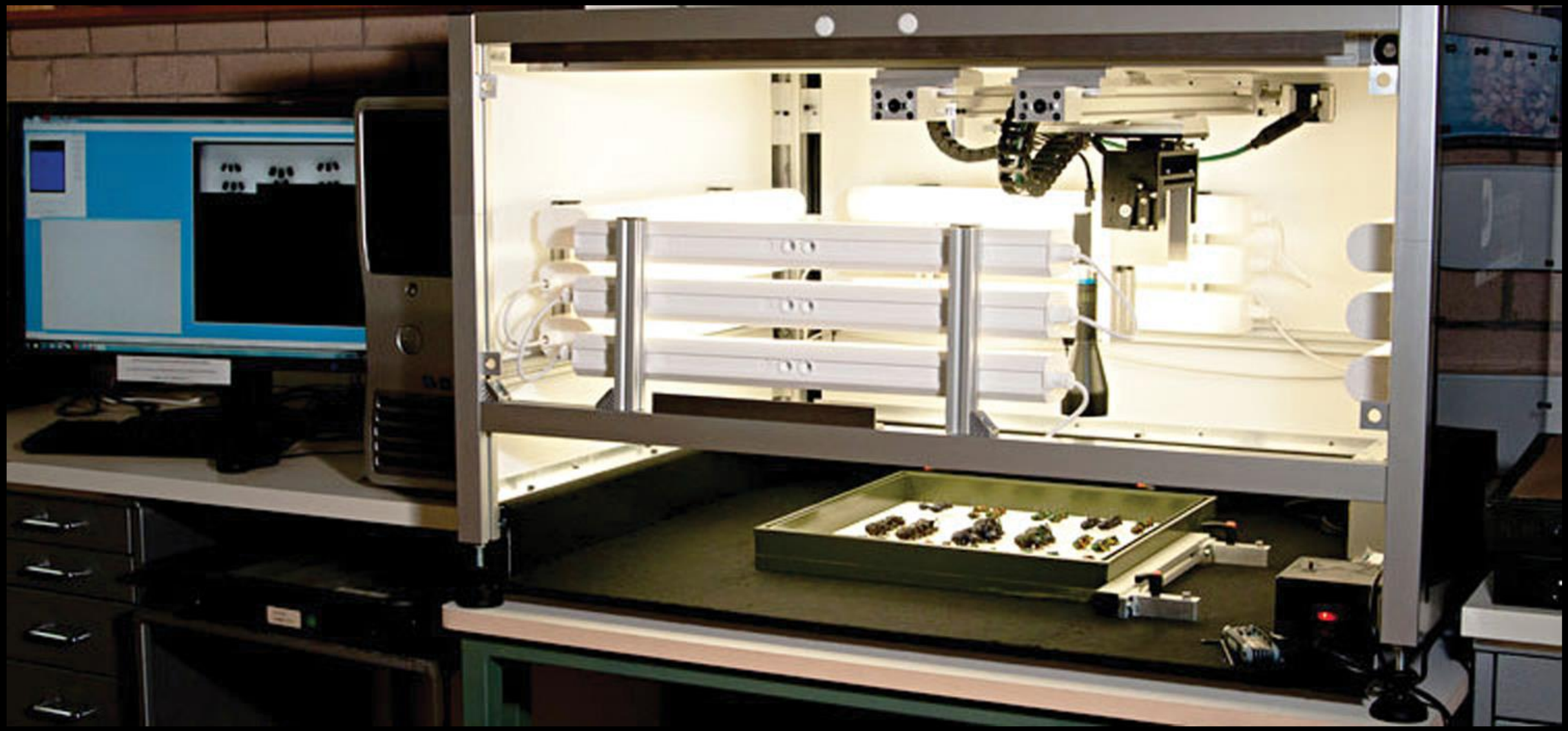
Moving beyond the box:
automating the digitisation of insect collections

Pieter Holtzhausen, Stéfan van der Walt, Alice Heaton, Laurence Livermore, Vladimir Blagoderov, Ben Price, Lawrence Hudson, **Vincent Smith**

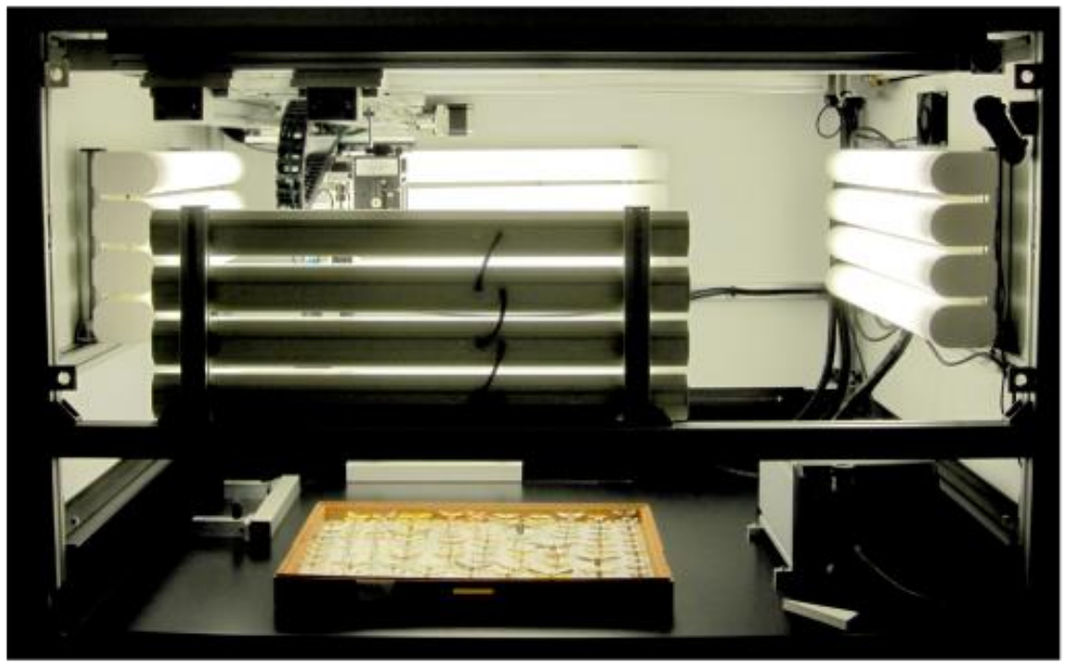
NATURAL HISTORY MUSEUM

SYNTHESYS
Synthesis of systematic resources



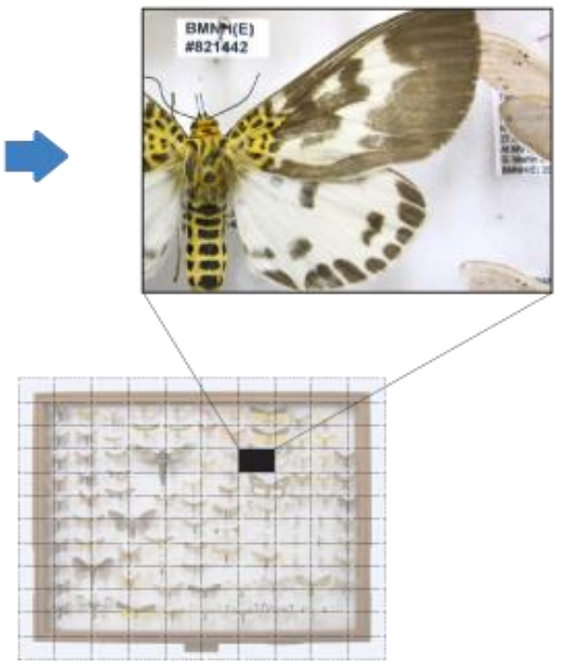
Blagoderov et al (2012) No specimen left behind: industrial scale digitization of natural history collections.
***ZooKeys* 209: 131–146, doi: 10.3897/zookeys.209.3178**

Drawer level imaging is (mostly) a solved problem



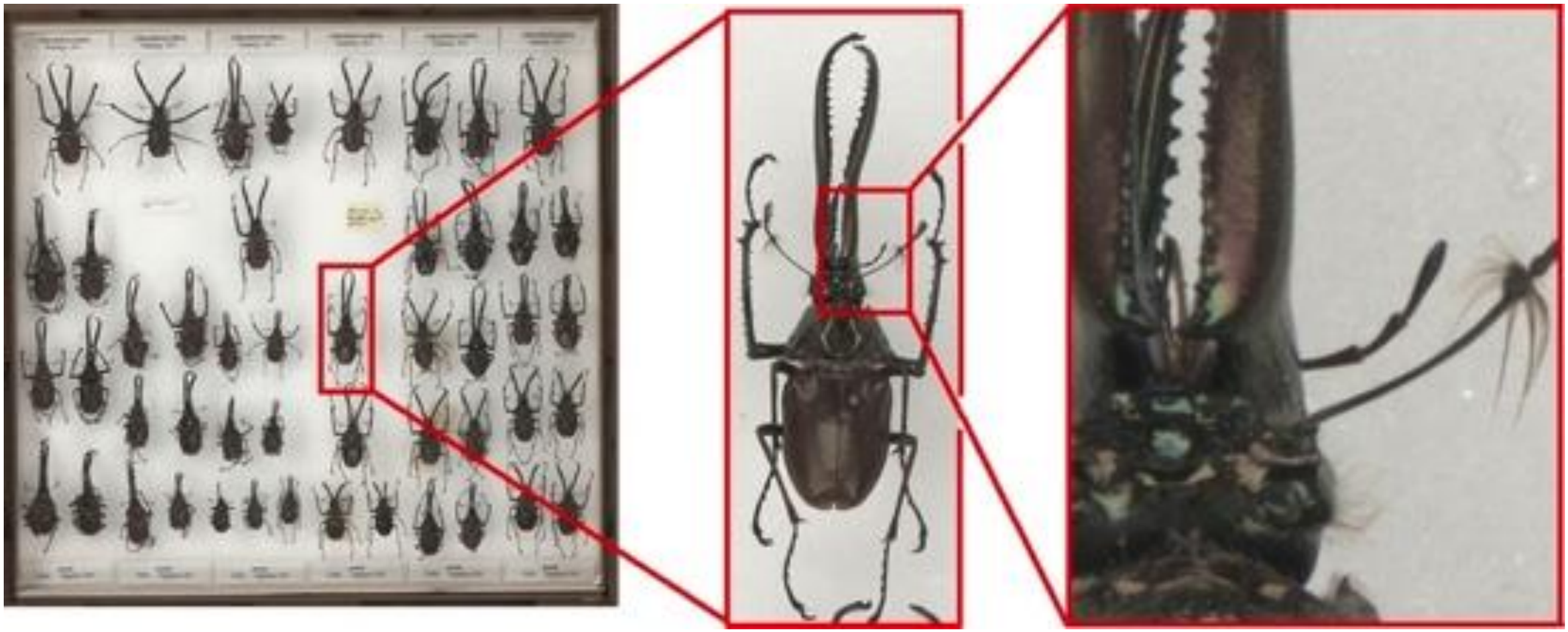
1. Place drawer

2. Scan



3. Stitch

Drawer level imaging is (mostly) a solved problem



Result: High resolution composite image

- Fast (5 mins per drawer)
- High resolution (circa 500MB per image)
- No specimen handling

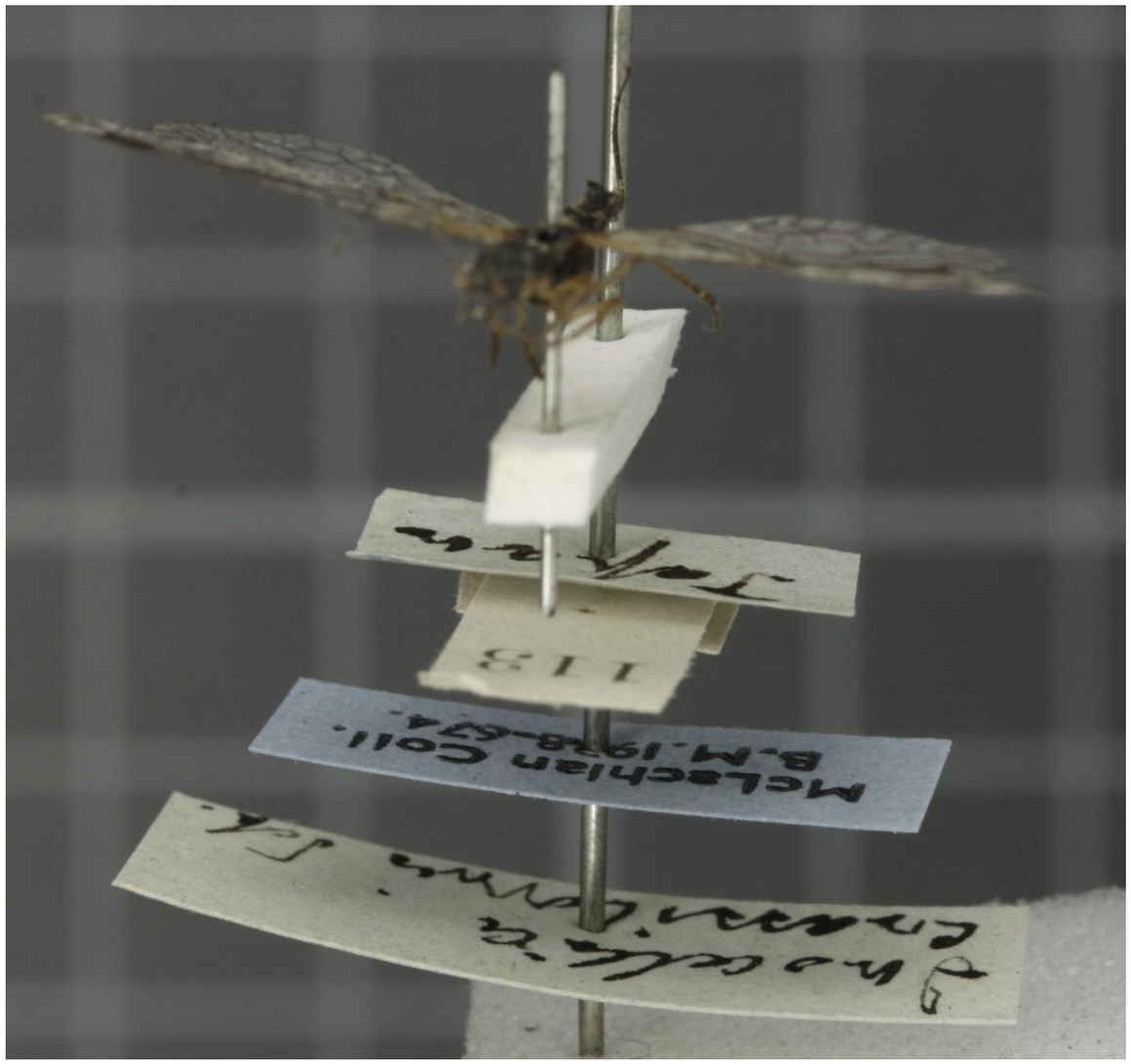
NATURAL HISTORY MUSEUM

But, two key problems remain…

Synchronisation

Keeping the physical & digital copies in sync

Label data



Capturing data from multiple pinned labels

NATURAL HISTORY MUSEUM

Approaches to the synchronisation problem

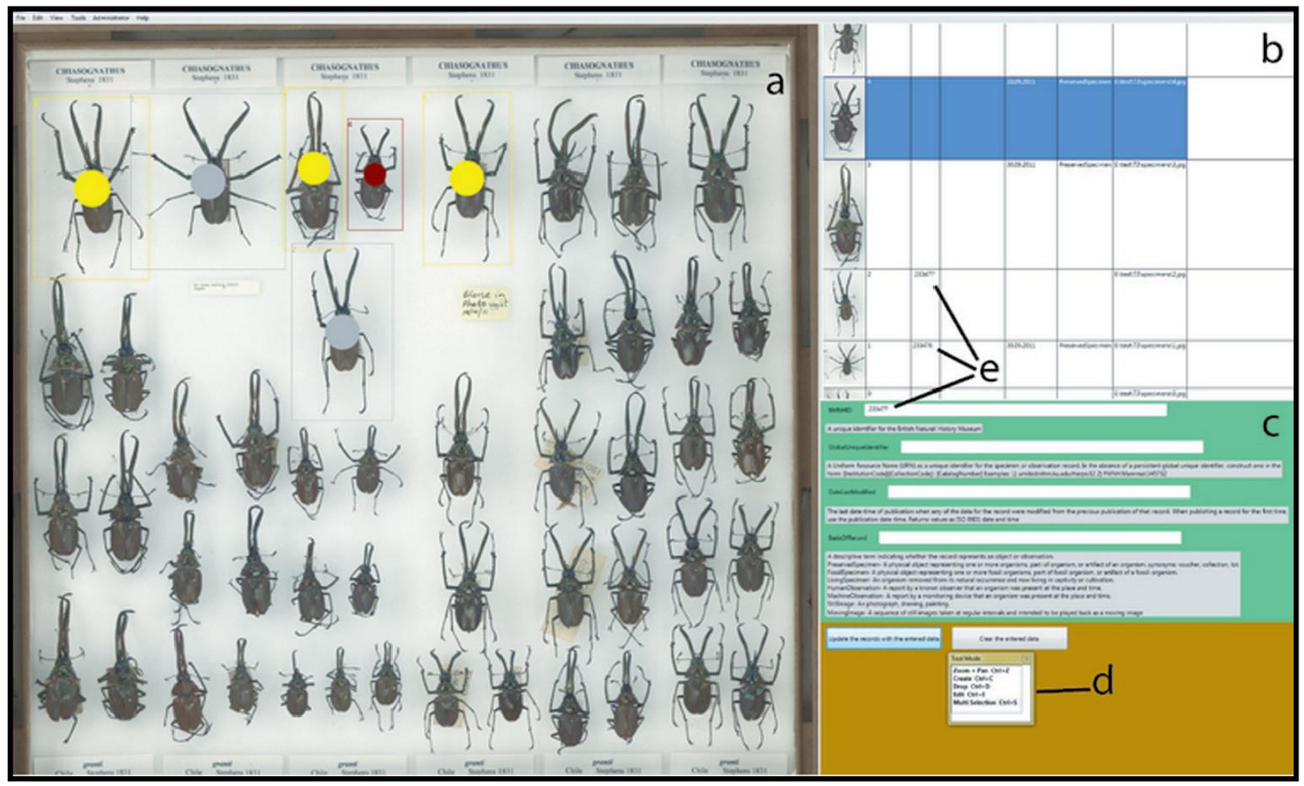
- Don’t worry about it, re-image as required
- Lock down the drawers
- Crop-out each specimen image
  - Automate the cropping process
  - Link each specimen to its digital image
  - Make it easy to collect label data

The only practical solution, but a new rate limiting step

Initial supporting software

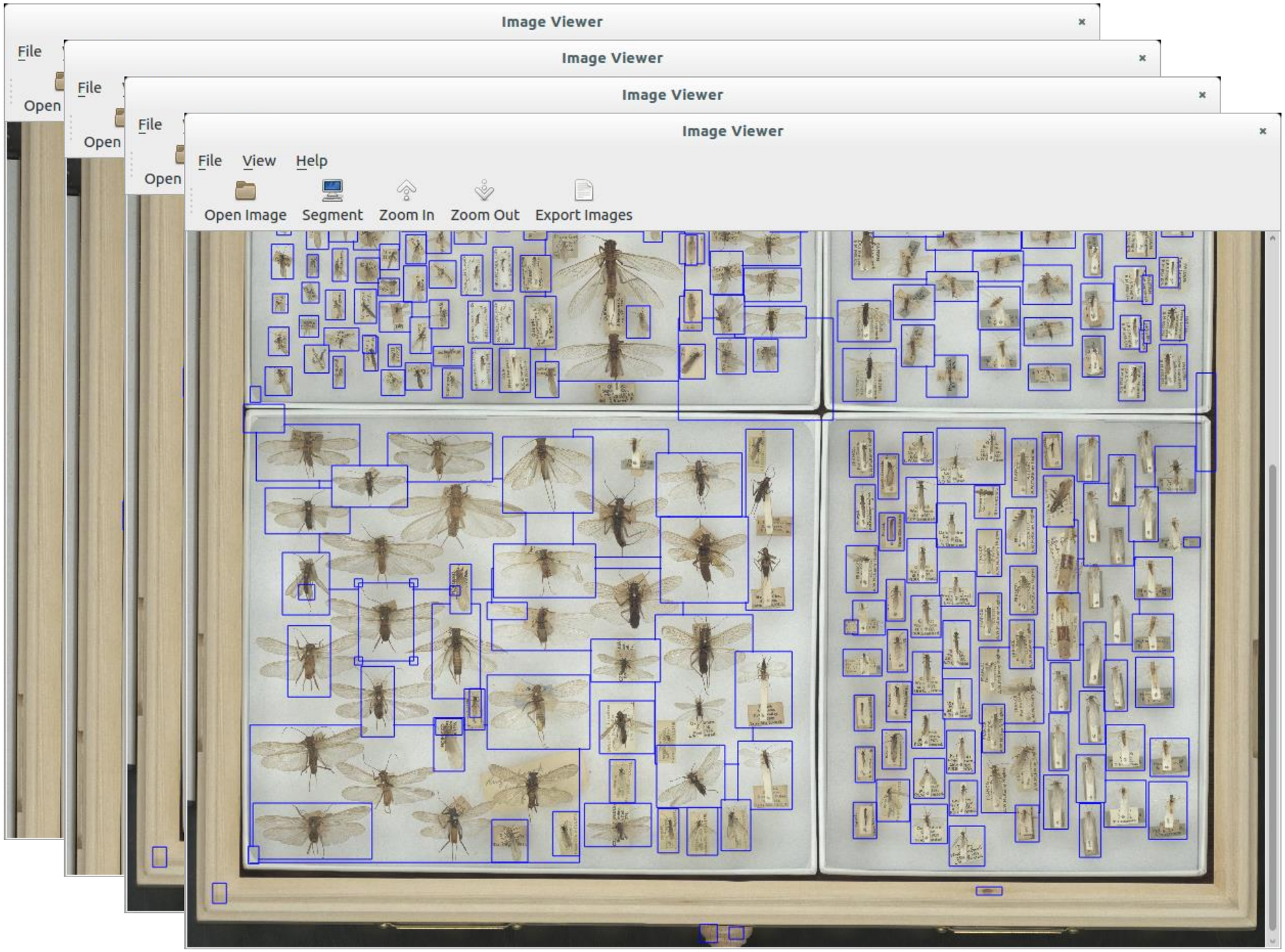
Annotation software, NHM working with SmartDrive:

- No automated cropping
- Manually link images & specimens
- Poor UX/UI
- Closed source and proprietary
- Not cross-platform (Windows only)



A good first step to understanding the problems

Automating specimen segmentation



Starting image

↓

Auto-segment

↓

Mark errors

↓

Correct

Work with Pieter Holtzhausen and Stéfan van der Walt (Stellenbosch University)
Software: Inselect, written in Python

Segmentation methods

Original

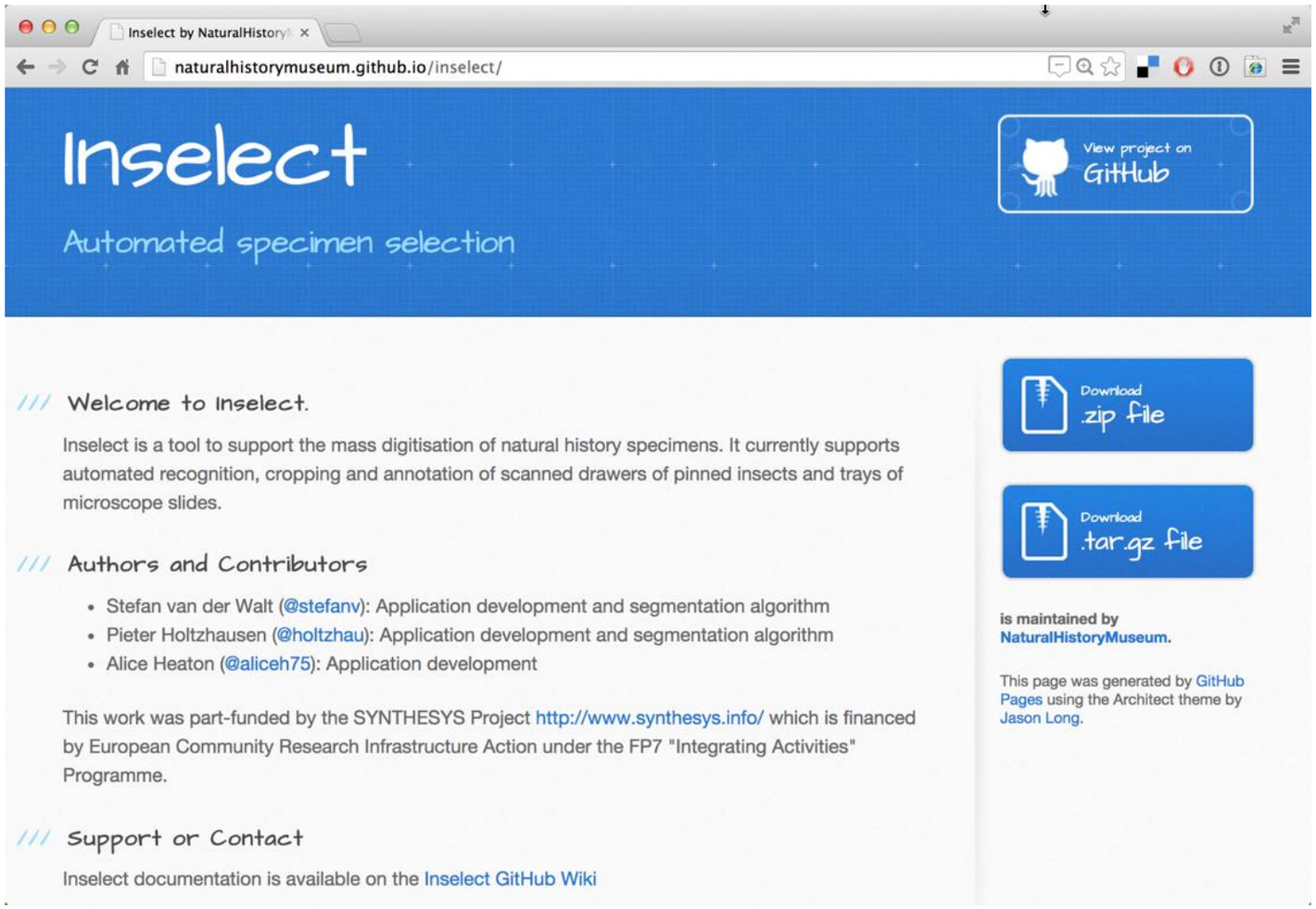
Secondary
(seed growing)

Primary segmentation
(contrast based)

NATURAL HISTORY MUSEUM

Inselect

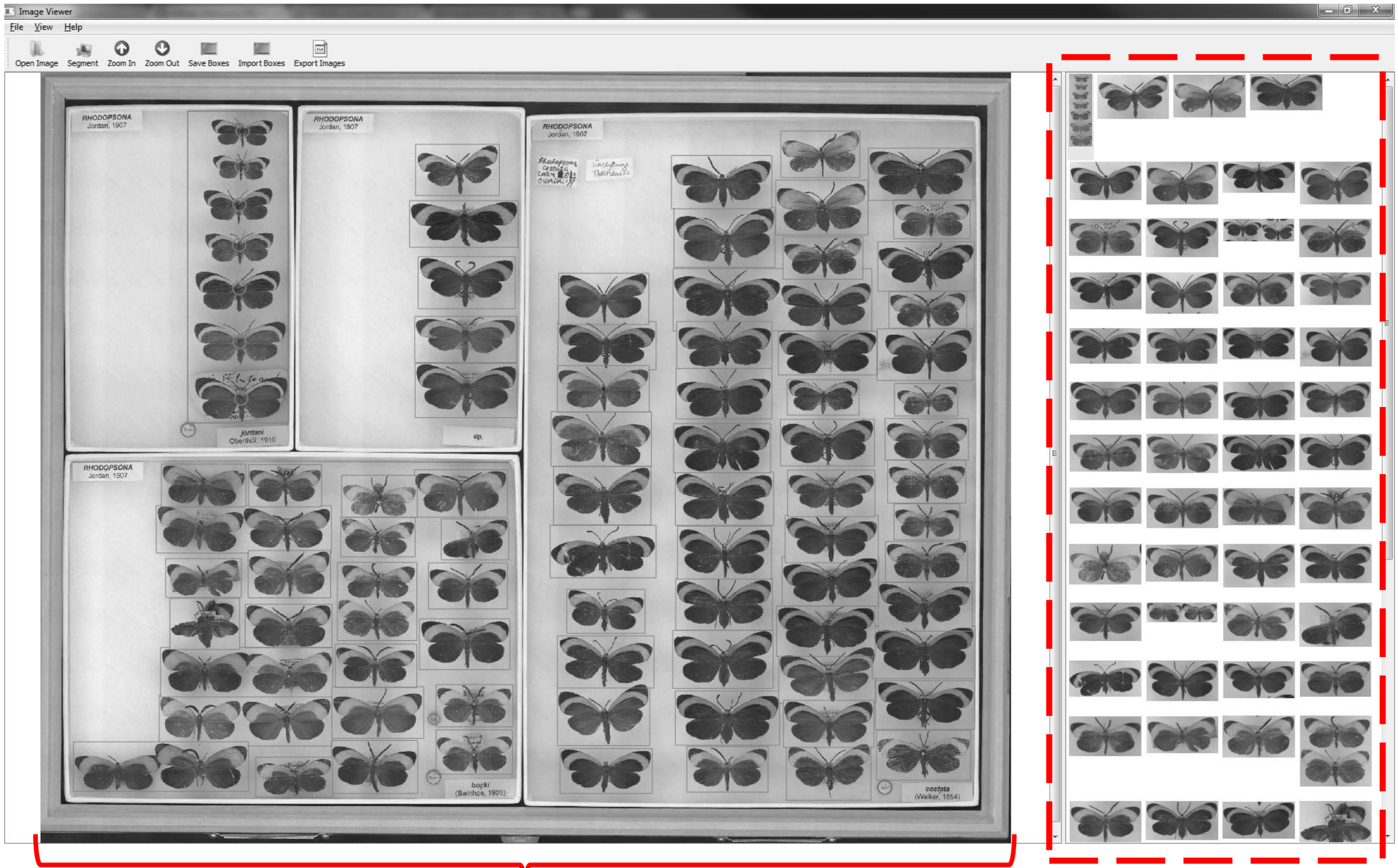
Automated recognition, cropping and annotation of specimens



- Currently pre-release (alpha)
- Automatically detects specimens
- Creates bounding boxes for cropping and exporting images
- Rapid annotation interface
- Persistent settings & keyboard shortcuts
- Data export in JSON format
- Open source & modular
- Python based (OpenCV, scikit-image libraries)
- Windows, OSX & Linux

http://naturalhistorymuseum.github.io/inselect/

Inselect: segmentation of specimen images

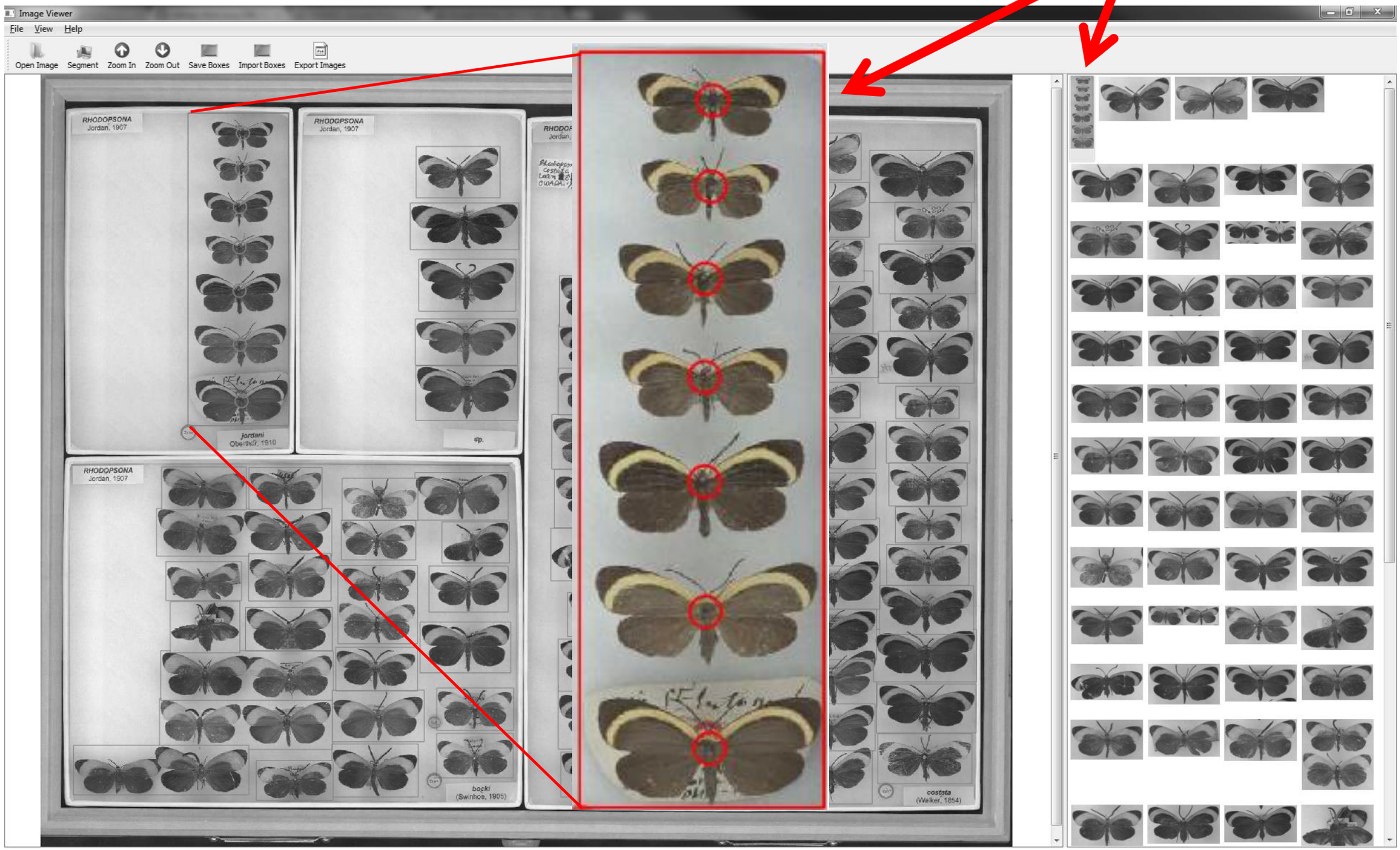


Whole Drawer image

Auto-segmented images in sidebar

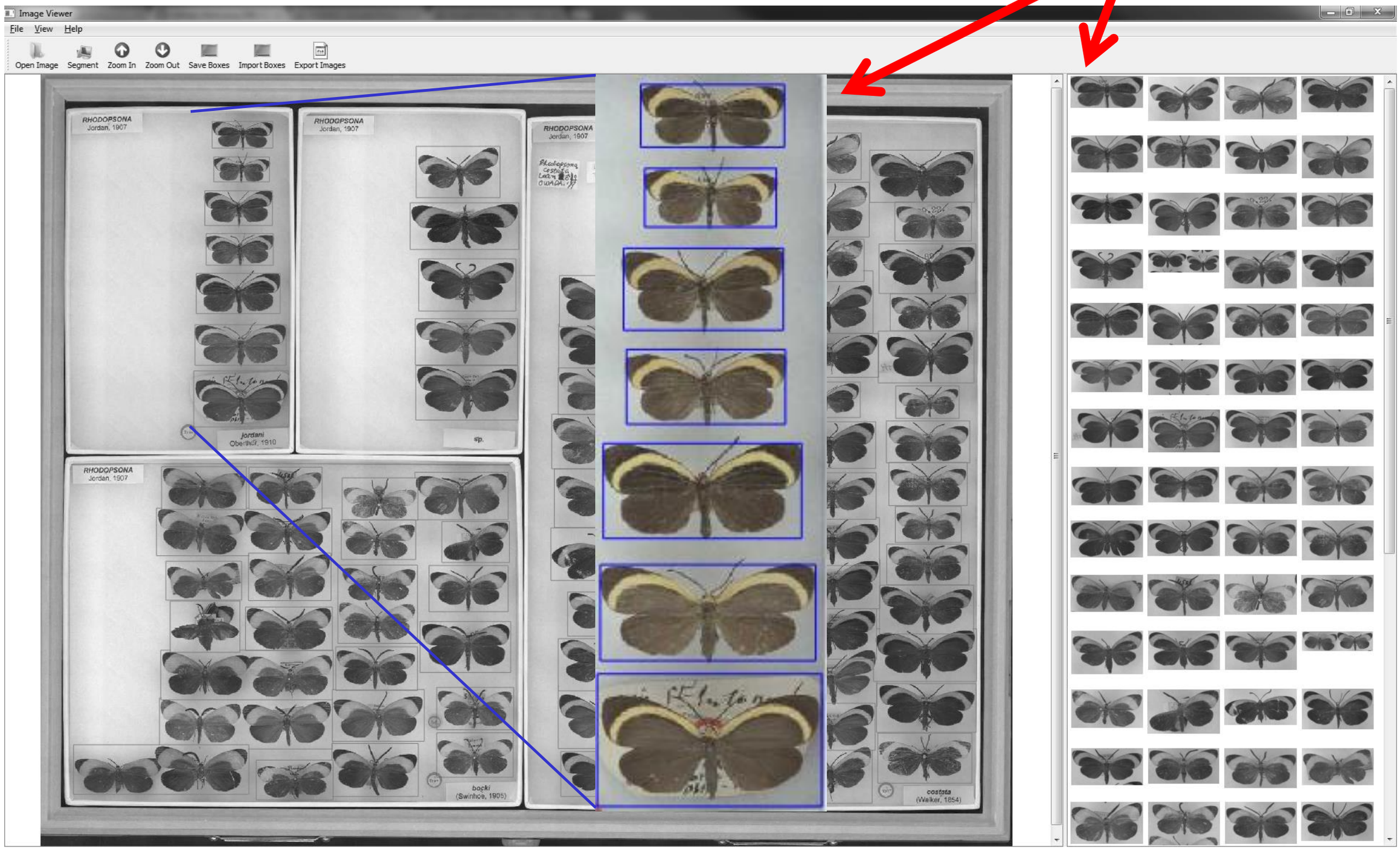
Secondary re-segmentation (seed growing)

Easy to spot & correct errors

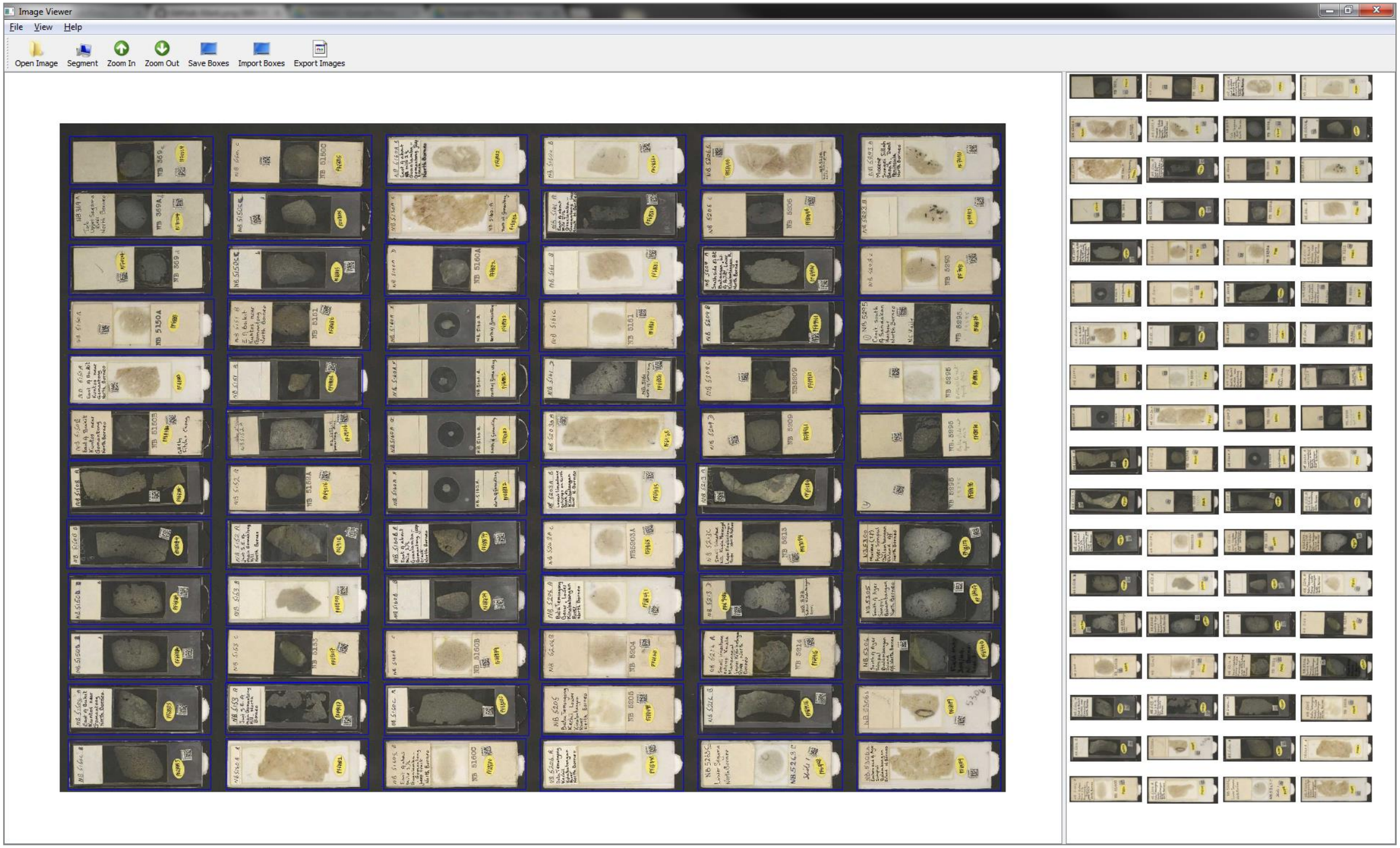


Secondary re-segmentation (seed growing)

Easy to spot & correct errors

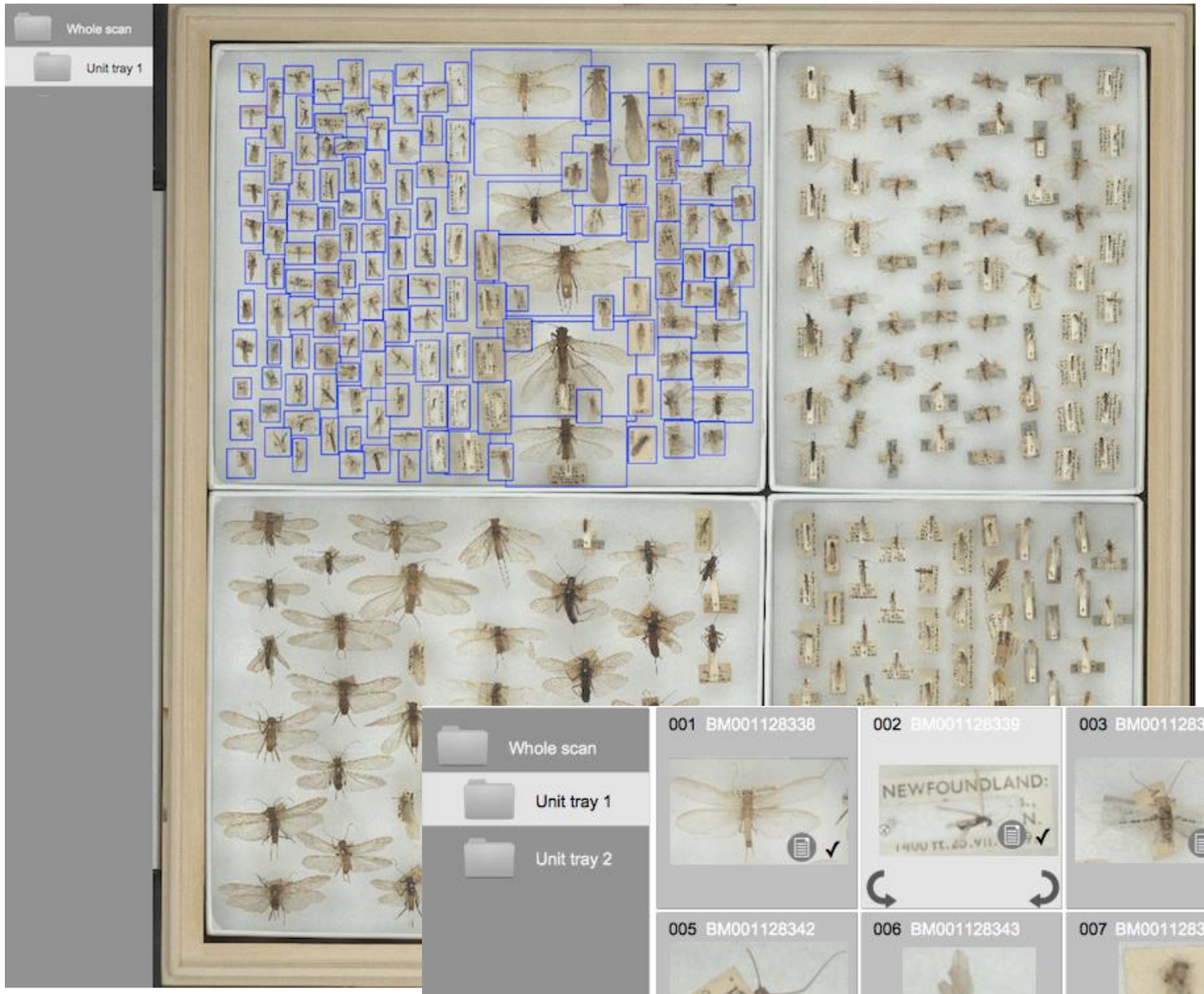


Works on slides & pinned insects



Also testing mineral & fossil samples

Planned UX / UI Enhancements



Unit tray recognition

Plug in controlled vocabulary services

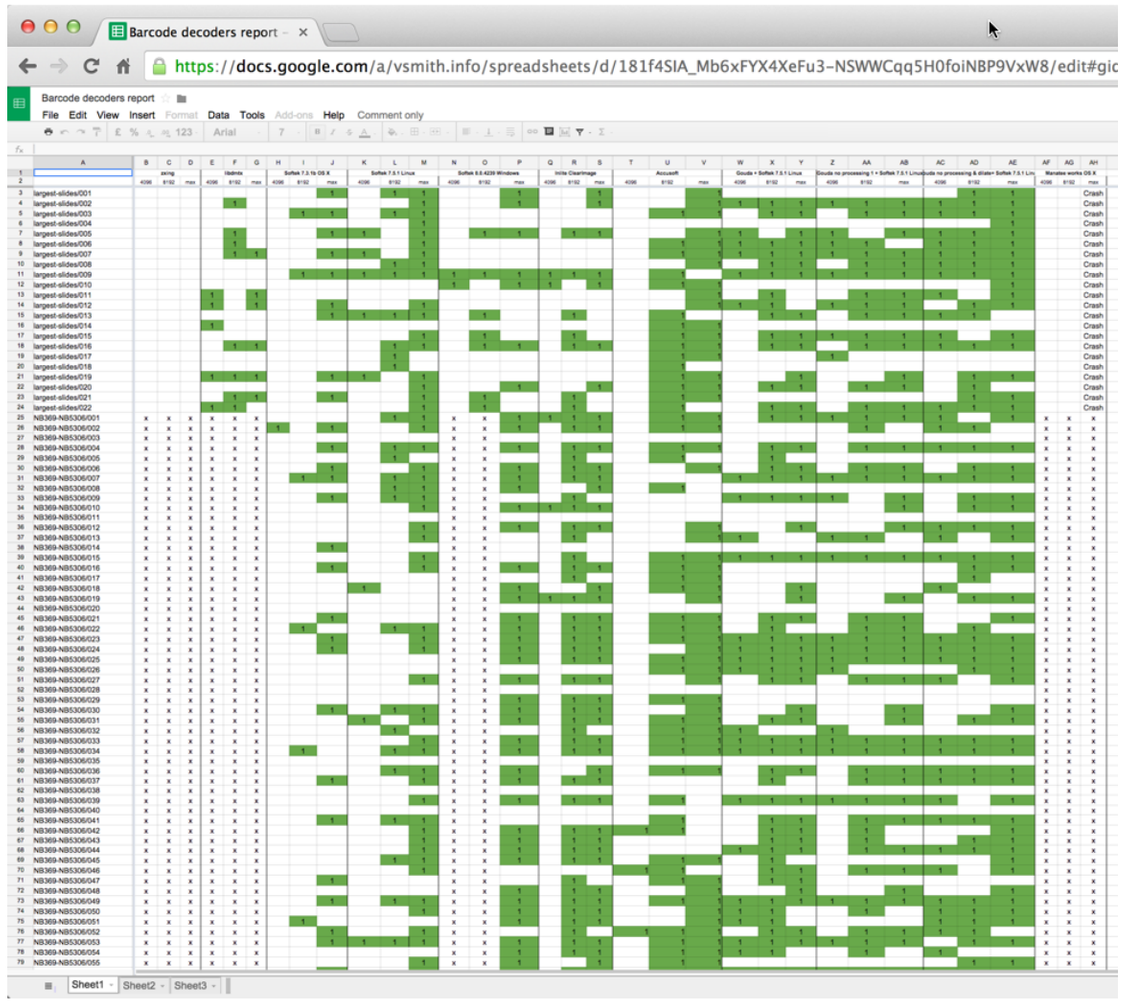
Multi-specimen annotation

1D & 2D barcode recognition

- Recognition and reading of 1D & 2D matrix barcodes from images
- Different physical requirements (smallest 6x6mm matrix, readable via handheld scanners)
- Testing open source & commercial libraries at different scan resolutions

Initial results

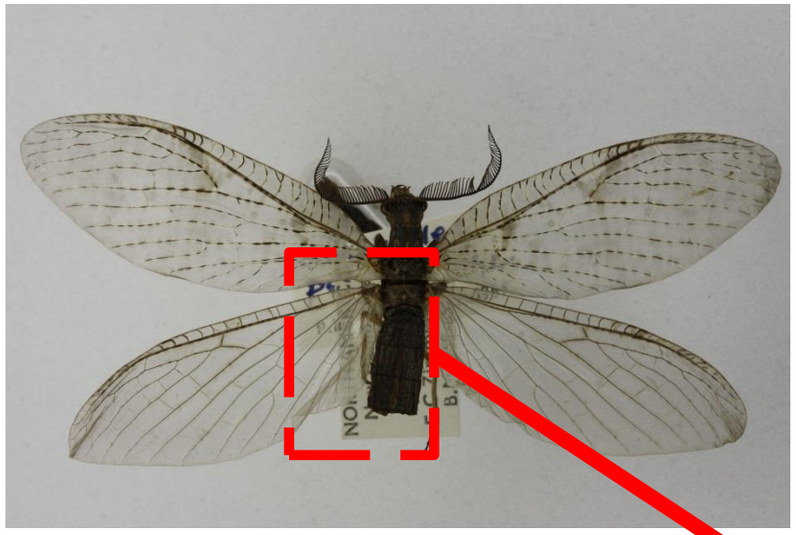
- Commercial solutions outperform open source
- Max. read success 94%
- Idiosyncratic results (different results on different OS)
- Testing continues…



The Holy Grail: label imaging & text recognition

Chauliodes pectinicornis

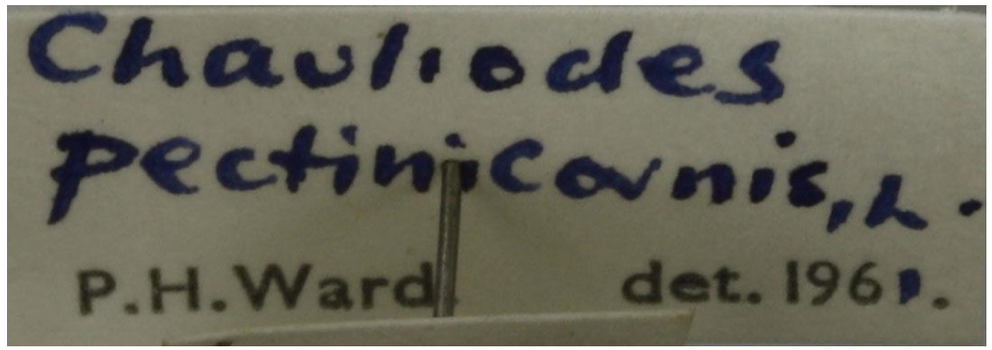
Dorsal

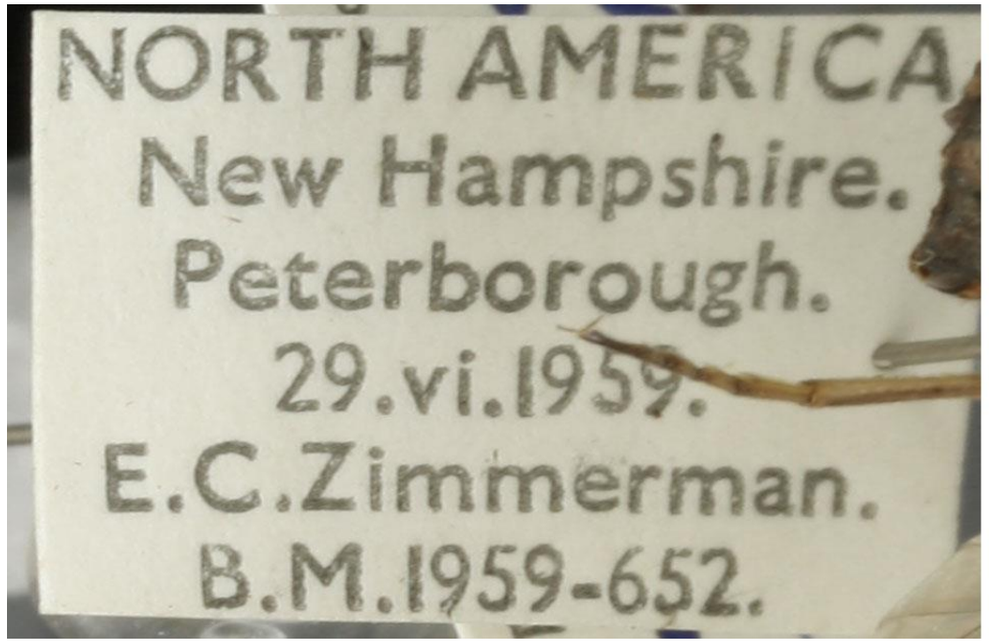


Caudal

Frontal

Reconstructed labels

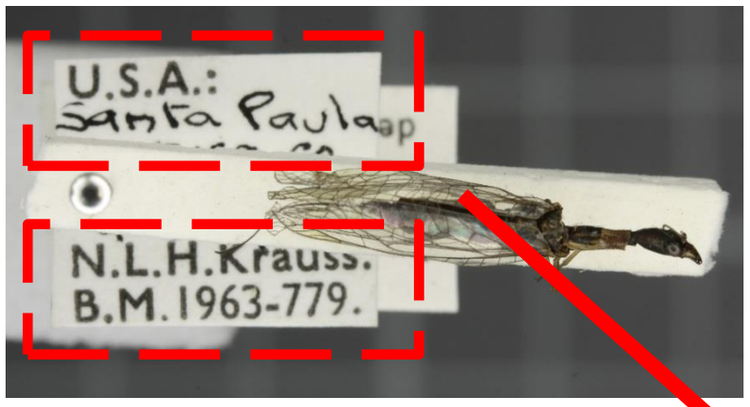




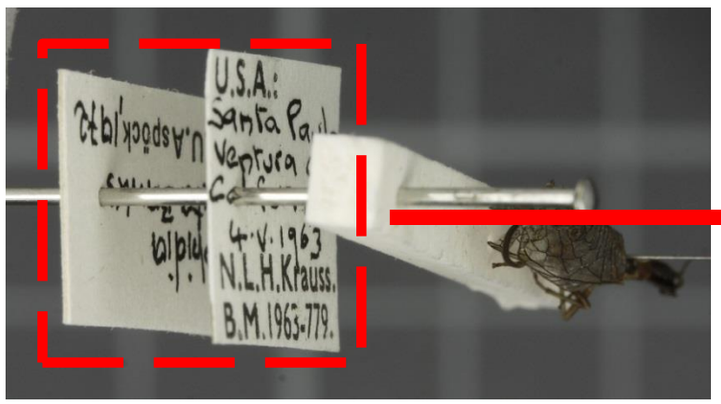
The Holy Grail: label imaging & text recognition

Agulla astuta

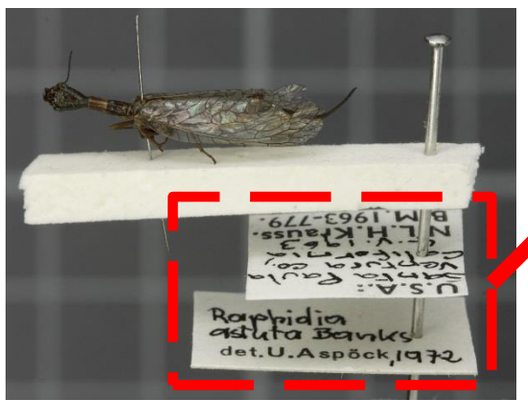
Dorsal



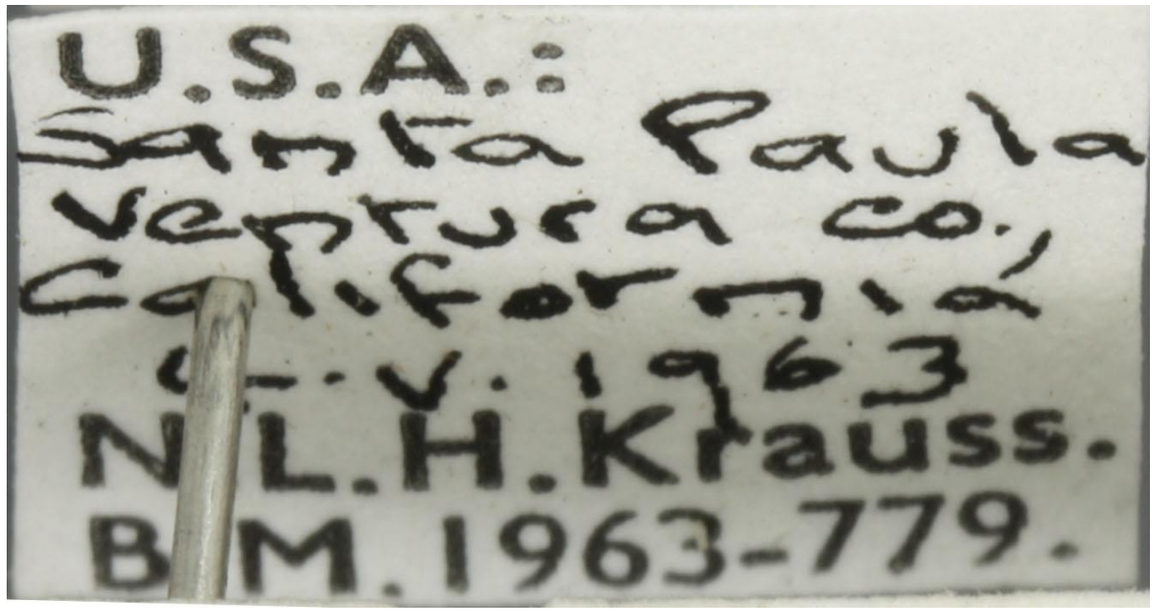
Caudal

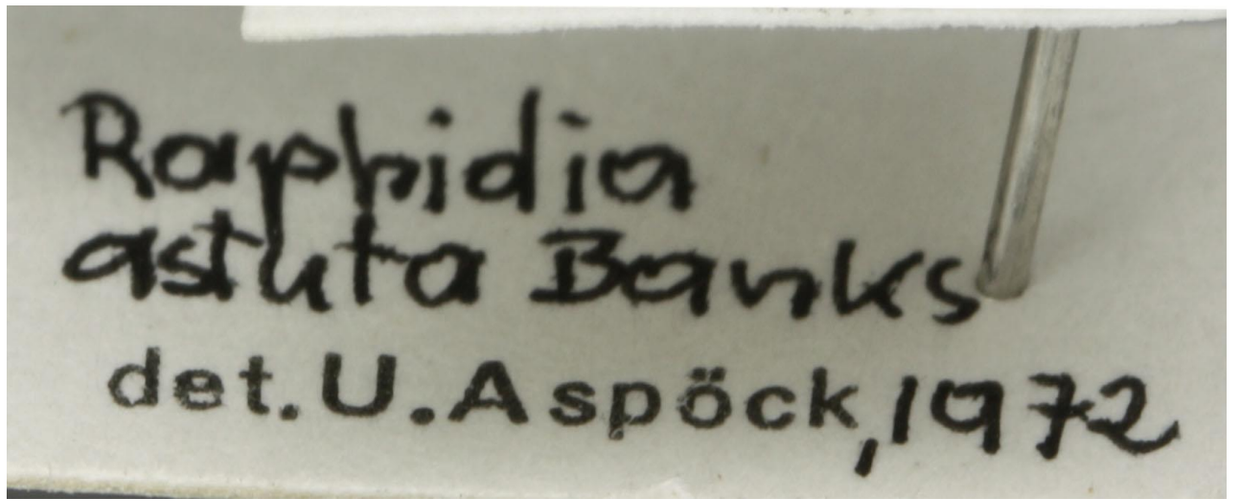


Lateral



Reconstructed labels

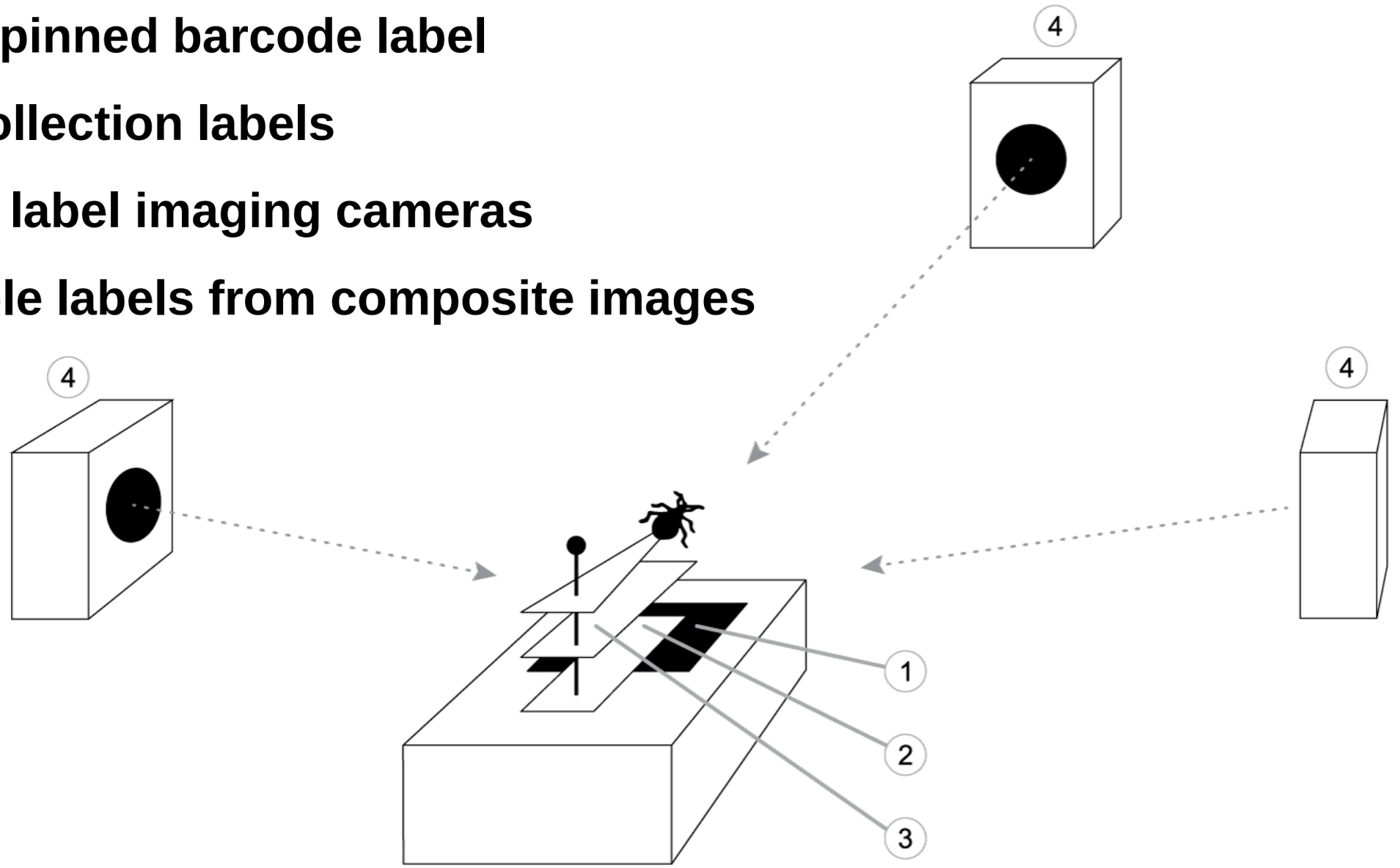




Approaches to label imaging pinned specimens

Could be incorporated as part of the barcode dispensing process

1. **Barcode dispenser & scanner**
   (two sides barcode labels)
2. **Freshly pinned barcode label**
3. **Other collection labels**
4. **Multiple label imaging cameras**
5. **Assemble labels from composite images**



Acknowledgements

Segmentation algorithm & app. development
Stefan van der Walt and Pieter Holtzhausen

Application development
Alice Heaton

Barcode recognition & testing
Lawrence Hudson

Analysis & testing
Laurence Livermore, Vladimir Blagoderov and Ben Price

Initial specification and funding
Vince Smith

UNIVERSITEIT STELLENBOSCH UNIVERSITY

NATURAL HISTORY MUSEUM

SYNTHESYS
Synthesis of systematic resources