



## Best practises in managing and digitizing scientific collections

### 1. Insects on pins

### 2. Automation of digitization activities

*Digitization Workshop at SiBBR, Brasilia, Brazil, 25.11.2014*

Digitarium / Riitta Tegelberg  
[www.digitarium.fi](http://www.digitarium.fi)

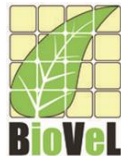




Service centre for high performance digitisation

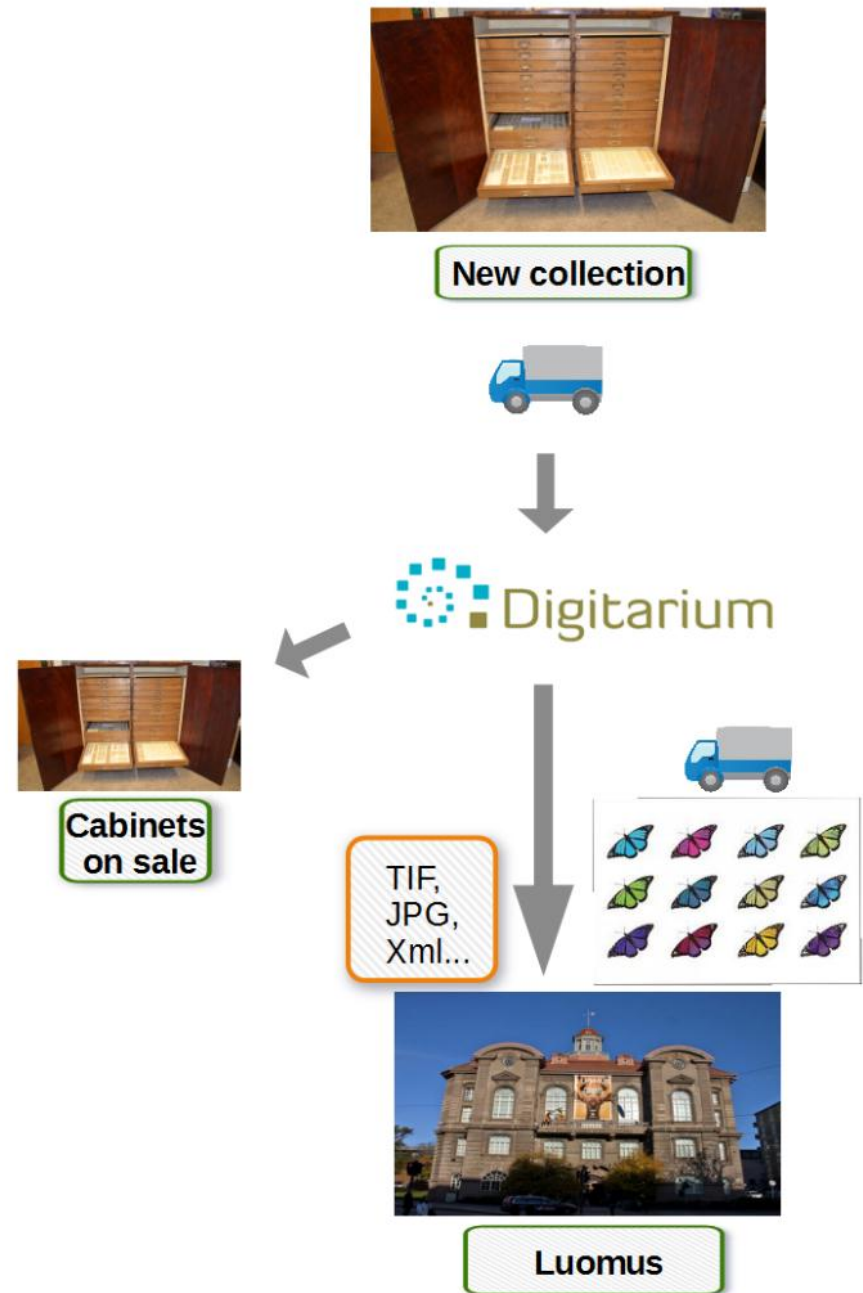
## Finnish digitization service centre

- Development of methods and processes
- Research
- Training, knowledge dissemination
- Outsourcing services for museums, including mass digitization



# 1. Imaging process at Digitalarium: insects

- Only specimens (and images & metadata) are delivered to Finnish Museum of Natural History (=Luomus), not the cabinets
- Imaging of individual specimens



# Before imaging

- Inspection of damages (photos, report)
- Possible freezing (also post -)
- Printing of unique identifier (ID) labels, 12 x 6 mm, printed on both sides

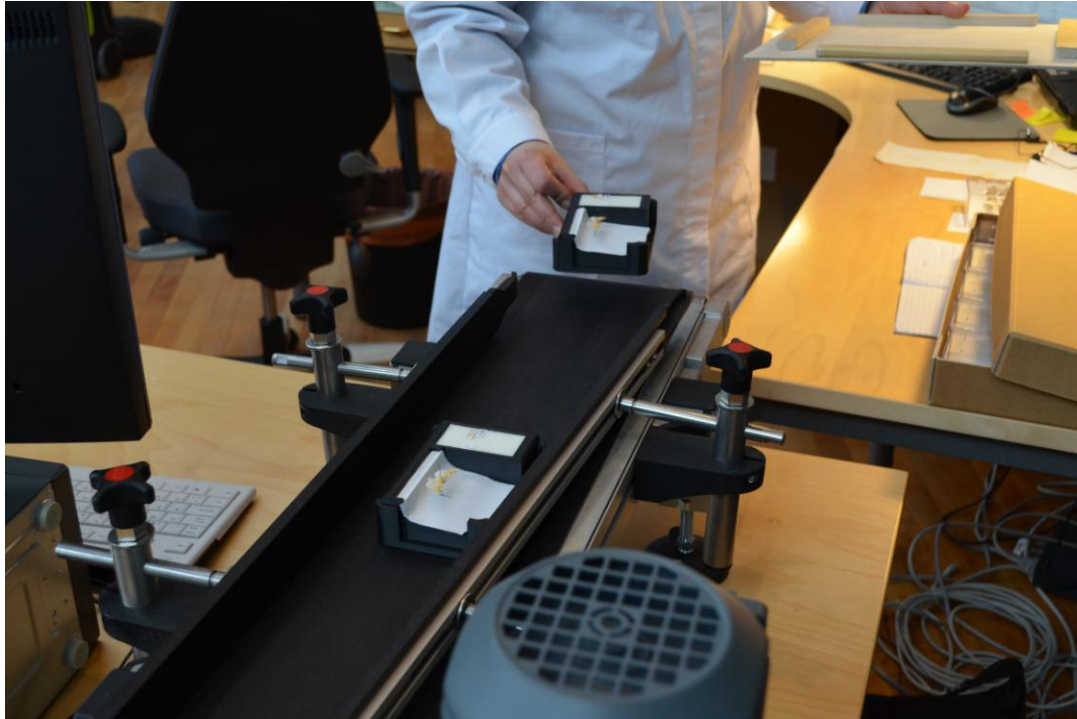


# Pinned specimens and labels are attached to pallets



- Position of a specimen is checked using web camera
- Pallets are made at Digitarium by using a 3D-printer

# Conveyor belts



- Insects on pallets are transferred through conveyor belt system
- On pallet, not yet pinned, is also the ID label (and possible labels from drawers)

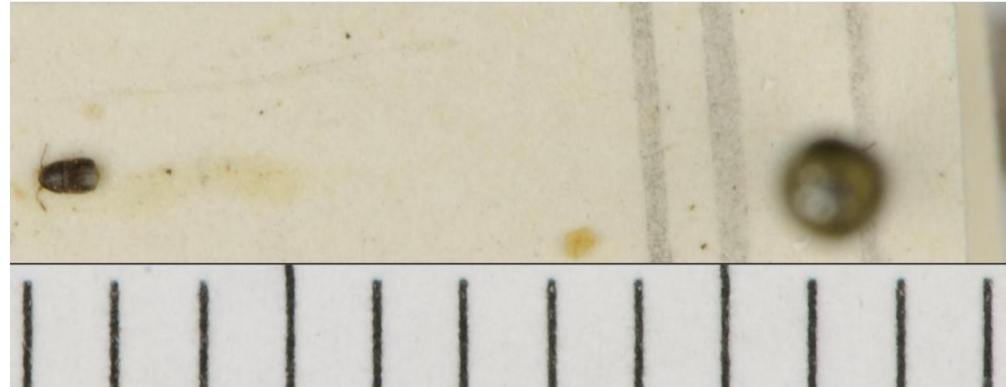
# Automatic imaging

- Transfer of pallets to imaging unit
- Nikon cameras, 12 and 24 Mpix
- Top image: insect
- Side image: insect, labels in the pin and on pallet
- Fixed positioning and focus



# Images

- Imaging of pinned labels is presently possible for many insect groups – but not e.g. for large butterflies (that cover labels)





# Packing

- ID label is pinned with the specimen
- The specimens are packed in taxonomical order and transferred to museum



# Quality check

- Checking of images produced, every day
- Before transporting the collection to museum, final quality check using ISO 2859

<https://www.youtube.com/watch?v=Z2tW1NkxEdU&feature=youtu.be>

## 2. Automation of digitizing activities

---

### Digitarium:

- Has designed and built automatic imaging lines for mass digitization of natural history specimens
- On-going development of tools for speeding up data entry work



# Transportation

---



# Transportation

---



# Task clusters: Inspection

---

- Digitization hall is pest-free
- If any damages are found -> photo and report to the collection owner -> immediate freezing of the specimen



# Freezing

---

- Freezers, freezing containers
- Also post-imaging



# IDs

---

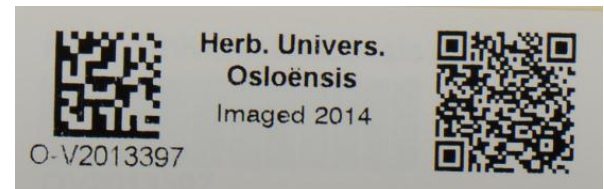
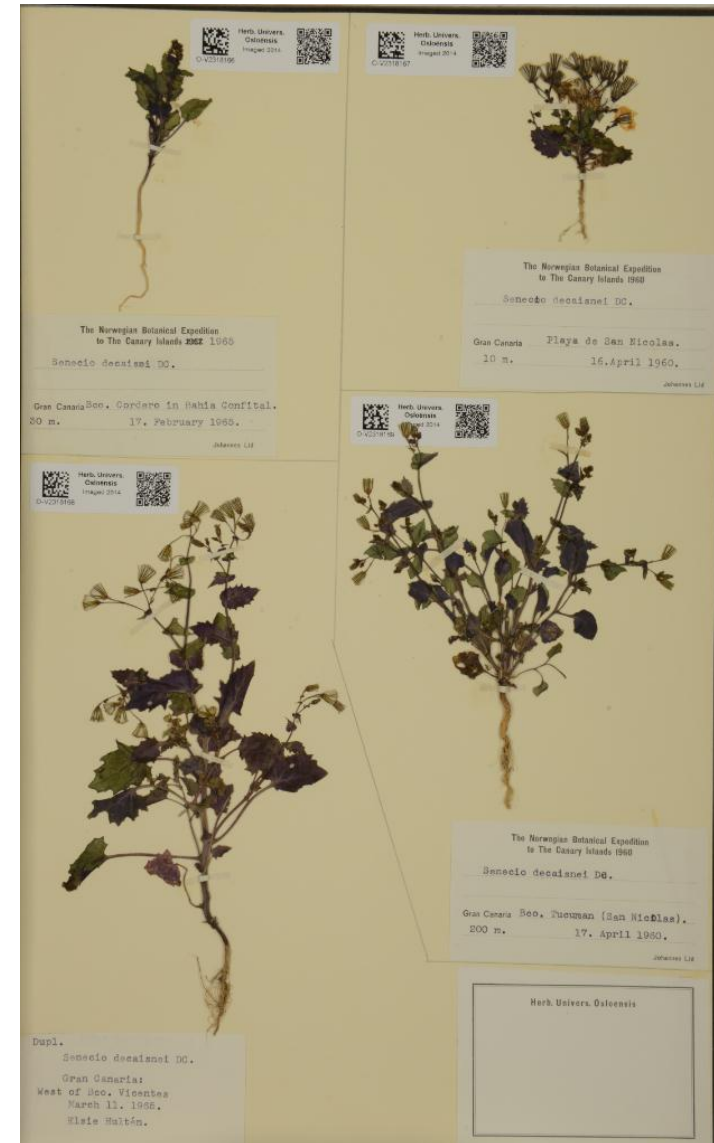
- 2-D barcodes; unique identifier is created according to the collection owner's demands





# IDs

- herbarium ID labels are attached to the sheets / capsules / etc.
- zoological ID labels: information printed on both sides, pinned



# Imaging, automated

---

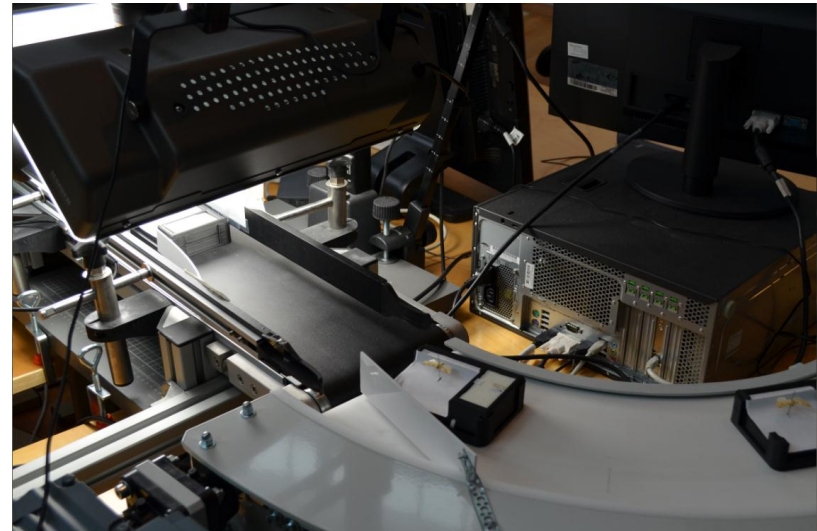
## Imaging lines

- One designed for herbarium sheets; also used for large butterflies, lichens, mosses...
- Smaller one for insects; beetles, small butterflies, two-winged, ...



## Cameras

- Not fixed, depending on demands
- 36 - 86 (108) Mpix (herb)
- 12 - 24 Mpix (insect)



# Imaging, herbarium specimens

- Sheets are carried through conveyor belt system on trays
- Automatic imaging
- Images accessible through DigiWeb minutes after imaging
- Presently 2000 – 3000 specimens / day



# Insect imaging line: Performance test, Coleoptera



- Specimen size 0.5 – 35 mm
- New specimen imaged every 14 s (max)
- Max 250 specimens per hour, 1750 per day
- In practise, process = 4 min, 2 operators:  $\pm 500$  specimens / day / line
- 2 images: top (specimen), side (+labels in pin)
- Metadata: collector's name, taxon, (date)

# Image processing

---

- Quality control: images are checked using ISO 2859
- Barcode capture: IDs in names of directory and xml-files
- File conversion: tif -> jpeg, dzi
- (Image stacking: 3D objects)
- (Image cropping: rarely with insects: dates, animals)
- (OCR: rarely)
- No photo manipulation

# Electronic data capture: DigiWeb

The screenshot displays the DigiWeb interface in a Google Chrome browser window. The page features a navigation menu with 'Main', 'Browse', 'Transcribe', and 'Administration'. The 'Transcribe' tab is active, showing a specimen record for 'Friaria ventricosa Mart.' collected by Otto Buchtien in 1907. The record includes a scanned image of the specimen label, a 'Re-finalization' form with fields for Year, Month, Day, and Collector's record number, and a 'Tools' section with buttons for 'Collector lookup', 'Georeference', and 'Transform'. A 'Save to state' section is also visible. On the right, a 'Tools' panel shows a list of reserved samples (20 total), with the current specimen (O-V2100008) highlighted.

**Specimen Data:**

Term	Data
Year	1907
Month	09
Day	11
Collected by	Buchtien, Otto
Collector's record number	

**Reserved samples (20):**

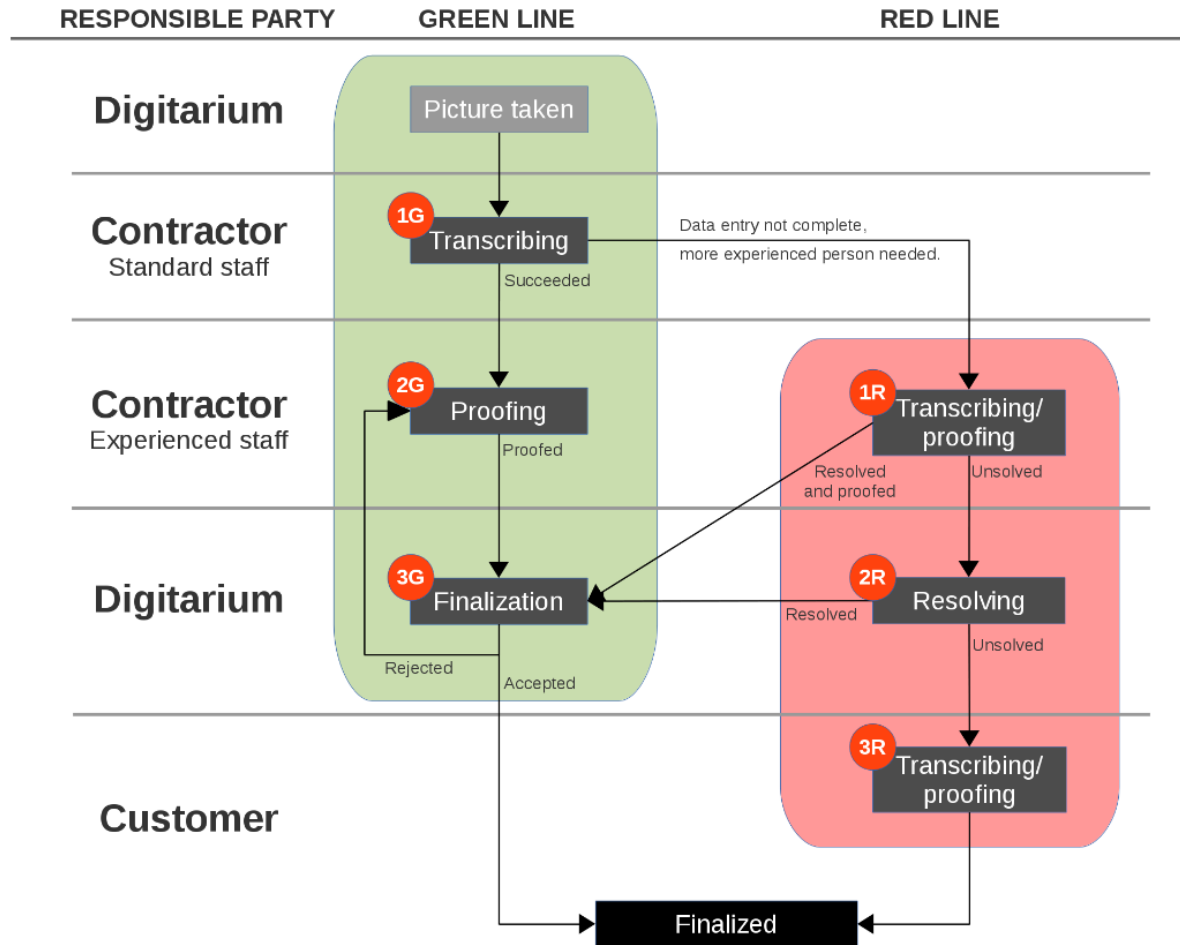
ID	Group
O-V2100020	UiO Samples 1
O-V2100017	UiO Samples 1
O-V2100016	UiO Samples 1
O-V2100015	UiO Samples 1
O-V2100014	UiO Samples 1
O-V2100013	UiO Samples 1
O-V2100012	UiO Samples 1
O-V2100011	UiO Samples 1
O-V2100010	UiO Samples 1
O-V2100009	UiO Samples 1
O-V2100008	UiO Samples 1
O-V2100007	UiO Samples 1
O-V2100006	UiO Samples 1
O-V2100004	UiO Samples 1
O-V2100003	UiO Samples 1
O-V2100002	UiO Samples 1
O-V2100001	UiO Samples 1
O-V2100005	UiO Samples 1

# External look-ups: better data quality, faster data entry

---

- Auto-complete input of taxonomic name (The Plant List) and hierarchical geographical area (e.g. write Florida: automatically Florida (StateProvince), United States (country), North and Central America (continent))
- Look-up function of collector's visited locations (earlier entries) – useful if hand writing is poor
- Shows errors made compared to standard: date (ISO) and person's name (DwC)

# Data capture, workflow example





# Data

---

- Possible: Data entry in phases; first data entry of names of collectors, then later the rest of data
- Possible: Groups of files based on level of difficulty: e.g. printed and typed vs. hand written
- Text files: xml, json
- Quality: must pass ISO 2859

# Georeferencing

---

- Not part of every day workflow
- Geolocate
- Automatic conversion of coordinates to WGS84
- Standardisation of coordinates of repeated locations, automatic data entry