Mass Digitisation of Individual Pinned Insects Using Conveyor-Driven Imaging

Riitta Tegelberg, Jere Kahanpää, Janne Karppinen, Tero Mononen, Zhengzhe Wu, Hannu Saarenmaa

Leverage from the EU 2007–2013

LUOMUS
LUONNONTIETEELLINEN KESKUSMUSEO
NATURHISTORISKA CENTRALMUSEET
FINNISH MUSEUM OF NATURAL HISTORY

Digitarium

European Union
European Regional Development Fund
European Social Fund

FIRI
FINNISH RESEARCH INFRASTRUCTURES

UNIVERSITY OF EASTERN FINLAND

Digitarium

Digitisation Centre of the Finnish Museum of Natural History (LUOMUS) and the University of Eastern Finland

http://digitarium.fi/en

- Know-how centre
- **Design, building and use of mass-digitisation equipment**
- Create and "tune" e-services
- Open science

Activities

On-going: Academy of Finland, Finnish Research Infrastructures FIRI *"FinBIF"*

- Building of image data services on IDA and EUDAT for https://laji.fi/en

Future activities

Starting funding 01-2018: H2020-INFRADEV "*ICEDIG*, Innovation and Consolidation for large-scale Digitisation of Natural Heritage"

- (-> ESFRI roadmap application DiSSCo: http://dissco.eu/)

Why mass digitisation?

Because

- On-going formation of e-repositories of collections (museum, national and European level…)
  - Scientists, collection managers, students, authorities, public
  - Data, images
- Urgent scientific needs for ”big” data: biodiversity, environmental changes …
- Requirements of open access

Reality

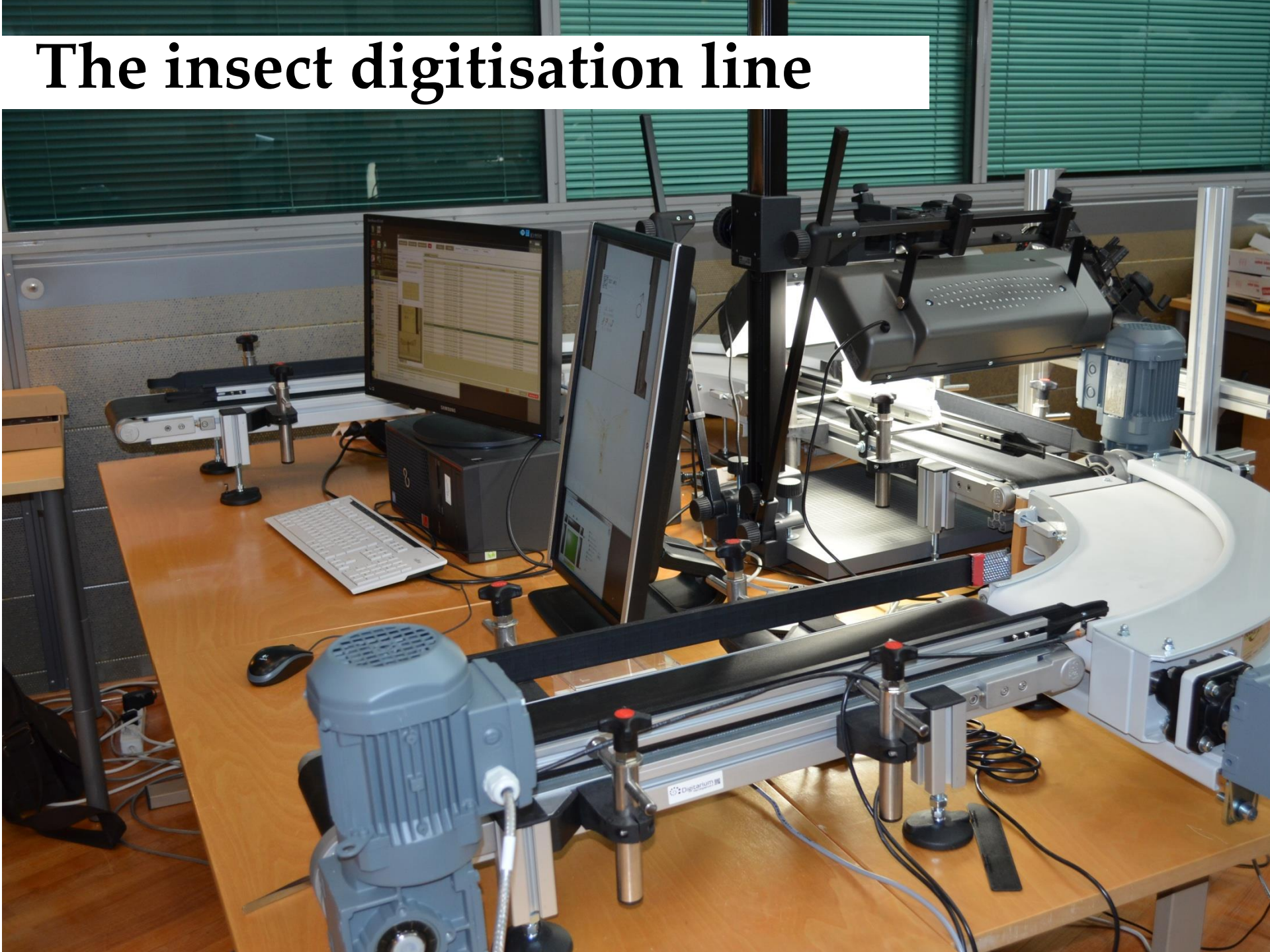
- In Finland, natural history collections store **25** million specimens
  - **15 – 18** million are insect specimens
  - Majority not in e-systems
- Original idea was to digitise new collections arriving the museums using digitisation line
- Now, also in-house collections are being digitised with a line at LUOMUS

Senior museum technician Jere Kahanpää working in the new insects collection vault, where millions of insect specimens are stored (Photo: Pekka Malinen/Luomus)

Conveyor belt driven automatic digitisation lines

- First, system for herbarium sheets
- Later, smaller system for pinned insects

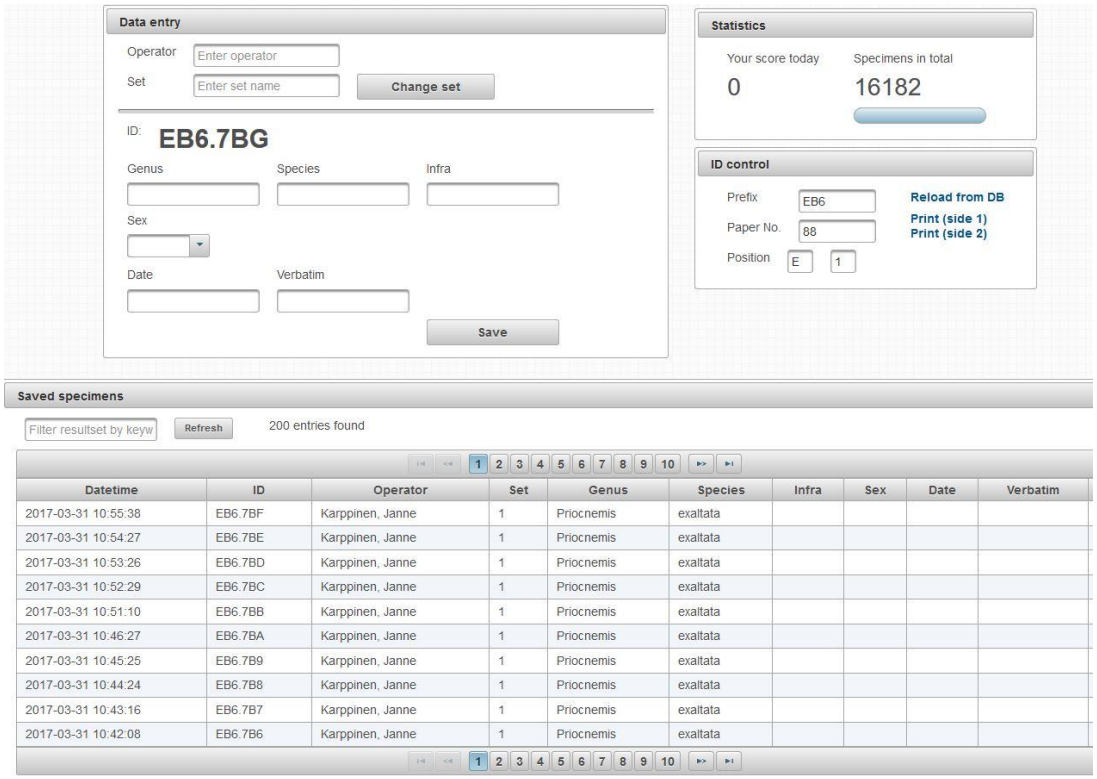
The insect digitisation line



Process: Preparation

- Unique code label
  - 18 x 7 mm <

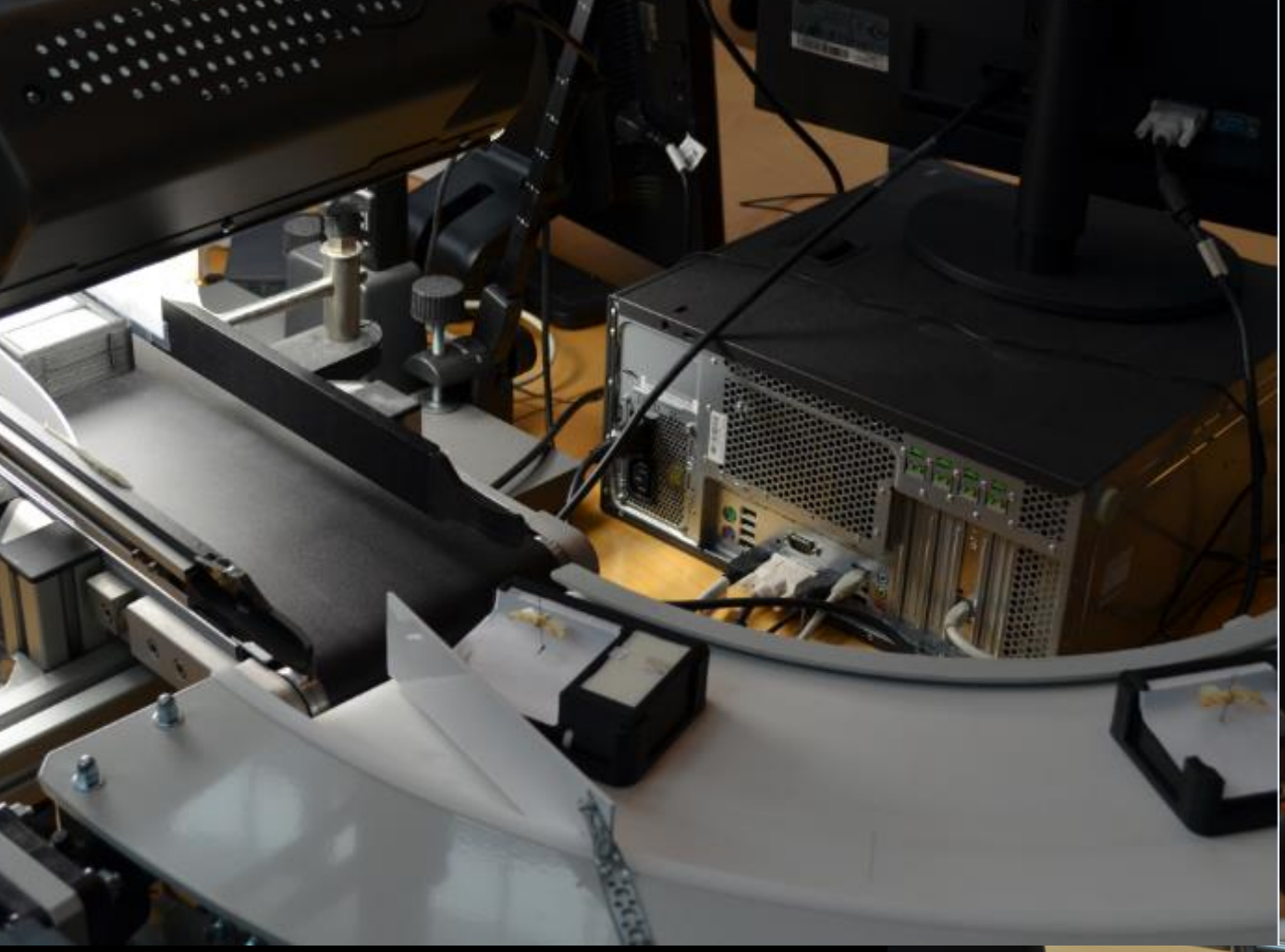
- Transcription



Datetime	ID	Operator	Set	Genus	Species	Infra	Sex	Date	Verbatim
2017-03-31 10:55:38	EB6.7BF	Karppinen, Janne	1	Priocnemis	exaltata				
2017-03-31 10:54:27	EB6.7BE	Karppinen, Janne	1	Priocnemis	exaltata				
2017-03-31 10:53:26	EB6.7BD	Karppinen, Janne	1	Priocnemis	exaltata				
2017-03-31 10:52:29	EB6.7BC	Karppinen, Janne	1	Priocnemis	exaltata				
2017-03-31 10:51:10	EB6.7BB	Karppinen, Janne	1	Priocnemis	exaltata				
2017-03-31 10:46:27	EB6.7BA	Karppinen, Janne	1	Priocnemis	exaltata				
2017-03-31 10:45:25	EB6.7B9	Karppinen, Janne	1	Priocnemis	exaltata				
2017-03-31 10:44:24	EB6.7B8	Karppinen, Janne	1	Priocnemis	exaltata				
2017-03-31 10:43:16	EB6.7B7	Karppinen, Janne	1	Priocnemis	exaltata				
2017-03-31 10:42:08	EB6.7B6	Karppinen, Janne	1	Priocnemis	exaltata				

Positioning of the specimens by using web camera

Custom pallets carry the specimens



Operator lays the pallets on and takes them from the belts. Rest is automatic

End products

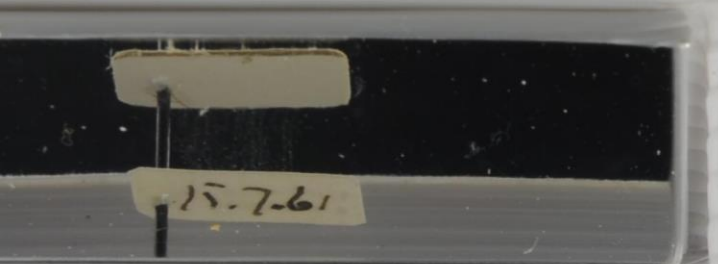
- Specimen with unique identifier
- Images, tif
- Metadata
- Specimen data, xml

Organism groups benchmarked

- **Coleoptera (beetles)**
- **Lepidoptera (butterflies and moths)**
- Hymenoptera (sawflies, wasps, bees, ants)
- Diptera (true flies: horse-flies, hoverflies, crane flies …)

Salpingus Gyll.

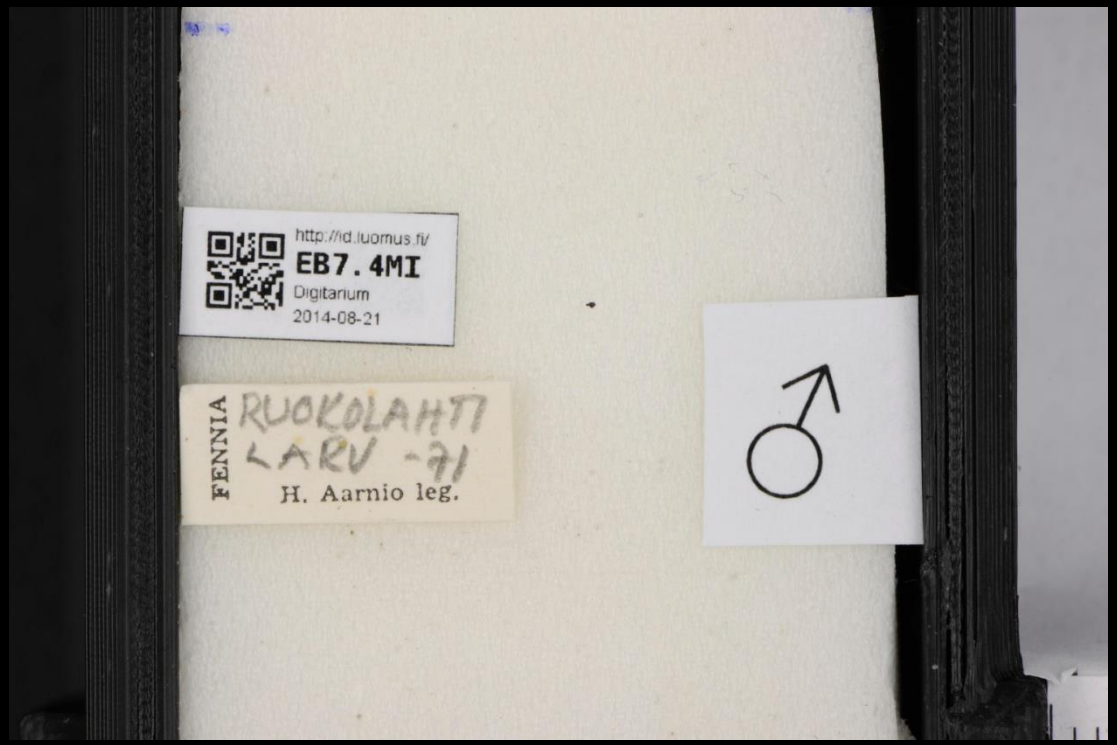
ater Gyll.



Coleoptera benchmark

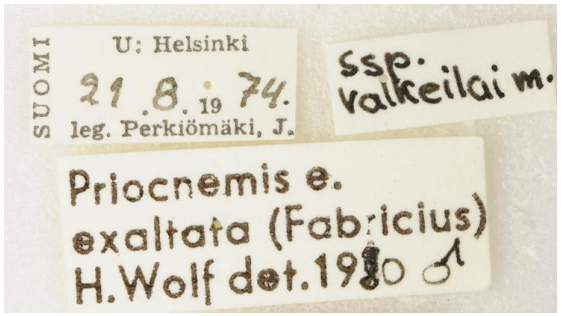
- The entire collection of Gunnar Blomqvist, 14,000 specimens, processed in 28 working days by two people
- Two images: lateral and side (labels)
- Labels stay intact
- Transcription of species name
- 80 /hour = 500 /day = 100,000 /year
- Basic rate in the range of 50,000 specimens /year /person

- Field notebooks of Blomqvist have also been digitised in a different project: Linking through the field number


http://id.luomus.fi/
EB7.4MI
Digitarium
2014-08-21
FENNIA
RUOKOLAHTI
LARV -71
H. Aarnio leg.

Lepidoptera benchmarks

- Macrolepidoptera 70 specimens /hour maximum speed by one operator. Transcription of species name.
- Microlepidoptera 57 specimens /hour, sustained speed over two months by two operators. Maximum speed 120 specimens /hour. Transcription of all specimen data from labels
- Labels removed for imaging
- Two lateral pictures (of specimen and labels)
- Macrolepidoptera: Basic rate in the range of 70,000 specimens /year /person
- Microlepidoptera: Basic rate in the range of 40,000 specimens /year /person (includes transcription)


SUOMI
U: Helsinki
21.8.19 74.
leg. Perkiömäki, J.
ssp. valkeilai m.
Priocnemis e.
exaltata (Fabricius)
H.Wolf det.19 80 ♂


Winter
Fennia
Terijoki
24.8.1938


Vespula
vulgaris (L.)
A.Pekkarinen det.

Bottlenecks

- Transports, packaging etc about 5% of project cost: Not a bottleneck
- Handling of specimens: 250-500 /day/person
  - Depends on organism group and need from removal of labels
  - More experience and fine-tuned workflows are needed
- Data transfers
  - Camera to workstation: Not a bottleneck in USB3. 6 seconds/image.
  - Workstation to image repository
- Transcription
  - Can be done later, using images

Results

- By using two lines for two years (one line full time), by 2-5 persons, we have digitised

150,000 insect specimens

- Cost of digitisation (at Digitarium) /specimen (full cost model with 60% flat overhead rate; service)

3.02 €

Conclusions

- Digitisation lines are small-sized, and can be part of every-day actions
- Given the huge size of collections, in-house systems make sense. Small/incoming collections can be digitised as an outsourced service

- Recently, IPR was sold to Sertifer Consulting Co., which is now responsible for the delivery of new lines to customers and service-based digitisation (http://www.sertifer.fi/eng/digitointipalvelut.html)