Digitization of Herbarium Specimens, a Collaborative Project

Larry Schmidt

Abstract
Digital collections in museums, libraries, and other agencies are expanding rapidly. Digitization of books, maps, photographs, and other objects for both preservation and dissemination are changing the way libraries work. Librarians who have, historically, worked with collections in print formats are anxious to move toward digital content as the wave of the future. Rather than stand on the sidelines and watch as the digital movement passes by, librarians need to be proactive and create digital content using both collections within the libraries' walls and proactively seek opportunities to work collaboratively with stakeholders outside of the libraries' hallowed halls. The creation of the Grand Teton National Park Digital Herbarium is an example of a collaborative effort between librarians, botanists, and national park personnel. All three groups needed to work together to create a digital herbarium useful to botanists and other individuals interested in the flora of Grand Teton National Park. This paper describes the project and suggests how other institutions may emulate the collaborative effort where more than two disparate groups have a stake in the final outcome of the digital project.

Introduction
Collaborative digital projects in academic libraries are fast becoming the norm as universities move toward the creation of digital institutional repositories.[1,2] The importance of digital collections in science are apparent.[3,4] Many libraries who have historically collected material for both research and education purposes are naturally drawn to these digital collections and are taking a lead in this arena. Efforts in digitization range from the creation of digital images of photographs to historical documents. An example of this type of digital collection is the Oregon State University Libraries' Linus Pauling Research Notebooks, a fully indexed digital collection of Linus Pauling's 46 research notebooks.[5] Collaboration exists on a number of levels within and without libraries. Collaborative projects between libraries and other departments and college campuses exist but they are not as common as digital projects with materials owned by libraries. Large statewide collaborative projects continue to increase in numbers and will likely do so as long as the funding and interest in these collaborative efforts continue. Examples of statewide and regional digital collections are Online Archive of California,

Larry Schmidt is Assistant Librarian, Reference and Instruction Services, University of Wyoming, email: lschmidt@uwyo.edu.

Northwest Digital Archives, Virtually Missouri, and Western Waters Digital Library.

Libraries are not alone in the digitization of their collections. Museums and herbaria have also entered into digitization efforts for purposes of both preservation and dissemination. One of the reasons that these collections have moved online is that they are all too often hidden from the public eye.[6] Herbarium specimens are stored in a manner that makes them hard to access and are often of limited use to the public. These digital collections have become more important as users see their potential for their research. Some examples of digital herbaria are listed below.

- MBL/WHOI Library Digital Herbarium[7]
- Utah Valley State College Digital Herbarium[8]
- vPlants A Virtual Herbarium of the Chicago Region[9]
- The Virtual Herbarium of The New York Botanical Garden[10]
- OSU Type Specimen Images and Original Descriptions[11]
- WTU Herbarium Image Collection[12]

Examples of the importance of herbaria collections to research are plentiful and digitization of herbarium collections can only add to the potential for new research. Digital as well as traditional collections have been used to document the decline of species, the range of rare species, and the spread of invasive plants.[13] Researchers are interested in the establishment of databases to record biodiversity and specimen inventories, which will allow regional and global inventory and monitoring of plant species.[14,15] To meet this need, herbaria around the world have begun to digitally image their collection and create searchable databases to maintain them.[16,17]

Project

The idea for the Grand Teton National Park Digital Collection started innocently as a possible "Weeds of Wyoming" digital collection. The plan was to work with the University of Wyoming Agriculture Experiment Station to develop a site illustrating common weeds of Wyoming. Initially both herbarium specimens and natural images would have been captured digitally for the collection. Instead, the current Principal Investigator (PI) for the project, Mary Ann Harlow, considered a collection that was more manageable and would benefit researchers. The Grand Teton National Park (GTNP) herbarium is a historical record of plant specimens collected in the park. This collection has provided researchers with a physical record of vascular flora since the 1920s. Herbarium specimens in GTNP are currently housed in a converted walk-in freezer in an old wooden building located near park headquarters. The collection is integrated with other artifacts of archeological and historical significance, and little space is available to work with the collection on site. Access to the collection is by appointment only and requires a park employee escort. Because specimens are not available for loan, individuals who wish to work with the collection have to transport them to another park facility in order to use them. This is a collection ripe for digitization. Not only would digitization of the collection make it easier to access, it would also allow researchers to look at the specimens before a visit to the park. They could decide which, if any, specimens they need to handle. The digital collection would therefore:

- reduce the need for physical handling,
- create access to a previously hidden collection,
- increase visibility through the WWW,
- support access regardless of user skill,
- create an archive of GTNP herbarium specimens,
- add to the Wyoming distributive database of vascular plants, and
- could be used as an inventory of specimens in the collection.

Harlow and two other librarians, Larry Schmidt and Paula Muñoz, captured digital images of herbarium specimens during the summer. Throughout the year, librarians cleaned up the database and ported metadata to the digital image and management software. This project began the summer of 2005 thanks to the generous support from the University of Wyoming Libraries administration and a grant from the National Park Service (NPS).

Procedures

The first step taken in preparation for this collaborative project was a meeting with National Park personnel. Investigators discussed protocol and procedures for the digitization workflow. This involved determining how specimens would be transported to and from the University of Wyoming-National Park Service Research Center (UW-NPSRC) where digitization of the specimens takes place. The visit also included instruction from the park botanist on handling herbarium specimens. After meeting with park personnel, the librarians met with the Rocky Mountain Herbarium (RMH) botanists to discuss research needs from a digital collection of specimens. As with any digital project, choosing hardware and software capable of delivering a high quality prod-

uct is one of the first priorities. In this case, a portable system was required due to space considerations and the fact that the camera would only be used seasonally. The camera equipment was set up at the UW-NPSRC in early summer soon after the facility opened.

The librarians transported specimens from the park location to a dedicated photography room, photographed them and returned them. Data about each specimen was entered into an Access database that was supplied by the RMH and reconfigured by a library systems programmer. At the UW-NPSC, images were saved on a hard drive and backed up on CD-ROM as well as an external drive. Once the camera and computer hardware were returned to the University of Wyoming Libraries (UW Libraries), the images were transferred from the external drive to a server and the CD-ROMs were stored in a temperature-controlled location. Currently over 6,000 images have been taken of GTNP herbarium specimens.

Images of each specimen were matched by field label and added to the Access database. This allows corrections and additions to be added for each specimen at a later time. Currently there are records for each specimen in an Excel spreadsheet and the Access database. Once the project was completed there were three records associated with each specimen: the Excel spreadsheet supplied by the Park Service, the RMH Access database, and the database associated with the digital imaging software Insight, from LUNA Imaging. The UW Libraries will maintain only the Access database and the digital imaging software database. In the end, the Park Service, the RMH and the libraries will have an accurate list of herbarium specimens complete with information required for botanists.

Collaboration

Herbarium collections represent the work of botanists who collect specimens in the field, press and dry them, and bring them back to the herbarium to mount on paper following specified procedures.[18] Librarians who take on digitization of herbarium specimens must know how to handle the specimens as well as how botanists catalog and organize them. Education or experience in biology facilitates discussions with scientists or botanists through the use of a common language. In order to create a digital herbarium, or for that matter, any digital collection, librarians should note how the material is originally organized and displayed.

Collaboration between the stakeholders is essential for this type of endeavor if the needs of all partners are to be met. In the case of the digitization of the GTNP herbarium there are three stakeholders involved in the project: the UW Libraries, the NPS, and the Rocky Mountain Herbarium (RMH). It is also important to look at how the digital collection will be formatted as well as what will be the expectations of the users of the collection. Librarians, NP personnel, and RMH botanists established how each member accessed the project and what parameters were needed. Botanists have their own protocols of cataloging and displaying field information pertinent to the object, a herbarium specimen. Librarians wish to address user needs and record historical content and information about the digital image itself. The NPS wants historical information as well as a link to the material from their website.

Table 1 illustrates the different types of metadata that will be important to botanists, park personnel and librarians. The metadata for the digital collection will be a combination of the herbarium data and Dublin Core metadata, including information for botanists and the general public. The archival camera data will follow Dublin Core specifications but will be hidden from public view. All the stakeholders and projected users will have the necessary information at their fingertips. It is expected that the digital collection will be available at the UW Libraries' web site with links to it provided at the RMH and GTNP websites.

There are many ways agencies and departments maintain inventories of their collections. The NPS uses a filing system that consists of an Excel spreadsheet which inventories their botanical specimens along with museum and historic artifacts. The Park Service modified the original Excel spreadsheet so it included only

Table 1: Metadata Types

Herbarium Data	Dublin Core Metadata	Archival Camera Data
Scientific Name	Scientific/Common Name	Camera Type
Specimen Number	Accession Number	Lens Type
Collector	Same	Aperture
Collection Data	Same	Image Date
Habitat Description	Habitat Description	Photographer
Geographic Location	Location/Location Withheld	Photo Number
Logitude, Latitude	Same	Shutter Speed
Annotated	Same	ISO

	A	B	C	D	E	F	G	H	I	J	K
1	Catalog	Class	Class	Class 3	Class 4	Sci. Name	Genus [Sci. Name.	Description	Locality	Collection Da	Change Da
5440	GRTE 11293	BIOLOGY	PLANTAE	LILIOPSI	CYPERACE	Carex paysonis Clokey	Carex pays	plant specimen	AMPHITHEATER CIRQUE	7/10/1962	1/7/1999
5441	GRTE 11294	BIOLOGY	PLANTAE	LILIOPSI	JUNCACEA	Luzula piperi (Cov.) Jone	Luzula pip	plant specimen	RIDGE NORTH OF AMPHITHEATER LAKE	6/28/1961	1/7/1999
5442	GRTE 11295	BIOLOGY	PLANTAE	LILIOPSI	ORCHIDAC	Corallorhiza maculata Raf	Corallorhi	plant specimen	ONE MILE UP INDIAN PAINTBRUSH CANYON TRAIL	7/7/1966	1/7/1999
5443	GRTE 11296	BIOLOGY	PLANTAE	LILIOPSI	ORCHIDAC	Corallorhiza maculata Raf	Corallorhi	plant specimen	ON TRAIL FROM SQUARE G RANCH TO LEIGH LAKE	7/7/1956	1/7/1999
5444	GRTE 11297	BIOLOGY	PLANTAE	LILIOPSI	ORCHIDAC	Corallorhiza mertensiana	Corallorhi	plant specimen	WEST OF COLTER BAY AMPHITHEATER	7/16/1967	1/7/1999
5445	GRTE 11298	BIOLOGY	PLANTAE	LILIOPSI	ORCHIDAC	Curallurhiza striata Lind	Curallurhi	plant specimen, rare at site	NEAR THE BRIDGE AT BRADLEY LAKE	7/6/1967	1/7/1999
5446	GRTE 11299	BIOLOGY	PLANTAE	LILIOPSI	ORCHIDAC	Listera caurina Piper	Listera ca	plant specimen	ON TRAIL TWO MILES UP INDIAN PAINTBRUSH CNAYON	7/7/1966	1/7/1999
5447	GRTE 2721	Biology	Thallophy	Eumycete		Peridermium coloradense	Peridermiu	Fungi specimen on Picea National catalog Note: 8/2001: Ther	Hobach Canyon 0.5 miles South East mouth Cliff 640		8/9/2001
5448	GRTE 2722	Biology	Thallophy	Eumycete		Stagonospora foliicola	Stagonospo	Fungi specimen on Graminae National catalog Note: 8/2001: T	2 miles North Jackson Lake Lodge Junction 6780'		8/9/2001
5449	GRTE 2724	Biology	Thallophy	Eumtcete		Thelephora terrestris	Thelephora	Fungi specimen on rotting conifer log National catalog Note	Signal mountain Lodge, GRTE		8/9/2001
5450	GRTE 4988	BIOLOGY	PLANTAE	MAGNOLIO	APIACEAE	CYMOPTERIS LONGIPES S. WA	CYMOPTERIS	3 INCOMPLETE PRESSED SPECIMENS WITHOUT ROOTS	RENDEZVOUS PEAK, 1/4 MILE SOUTH OF TRAM	28-Jul-78	/ /
5451	GRTE 4989	BIOLOGY	PLANTAE	MAGNOLIO	ASTERACE	CHRYSOTHAMNUS NAUSEOSUS (	CHRYSOTHAM	1 INCOMPLETE SPECIMENT WITH NUMEROUS FLOWERS BUT NO	1/2 MI. WEST OF EAST BOUNDARY ON ROAD WHICH PARALL	9-Jul-79	/ /
5452	GRTE 4990	BIOLOGY	PLANTAE	MAGNOLIO	POLEMONI	LINANTHUS HARKNESSII (CUR	LINANTHUS	NATIONAL CATALOG NOTE 9/2001: CONDITION: COM/EX,	SOUTH SIDE OF RIBBON FALLS ON TEEWINOT MOUNTAIN	26-Jun-79	/ /
5453	GRTE 4991	BIOLOGY	PLANTAE	MAGNOLIO	SCROPHUL	CORDYLANTHUS RAMOSUS NUTT	CORDYLANTH	3 COMPLETE PRESSED SPECIMENS.	1/2 MILE WEST OF EAST BOUNDARY ON ELK RANCH ROAD	9-Jul-79	/ /
5454	GRTE 4992	BIOLOGY	PLANTAE	MAGNOLIO	HYDROPHY	NEMOPHILA BREVIFLORA	NEMOPHILA	5 INCOMPLETE PRESSED SPECIMENS WITHOUT ROOTS -	FIRST SWITCHBACK GLACIER TRAIL 500 YDS., ABOVE BRA	7-Jul-79	/ /
5455	GRTE 4997	BIOLOGY	PLANTAE	MAGNOLIO	RANUNCUL	THALICTRUM FENDLERI ENGEL	THALICTRUM	2 COMPLETE PRESSED SPECIMENS,	COTTONWOOD CREEK BANK, HIGHLANDS RESIDENTIAL AREA	17-Jul-78	/ /

Figure 1. Original GTNP Excel Spreadsheet

the herbarium specimens for use with this project. Figure 1 shows a screenshot of the original Excel spreadsheet. Needless to say, this type of collection management system would not meet the needs of the botanists using the digital collection. The curator and manager of the RMH looked at some of the early images and labels on the specimens and informed librarians of data that was needed in order to make the collection useful to botanists. They also worked with librarians to determine if the images produced by the digital camera would enable botanists to work with the specimens digitally.

The RMH database was used to clean up the spreadsheet information. Fields from the spreadsheet were exported to the database. The spreadsheet records were incomplete and the entries did not follow a standard formatting procedure. The database entries had to be edited, either as the digitization process took place or at a later date. As images of the specimens were attached to the database records, it became apparent that a modification of the RMH database would improve the workflow for management of the data. The modified database included a field for the image number and special attention notes to help with the editing and clean-up of data. Many of the specimens were originally misidentified and field label annotations had to be cross-referenced through the use of the RMH authority list. Figure 2 contains a screenshot of the modified database used for the project.

To determine which fields to include in the digital herbarium, it was again necessary to have an understanding among the collaborators. What fields are going to be necessary for the botanists? Do we need to create fields for the public such as common names, medicinal uses, or historical significance? In creation of a digital database, it is better too err on the side of too many fields than not enough.

Many herbariums across the United States and the world use their own software to organize digital collections for in-house use and these disparate tools often do not communicate with each other. Scientists are creating federated databases for biological data from DNA sequences to biological taxonomies, sometimes referred to as bioinformatics. This has led to the creation of integrated regional databases that can be accessed from outside of the herbarium. This trend may make it necessary to create digital collections that have the same functionality.

Further discussions with park personnel and RMH botanists resulted in attention paid to the issue of sensitive species. Some of the plant specimens are rare in GTNP and the NPS wants to keep locations of these

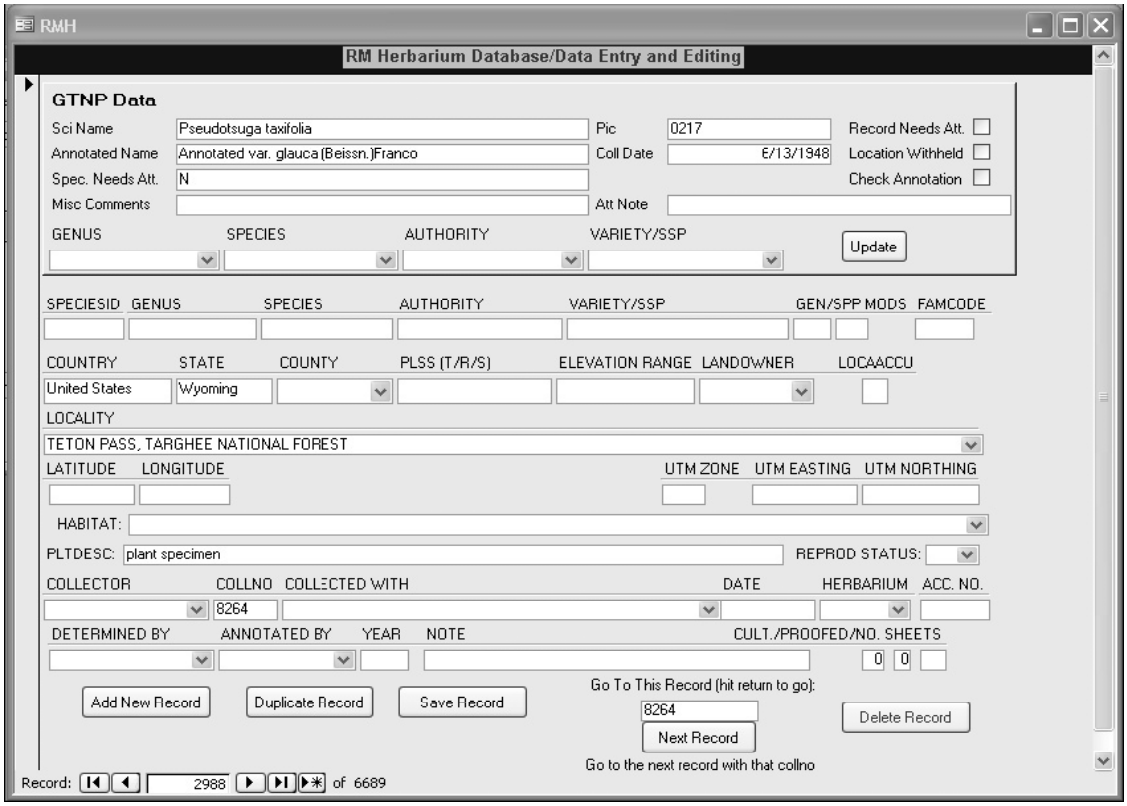

Figure 2. RMH Access Database Table

plants hidden. How do you provide botanists with the full descriptive information of specimens while hiding their locations from the public? The solution was to provide a copy of the database to the NPS and the RMH so that researchers will have access to the location information. However, the public interface, available on the web, will have sensitive information hidden from public view and researchers and park personnel will still have access when they need it.

For researchers and park personnel, the quality of the location information that will be available will increase. As records for new specimens are added, and old records modified, with Geographic Information Systems (GIS) enhanced location information, addition location data will become available. As before, the locations will be withheld on the public interface for rare or endangered species. The use of digital herbarium collections will only expand. In fact, GIS is being used today as an important method to track and prioritize the procedures for the preservation of rare specimens in natural habitats.[19]

Conclusion

In the future, there will be a call for digital projects to fulfill internal needs of the libraries and external needs of the university. Librarians need to be ready for these opportunities, whether they are in-house, at the university level, or multi-state collaborations; in fact librarians need to take the lead on these projects. Some things to think about for future projects:

• Pick your projects carefully and beware of the Google™ effect: Google™ Maps, Google™ Books, Google™ Earth, and Google™ Patents. If Google continues to digitize material of all sorts it will pay to keep track of their progress so duplication of effort does not occur.

• Know the standards for metadata and herbarium information.

• With the concern for global inventories of both plant and animal life, collections need to be federated so users can look at collections independent of location.

• With the advent of GIS, the need for records, including location information that can be pulled and imported into ArchView or other GIS software, is apparent.

Some issues to consider when building a cooperative project are:

1. When beginning a cooperative project take into account cultural issues of different institutions. Who has ownership of the digital collection, and how will

copyright issues be handled? Remember that each institution will put forward their needs and expect them to be met, even though it may not be possible to do so. In order for the project to move forward, everyone has to be convinced of the utility of the project.

2. Software can become a limiting factor. Database compatibility and user needs often force compromise that may not be ideal for all users. If database issues are not discussed early enough workflow may be impacted and the project will be delayed. There will always be tradeoffs between needs and wants.

3. Time and money constraints will also impact the project.

4. With limited funding and time you will not be able to do everything you set out to do. It pays to break up the project into parts that have clear objectives and outcomes. When more funding becomes available, additional pieces can be added onto the project.

Librarians need to be aware not only of changes digital publishing has had on libraries, but how researchers are utilizing digital content. New hardware and software continue to change how researchers do their work, and there are opportunities for librarians to become partners in this digital evolution of the WWW. Librarians should not be afraid to work on cross-disciplinary projects that may initially seem to have little to do with libraries. Who knows what the future holds for librarians, but, rest assured, a digital world will increase opportunities for collaborative projects.

Notes

1. Miriam A. Drake, "Institutional Repositories: Hidden Treasures," *Searcher* 12 (5) (2004): 41–45.

2. Clifford A. Lynch, "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age," *ARL* 226 (2003): 1–7.

3. C. Petersen, "Linus Pauling and the Nature of the Chemical Bond: A Documentary History," *D-Lib Magazine* 11 (1) (2005): 1.

4. H. Roes, "Digital Libraries in Education: Trends and Opportunities," *D-Lib Magazine* 7 (7/8) (2001).

5. Oregon State University, "Linus Pauling Research Notebooks," http://osulibrary.oregonstate.edu/specialcollections/rnb/index.html. (accessed November 22, 2006).

6. Matthew S. Dunn and P. Christopher, "vPlants: A Virtual Herbarium of the Chicago Region," *First Monday* 5 (2002): 3.

7. MBLWHOI Library, "MBLWHOI Library Digital Herbarium," http://www.mblwhoilibrary.org/archives/herb/dindex.php (accessed November 11, 2006).

8. UVSC Herbarium, "The UVSC Virtual Herbarium," http://herbarium.uvsc.edu/Virtual/ (accessed on January 3, 2007).

9. The vPlants Project, "Vplants: A Virtual Herbarium of the Chicago Region," http://www.vplants.org (accessed on May 28, 2005).

10. The New York Botanical Garden, "The Virtual Herbarium," http://sciweb.nybg.org/science2/VirtualHerbarium.asp. (accessed on October 21, 2006).

11. Oregon State University, "OSU Herbarium Type Specimens," http://digitalcollections.library.oregonstate.edu/cdm4/client/herbarium/index.php?CISOROOT=/herbarium. (accessed on November 12, 2006).

12. Burke Museum, "WTU Image Collection: Plants of Washington," http://biology.burke.washington.edu/herbarium/imagecollection.php. (accessed on December 6, 2006).

13. H. B. Shaffer, R. N. Fisher, and C. Davidson, "The Role of Natural History Collections in Documenting Species Declines," *Trends in Ecology & Evolution* 13 (1) (1998): 27–30.

14. A. F. O'Connell, A. T. Gilbert, and J. S. Hatfield, "Contribution of Natural History Collection Data to Biodiversity Assessment in National Parks," *Conservation Biology* 18 (5) (2004): 1254–61.

15. G. F. Smith, Y. Steenkamp, R. R. Klopper, S. J. Siebert, and T. H. Arnold, "The Price of Collecting Life—Overcoming the Challenges Involved in Computerizing Herbarium Specimens," *Nature* 422(6930) (2003): 375–76.

16. M. Begnoche, "Specimens at Herbarium Get Digital Makeover," http://www.umich.edu/~urecord/0102/Nov11_02/21.shtml. (accessed December 16, 2006).

17. T. J. Ong, J. J. Leggett, H. D. Wilson, S. L. Hatch, and M. D. Reed, "Interactive Information Visualization in the Digital Flora of Texas," *Visual Interfaces to Digital Libraries* 188–98, 2002.

18. Rocky Mountain Herbarium, "A Concise Guide to the Collection, Preparation, and Preservation of Herbarium Specimens," http://www.rmh.uwyo.edu/prelude/intro/rmcoll.htm. (accessed January 5, 2007).

19. A. Jarvis, K. Williams, D. Williams, L. Guarino, P. J. Caballero, and G. Mottram, "Use of GIS for Optimizing a Collecting Mission for a Rare Wild Pepper (Capsicum Flexuosum Sendtn.) in Paraguay," *Genetic Resources and Crop Evolution* 52 (6) (2005): 671–82.