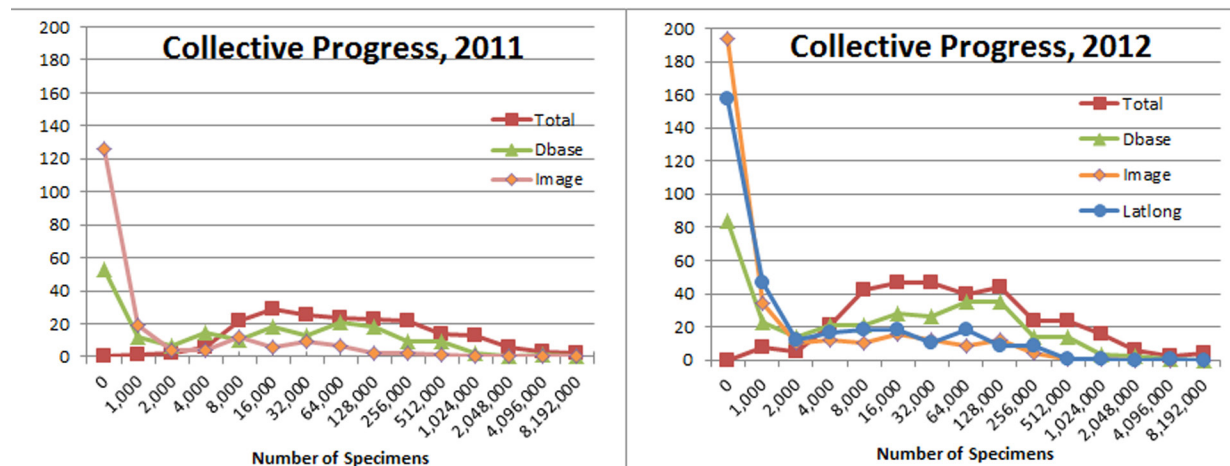


Fig. 2. Number of responding herbaria with indicated number of specimens (total), databased specimens (Dbase), imaged specimens (Image), and lat/longed specimens.

There was no question about latitude/longitude data in 2011. In 2012, herbaria had more specimens with such information than had been imaged: 163 (50%) of the responding herbaria stated that they had at least some specimens with latitude/longitude information, the total number being more than 7,581,513 (10.7% of the total) in the responding herbaria. The figure given is undoubtedly an underestimate because, if the respondent stated (as many did) that they did not know how many specimens they had with such data, a zero was entered in the appropriate field. If the statement made was something like “No more than 10% of the total”, a number somewhat less than 10% of the total was entered.

The survey did not ask for the number of specimens that included an estimate of uncertainty and/or datum, nor on how or by whom the latitude/longitude information was obtained. Another unknown is whether the respondents counted specimens with UTM data or township, range, and section (TRS) data as having latitude/longitude data. UTM data enables automatic calculation of geographic coordinates but not of an uncertainty; TRS data for the western states can be used to generate geographic coordinates with an uncertainty of ~1000m but the process is harder to automate than the calculation for UTM data. Geolocate (<http://www.museum.tulane.edu/geolocate/>), a program for assisting georeferencing, can now use TRS data to calculate latitude/longitude values but was not able to do so until the second half of 2012.

Progress within Herbaria

To examine progress within herbaria, the percent of specimens in each herbarium that had been through each of the three processes (databasing, imaging, latlonging) was calculated and the percentages added together. In most herbaria the greatest progress was in databasing but 12 herbaria had imaged more specimens than they had databased and there were 104 herbaria in which the number databased equaled the number imaged. This might be because, in many digitization workflows, a record is created as soon as a specimen is imaged, whether or not any of the label data have been captured. In 2012, 33 herbaria had databased all their specimens. Only one herbarium had latitude and longitude data for all its specimens but 10 of those that had started

databasing had such data for all their databased specimens. At the other end of the scale, 85 had not started any of the three processes), 195 had not started imaging, and 158 gave no estimate for the number of specimens with latitude/longitude data and were, therefore, counted as having no records with such data.

Comparison of the frequency curves for percent complete in 2011 and 2012 (Fig. 3) show the progress that was made in databasing and imaging between these two years. The number of herbaria reaching each level of completeness increased from 2011 to 2012, with the increase at the two ends of the scale being most striking. The increase in number of herbaria that have made a little progress reflects the larger number of herbaria responding in 2012. The increase at the high end of the scale demonstrates the increase in number of herbaria that are fully, or almost fully, databased and/or imaged.

The progress made in digitization by individual herbaria is best appreciated by comparing the curves for the two activities reported in both years (Fig. 4). The number of herbaria at each level



Fig. 3. Progress towards completing core digitization activities within individual herbaria. Vertical axis: number of herbaria at the percent completion shown on the horizontal axes.

of completeness has increased, with the greatest increases being for herbaria less than 10% complete and those between 50-60% complete. Adding progress in latlonging to the sum decreased the number of herbaria that were more than 50% digitized, dramatically so for herbaria in the 50-60% complete range, but it increased the number of herbaria in the 10-50% range.

One herbarium, HALE, is completely digitized. Five more have scores of 200 or above (UVSC, ALA, DES, TROY, UMFK, and UVSC) but 84 (25.8%) of the respondents had not yet started any of the processes and many more that were less than 10% complete. Several respondents stated that

they would like to start digitizing their collections but had neither the time nor resources needed. A few respondents stated that digitization was not a high priority, given their institution's goals.

Distribution of digitization

The two digitizing activities that require no special equipment, databasing and providing latitude/longitude data, are taking place across the country (Fig. 5). Most of the herbaria shown as databasing in 2011 but not in 2012 are shown that way because they failed to respond to the 2012 survey. No respondent stated that databasing had ceased in his or her herbarium. Imaging, which requires special equipment, is unevenly distributed and there were eleven states (Colorado, Delaware, Iowa, Nebraska, New Jersey, North Dakota, Oklahoma, Rhode Island, South Dakota, South Dakota, and Virginia) in which no responding herbarium had started the process. In other states, many herbaria had done so (Fig. 5).

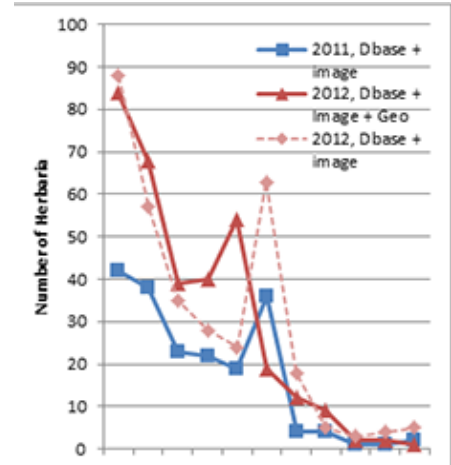


Fig. 4. Progress of individual herbaria towards completion of digitization.

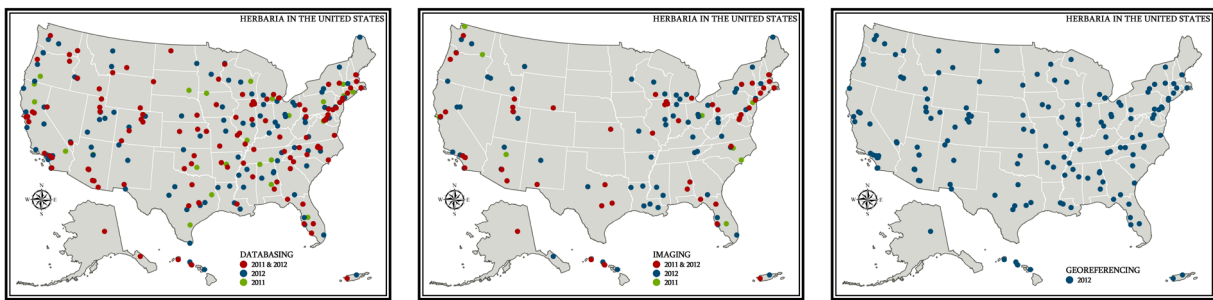


Fig. 5. Geographic distribution of databasing, georeferencing, and imaging. There were no questions about georeferencing on the 2011 survey.

Access to Information

One of the reasons for digitizing collections is to make their information widely accessible by enabling searchable web access. The USVH project's goal is to enable construction of web sites that enable a search to draw on all the information herbaria are making available because, from the users' point of view, the fewer sites that have to be visited in order to obtain all available data, the better. Building a single database would achieve this goal; it would also enable herbaria to make better use of tools for collaboration.

The Global Biodiversity Information Facility (GBIF) would, in theory, provide the single data pool but, in the two surveys, only 16 US herbaria stated that they provided data to it. In 2011, 10 of these stated that they also provided information to a regional network, a number that had increased to 14 in 2012.

In the US, the level below GBIF comprises regional and taxonomic networks. The survey respondents identified eight regional networks within the US: for Alabama (AL), California (CA), Florida (FL), Northeast (CNE), Pacific Islands (Pacls), Pacific Northwest (PNW), Southeast (SERNEC), Southwest, Intermountain, and Southern Rockies (SEINet), and Wisconsin (WIS) (Fig. 6). There are

large parts of the country outside the range of the scope of these networks as well as some herbaria that had not, at the time of responding, started providing records to the network for their region. Some networks are partially overlapping, e.g., CA and SEINet, AL and SERNEC. Three herbaria, NY, RM, and UCR, contribute records to more than one regional network.

The number of herbaria providing access through a regional network more than doubled between

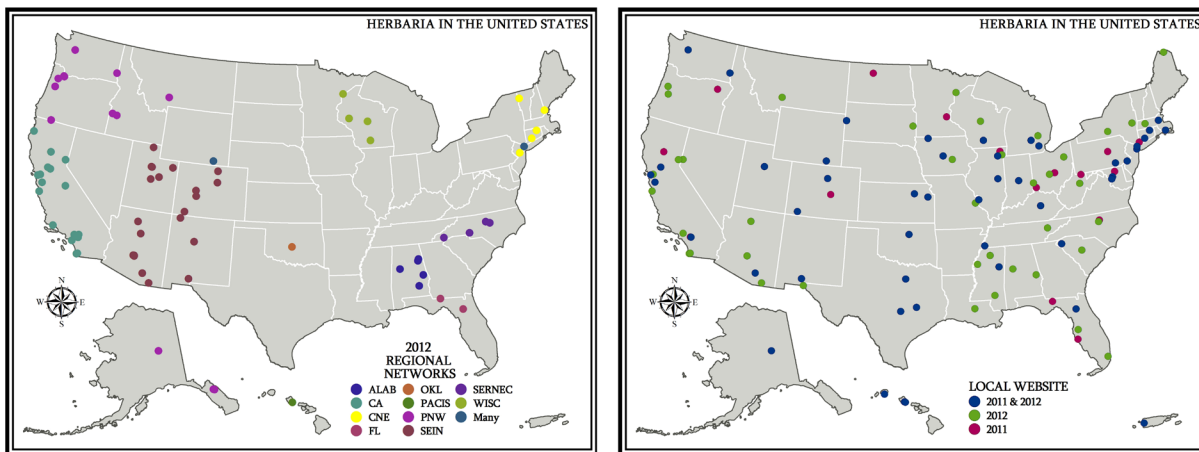


Fig. 6. Access to herbarium records in 2012. Left: herbaria providing access to some or all of their records through a region website. Right: herbaria providing access through a local web site.

the two surveys, from 33 to 71, and the number of herbaria whose records were only searchable through their own web site fell from 61 to 59. Many of the herbaria that were searchable only through their own web site were outside the range of existing regional networks; others expected to be providing data to a regional network in the near future. Some herbaria maintain their own web site in addition to providing records to a regional network. In some instances, they only contribute regional records to the regional web site but make all their records available on their own web site. In other cases, the local web site is updated on a daily basis whereas the regional network is updated less frequently.

The number of herbaria with databases that are not web-accessible increased from 74 to 118 between the two surveys. This reflects the increased number of respondents in 2012, not an overall decrease in access. In 2012 these inaccessible herbaria held 3,807,181 database records, 646,071 images, and 661,844 records with latitude/longitude data. Some stated that they planned to provide access in the future; others indicated that providing access was not a priority. Some provide access to more of their records through their own web site than they do via their regional web site.

In 2012, there were three national herbarium networks with a taxonomic focus: bryophytes (27 herbaria), lichens (38 herbaria), and macrofungi (24). Only 20 of the responding herbaria stated that they provided information to one of these networks. All but two of the 20 also contributed records to a regional network. A few of the herbaria contributing to these networks are from out-

side the US. In addition, several herbaria reported providing image and specimen data for type specimens to the Global Plants Initiative (<http://gpi.myspecies.info/>).

Discussion

Limitations of the surveys

The intent of the US Virtual Herbarium surveys is to obtain an annual snapshot of the progress US herbaria are making towards going digital. So far, only two surveys have been conducted. There are limitations to the data. First, they do not reflect a random sampling of US herbaria. In each year, after the initial number of returns received was unsatisfactory, steps were taken to obtain more returns but the steps taken were not the same in the two years. It is also probable that herbarium curators who are or would like to digitize their collections are more likely to respond to a survey from the US Virtual Herbarium project than those not interested in doing so. Nevertheless, the proportion of herbaria represented in the 2012 survey is high enough to give value to the results and the similarity of the data patterns from the two years suggests that the trends revealed are real.

It was evident that the intent of some of the questions should be clarified. For instance, informal conversations showed that some respondents had interpreted “Total number of specimens” as excluding unmounted and/or uncatalogued specimens, some as meaning only databased specimens, or only vascular plants, or only the plants from their state. Other limitations of the data include the brevity of the surveys and the narrow focus of the questions. As discussed earlier, both were deliberate choices.

Despite these limitations, the survey results clearly demonstrate that US herbaria had made substantial progress in digitizing their collections by 2012. This progress would not have been possible without the contributions from the various funding sources, primarily the National Science Foundation but also the Institute for Museum and Library Services, several state governments, and the various agencies, both public and private, that have provided for collecting and digitizing the collection data arising from projects they fund. In addition, some herbaria have started to digitize their collections without any external support because the person involved saw value in doing so.

Accelerating progress

All those involved should be complimented on the amount of progress that has been made in digitizing US herbaria. Nevertheless, it is appropriate to consider how progress could be accelerated. The comments that follow are based on data from the surveys, conversations with numerous individuals, and our personal experiences. They reflect a conviction that all herbaria must “go digital” in order to provide better resources for those interested in studying the organisms housed in

herbaria; more time for those working in herbaria to pursue research, teaching, and/or outreach activities; and to ensure that the next generation of biologists gains the knowledge and experience needed to contribute to and take advantage of the emerging resources.

The biggest obstacle to accelerating digitizing, both nationally and locally, is undoubtedly finding the resources for the work involved. Ultimately this usually comes down to money. This is particularly true of imaging, a process that requires relatively expensive equipment. There is no magic solution to this problem. Once the equipment has been purchased, however, imaging is the least expensive aspect of digitization. Options in use for minimizing the equipment cost include sharing equipment, either by sending it to different institutions or by having smaller herbaria take or send their specimens to an imaging center. An advantage of sharing equipment is that more institutions and individuals become familiar with its use; an advantage of using an imaging center is that few people need to be trained in using the equipment. Whichever approach is adopted for processing existing specimens, plans must be made for incorporating imaging into the processing of newly acquired specimens.

Many approaches are being taken to minimize labor costs. Some are technological, e.g., incorporating Optical Character Recognition into the workflow. Others are more people oriented, encouraging volunteers to complete data capture onsite or online, including via crowdsourcing. One program that enables crowdsourcing is *Notes from Nature* (Denslow et al. 2013). Enabling both crowdsourcing and duplicate discovery would be even more advantageous.

Some networks offer browser-based data entry directly into a network's database (direct data entry or DDE). This eliminates the need for a herbarium to install and maintain a local database while ensuring that data are captured in accordance with international standards. The inclusion of DDE as an option in Symbiota has persuaded many smaller herbaria to contribute to SEINet which, as of January 2013, had 43 contributing herbaria. According to Gilbert, developer of Symbiota and manager of SEINet, it also decreases the work required to maintain the network. Direct data entry is not necessarily the best option for a herbarium; it is a valuable option to have available when seeking to expand the number of participating herbaria without unduly increasing the workload of those in charge of them.

One of the most effective steps for accelerating digitization in US herbaria would be development of national networks for vascular plants, algae, and microfungi. These would make it possible to maximize the advantages offered by existing and developing tools for collaboration and, if they include images of the specimens, for on line review by experts. Two processes that would immediately benefit from a national network are databasing and georeferencing (adding latitude longitude information together with the datum used, an estimate of uncertainty, and the protocol followed).

Databasing would benefit because the common herbarium practice of exchanging specimens of duplicates means that there may be specimens with identical collection data at multiple institutions. When all the data from these records are in a single data pool, it is possible to copy (and modify if appropriate, e.g., to reflect new annotations) information from the initial record entered into the records from other herbaria, thereby eliminating the need for re-entering the data at each institution. Regional networks permit benefiting from duplicate discovery within a region but most US herbaria have some specimens, often a large number of specimens, from outside their own region.

Georeferencing is another time-consuming activity that would benefit from creation of national networks because it would enable qualified individuals to georeference (or review the georef-

erencing) of all collections from a particular area, not just those from their own herbarium or herbaria in their own regional network. Both Specify and Symbiota incorporate Geolocate's tools for collaborative georeferencing but the benefits from collaboration are decreased by lack of a national network. There are already national networks for bryophytes, lichens, and macrofungi, but not for the other organismal groups stored in herbaria – vascular plants, microfungi, and algae. Regional networks have been beneficial in making vascular plant data available and are effective in promoting the personal interactions that improve the dissemination of information and encourage participation in the digitization efforts. This is evident from the fact that the oldest regional networks also have the highest proportion of the herbaria in their region engaged in digitization. But for maximum efficiency in digitization, national networks are needed.

Another aspect that has aided herbaria in their efforts to digitize has been the decision of those developing the programs used to incorporate tools for displaying the information in a way that benefits multiple user groups. If herbaria are to continue or expand their ability to digitize their collections, they must make sure that the benefits from doing are rapidly evident to potential users. Most web sites serving specimen information from herbaria demonstrate an awareness of this need. Regional networks, for example, offer the ability to view distribution maps, images of living plants, and descriptions of the taxa as well as specimen images in addition to enabling downloading of data. There are differences among the networks (and individual web sites) but all offer more than just the ability to download records belonging to particular taxon. Some regional networks,, such as those using Symbiota software, also offer collectors the opportunity to record collection information into the central database. In so doing, collectors become familiar with the standards required to make their information maximally informative and facilitate incorporation of their specimen data into the databases of the recipient herbaria.

Data quality

The surveys focused solely on progress in digitization, not on data quality, nor have we ourselves made any attempt to assess the quality of the data being provided. Once data are online, people will send in corrections and comment. If the number of corrections Barkworth has received for UTC (~250,000 records online) is typical, the data are considered by most users to be fit for use but it could be that users are downloading data and making corrections in their own file without informing the data provider.

Some of those advocating digitization argue that the most important task is to make the records and associated specimen images available so that they can be annotated online; others think that those in charge of herbaria should verify the quality of the data they are making available before it is posted to the web. Informal discussions suggest that the attitudes of those in charge of herbaria vary from one extreme to the other. Realistically, if those in charge of herbaria were required to review all specimens before data were posted to the web, there would be very few records available.

Network databasing protocols and many stand-alone herbarium databases have some built-in tools for quality checking, e.g., verification of taxon names, names of collectors, the validity of an elevation (plants do not grow above 23,000 ft), and dates (there are no specimens collected in 216 or 2016) but some herbaria are capturing data into spreadsheets that have no such checks. Other tools that would be helpful at both the herbarium and network level include scripts for checking the consistency of geographic coordinates with the locality information, for checking the years during which a particular collector was active, and for identifying records that suggest an

ecological outlier. Comparison of data from duplicate specimens is an activity best conducted on a network, preferable a national network.

Conclusions

The progress US herbaria have made in digitizing their collections and making the results available in a manner that benefits multiple user groups reflects a major commitment on the part of all those in charge of herbaria. The demands of this commitment cannot be underestimated. Moving a herbarium into the digital era requires changing workflows and attitudes, educating those who contribute specimens to herbaria, and explaining how digitization enhances the value of a herbarium to its owners and its users. All those involved with herbaria already have many obligations; for many, the herbarium is only a small part. All involved are to be complimented on what they have achieved. One of the purposes of this paper is to draw attention to their efforts while encouraging development of resources to make the task easier and progress faster while ensuring that the next generations of collector learns to provide information in a way that maximizes the value of their work by enabling it to be easily and rapidly incorporated into the national biological infrastructure.

References

- Alpha Software (2012) Alpha Five, v. 11. Alpha Software, Inc, 70 Blanchard Road Burlington, MA 01803-5100. <http://www.alphasoftware.com/company>
- Denslow, M., Ball, J., Gillespie R.G., Gross, J, Guralnick, R., Hill, A., Lynn, S., Murrell, Z., Oboyski, P., Sallans, A., Smith, A., Snyder, C., Kipling, W. (2013) Notes from Nature - Involving Citizen Scientists in Transcribing Natural History Specimen Records. Presentation to Annual Meeting of the Society for Preservation of Natural History Collections, 2013. [Abstract].
- Kelvin (1883/05/03). Electrical Units of Measurement. Pp. 73-136 *in* Kelvin, W.T. (1889) Popular lectures and addresses. vol. 1. MacMillan and Company, London. <http://ia600308.us.archive.org/17/items/popularlecturesa01kelvuoft/popularlecturesa01kelvuoft.pdf>
- Microsoft (2010) Excel. Microsoft, Redmond, Washington.
- Thiers, B. [continuously updated]. Index Herbariorum: A global directory of public herbaria and associated staff. New York Botanical Garden's Virtual Herbarium. <http://sweetgum.nybg.org/ih/>

Appendices

1. 2011 Survey
2. 2012 Survey, hard copy version.
3. Herbaria that had been transferred, closed, or discarded by December 31, 2013. (In preparation).

Survey of Digitization in US Herbaria - 2011

This shows the questions asked. It is not the original form; that had a lot more blank space. The survey was kept short out of respect for the respondent's time. It was sent by the US Postal Service to all herbaria listed in Index herbariorum and not described as transferred, closed, or ddefunct.

Measuring Digitization Progress

Herbarium Code: _____

Specimen total (estimate): _____

Number of specimens databased: _____

Number of specimens imaged: _____

URL for searching database: _____

URL of regional node through which data are available: _____

Other nodes through which your specimen data are available: _____

Basic information

Herbarium Name: _____

Department: _____

Address 1: _____

Address 2: _____

City: _____ Zip Code: _____

Phone: _____

Taxonomic focus: _____

Geographic focus: _____

PO Box: _____ Mail Stop: _____

Lat.: _____ Lon.: _____

Name of contact person: _____

Email of contact person: _____

Appendix 2

In April, I sent out the following survey electronically on behalf of the US Virtual Herbarium executive. The response was rather poor so I am sending out a slightly modified version via regular mail. The information really is important. We need to know how much progress is being made, where it is being made, and where it is not being made. Please, take the time to answer the questions and mail it back to me, preferably before June 30 (but better late than never). A summary of the results will be posted to <http://herbarium.usu.edu>. There is space for your comments and suggestions at the end.

Return address

Dr. Mary E. Barkworth, Co-Chair
US Virtual Herbarium Project
Department of Biology
Utah State University
5305 Old Main Hill
Logan, UT 84322-5305

CONTACT INFORMATION

1. Who is completing this survey? _____
2. What is your email address? _____

ABOUT YOUR HERBARIUM

3. About how many specimens are there in your herbarium? Please provide a single number, not separate estimates for different kinds of specimens. _____
4. Is your herbarium listed in *Index herbariorum*? Circle the appropriate answer. (Check its web site, <http://sweetgum.nybg.org/ih/>, if you are not sure). YES NO

If your answer to #4 was "YES":

5. What is its code? (Please make sure that the information in *Index herbariorum* is current) _____

This is in red because several people did not answer it – which meant searching IH for the information.

Go to question 9

If your answer to #4 was "NO":

6. What is the name of the institution that owns your herbarium?

7. What is the street address of your herbarium, if different from the above?

8. What is the mailing address of your herbarium, if different from the above?

ABOUT YOUR COLLECTION

9. Have you started databasing, imaging, or georeferencing the specimens in your collection?

YES NO

If your answer to question 9 was “NO”, you have finished the survey BUT, we would like to know whether you are interested in digitizing your collection (digitizing = databasing, imaging, and georeferencing its specimens) and, if so, what you consider the biggest obstacle to do so. If you are not interested in doing so, we would be interested in knowing why. There is the space at the end for comments.

If your answer to question 9 was “YES”, Please answer the questions below.

DIGITIZATION PROGRESS

Databasing. Some herbaria are entering data for a few fields when imaging, then completing data entry later. For that reason, there are two questions concerning databasing.

10. How many specimens in your collection have been at least partially databased? _____

11. How many specimens have been fully databased (you may answer unknown)? _____

Imaging. The questions below distinguish between imaging specimens (biological material) and imaging labels. If you do not distinguish between the two, put an asterisk by the answer for specimens.

12. How many of your *specimens* have been imaged? _____

13. How many of your *labels* have been imaged? _____

Georeferencing.

14. How many of your specimens have latitude and longitude information? _____

ACCESS

The next questions ask about the web site(s) through which your specimen information is available. If your database cannot be searched via a web site, you have finished the survey. Thank you for taking the time to complete it. If you wish to make a comment or suggestion, please use the space at the end. Hand written comments are welcome.

15. If your records are searchable via an institutional web site, what is its URL?

16. If your records are searchable via one or more regional websites, what are their URLs?

17. If your records are available through one or more taxonomically focussed web sites, what are there URLs?

18. If you provide searchable access to your records through a regional web site that lies primarily outside the US, please indicate the focus of the site and its URL.

19. If your records are searchable via an institutional web site, what is its URL?

20. If your records are searchable via one or more regional websites, what are their URLs?

21. If your records are searchable through on or more taxonomically focussed web sites, what are there URLs?

22. If you provide searchable access to your records through a regional web site that lies primarily outside the US, please indicate the focus of the site and its URL.

YOUR COMMENTS

A summary of the survey results will be posted to <http://herbarium.usu.edu/USVH> in July.

THANK YOU FOR TAKING THE TIME TO COMPLETE THIS SURVEY

