

Minutes of annual meeting of Consortium of Northeastern Herbaria, June 8-9, 2009, Durham, N.H.

Janet Sullivan and Chris Neefus of the University of New Hampshire, who were hosts for the meeting, welcomed representatives of regional herbaria, and Patrick Sweeney, chair of the consortium's steering committee, provided a brief introduction on consortium activities. He pointed out that the region covered by the consortium includes 138 institutions with more than 24 million specimens; many of the collections are small (more than 50 have fewer than 10,000 specimens).

Sweeney reported that several goals set during the 2008 meeting had been achieved, including establishment of a web page and creation of subcommittees that had been busy considering governance, membership and other issues. The consortium now has 36 member institutions in eight states or provinces, representing 45 collections. Sweeney led a preliminary discussion on key issues that will be discussed more fully later in the day, including goals for the next year, governance, funding for projects, how the consortium should share data online (whether through GBIF only or through a portal established by the node to serve regional data), and how to increase membership.

Janet Sullivan presented a report on the status of digitization at UNH's A.R. Hodgdon Herbarium (NHA). The herbarium, which has about 200,000 specimens, began scanning specimens in 2007, entering specimen data in a Specify database. More than 25,000 specimens have been scanned in the past two years, and more than 31,000 specimens have been entered in the database, although the information on specimens will not be available online until the internet interface is released for the new version of Specify (Specify 6) later this year.

Bob Capers reported on databasing and imaging activity supported by a recently awarded NSF grant at the G.S. Torrey Herbarium (CONN) at the University of Connecticut. Students are using three scanners to image specimens and to enter information into a BG-Base database. About one-third of the 180,000-specimen collection has been databased in the past 10 years, and information on those specimens is available online. The grant will allow the entire collection to be databased and imaged in the next three years. Capers also reported on an initiative involving high school teachers, who have been writing lesson plans so their students can use the online specimen data as part of biodiversity, botany or ecology instruction.

Paul Harwood of the Brooklyn Botanic Garden (BBG) reported on an effort to database specimens from the Tri-State area, supported by an NSF grant, and to image all types, supported by a Mellon Foundation grant. He also reported on the Metropolitan Flora Project, which is documenting the flora that occurs within 50 miles of New York City. About 3,000 species – 15% of the total for the United States – are listed in the searchable database. Harwood said the project has been able to document change in the regional flora over time, showing, for instance,

that the exotic *Celastrus orbiculatus* has increased in frequency while the native *C. scandens* has declined.

Kat Dewitt of the New York Botanical Garden (NY) presented a talk on the garden's Latin American Plants Initiative, of which she is the coordinator. Through the project, all types from Latin America are being imaged, as previously was done with types from Africa; both projects have been financed by the Andrew W. Mellon Foundation. In addition to establishing a permanent record of types, the project has created an international network of herbaria and makes possible a form of repatriation because images of the types can be sent back to the countries where the type specimens were collected. One of the major benefits of the project is that it captures images of specimens in foreign countries that might not be able to image the specimens themselves. LAPI has 68 participating institutions in 29 countries; approximately 605,000 types (more than half of them in European institutions) will be imaged through the project eventually. Types are being scanned at 600 dpi with an inverted scanner, and the resulting TIFF format images, which are about 200 MB each, are saved. African type images are available through the Aluka web site, and the LAPI project images will be made available through JSTOR. Five people are working full-time on the project.

Rodger Evans of the E.C. Smith Herbarium (ACAD) at Acadia University reported on databasing and imaging work being done there. Databasing of the herbarium, which has 200,000 specimens, dates from 1995, and about 25,000 specimens have been databased and imaged to this point. Scans are made at 300 dpi, using an Indus 5002 book scanner, which creates a 60 MB digital image in about eight seconds. The resulting image files are saved on a Sun X4540 storage server, which can hold up to 100 TB of data. The database includes specimen record information in 24 fields. Data on specimens and images are served to the web using Mr. Sid software.

Peter Desmet spoke about Canadensys, a consortium of biological collections in Canadian universities, including 12 herbaria, nine of which are shared with the CNH. Canadensys hopes to database 3 million specimens during the next five years, an estimated 20% of the collections held by Canadian universities. The Montréal Biodiversity Centre is coordinating the databasing work, ensuring that data are compatible with GBIF standards, and is providing a web portal for the data. The center will maintain a data server as well as an image server, providing information to users and member institutions. Desmet spoke about challenges that Canadensys is dealing with, which are similar to those of other institutions and agencies involved in databasing – questions about how to deal with duplicates, how to decipher hand-written labels and the difficulty in georeferencing specimens. Canadensys plans to make available guidance on best practices, resources on georeferencing and to offer training workshops for member institutions.

CNH business

Sweeney then led a discussion on CNH business, including the location and format of next year's annual meeting. The group agreed that the next annual meeting will be held June 1-5, 2010, in

Ottawa in conjunction with those of the Canadian Botanical Association and the Society for the Preservation of Natural History Collections. Several people suggested that a three-day meeting be considered so a field trip could be organized either before or after the formal program; the idea was greeted favorably, but no final action was taken. Jennifer Doubt, a consortium member from the National Herbarium of Canada (CAN), and Luc Brouillet from the Herbarium Marie-Victorin (MT) at the Université de Montréal agreed to help coordinate planning for the meeting; James Macklin from Harvard agreed to help coordinate with SPNHC.

In discussing membership, Sweeney noted that only 25% of the herbaria in the consortium's area are members. Individuals volunteered to make personal contact with collections in each of the states and provinces to encourage additional herbaria to participate in consortium activities. These people and their states/provinces are: Karen Searcy, Massachusetts; Patrick Sweeney, Connecticut; Bob Capers, Rhode Island; Alina Freire-Fierro, Pennsylvania and New Jersey; David Barrington, Vermont; Rodger Evans, Maritime provinces of Newfoundland, New Brunswick, Nova Scotia and Prince Edward Island; Luc Brouillet, Quebec and Ontario; Robert Naczi and Paul Harwood, New York; Janet Sullivan, Maine and New Hampshire. It was agreed that these contacts would be made by Sept. 1. Capers agreed to revise an earlier letter that had been emailed to herbaria in the region, inviting participation in the consortium and providing information on next year's meeting.

James Macklin said he remains active in national informatics efforts and will be attending the national meeting of the Botanical Society of America, where he and others attending (Janet Sullivan, Luc Brouillet, Alina Friere-Fierro and others) will report on the CNH's activity.

Several people made suggestions for additional information that could be placed on the consortium's web page, including information comparing the merits of cameras vs. scanners to create digital images; detailed instructions on how to use GBIF's new Integrated Publishing Toolkit; information on best-practices in terms of resolution of digital images and the number of fields needed to meet GBIF standards.

There also was a brief discussion about what projects the consortium should be thinking about in the short-term, such as finding ways to help small herbaria database their specimens. Luc Brouillet suggested that the consortium should consider a new project to do after finishing the three-species demonstration project – a vascular flora of the region, possibly, or a project involving alpine and arctic plants. It was agreed that the consortium should finish the demonstration project first and set deadlines by which all participating herbaria will submit their data – *Celastrus scandens* by Nov. 1, *C. orbiculatus* by Feb. 1 and *Epipactis helleborine* by May 1.

On behalf of a subcommittee that also included Karen Searcy and Janet Sullivan, Sweeney led a discussion of recommendations for governance of the consortium. In a voice vote, those attending the meeting accepted the committee's recommendation that the steering committee be

retained. Peter Desmet was added to work on Canadian membership. David Barrington reported that Julie Dragon had left the University of Vermont (VT) and the region and Sweeney reported that Kathie Hodge (CUP) was no longer helping with the development of the web site. Dragon and Hodge were both retired from the steering committee. Those attending the meeting also approved a recommendation that a decision be deferred on whether an executive committee should be created to make major organizational decisions. It was suggested that the steering committee chairman should serve a two-year term so the outgoing chair can assist the new chair during the third year of his or her term on the steering committee. By consensus it was agreed that the three-year terms of steering committee members should be staggered to ensure continuity. It also was agreed that subcommittees could be created of consortium members and that only the chairman of the subcommittee needs to be a member of the steering committee. Subcommittees were approved for: informatics; national coordination; membership; organizing the next meeting; funding and grants; and the web site. The government subcommittee, having completed its work, was abolished. Elizabeth Zacharias of the Harvard University Herbaria agreed to serve on the steering committee to help coordinate the consortium's work on selecting projects. Those at the meeting agreed that the steering committee should be empowered to make routine business decisions on behalf of the consortium but that major decisions should be made only after seeking the advice of other members during the annual meeting. It was agreed that a decision would be deferred on development of an agreement of understanding among institutions that are members of the consortium.

At the conclusion of the business meeting, Chris Neefus spoke about a funding opportunity available to land-grant universities through their agricultural experiment stations. This money should be more readily available than the highly competitive grants from the NSF or the Mellon Foundation. The money is available because a project entitled Developing a National Virtual Herbarium already has been approved by the USDA. Because that program has been approved, individual state experiment stations can designate resources to support the goals of that WERA1015 program. In fact, individual experiment stations must spend 25% of their budgets each year for multi-state programs such as the Virtual Herbarium project, Neefus said, and anything related to databasing or imaging of herbaria specimens would be eligible for funding under this multi-state funding obligation.

Macklin reported that the NSF's Improvement of Biological Research Collections program supports digitization of biological collections and now encourages creation of regional networks linking large and small institutions in the interest of making more data available online. A total of \$6 million is available this year for such efforts. He said even more money is available through the cyberinfrastructure program. Sweeney said that Mike Donoghue is involved in the NEON program and that, although the program is well-funded, it is unclear how the consortium might benefit from that resource.

Les Mehrhoff presented a brief talk on the National Early Detection Network for Invasive Plants, which has been organized by the US Geological Survey, the National Biodiversity Information

Infrastructure and the Invasive Plant Atlas of New England. As planned, the network will involve people of varying skills to report non-native plants, including volunteers who are not professional trained botanists; they will work with herbaria to confirm the identification of newly recorded species and with government agencies who would be able to respond and possibly eradicate potentially troublesome species before they become well-established. Mehrhoff said one of the initiatives of the network is to distribute specimens of newly arrived species to herbaria to facilitate identification of the species by botanists who may not have seen them before.

Richard Primack presented an after-dinner talk on his research using historical records of plant phenology, including notes from Henry David Thoreau, to explore changes in flowering time during the past 150 years.

The second day of the meeting was devoted to workshops. The morning workshops included a demonstration of the new Integrated Publishing Toolkit developed for GBIF to make it easier to upload collection databases and publish them on the web. The workshop was led by Paul Morris and James Macklin. The software is available for free download at <http://code.google.com/p/gbif-providertoolkit> and the demonstration site is at <http://ipt.gbif.org>. The toolkit makes it possible to upload tab-delimited text files. It assigns globally unique identifiers and assigns data to Darwin Core fields. The application includes basic quality control tools and a mapping server, and it publishes the data to the web, using Apache Tomcat software. Morris provided a demonstration of the IPT, using a 1,000-specimen dataset from Acadia University.

Nestor Pérez-Molière led a workshop on digital imaging technology, discussing both cameras and scanners. Pérez-Molière, who manages the digital imaging center at the New York Botanical Garden, stressed the answer to the question of whether cameras or scanners are better for obtaining digital images depends on the material being imaged, the intended uses of the images and the amount of money available. NYBG has more than 250,000 digital images in its archive. Individual specimens can have as many as 20 images associated with them. Pérez-Molière predicted that there will be increasing demand in the future to associate photographs taken in the field with images of pressed specimens. The advantage of scanners, he said, is that they are simple to configure and use, and they produce excellent images; however, they produce images slowly, so the number of specimens that can be imaged in a given time is less than with a digital camera. Cameras also are able to image bulky material that cannot be scanned, but using a digital camera requires a more sophisticated understanding of light quality and photographic technique to obtain good images, he said, plus it is more expensive to obtain all the equipment needed to take good digital images with a camera. Pérez-Molière also discussed the various formats in which image files can be stored and the importance of recording metadata on images.

Reed Beaman of the Florida Museum of Natural History led a workshop on georeferencing. Beaman said it is essential to have latitude and longitude information associated with specimens

in natural history collections. Having location information makes possible research on why an organism occurs where it does, including predictive modeling. Beaman said that having accurately georeferenced specimens is important to identify biodiversity hotspots, to explore changes in phenology within a particular area, to investigate changes in species' frequency over time (useful for studying increases among invasive species and declines in rare taxa), and to manage pests and pathogens. Beaman provided information on georeferencing resources, including BioGeomancer, GeoLocate, Google Earth and on best-practices guidelines such as the [MaNIS/HerpNET/ORNIS Georeferencing Guidelines](#) and those available through GBIF. Beaman spoke about factors that contribute to coordinate uncertainty, including the extent of a locality, lack of information on the geographic datum that is used, imprecision in distance measurements and imprecision in direction measurement. The BioGeomancer software includes a procedure that allows for managing uncertainty in location information.

Andy Bentley demonstrated the Specify 6 open-source database software developed by the University of Kansas. Bentley said that Specify, which has been supported by the NSF since 1987, already is used to catalog about 20 million specimens in more than 230 collections but that it is hoped the new version will advance digitization of collections because it will permit better collaboration and because of the improved support and training being offered. Specify 6 is a more capable program than earlier versions, recording not just routine specimen data but also DNA sequences and field notebook entries and images; it also makes it possible for collections to track loans, map specimen information and produce specimen labels with barcodes printed on them. Specify will provide (later this summer) web-posting capability for natural history collections, using GBIF's new Integrated Publishing Toolkit. Support is available for all non-profit research institutions, and data entry forms can be customized to satisfy individual client's needs, Bentley said. Because it has been developed as an open-source product, it is hoped that the user community will develop plug-ins to meet specialized needs and to expand the capabilities of the program. Plugins already are available for ESRI products like ArcGIS and for Photoshop; the software comes with Google Earth and BioGeomancer built in. Bentley also talked about Specify Workbench, which is a simplified version of Specify that is designed for easy data entry and for use in the field. Data can be entered in spreadsheet form, which may make the process faster, and the software permits georeferencing before the data are uploaded in batches to the full database.

Submitted by Robert Capers on June 17, 2009 with slight modifications made by Patrick Sweeney on July 30, 2009. Posted on the CNH website on July 30, 2009.