

A framework and workflow for extraction and parsing of herbarium specimen data

Jason Best

Head of Biodiversity Informatics, BRIT

Amanda Neill

Director of the Herbarium, BRIT

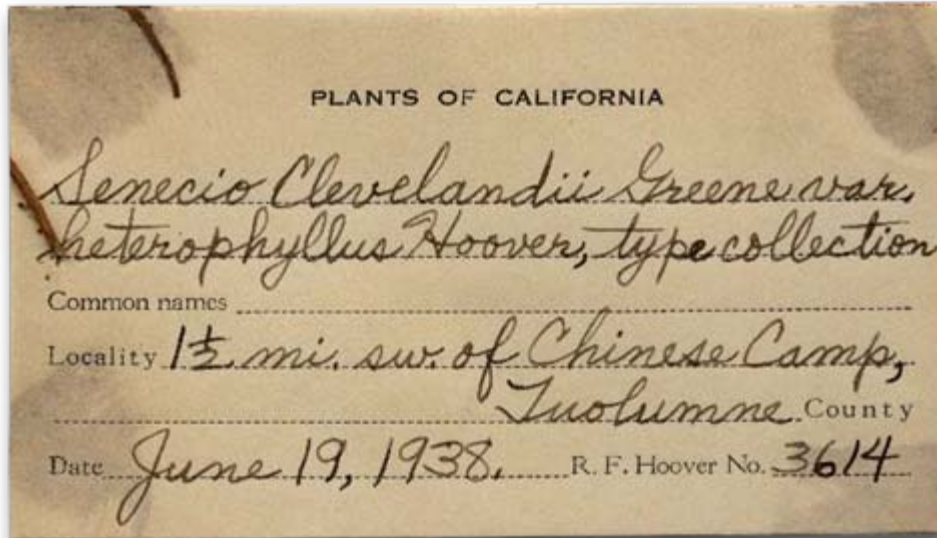
William Moen

Director of The Texas Center for Digital Knowledge and Associate Professor, UNT

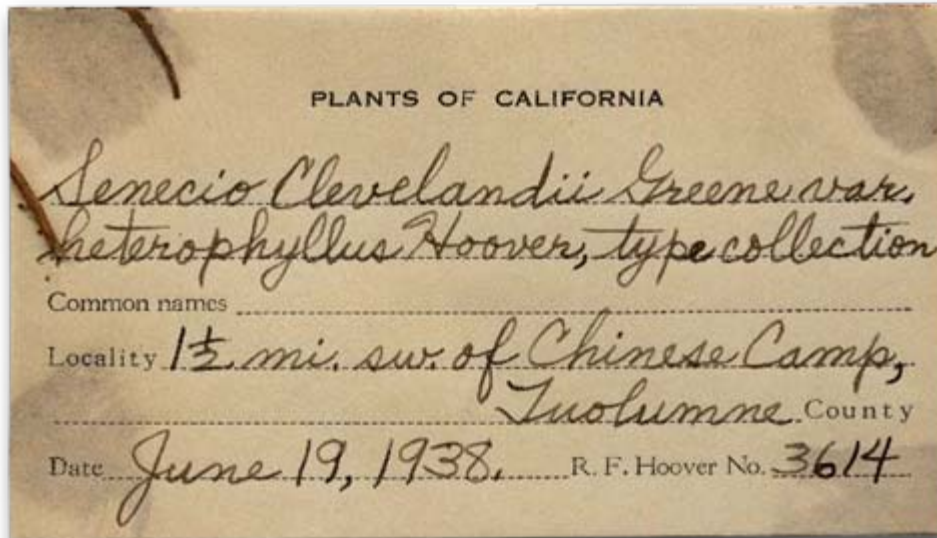
TORCH - 2011



The BRIT Collection



The BRIT Collection



A wealth of data that
are difficult to access

Apiary Project

PLANTS OF BRITISH HONDURAS

DISTRIBUTED BY C. L. LUNDELL, UNIVERSITY OF MICHIGAN HERBARIUM

Saurauia belizensis Lundell
"limonaro"

Tree, 7 in. diam., fls
white, fragrant.

TOLEDO DISTRICT: *between Monkey River
and Cockscomb, Monkey River*

PERCY H. GENTLE

No. *4439*

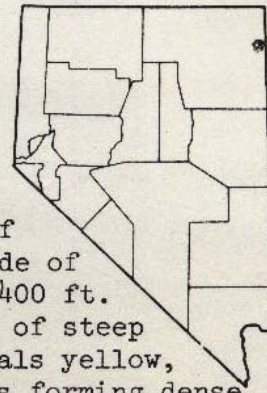
Date

July 9, 1943

FLORA OF NEVADA

Potentilla cottamii

N. Holmgren



ELKO CO., Pilot Range, Pilot Peak, aqueduct canyon east of the Pilot Mt. Ranch, west side of the peak, T36N, R70E, S6. 10400 ft. Plants growing in the cracks of steep quartzite rock outcrops, petals yellow, notched, fading white, plants forming dense clumps in open areas and loose clumps to 4 dm across in protected areas.

5 Jul 1986

Arnold Tiehm 10738

Apiary Project

PLANTS OF BRITISH HONDURAS
DISTRIBUTED BY C. L. LUNDELL, UNIVERSITY OF MICHIGAN HERBARIUM

Saurauia belizensis Lundell
"limonaro"

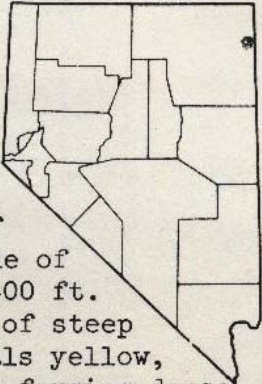
Tree, 7 in. diam., fls
white, fragrant.

TOLEDO DISTRICT: *between Monkey River
and Cockscomb, Monkey River*

PERCY H. GENTLE
No. *4439* Date *July 43*

FLORA OF NEVADA

Potentilla cottamii
N. Holmgren



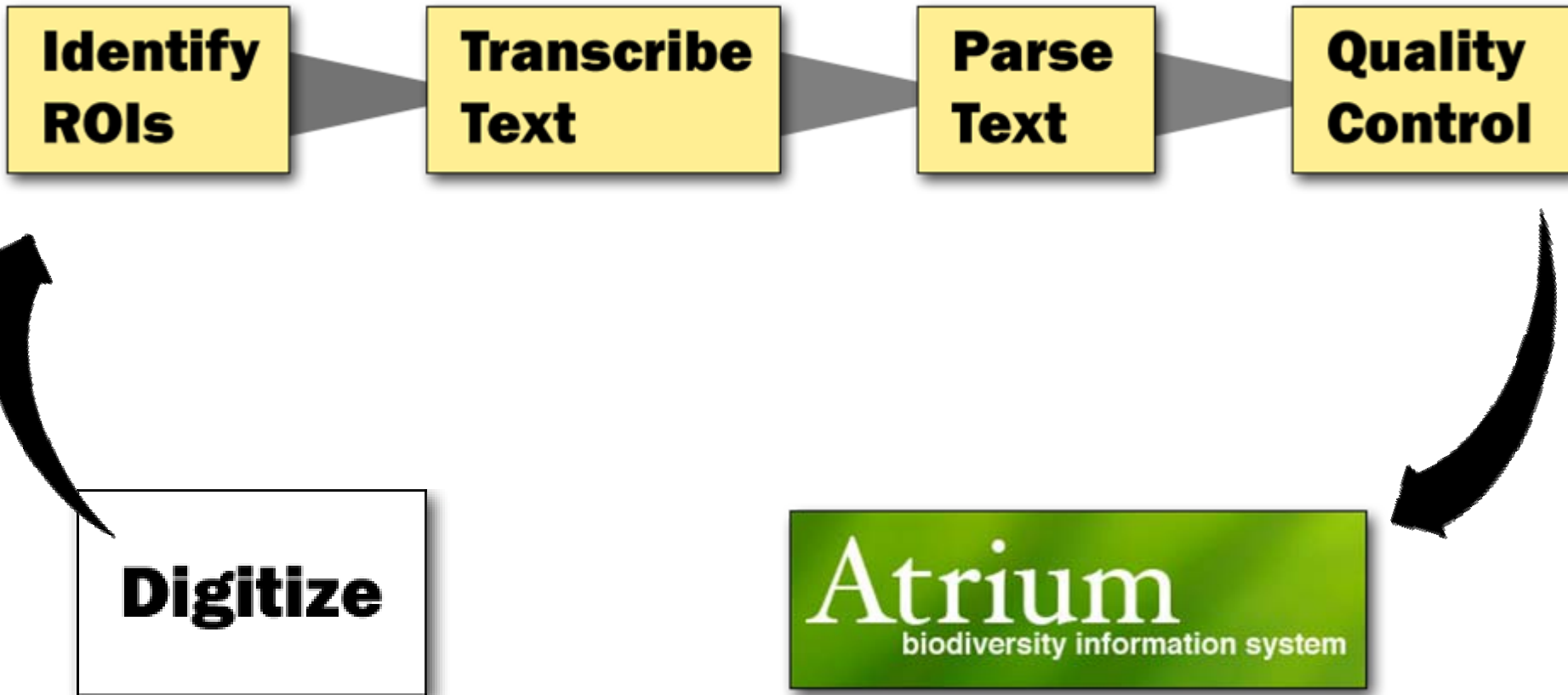
ELKO CO., Pilot Range, Pilot Peak, aqueduct canyon east of the Pilot Mt. Ranch, west side of the peak, T36N, R70E, S6. 10400 ft. Plants growing in the cracks of steep quartzite rock outcrops, petals yellow, notched, fading white, plants forming dense clumps in open areas and loose clumps to 4 dm across in protected areas.
5 Jul 1986
Arnold Tiehm 10738



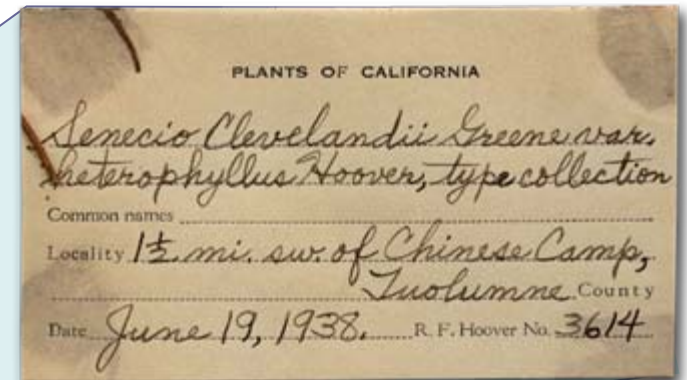
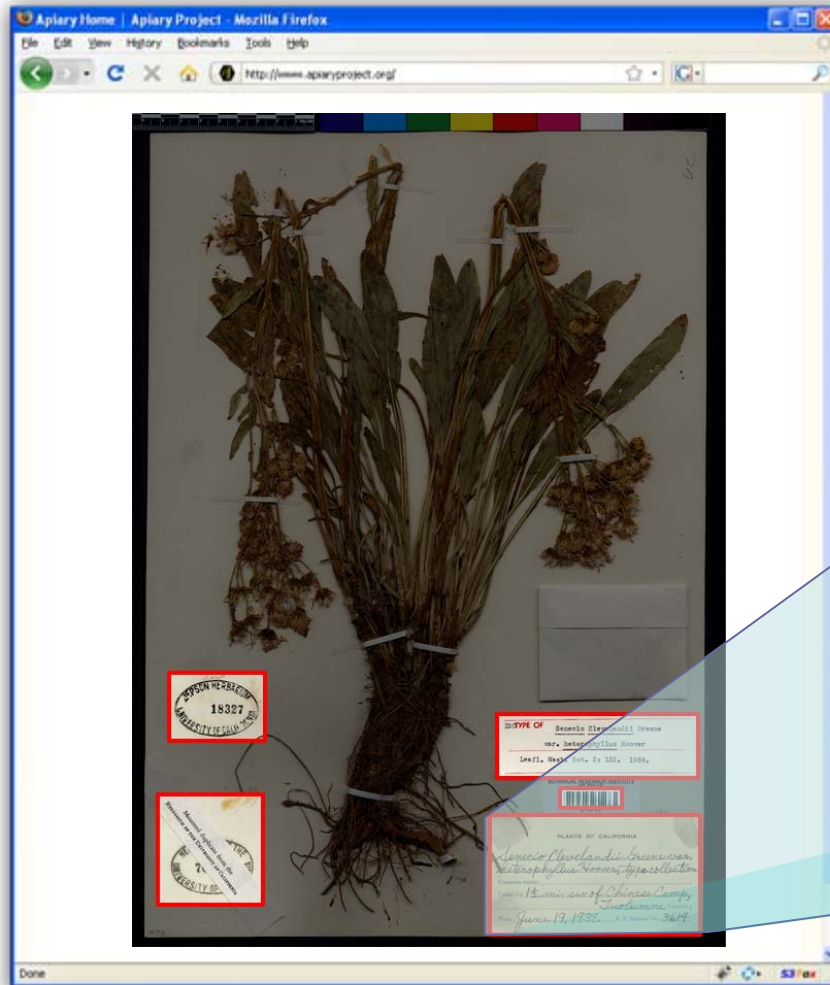
Apiary Project



The Technology and Workflow



Identify Regions of Interest



Transcribe text

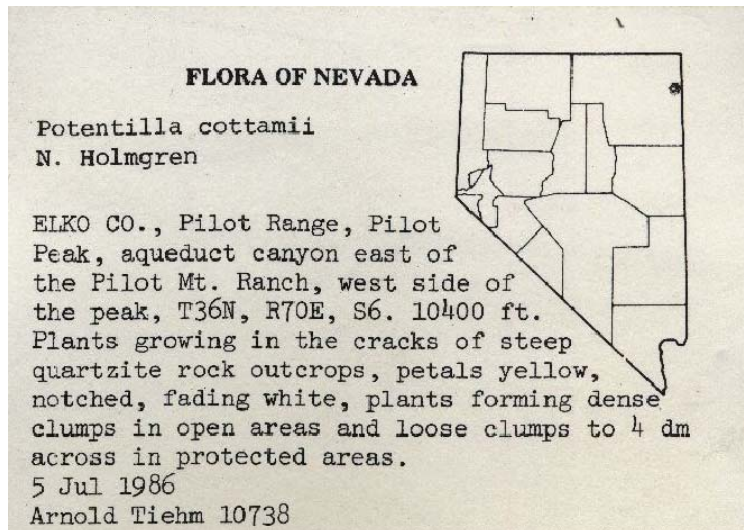
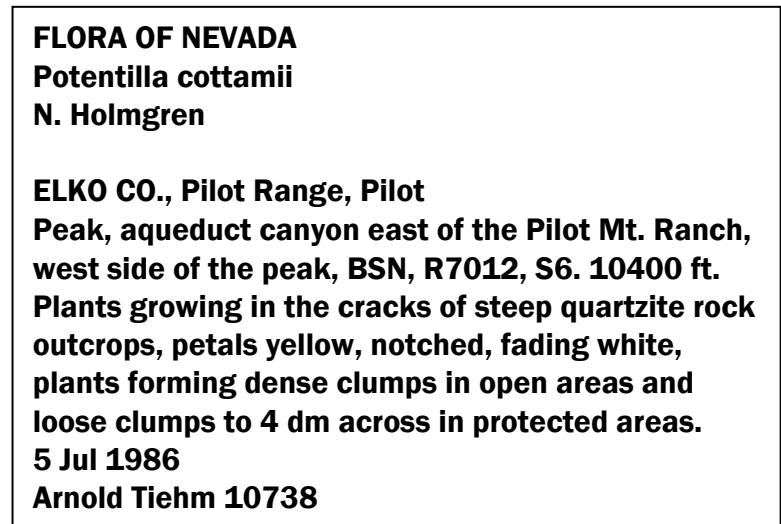


Image with imprinted text



Digital text

Transcribe text

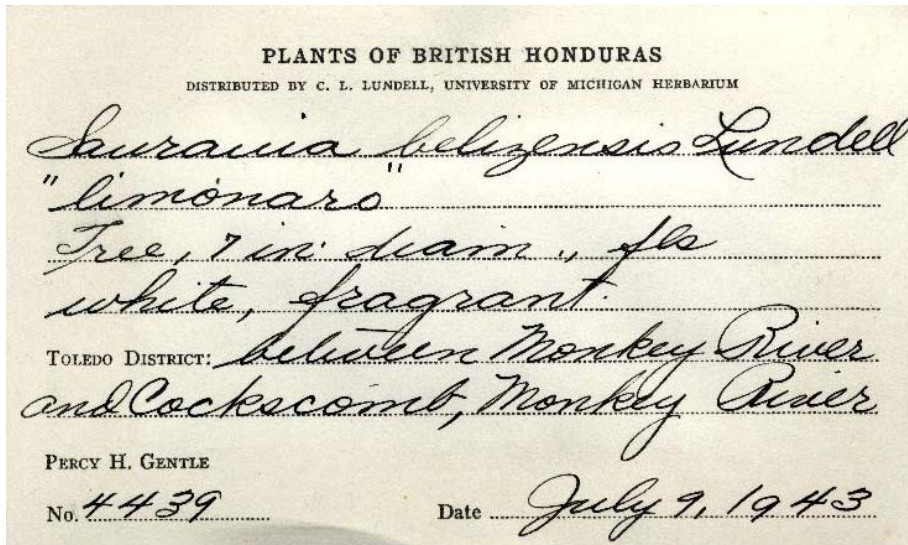


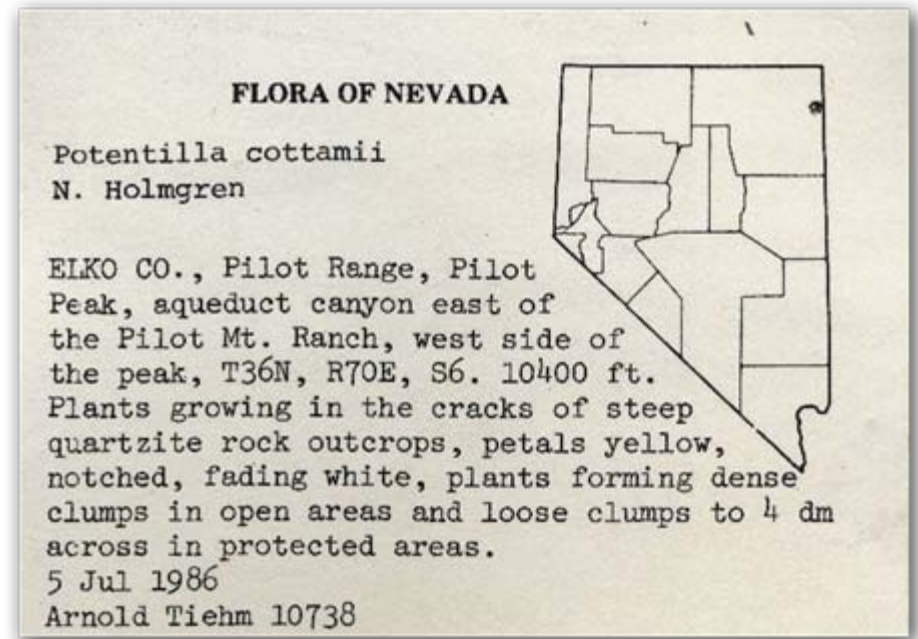
Image with imprinted text

PLANTS OF BRITISH HONDURAS
DISTRIBUTED BY C. L. LUNDELL, UNIVERSITY OF
MICHIGAN HERBARIUM

Saurauia belizensis Lundell "limonaro"
Tree, 7 in. diam, fls white, fragrant
Toledo District: between Monkey River and
Cockscomb, Monkey River
Percy H. Gentle
No. 4439 Date July 9, 1943

Digital text

Parse text



Parse text

CLASSIFICATION

Scientific Name: *Potentilla cottamii*
Author: N. Holmgren

LOCALITY

Region: Nevada
Subregion: Elko County
Locality Summary: Pilot Range, Pilot Peak
Site Specific: aqueduct canyon east of the Pilot Mountain Ranch, west side of the peak
Coordinates: T36N, R70E, S6
Elevation: 10400 feet

HABITAT

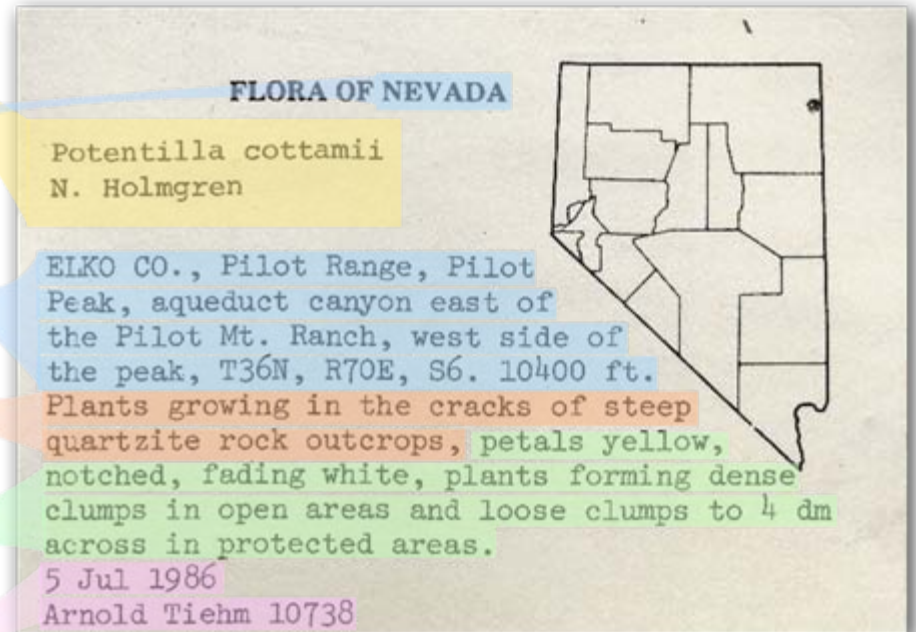
Plants growing in the cracks of steep quartzite rock outcrops

PLANT NOTES

petals yellow, notched, fading white, plants forming dense clumps in open areas and loose clumps to 4 dm across in protected areas.

COLLECTION EVENT

Date: 1986/07/05
Collector Name: Arnold Tiehm
Collection Number: 10738

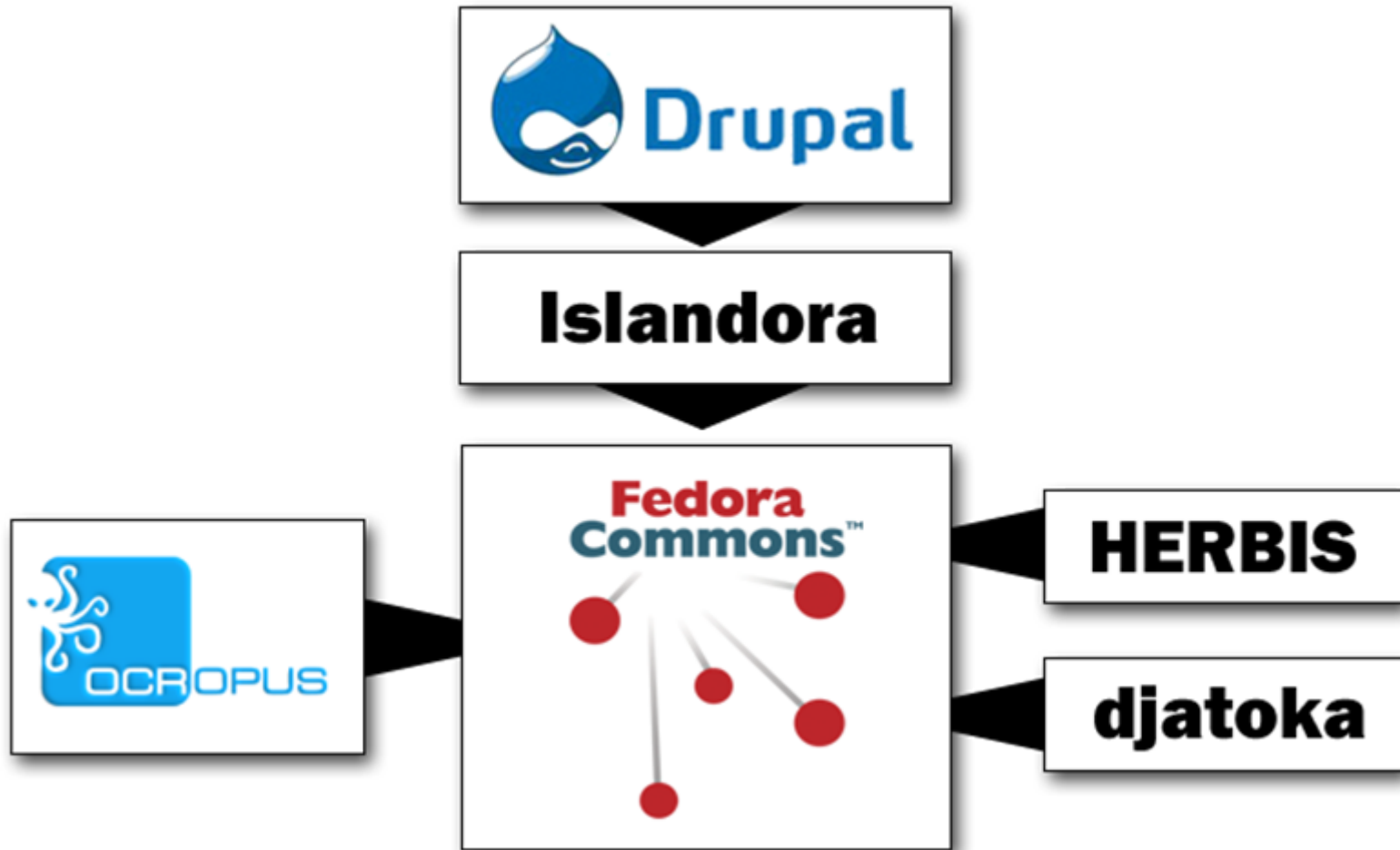


Quality Control then Publish

- Humans can view the extracted data and compare to the source image
- Machine processes can check spelling, validate coordinates etc.
- When all data have been checked and approved, they are ready to be published



The Technology and Workflow



The Project

- Open source
- Demo at demo.apiaryproject.org
- In active development (still rough in some areas)
- More information at www.apiaryproject.org

Thanks

Institute of Museum and Library Services

Dr. Bill Moen, Amanda Neill

BRIT Team

**Dr. Brooke Byerley, Tiana Franklin, Keri McNew, Sean
Murphy**

UNT Team

Dr. Miguel Ruiz

**Zainab Arsiwala, Moiz Bhukhiya, Jane Huang, Melody
McCotter, Hong Xu**

www.apiaryproject.org - www.txcdk.org - www.brit.org - www.unt.edu