Digitization Workflow
Consortium of Pacific Northwest Herbaria

- Collaborative NSF BRC grant (similar to a TCN) – 2010 to 2014.
- Digitization workflows have been used at 26 herbaria in PNW.
- Captured 518,000 images.

Ben Legler, University of Washington, blegler@u.washington.edu

Image Capture:

Pre-digitization:
Mount/repair?
Pull folders from cabinets.

Imaging workstation:
Lightbox,
dSLR,
Laptop,
Software

Metadata (per folder):
Storage location,
Personnel name,
Timestamp, etc.

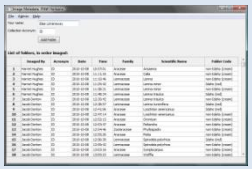
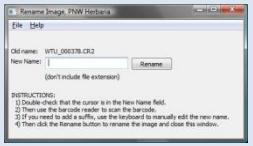


Image Capture (per specimen):
Affix barcode?
Capture image.
Rename image.



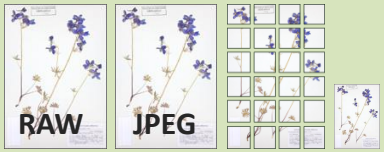
Backup to external hard drive.

Image Processing:

Convert RAW to JPEG:
Using camera software.

Upload to Server:
FTP
Hard Drive
Browser

PNW Herbaria Server



MySQL

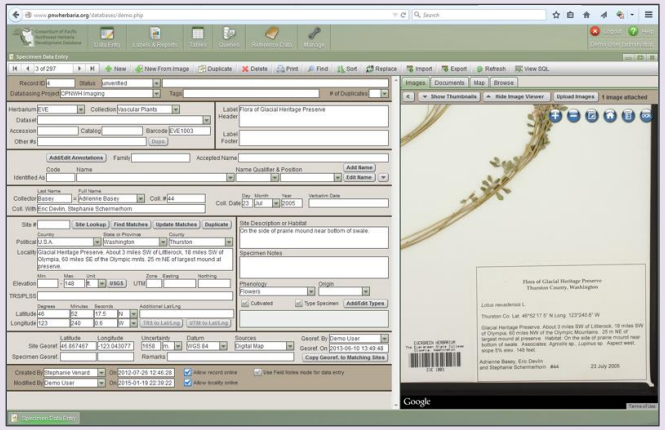
Image Processing Scripts:
Import folder metadata into database.
Add image metadata records to database.
Create tiled images and thumbnails.
Store images for online access.

Backup to remote server

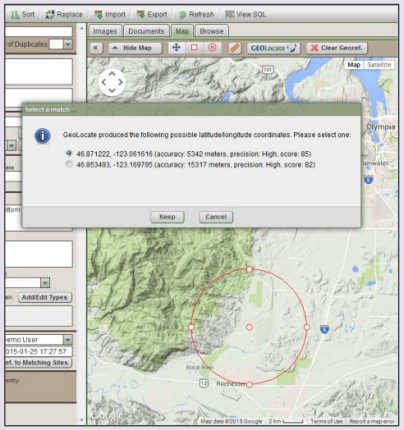
Harvested by PNW Herbaria for online access (separate database)

Exports for other data aggregators

Data Entry From Image:



Georeferencing:



Imaging & databasing comments:

- Lightbox works best. Have also used light tents with flash units, tripods with fixed lights, and flat-bed scanners.
- Rate of imaging with lightbox: 50 to 150 specimens/hour. Cost: ~ $5,000.
- Imaging workflow is designed to minimize post-processing. (exception: converting RAW to JPEG is done using camera software for best quality)
- Workflow includes some customization to imaging equipment, and several custom-built software components (e.g., metadata capture, image rename & crop, processing scripts, database software).
- Data entry rate for key-stroke entry of complete records averages 20-30 specimens per hour. Have tried OCR, with minimal efficiency gains.

Workforce:

- Rely mostly on paid hourly personnel.
- Mostly students and recent graduates (some stay long-term); some workstudy.
- Very little use of volunteers. Have not tried crowd-sourcing.