

THE MBLWHOI LIBRARY DIGITAL HERBARIUM

Joseph M. deVeer
Technical Services Coordinator
jdeveer@mbi.edu

Amy Stout
Digital Systems & Services Coordinator
astout@mbi.edu

Margaret Rioux
Systems Librarian
mrioux@whoi.edu

MBLWHOI Library
Marine Biological Laboratory
7 MBL Street
Woods Hole, MA 02543-1015

ABSTRACT: The MBLWHOI Library Digital Herbarium is a project to digitize the Library's herbarium collection of 8,000 vascular and non-vascular plant specimens. Images along with species and specimen-related data will be freely available on the Internet via a fully searchable database. This paper discusses the value of the project, and the technical and botanical background of the database, highlighting use of Endeavor's Voyager and ImageServer software.

KEYWORDS: MBLWHOI Library Herbarium; herbaria; digital herbarium; marine algae; plants; Cape Cod

Introduction

The MBLWHOI Library Herbarium (SPWH), originally part of the George M. Gray Museum at the Marine Biological Laboratory, documents the flora of Cape Cod, Massachusetts and associated coastal islands. The collection, dating back to the 1870's, consists of approximately 7,000 vascular and 1,000 non-vascular plant specimens. The non-vascular collection consists primarily of marine algae of Cape Cod and the Islands, though other regions of the world are represented. Several eminent botanists such as F. W. Pennel, F. S. Collins, W. G. Farlow and W. R. Taylor have contributed to the herbarium.

In 2002 the MBLWHOI Library assumed ownership of the herbarium collection that is now housed at the Data Library and Archives facility. Supported by a grant from the Woods Hole Oceanographic Institution Sea Grant Program, the Library is digitizing the collection and creating a fully searchable database of specimen images and associated

data that will be freely available on the Internet [<http://www.mbl.edu/herbarium>]. This paper discusses why we felt such a project was worthwhile, and methods, equipment, software and metadata standards used. Use of the Voyager integrated library system from Endeavor Information Systems, Inc. is highlighted.

Why a Digital Herbarium?

Though modest in size the MBLWHOI Library Herbarium is an invaluable resource because Cape Cod is a botanically unique and interesting region. Most of Massachusetts' rare plants occur on Cape Cod or in the Connecticut River valley. With regards to marine algae, Cape Cod is widely recognized as a significant phytogeographic boundary between the cold and warm water flora of the eastern coast of the United States. Cape Cod is a barrier to the northward movement of the Gulf Stream, and the southward flow of cold Gulf of Maine water. As such, it is the effective northern limit of many warm-water, even tropical species, and the southern limit of several cold-water species (Hooper, et al. 2002). The Cape is thus unique in the diversity of marine algal species found there. The MBLWHOI Library Herbarium documents this important and diverse flora. Given the significance of the collection, we felt that digitization would certainly be a worthwhile effort.

The benefits of digitizing such a collection and making it available via the Internet are many. Most obviously, it provides worldwide access to specimen images and concomitant data. Collection data, previously accessible only via specimen labels and a local paper catalog, will now be accessible in a database with rich searching capabilities. This will greatly facilitate and enhance studies of floristic and environmental changes over time. Filtering name searches through uBio©, a Taxonomic Name Server, will assure retrieval of desired results, even if an investigator searches with antiquated nomenclature, or obscure or foreign common names. Another anticipated benefit is that electronic access will reduce unnecessary handling, thus contributing to long-term preservation of the specimens.

Digitization of the marine algal collection is a joint project with the Northeast Algal Society (NEAS). One goal of NEAS is "to bring together information on the diversity of algal species and their distributions along the northeastern coast of North America" (Sears 2002). To facilitate this, the NEAS has divided the northeastern coast into twelve regions. As stated in the Introduction to the *NEAS Keys to Benthic Marine Algae of the Northeastern Coast of North America from Long Island Sound to the Strait of Belle Isle*, "to document floristic changes with a degree of confidence it is imperative to have a broad database of the taxa present in each of the twelve regions represented. With such a database, introductions of non-indigenous species are determined with confidence, as are changes in geographic ranges of our taxa." Each marine algal record in the database contains a field indicating from which of the twelve regions the specimen was collected, allowing for searches based on this NEAS scheme. This will enable existing and future collections of the MBLWHOI Library Herbarium to contribute toward NEAS' goal of understanding species diversity and distribution along the northeastern coast.

NEAS has kindly allowed us to use data from their *Keys* to describe each species. Current taxonomy, authorities, geographic range, habitat and seasonality are taken from the *Keys*. A similar source of information for the herbarium's vascular plants will be sought for these specimens. Thus another benefit of the project is the creation of a content-rich, online compendium of the vascular and non-vascular flora of Cape Cod and the Islands.

Digitizing the Specimens

One goal of this project is to create digital images of the herbarium specimens that are of high enough quality to serve for scientific study and identification where possible. Many similar projects such as the digitization initiative at the New York Botanical Garden [<http://www.nybg.org/bsci/hcol/>] are making use of a digital camera and copystand with excellent results. We soon realized that obtaining the desired image quality with a digital camera would require an investment beyond the reach of our funding. Inspired by vPlants™ a virtual herbarium of the Chicago region [<http://www.vplants.org/>], we decided to use a flatbed scanner as our primary means of digitization. Following methods established by the vPlants™ project, specimens are scanned with an EPSON Expression 1640XL Graphic Arts scanner. The unit is inverted and suspended 17 inches above the support bed by means of a custom built stand (Fig. 1). Inversion of the scanner allows the herbarium sheets to be placed right side up beneath the scanner to minimize damage to the specimens. A borrowed digital camera and copystand will be used for bulky or fragile specimens that would not withstand scanning.

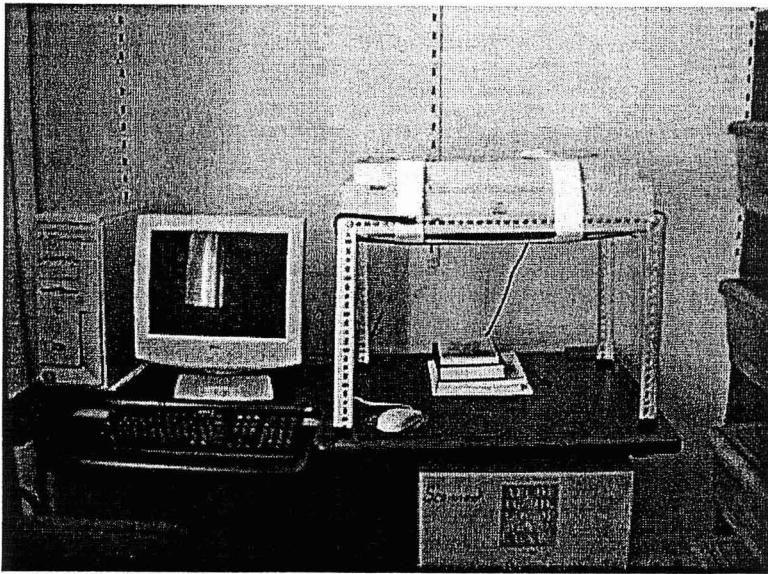


Fig. 1. Specimen scanning workstation showing inverted scanner and custom-built stand

Adobe Photoshop 5.5 (Educational Version) is used for image manipulation. The entire herbarium sheet is scanned at a resolution of 600 ppi (pixels per inch). This allows for the creation of high-resolution images, showing fine detail. For each specimen we present an image of the full sheet and enlargement images of small specimens or where greater detail is required. The original scan results in the creation of a ~200 MB TIF, approximately 7000 pixels in width. This full-size TIF is cropped to create high-resolution images of small specimens or of specific portions of a plant. Once these have been prepared, the full-sheet TIF is resized to 2000 pixels in width (~17 MB) for archiving. The detail TIF images are archived along with the image of the full sheet. A 960 pixel JPEG of the full herbarium sheet is created for web presentation. When viewed at a screen resolution of 1024x768 pixels this appears roughly life size, giving a good overall impression of the entire specimen.

Completed JPEGs are filed on the server for web presentation via Endeavor's ImageServer software. ImageServer is used to associate the images with their concomitant data from the herbarium sheet labels. These data include scientific name, accession number, collector, determiner, collection date and location. ImageServer communicates with the Voyager Cataloging client on the same workstation computer, creating a skeletal MARC record that includes a link to the associated images.

Workflow

The vast majority of scanning is accomplished by volunteer workers from the MBL Associates and the Botanical Club of Cape Cod and the Islands. Volunteers, working in teams of two, scan the specimens, creating the original TIF, a 960 pixel JPEG of the full herbarium sheet, and a thumbnail. The original TIF is saved at full size (~200 MB, ~7000 pixels in width) in a desktop folder. Volunteers next use ImageServer to file the JPEG images on the server and to enter associated data from the herbarium sheet labels, creating a brief MARC record in the cataloging module that contains a link to the image. Project staff follow up by examining the images to determine whether enlarged detail images of small specimens or portions of larger ones are required for web presentation and archiving. These are created from the original 7000 pixel TIF, which is subsequently reduced to 2000 pixels for archiving on CD's. Project staff also perform full cataloging of the original brief record created with ImageServer. To keep track of the CD archive, each MARC record has a field to record which CD contains the associated image(s).

Cataloging and Metadata

The Voyager integrated library system from Endeavor Information Systems, Inc., is the database platform used for this project. The herbarium database is set up as its own catalog – separate from the main MBL/WHOI Library catalog. Each specimen is fully cataloged in MARC format. Within the cataloging client, the MARC tag definition tables were edited for the herbarium database so that each field and subfield is defined for herbarium-specific data. While remaining consistent with standard MARC format for

bibliographic data, the fields and subfields were re-labeled to make it easier for the cataloger to put the appropriate specimen data in each field. Also, in some cases, additional indicator codes were added in order to allow differential labeling of the display in the public online catalog. In the brief and detailed displays of the public interface, the headings are tailored to herbarium data. Below is a sampling of bibliographic fields and subfields, including indicators, defined for herbarium data, along with the display heading as it appears in the public interface:

090: : |a SPWH 3849 – [*Specimen Accession Number*]
100:1 : |a Hehre, Edward J. – [*Collector*]
245:00: |a *Coccotylus truncatus* |b (Pallas) M. J. Wynne et Heine – [*Current Species Name*]
247:00: |a *Phyllophora brodiaei* |b (Turner) J. Agardh – [*Synonyms*]
520:0 : |a Drift. – [*Ecological Data*]
561: : |a 25 Jan 1969 – [*Collection Date*]
561: : |b Nobska Point, Woods Hole, MA – [*Collected at*]
700:12: |a Sears, James R. – [*Determiner*]
700:15: |a Conway, J. R. – [*Specimen Preparator*]

In addition to data from the specimen labels, species-specific data are added to each record. Following is a sampling of the fields used to describe the species in question:

522:1 : |a Long Island Sound northward. – [*Range (Northeastern North America)*]
522:2 : |a N/S – [*Relation to Cape Cod*]
522:3 : |a Sublittoral to 40 m; on rock, shell or other hard substrata. – [*Typical Habitat*]
522:4 : |a Perennial. – [*Seasonality*]
522:5 : |a Rare. – [*Status*]
754: : |a Rhodophyta (Red algae) – [*Division*]
754: : |b Florideophyceae – [*Class*]
754: : |c Gigartinales – [*Order*]
754: : |d Phyllophoraceae – [*Family*]

To enhance location searching, the collection location is described in detail with LC geographic name headings and latitude/longitude as follows:

034:0 : |d W0703900 |e W0703900 |f N0413100 |g N0413100 – [*Lat/Long coded*]
255: : |c (41°31'N, 70°39'W) – [*Lat/Long for display*]
651:0 : |a Barnstable County (Mass.) – [*Related Geographic Locations*]
651:0 : |a Nobska Point (Woods Hole, Mass.)
651:0 : |a Woods Hole (Mass.)
651:0 : |a Falmouth (Mass.)
651:0 : |a Cape Cod (Mass.)
651:0 : |a Massachusetts.
651:0 : |a NEAS Region 2.

A concise textual description of the specimen, drawing together the primary data elements, is displayed in the brief record portion of the public display. An example MARC field for specimen description is shown below:

520:1 : |a Herbarium sheet with 4 pressed, mounted specimens of *C. truncatus* collected 25 Jan 1969 by E. J. Hehre. Found in drift at Nobska Point, Woods Hole, MA. Identified by J. R. Sears; prepared by J. R. Conway.

Data elements have been associated with appropriate Access Points within the Darwin Core metadata standard for description of natural history collections (Table 1).

Herbarium Data Element	Relevant Darwin Core Access Point
Current Species Name, Synonyms, Original Identification	Scientific Name
Class	Class
Order	Order
Family	Family
Specimen Accession Number	Catalog Number
Collector(s), Determiner	Collector
Collection Date	Year, Month, Day
Collection Location, Location Description	Locality
Related Geographic Locations	Country, State/Province, County
Longitude, Latitude	Longitude, Latitude

Table 1. Sampling of data elements as mapped to Darwin Core

Links to/from External Resources

The Digital Herbarium is set up as a linkout provider for the National Center for Biotechnology Information (NCBI) Taxonomy Browser [<http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi>]. In addition, each record in the Digital Herbarium contains a direct link to the appropriate species record in AlgaeBase [<http://www.algaebase.org/>].

References:

- Hooper, R.G., Mathieson, A.C. & Wilce, R.T. 2002. Geographic distributions of marine algae along the northeastern coast of North America. In: *NEAS Keys to Benthic Marine Algae of the Northeastern Coast of North America from Long Island Sound to the Strait of Belle Isle* (ed. by J.R. Sears), pp. 133-136. Dartmouth, MA: NEAS.
- Sears, J.R. 2002. *NEAS Keys to Benthic Marine Algae of the Northeastern Coast of North America from Long Island Sound to the Strait of Belle Isle*. Dartmouth, MA: NEAS.