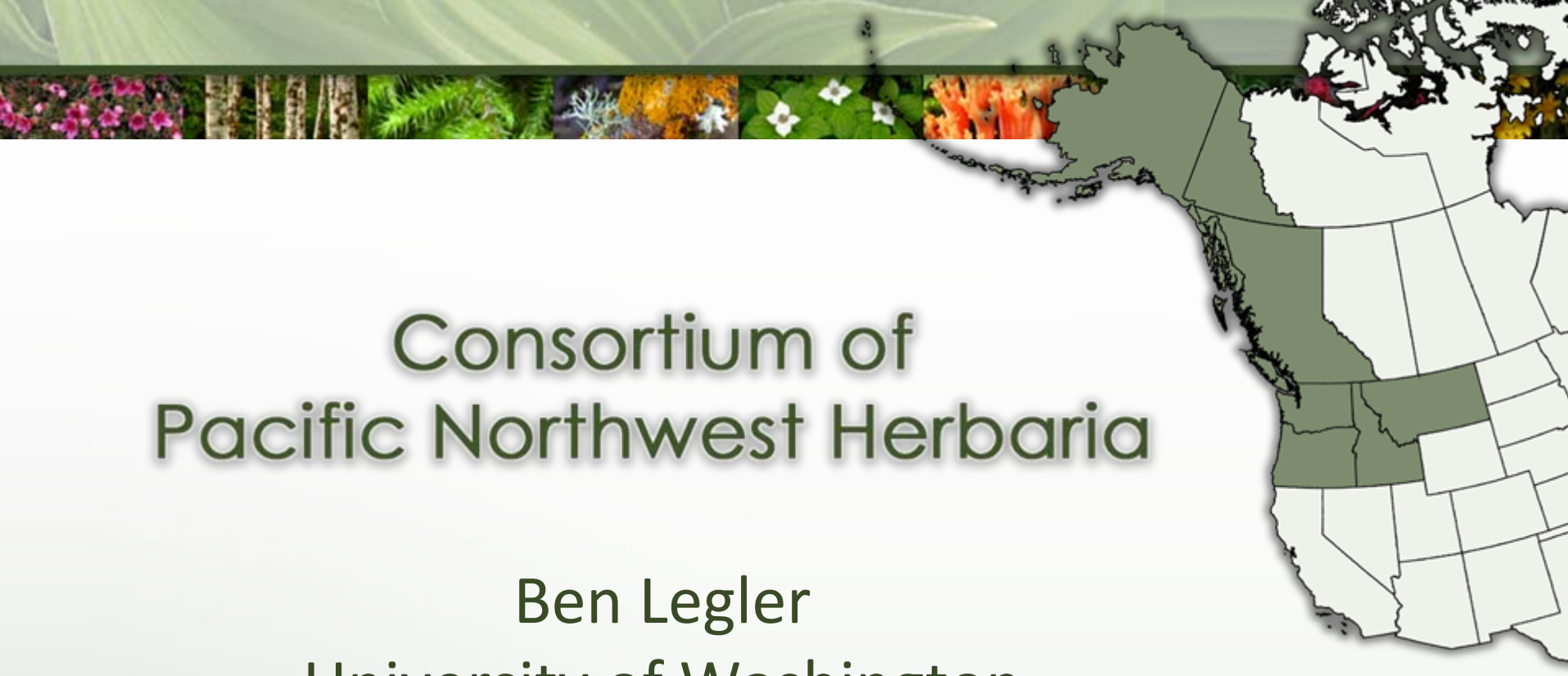
Consortium of Pacific Northwest Herbaria

Ben Legler
University of Washington

blegler@u.washington.edu
www.pnwherbaria.org



Alaska
British Columbia
Idaho
Montana
Oregon
Washington
Yukon Territory

About the Consortium (CPNWH)

How CPNWH began

Funding strategies & sources

What we're doing right now

Technology, software, & solutions

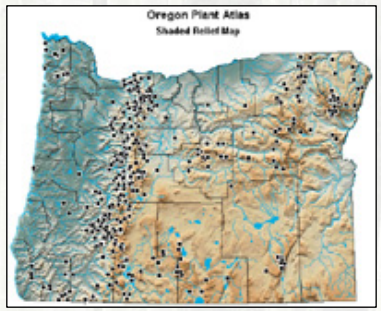
Challenges, problems, & recommendations

Rock Lake, Eastern Washington

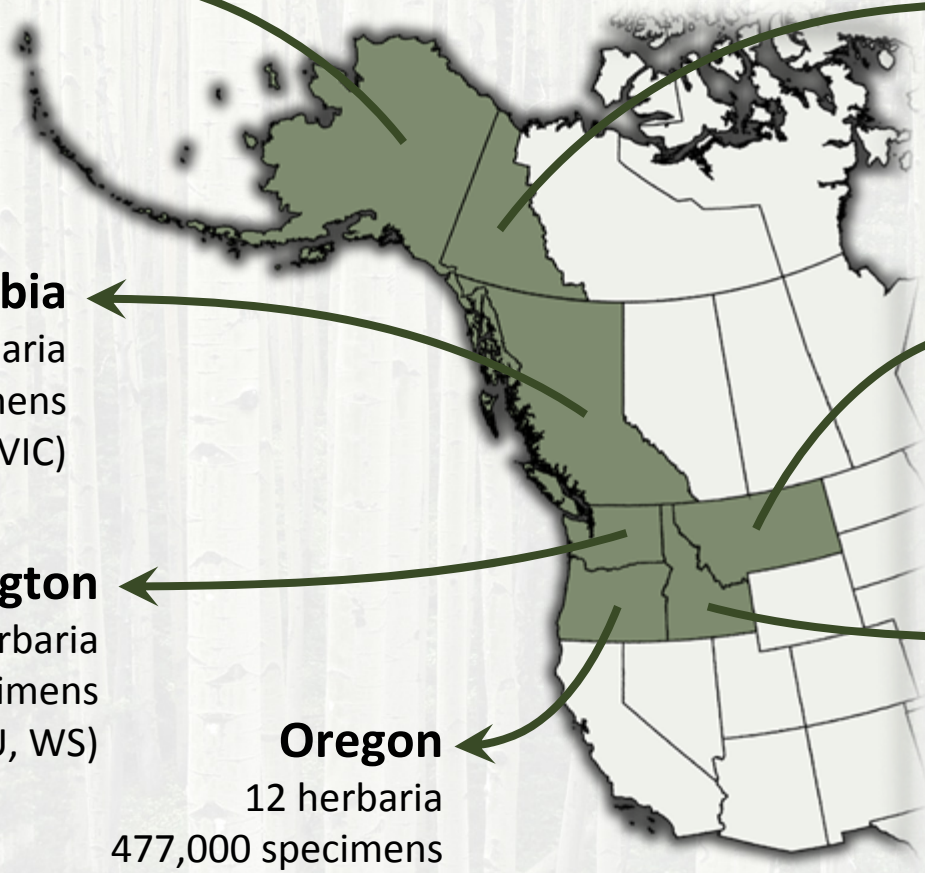
Consortium of Pacific Northwest Herbaria (www.pnwherbaria.org)

Consortium of PNW Herbaria (CPNWH):

- "Created in 2007 to bring together regional herbaria and provide an online portal to the wealth of existing and emerging information about the flora of Pacific Northwest North America."



56 public and private herbaria. 3,400,000 specimens.



Alaska
5 herbaria
250,000 specimens
(ALA)

Yukon Territory
1 herbarium
7,000 specimens
(private collection)

British Columbia
9 herbaria
900,000 specimens
(UBC, V, UVIC)

Montana
6 herbaria
279,000 specimens
(MONTU, MONT)

Washington
16 herbaria
1,150,000 specimens
(WTU, WS)

Idaho
8 herbaria
343,000 specimens
(ID, IDS, SRP, CIC)

Oregon
12 herbaria
477,000 specimens
(OSC)

Portal initiation:

- Discussions among regional curators was initiated in late 2006 by Dick Olmstead following early completion of WTU's previous NSF grant.
- Favorable support of regional curators led to submission of supplemental request to NSF. $32,000 was awarded in early 2007.

Stated goals in NSF supplemental request:

1. Link regional herbarium specimen records through an online portal.
2. Provide a unified access point for online resources associated with participating collections.
3. Develop data-sharing protocols to minimize redundant data entry.
4. Facilitate digitization efforts at regional herbaria.

Douglasia nivalis

Results of initial work:

1. Web site providing access to ~ 400,000 specimens from 3 regional herbaria (ALA, OSC, WTU), with infrastructure in place to add more.
2. Linked list of online resources hosted by regional herbaria.
3. Contact info and statistics for regional herbaria.

A few observations:

- Initial discussions were done by email among curators and staff.
- We did not hold any meetings or lengthy debates about the concept or goals.
- Instead, we just took an early idea and quickly turned it into a working result.
- One institution (WTU) started the idea and took on the initial work.

Pasayten Wilderness, Washington

Current NSF grant:

- In 2009, WTU re-initiated discussions among regional herbaria.
- We submitted a collaborative grant proposal to NSF in summer 2009.
- **$1.3 million** was asked for and awarded, with funds split between four institutions.
- We started work in June, 2010.

Institution	Funds
University of WA (WTU)	$642,611
Oregon State U. (OSC)	$325,141
U. of Idaho (ID)	$301,351
Montana State U. (MONT)	$71,776
TOTAL	**$1,340,879**

Grant goals:

1. Image and database 180,000 vascular plant specimens from the region's last large herbaria that lack databases (ID, MONT).
2. Image and database ca. 140,000 vascular plant specimens from about 12 small herbaria in Idaho, Oregon, and Washington.
3. Database 200,000 PNW non-vascular plant, fungal, and lichen specimens at OSC and WTU.
4. Set up new data access points for herbaria with existing databases (WS, UBC, MONTU, SOU, SRP, CIC) and hopefully replace old DiGIR connections (WTU, OSC, ALA) with newer technology.
5. Provide online access to all specimen images and data available for the region (ca. 2,000,000 specimens and 300,000 images).
6. Make all of this data available to GBIF, USVH, and other data aggregators.
7. Develop software solutions and expand the functionality of the Consortium web site.

Image and database smaller herbaria in ID, OR, and WA

Washington:	
Western WA U.	26,000
Whitman College	17,000
Central WA U.	25,000
Eastern WA U.	7,000
Pacific Lutheran U.	5,000+
TOTAL:	**78,000**

Idaho:	
Lewis & Clark College	10,000
Northern Idaho College	10,000
Forest Service Herbaria	5,000
TOTAL:	**25,000**

Oregon:	
Reed College	10,000
Portland State U.	11,000
Linfield College	2,000
Southern Oregon U.	14,000
HJA Experimental Forest	1,400
TOTAL:	**38,400**

GRAND TOTAL: 142,000+

Image and database smaller herbaria in ID, OR, and WA

- Small herbaria house valuable specimens not present in the larger herbaria. But they are often not examined by specialists.
- It is often challenging to justify the existence of these collections to university administrators. Can we increase their utility and prominence?
- Limitations in staff, budgets, and computer infrastructure make it difficult for these herbaria to manage a specimen database or provide online access.

Aquilegia jonesii

Image and database smaller herbaria in ID, OR, and WA

- Small herbaria house valuable specimens not present in the larger herbaria. But they are often not examined by specialists.
- It is often challenging to justify the existence of these collections to university administrators. Can we increase their utility and prominence?
- Limitations in staff, budgets, and computer infrastructure make it difficult for these herbaria to manage a specimen database or provide online access.

We've proposed a model that leverages the resources at larger herbaria to assist smaller herbaria with imaging and databasing.

Our solution for digitizing small herbaria

- Centralize the image storage, databases, and software/hardware infrastructure at the larger institutions:

1. We send imaging equipment to each herbarium and use hourly, work study, or volunteer help to image the entire collection.
2. Images are sent to WTU and stored on the Consortium's server.
3. We create a database for each herbarium. The database lives on the Consortium web server and is accessible through the internet via a web application interface.
4. Label data is captured from the images by staff and personnel at the larger institutions (using the web app).
5. The same web app is accessible to staff at each herbarium for use as their own database.

What we've accomplished thus far on the grant

- Held a meeting among regional herbaria in June, 2010.
- Configured and deployed imaging equipment to all four states.
- Have acquired ~ 110,000 images to date (of 300,000 projected).
- Set up a web server dedicated to the Consortium, and configured image processing and storage work flows.
- Created the online web application database interface.
- WTU has finished databasing nearly all of their PNW bryophyte, lichen, and fungi specimens (ca. 80,000). OSC is working on their collections.
- Deployed a new FileMaker database for the U. of Idaho Herbarium.
- I am currently developing data provider connections and rebuilding and expanding the Consortium database and web site.

Why we think we got funded

- Leveraged the success of prior grants and similar projects.
- Used small pots of money to initiate projects and develop proof-of-concepts. These may have helped convince reviewers we knew what we were doing.
- Innovation. We proposed a novel approach to assist smaller herbaria with imaging and databasing.
- Timing of grant coincided with initiation of NSF's new ADBC funding track, and with availability of federal stimulus funds to NSF.

Technology we've developed:

Image processing equipment & workflows

Specimen image viewer

Data entry interface

Online search interface

Technology, Software & Solutions

1. Image Capture:

Transfer RAW images and metadata on a portable drive

2. Image Processing:

Dropbox (images & metadata)

Image Processing scripts

RAW:

JPEG:

Tiles:

Participant databases

PNW Portal database

Web site scripts

3. Data Entry:

AJAX

Data entry form shows blank record. Adjacent image viewer shows label data.

4. Online Access:

Image viewer allows pan, zoom, and measure, with link to download JPEG.

Search results shows map, text data, and thumbnail images.

5. USVH & GBIF

Consortium of Pacific Northwest Herbaria (www.pnwherbaria.org)

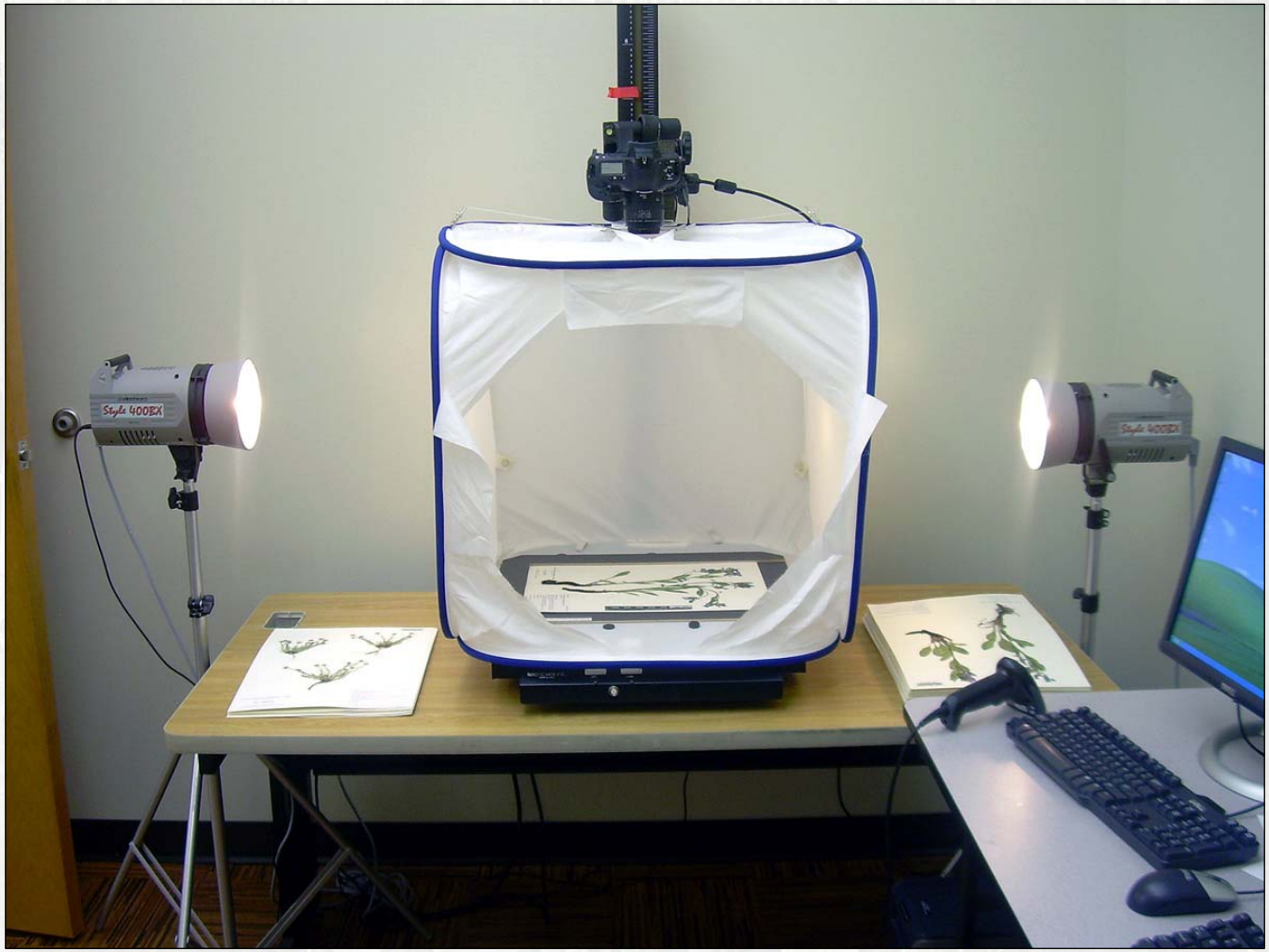
Imaging Equipment:

Lightbox
Canon 21MP dSLR
Custom camera mount
Laptop computer
Canon software
Custom metadata form

Justification:

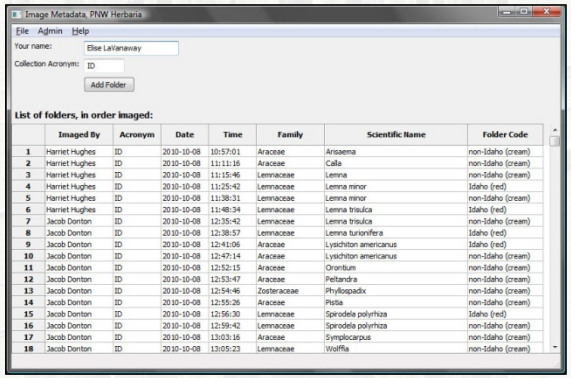
Easy to set up & use
Compact & portable
Reasonable price
Camera is much faster than scanner
Good image quality

A good alternative (that we don't use):

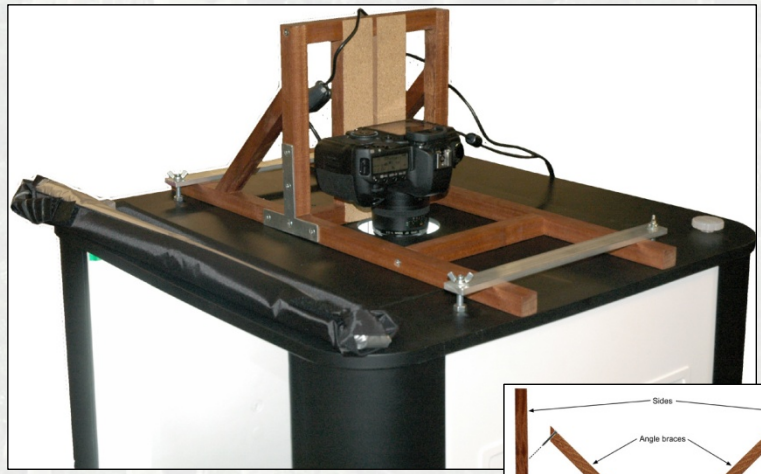


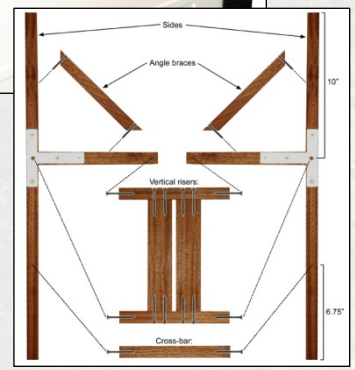
Imaging – what's available:

- Imaging documentation on the consortium web site: http://www.pnwherbaria.org/documentation/specimenimaging.php
- Descriptions of hardware and software we use.
- Python scripts for metadata capture, image processing, and tiling.



	Imaged By	Acronym	Date	Time	Family	Scientific Name	Folder Code
1	Harriet Hughes	ID	2010-10-08	10:57:01	Araceae	Arisaema	non-Idaho (cream)
2	Harriet Hughes	ID	2010-10-08	11:11:16	Araceae	Calla	non-Idaho (cream)
3	Harriet Hughes	ID	2010-10-08	11:15:46	Lemnaceae	Lemna	non-Idaho (cream)
4	Harriet Hughes	ID	2010-10-08	11:25:42	Lemnaceae	Lemna minor	Idaho (red)
5	Harriet Hughes	ID	2010-10-08	11:38:31	Lemnaceae	Lemna minor	non-Idaho (cream)
6	Harriet Hughes	ID	2010-10-08	11:48:34	Lemnaceae	Lemna trisulca	Idaho (red)
7	Jacob Donton	ID	2010-10-08	12:35:42	Lemnaceae	Lemna trisulca	non-Idaho (cream)
8	Jacob Donton	ID	2010-10-08	12:38:57	Lemnaceae	Lemna turionifera	Idaho (red)
9	Jacob Donton	ID	2010-10-08	12:41:06	Araceae	Lysichiton americanus	Idaho (red)
10	Jacob Donton	ID	2010-10-08	12:47:14	Araceae	Lysichiton americanus	non-Idaho (cream)
11	Jacob Donton	ID	2010-10-08	12:52:15	Araceae	Orontium	non-Idaho (cream)
12	Jacob Donton	ID	2010-10-08	12:53:47	Araceae	Peltandra	non-Idaho (cream)
13	Jacob Donton	ID	2010-10-08	12:54:46	Zosteraceae	Phyllospadix	non-Idaho (cream)
14	Jacob Donton	ID	2010-10-08	12:55:26	Araceae	Pistia	non-Idaho (cream)
15	Jacob Donton	ID	2010-10-08	12:56:30	Lemnaceae	Spirodela polyrhiza	Idaho (red)
16	Jacob Donton	ID	2010-10-08	12:59:42	Lemnaceae	Spirodela polyrhiza	non-Idaho (cream)
17	Jacob Donton	ID	2010-10-08	13:03:16	Araceae	Symplocarpus	non-Idaho (cream)
18	Jacob Donton	ID	2010-10-08	13:05:23	Lemnaceae	Wolffia	non-Idaho (cream)





Specimen Image Viewer:

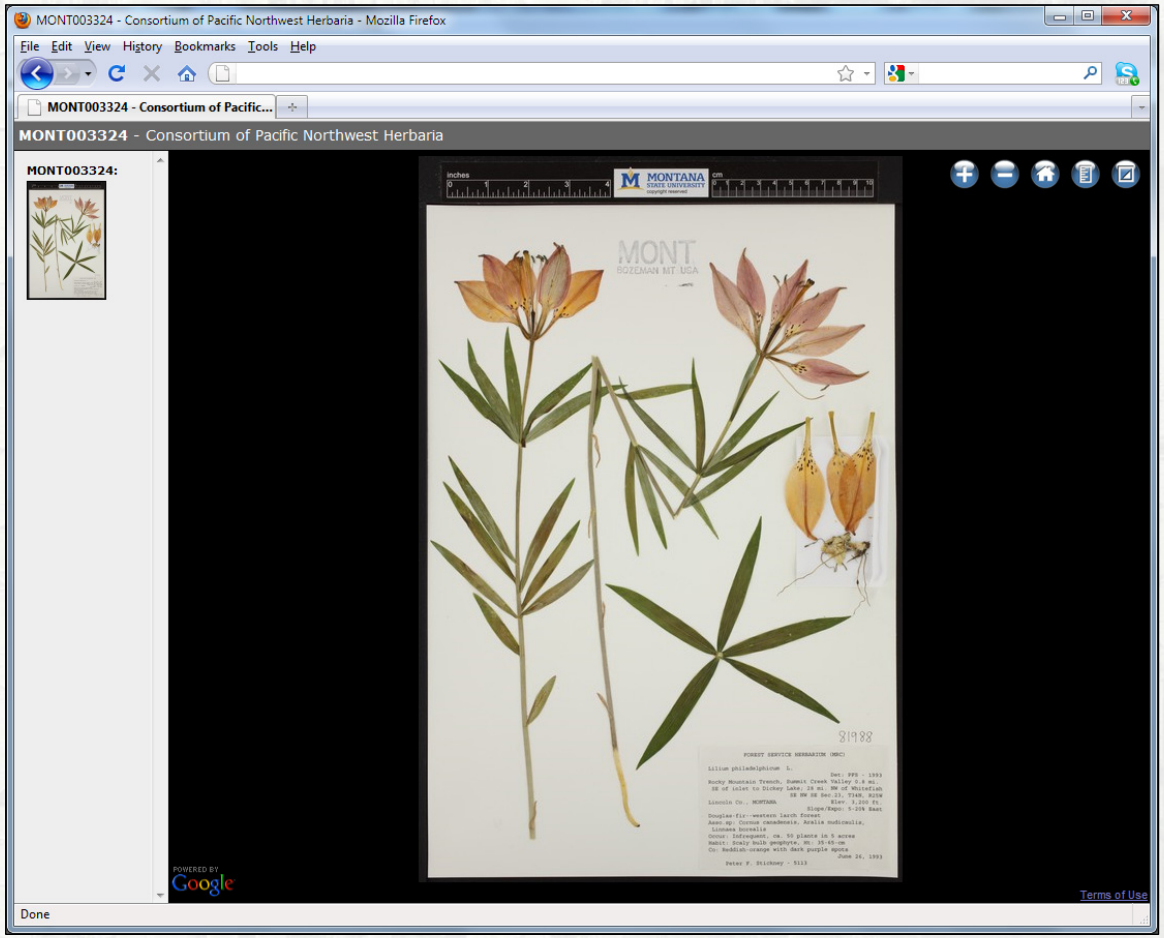
Open source

Based on Google Maps

Pure JavaScript

Fast!

Efficient storage of Images on server



http://www. rmh.uwyo.edu/gmapviewer/about.php

Data Entry Interface:

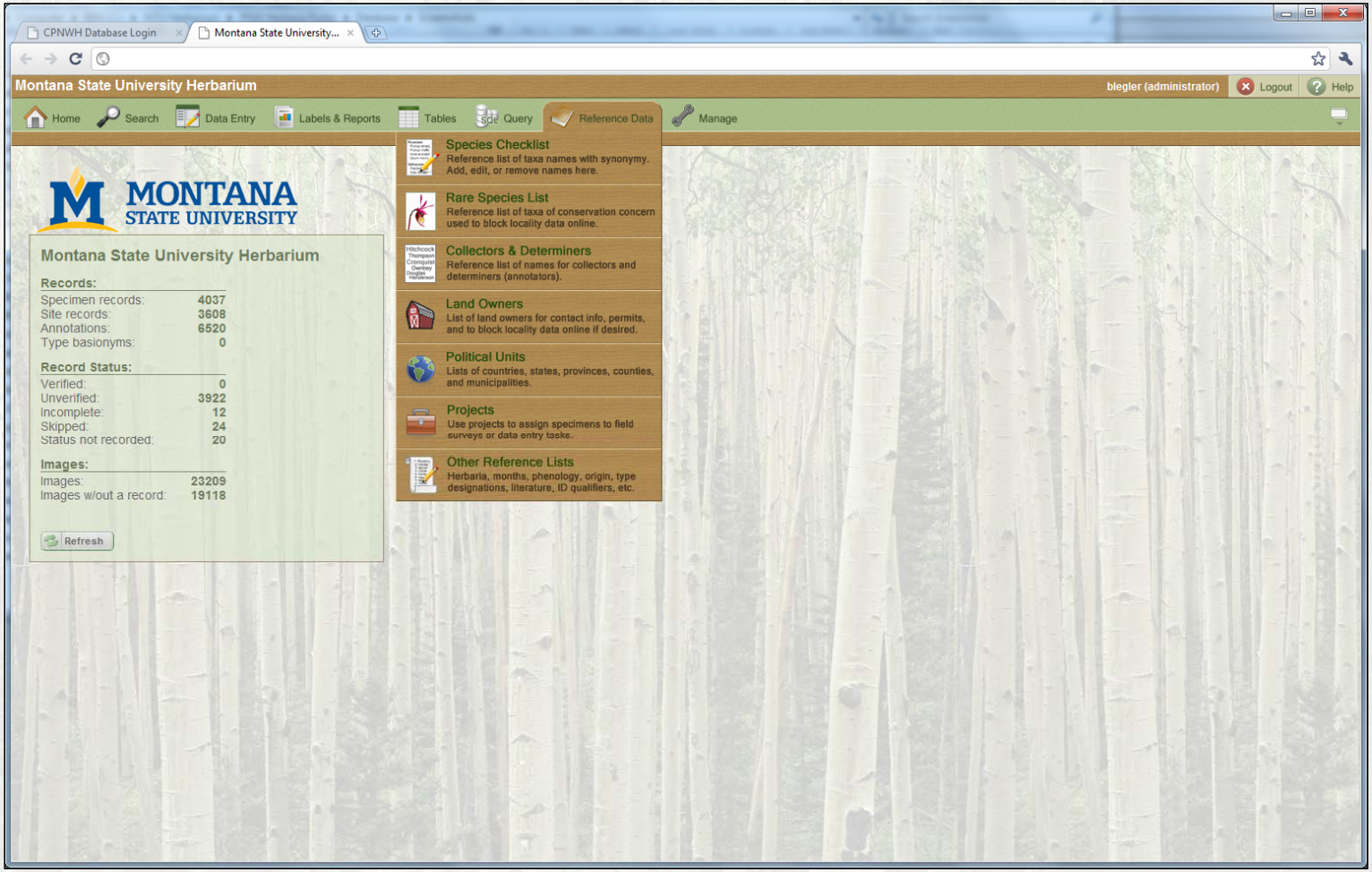
We needed a solution that satisfied the following requirements:

- Can be managed centrally by the Consortium.
- Accessible to small herbaria with minimal requirements.
- Tight integration of specimen images.
- Flexible enough to allow development of novel approaches (e.g., OCR).
- Preferably, based on open source software.

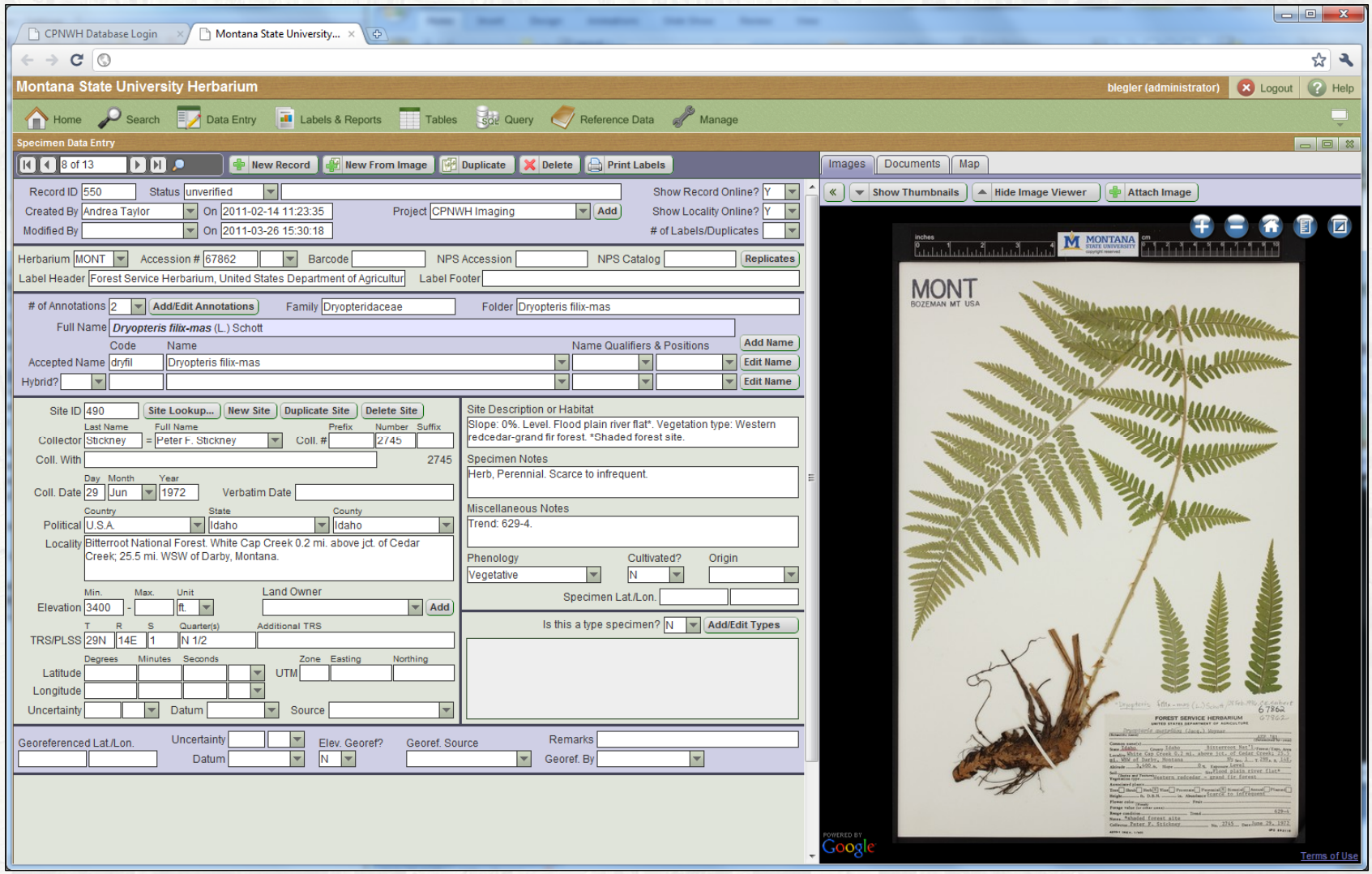
Harrimanella stelleriana

Our solution:

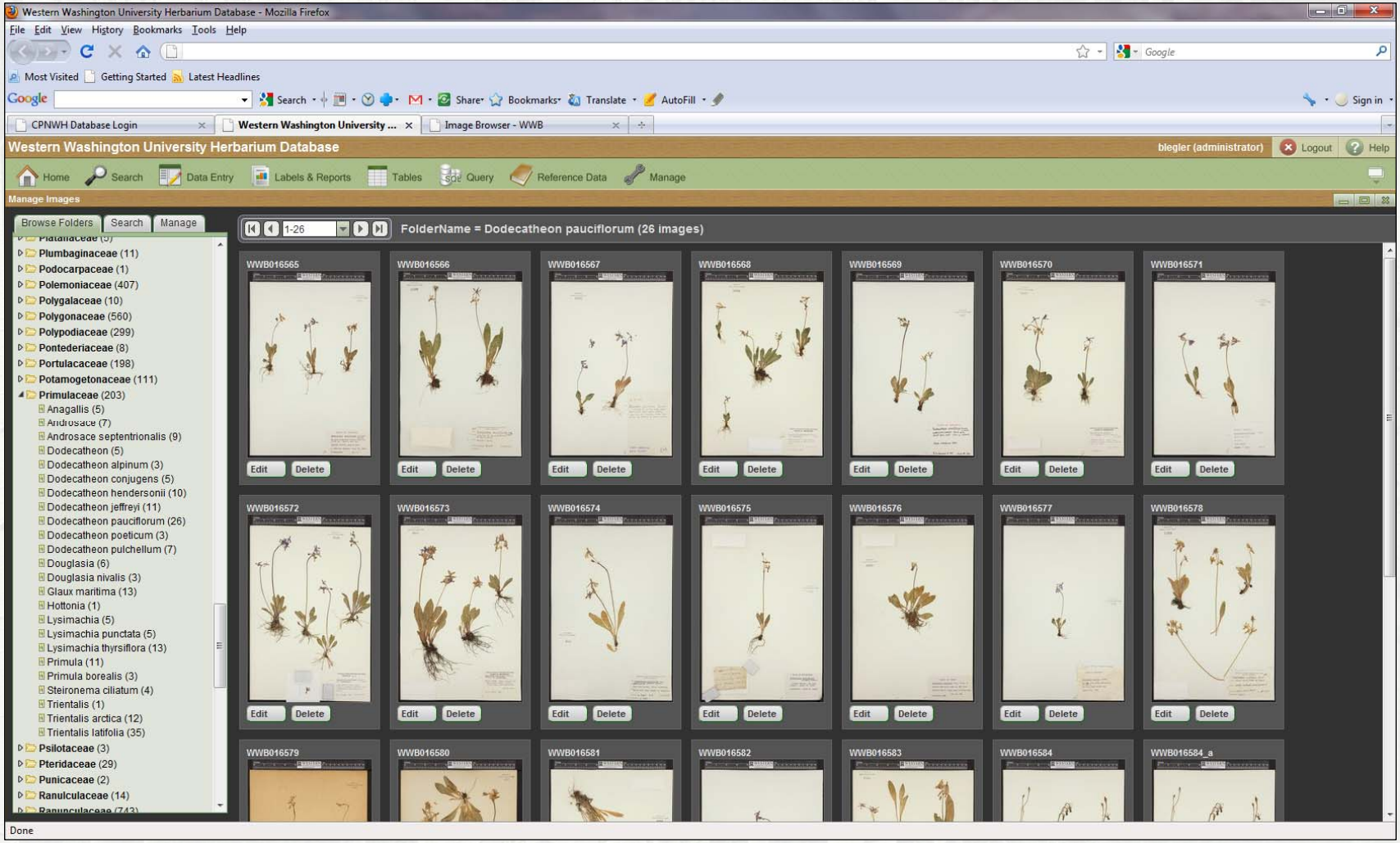
- Develop a new database system that runs as a web application based on Apache, PHP, MySQL, Python, JavaScript, and HTML.



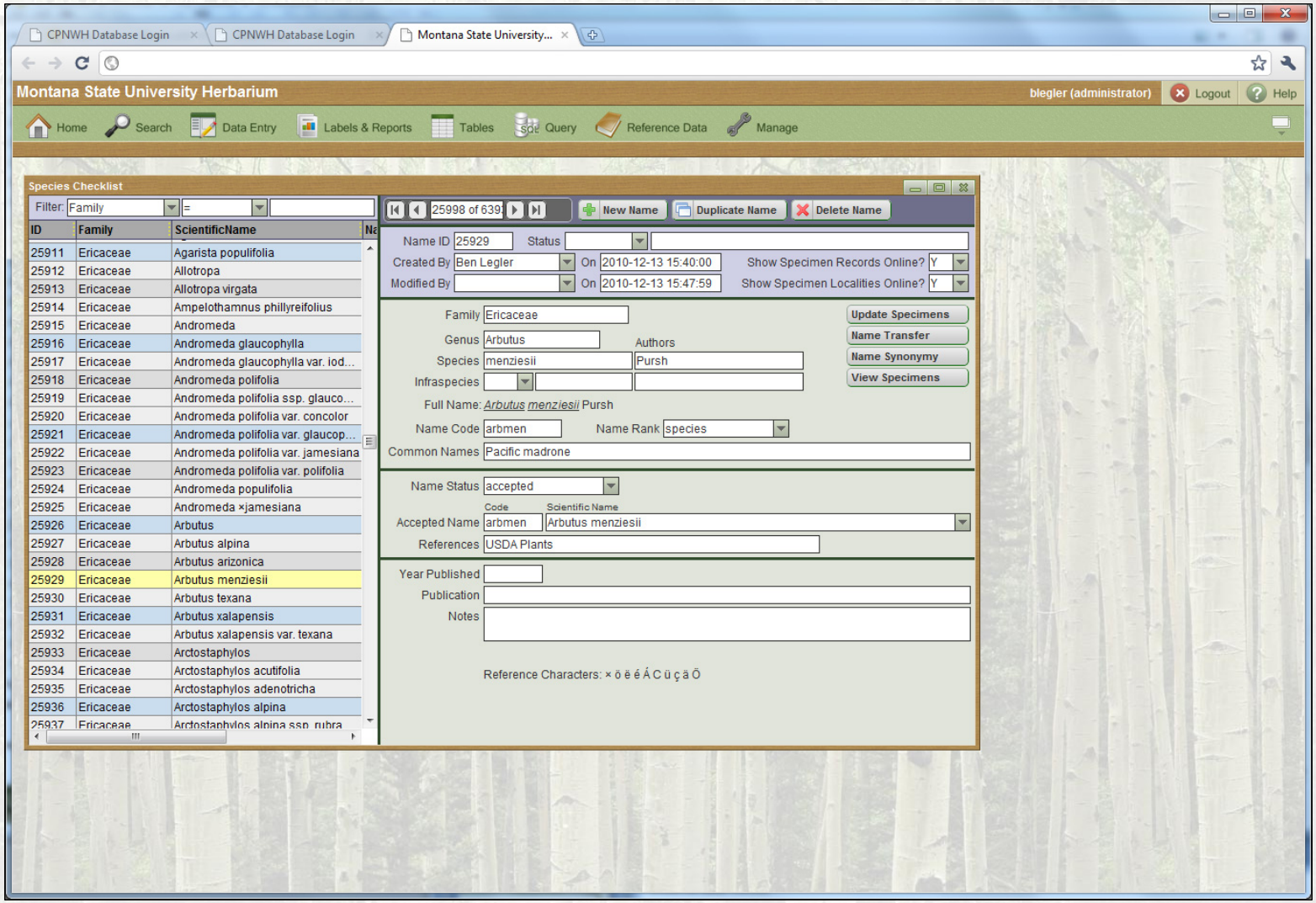
Data Entry Interface:



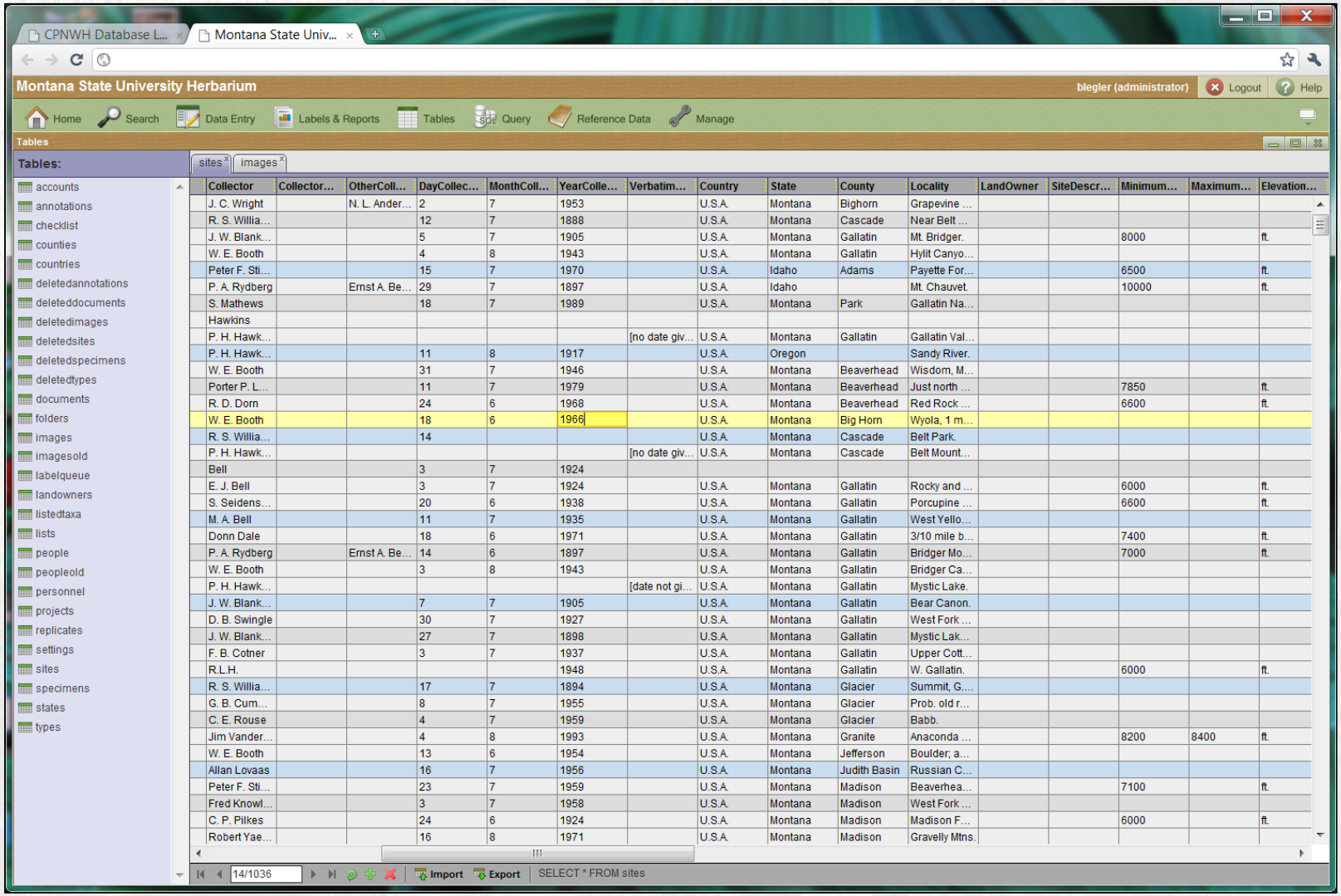
Data Entry Interface:



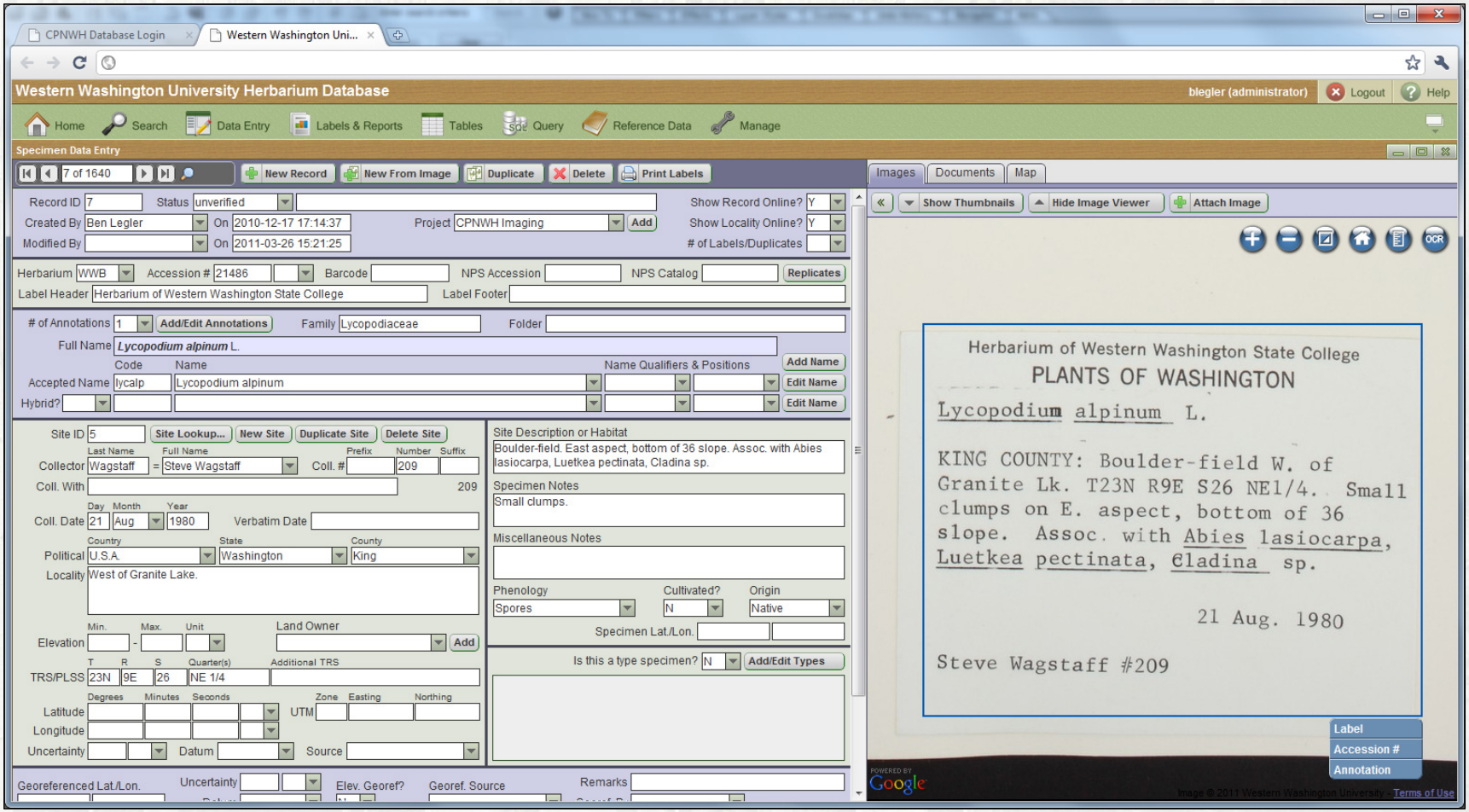
Data Entry Interface:



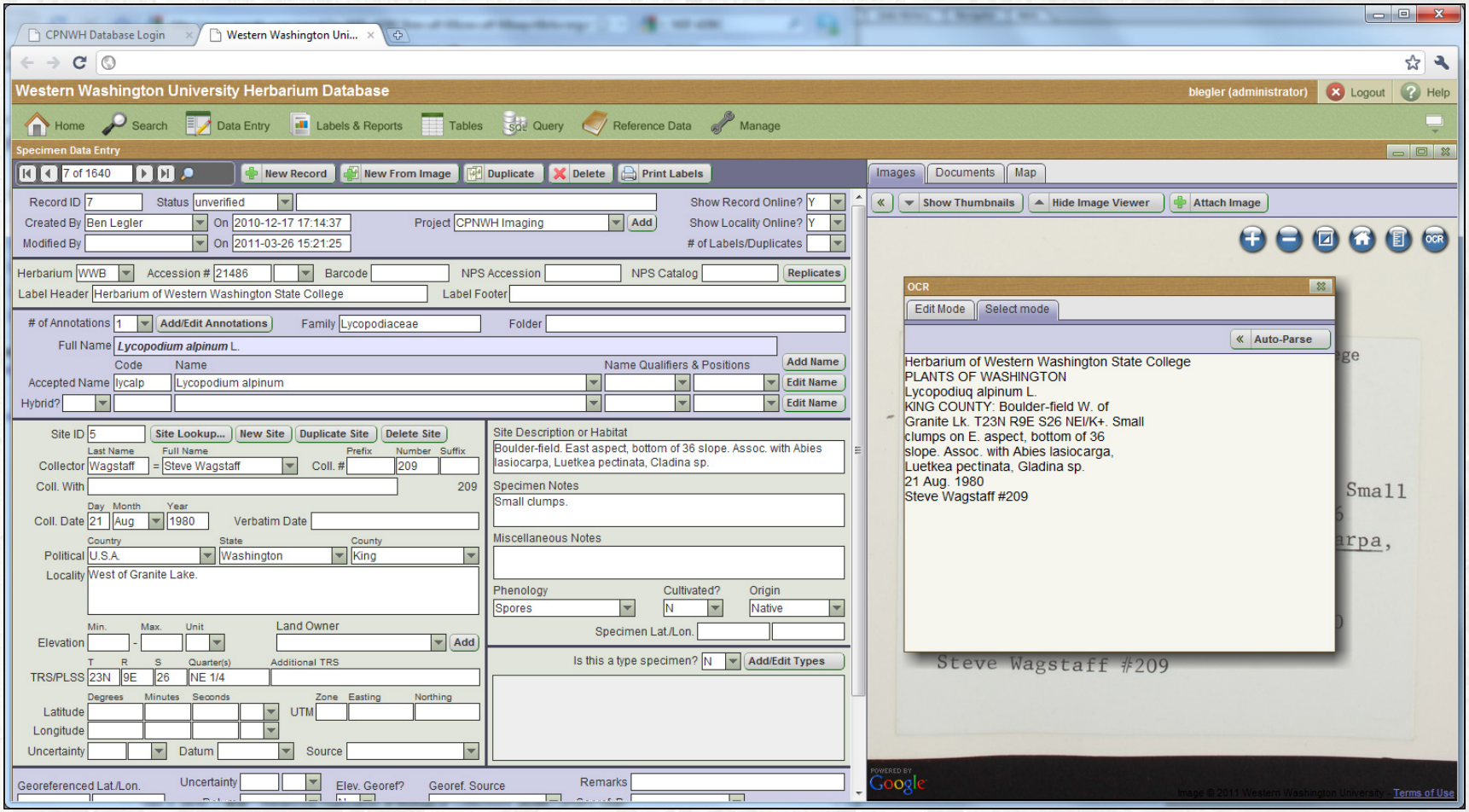
Data Entry Interface:



OCR-Assist (experimental):

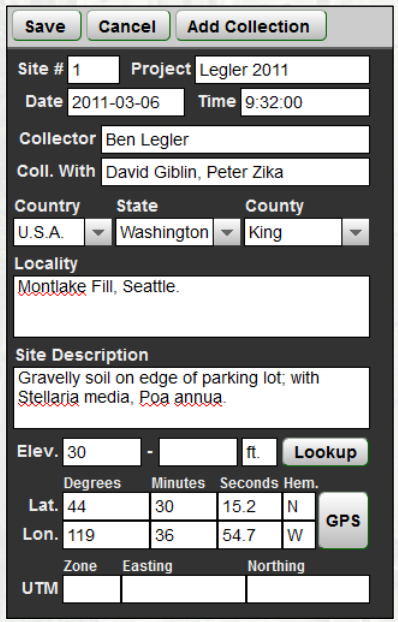


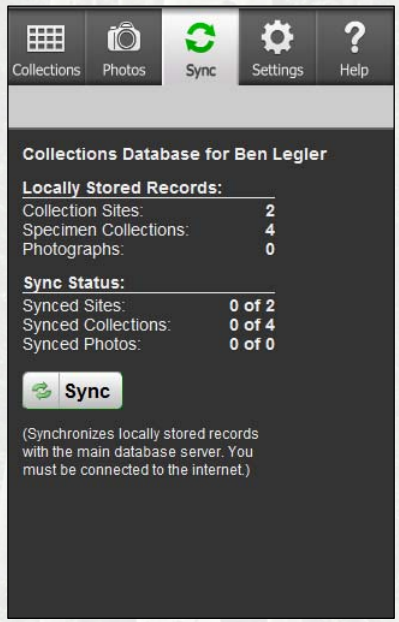
OCR-Assist (experimental):

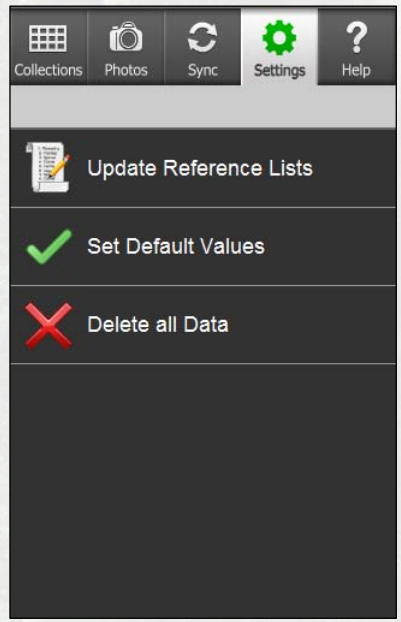


Data collection in the field (experimental):

- Intended for collecting field notes for herbarium specimen labels.
- Web app using HTML5 Local Storage and phone's GPS. Works even if wireless networks aren't available (e.g., in the mountains).
- Collect your data, then, when a network is available, send it to the database server. When you return from the field, the data is waiting and ready for you to clean it, enter annotations, and print labels.







Data Entry Interface:

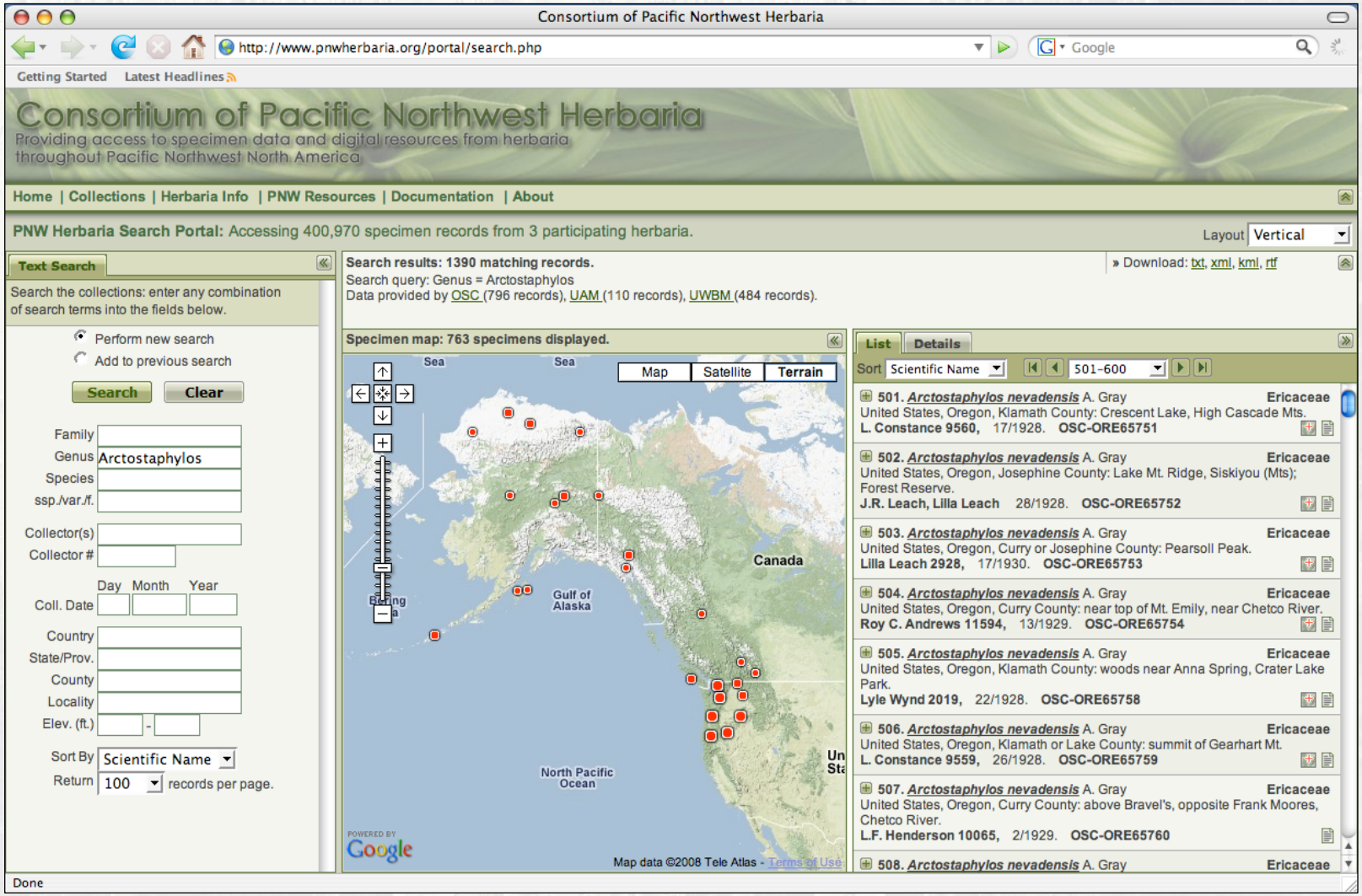
How well is this interface working?

- Core functionality used for the grant is operational.
- Additional features (label printing, etc.) are in development.
- Integrated OCR (optical character recognition) is in development.
- Core code base is still being refined & debugged.
  (code: 800 KB uncompressed JavaScript, 100 KB PHP, and some CSS and graphics)

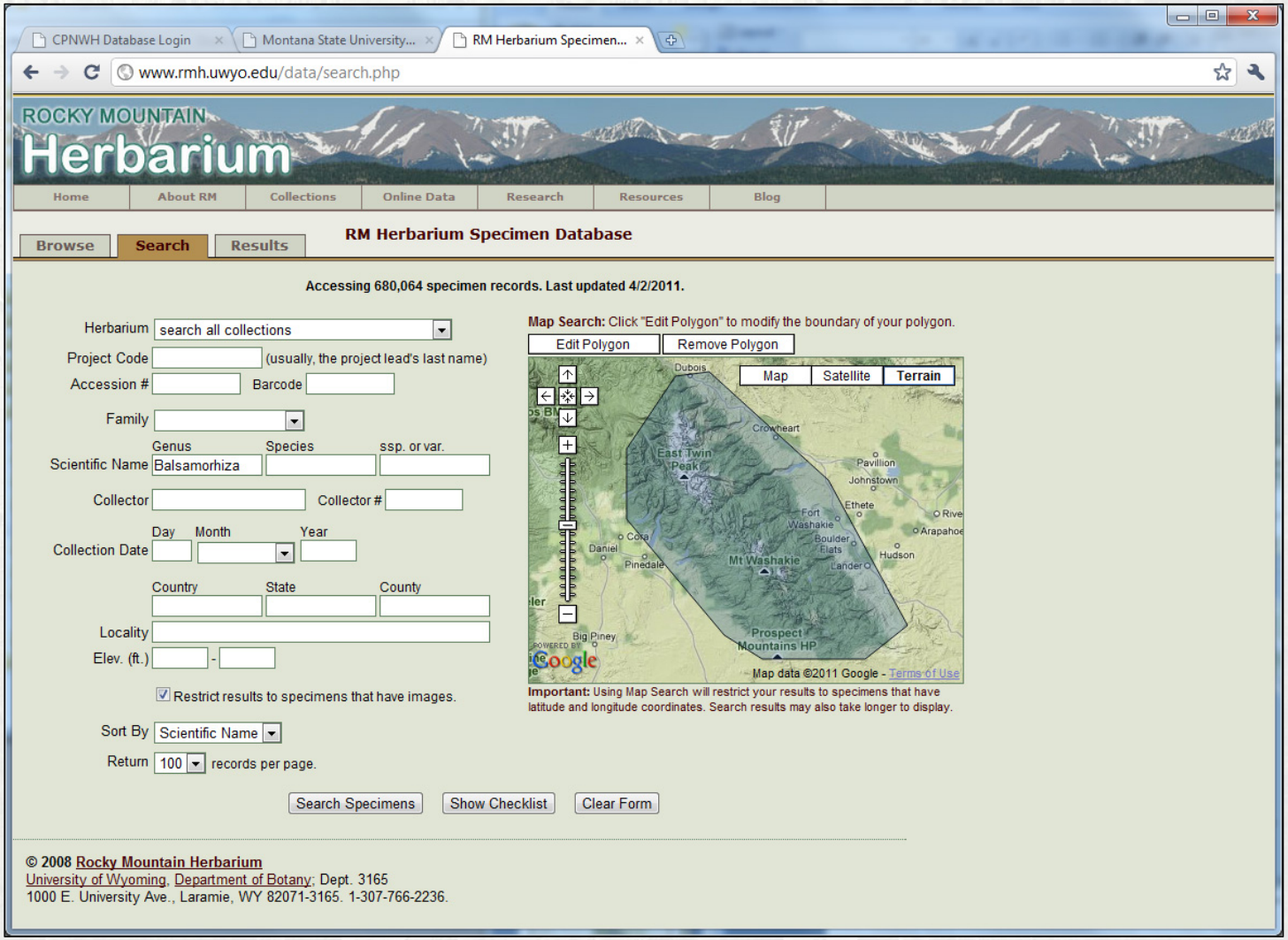
We are currently using this interface for five herbaria. Several other regional herbaria are considering using it independently from the Consortium.

At some point I'll release it as open-source code.

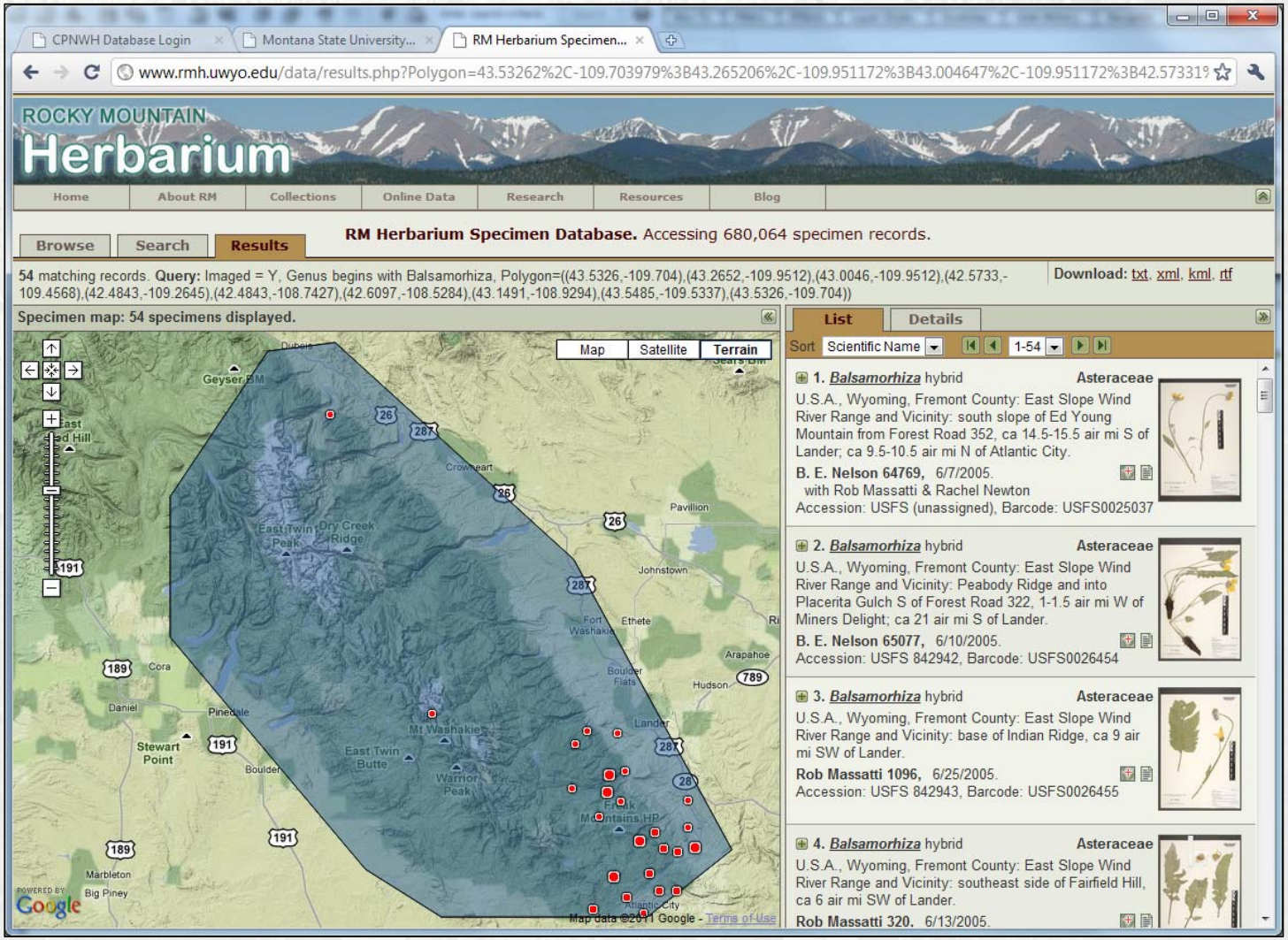
Online Search Interface:



Online Search Interface:



Online Search Interface:



Some lessons learned:

- Everything seems to take more time than anticipated.
- Managing several dozen hourly personnel (mostly students) leads to challenges with data and image quality and consistency.
- Mission creep – try to stay focused on the goals and minimize distractions or desire to add in more.
- It's a collaborative grant. Each partner has autonomy and can diverge from the grant goals to use the money on unrelated projects.
- Developing a new database system isn't a trivial task! Maybe we should've tweaked an existing system to "make it do."
- Image processing work flows are not fully automated like I had originally planned. Instead, I prioritized image quality and consistency.

Questions?

Ben Legler
(blegler@u.washington.edu)

Mt. Baker, Washington

Consortium of Pacific Northwest Herbaria (www.pnwherbaria.org)