

# Lecture 8

## Digital biological collections

PBIO 161 Biological collections

Jere Kahanpää & Kari Lahti (Luomus)

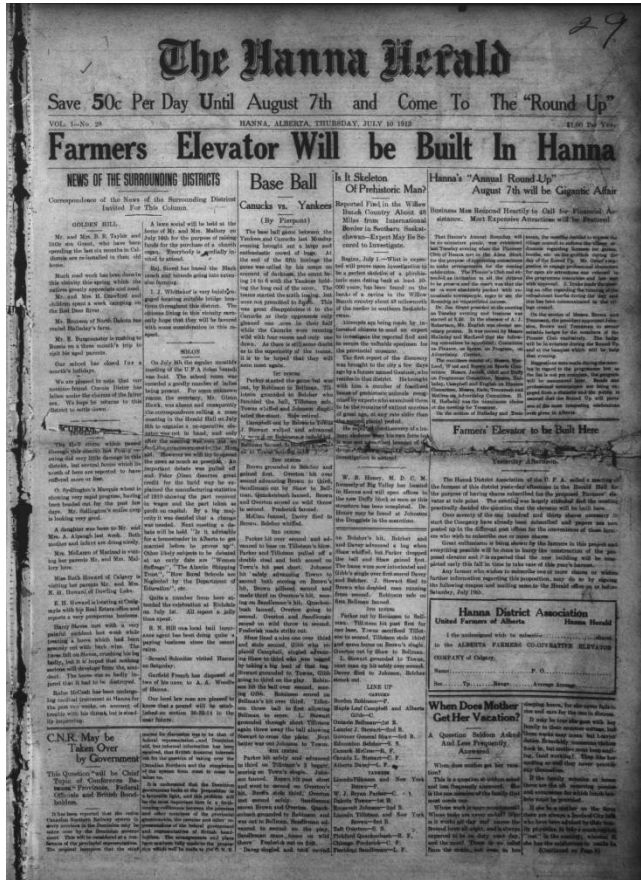


# Basic plan of the lecture

- Creating digital data from biological collections (*Jere*)
  - motivation
  - methods
- Acquiring specimen data (*Jere*)
- Distributing data (*Kari*)
  - Finnish Biodiversity Information Facility
  - Legal issues: licenses, restrictions (Nagoya Protocol etc)

# Definitions

- Digital data
  - In practice data converted to a binary, computer-accessible format
  - Data != meaningful content
- Digitization (*digitointi*)
  - Converting into digital data
  - Often, but not always, means imaging (or otherwise recording) a specimen in our context
- Digitalization (*digitalisaatio*)
  - Increasing digitization & use of digital data



# Why do we digitize natural history collections?

- research infrastructure
  - access to data
  - analysis
  - backup/security
- public relations

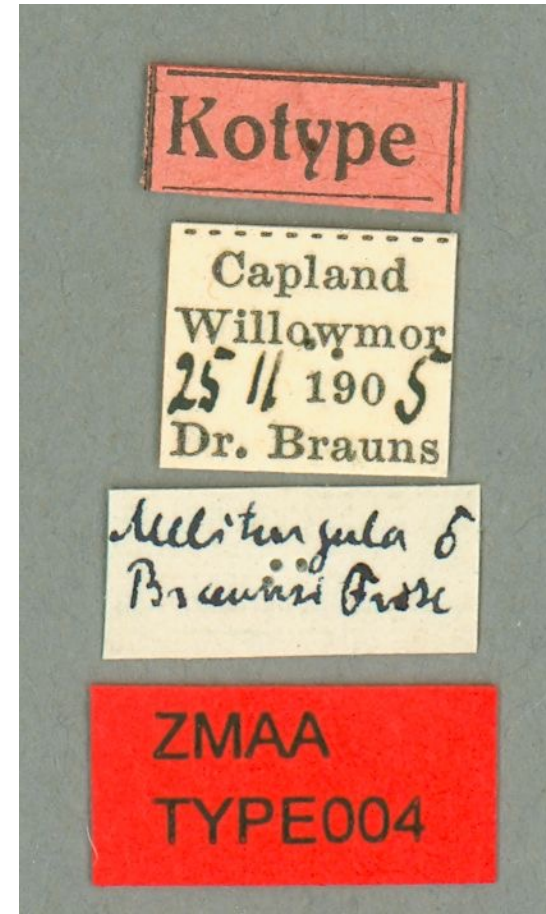


(By) Felipe Milanez CC-BY-SA 4.0

# Physical specimens

<http://mus.utu.fi/ZMAA.TYPE004>

cotype *Meliturgula braunsi* Friese,  
1903



by U. of Turku

# http://mus.utu.fi/ZMAA.TYPE004



**Collection:** [Type specimen collections of Åbo Akademi](#)

**Datasets:** [Coll. Forsius](#)

**Notes:** Collected after the description by the type collector from the type location

**Specimen location:** Åbo Akademi/Coll. Forsius/XVIII.7. APIDAE

Collecting event (gathering)

**Leg:** Brauns, J.

**Start date:** 25.11.1905

**Verbatim leg from original source:** Dr. Brauns

**Lat (N):** -33.287796

**Lon (E):** 23.489593

**Coordinate** WGS84 and ETRS89, decimal

**Created by** Hopkins, Tapani on 2017-12-19

**Edited by** Teräs, Anssi on 2019-03-05

**Current owner is** Zoological Museum, Biodiversity Unit, University of Turku

PE004\_W ing.jpg PE004\_S de.jpg PE004\_T op.jpg PE004\_H ead.jpg



ZMAA.TY  
PE004\_L  
abel.jpg



Field notes, manuscripts  
other papers

Meade River P.O., July 3, 1963

11

63-90 *Desmatrodon leucostoma*

On soil, vertical wall of deep  
ravine in <sup>high</sup> cut-bank bluff, above  
the Meade River

63-91 *Ditrichum*

with the preceding

63-92 *Pohlia cuneoides* (?)

with the preceding

discarded

63-93 *Prorissa quadrata*

with the preceding

63-94 *Prorissa quadrata* ?

# Experimental data

- Sonograms
- Chromosome slide preparations



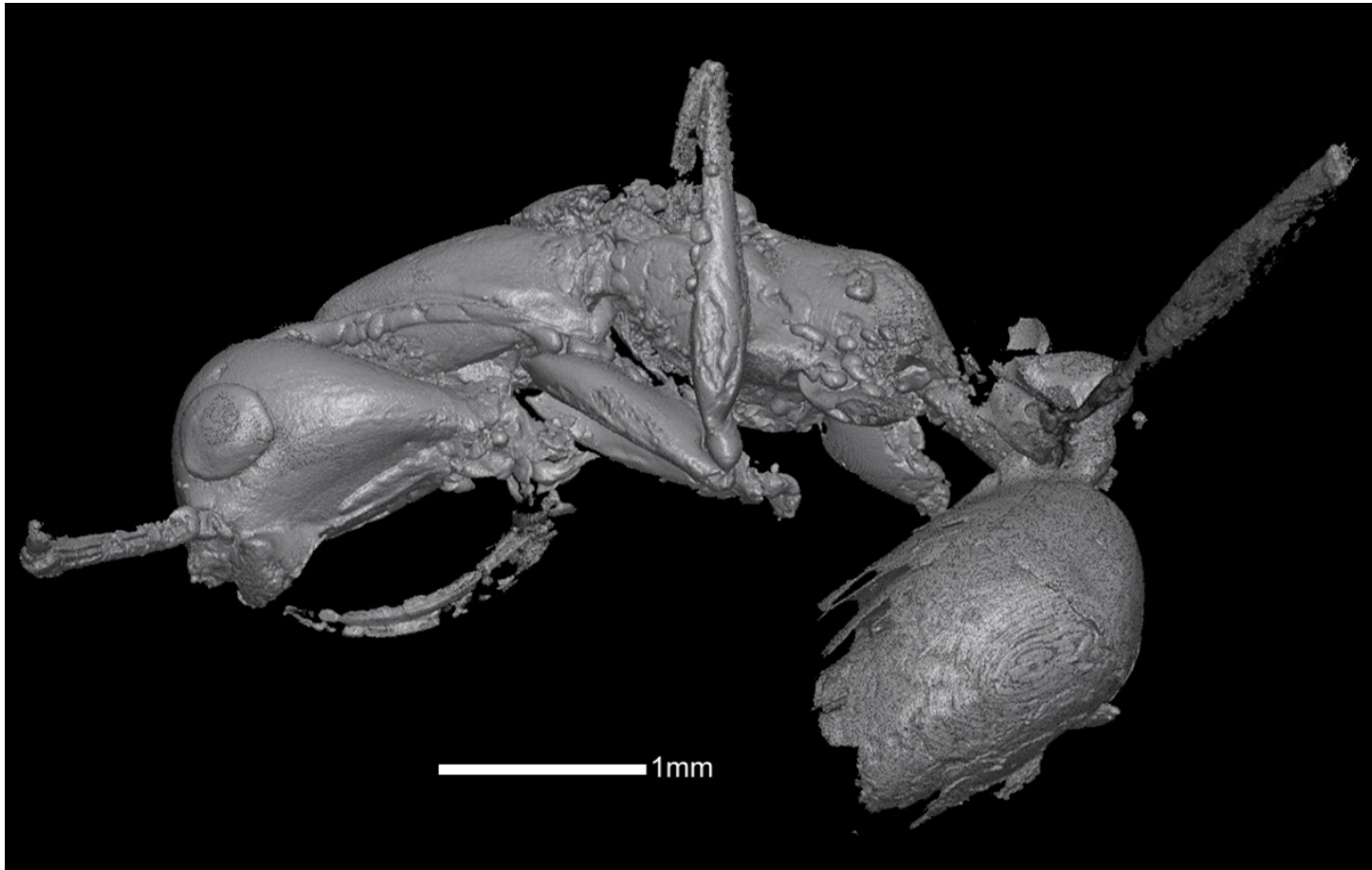
(By) Josef Reischig CC-BY-SA 3.0



# Other items digitized in natural history collections

- Chemical data: DNA
- Collection-level data
- Collection maintenance: loans, donations etc.
- Metadata
- Observations
- Literature

# Methods and tools



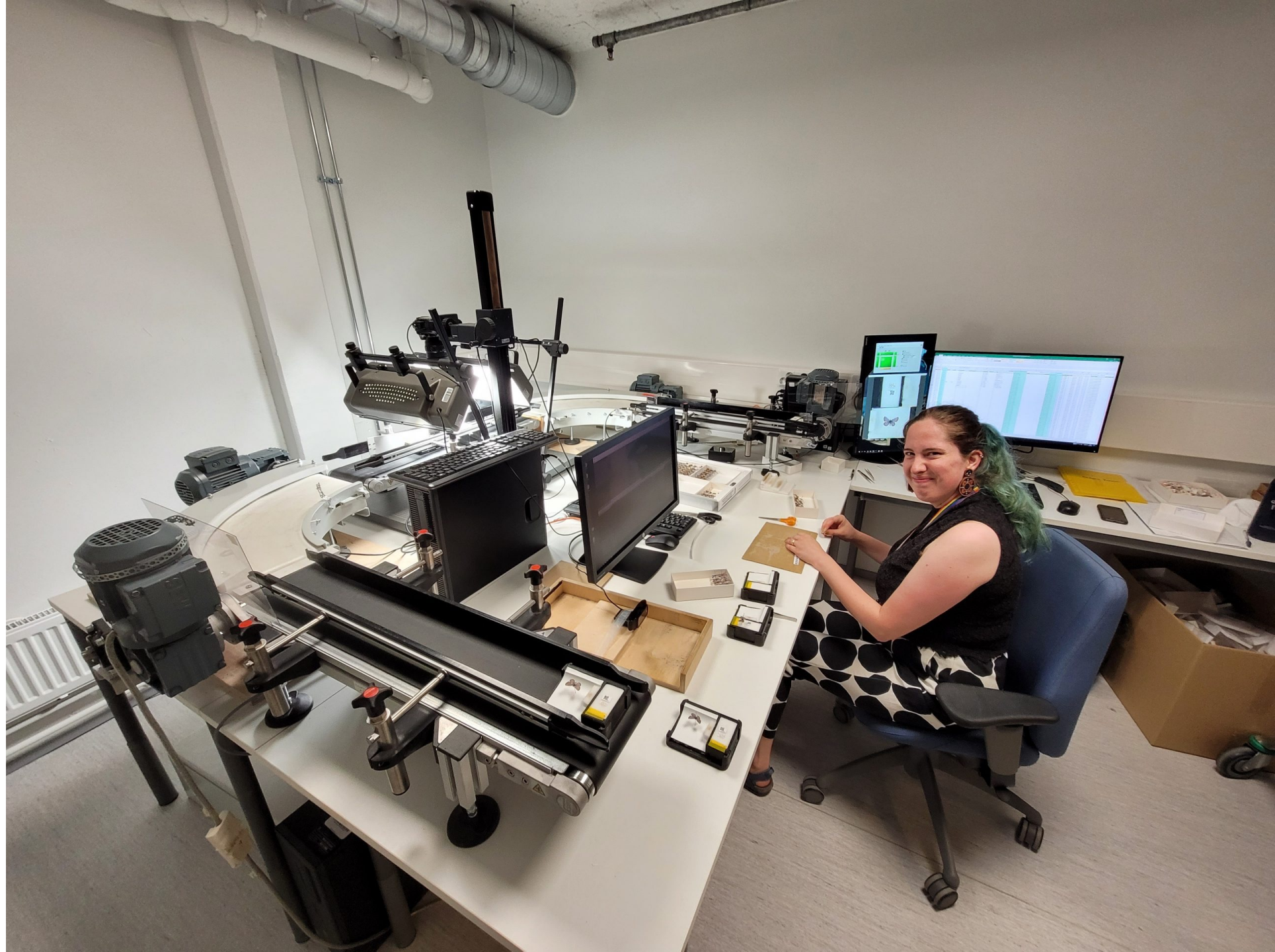
Profile computed tomographic scan of the *Haidomyrmex scimitarus* holotype. specimen AMNH-BUFB80 by Phillip Barden, [Creative Commons Attribution 4.0 International](#)

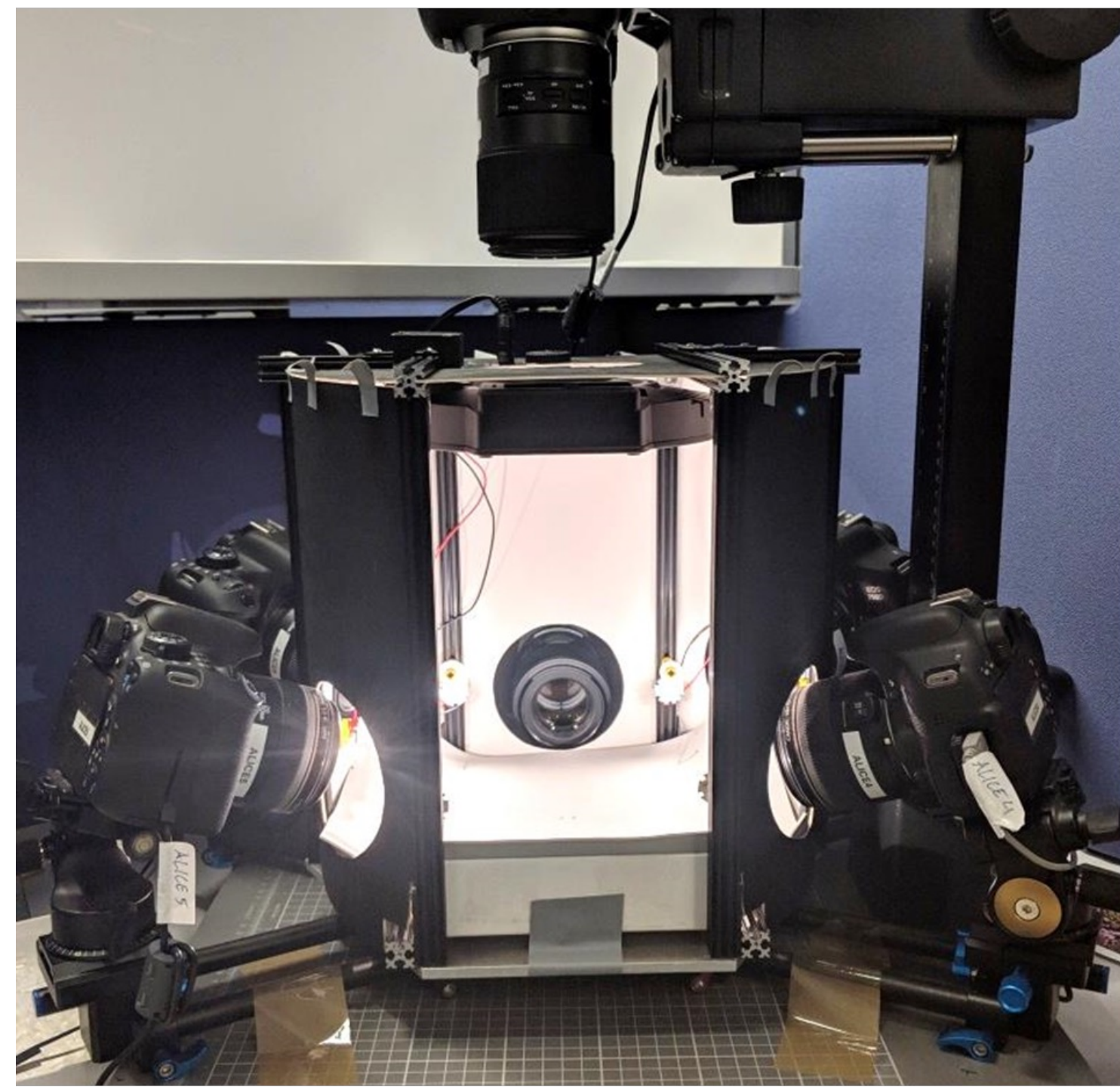
# Imaging: 2D

- Fast
- Cheap
- Can handle many types of material
- Limits on quality

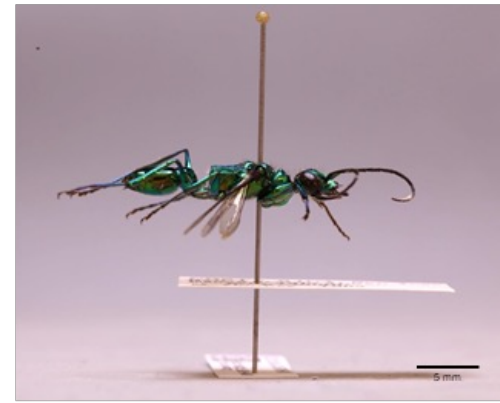
<http://plants.jstor.org/stable/10.5555/al.ap.visual.ma-ajb04-d-0812>



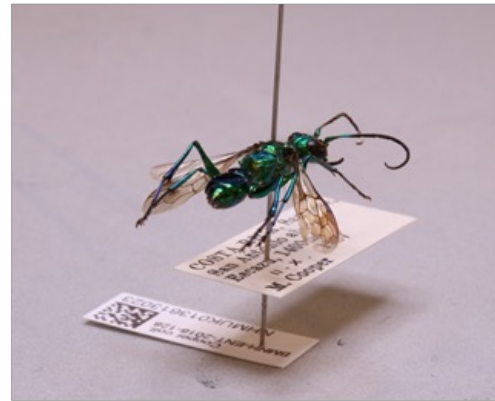




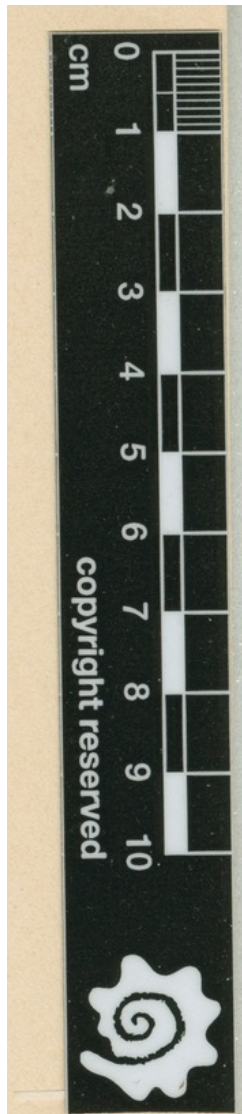
## Specimen Images



## Label Images



# Calibration data



# Imaging: 2D

- Coll. Seppo Karhula aquired & documented @ Luomus

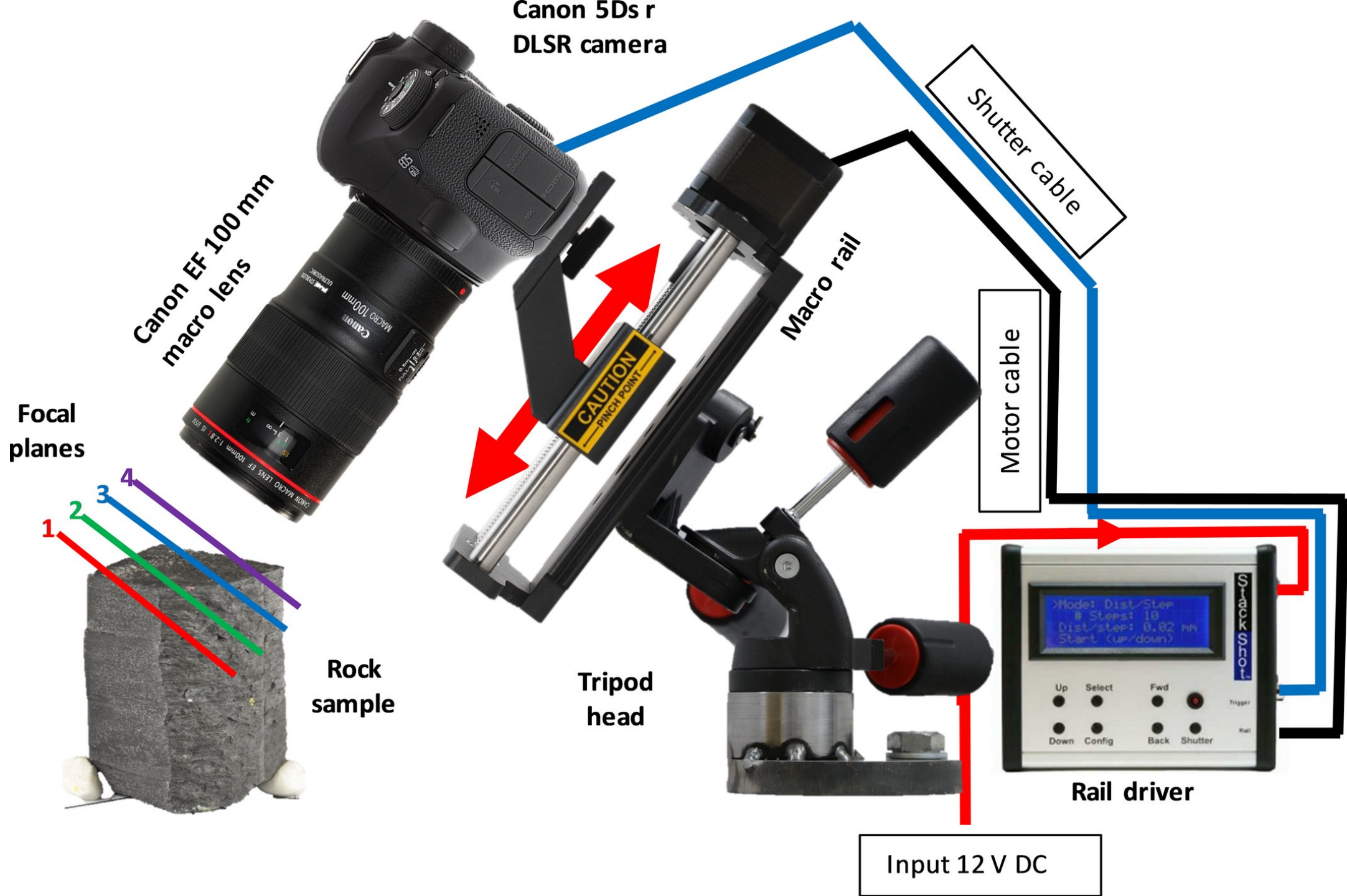


# Imaging: 2D stacking

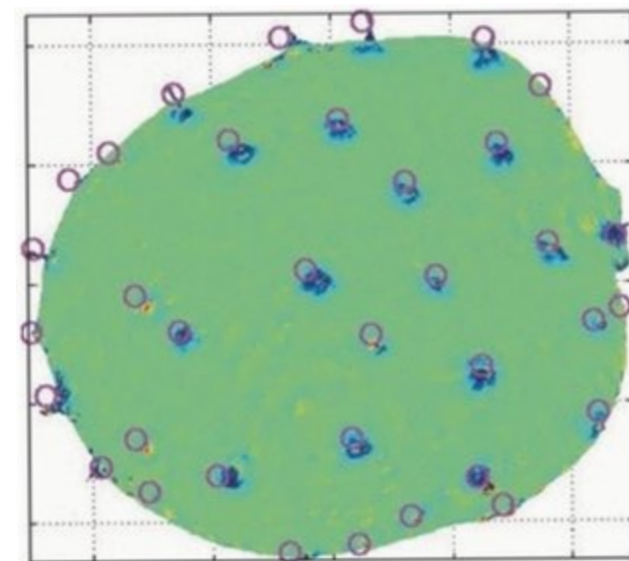
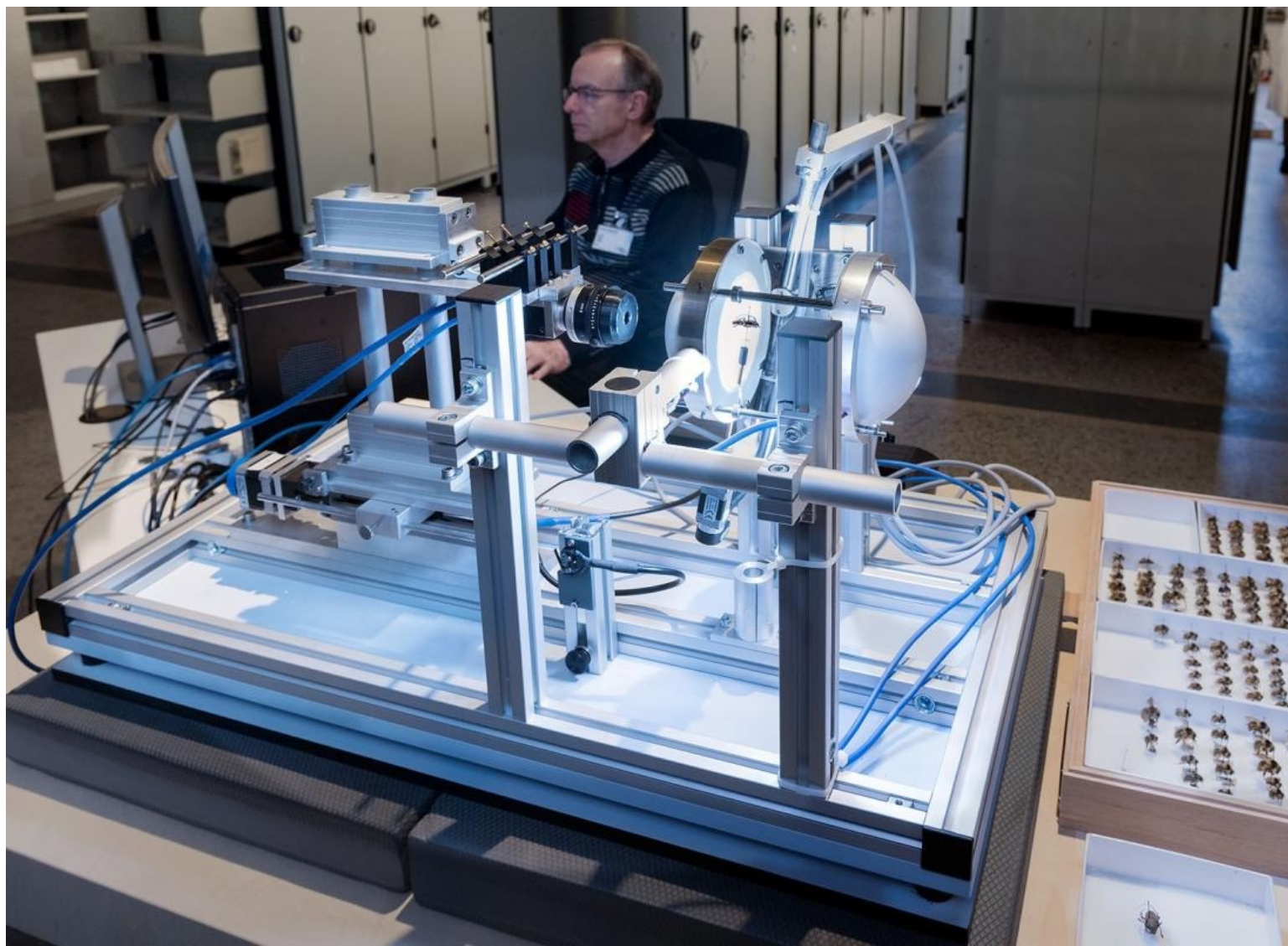


(By) Muhammad Mahdi Karim CC-BY-SA 3.0

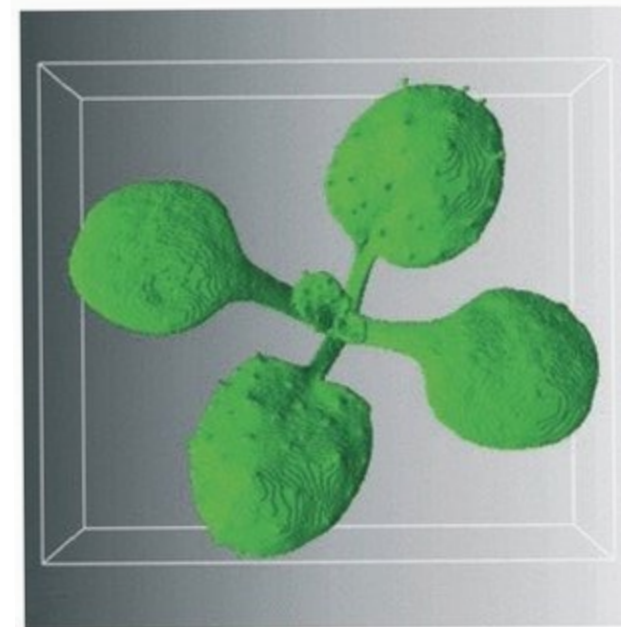




# Imaging: 3D surface



B



A

# Imaging: 3D internal (microCT)



**Verbatim leg from original source:** Dr. Brauns

**Lat (N):** -33.287796

**Lon (E):** 23.489593

**Coordinate system:** WGS84 and ETRS89, decimal degrees

**Verbatim locality from label or other original source:** Capland Willowmor

**Country:** South Africa

**Biogeographical province:** The Cape Provinces

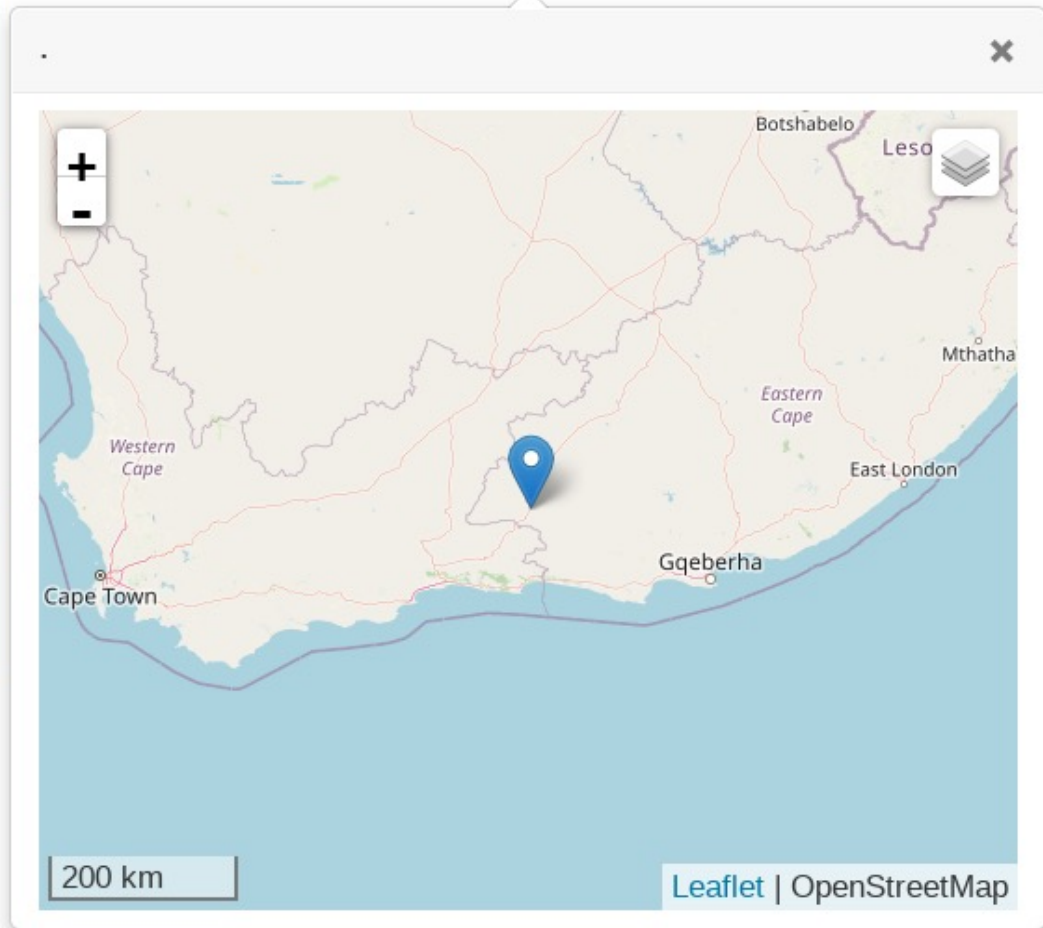
**Administrative province:** Eastern Cape

**Verbatim date from original source:** 25 11 1905

**Municipality:** Sarah Baartman District Municipality

**Locality names:** Willowmore

el.jpg



<http://id.luomus.fi/HA.H0769110>

*Orchis curvifolia* F. Nyl. 1844

MUSEUM BOTANICUM  
UNIV. HELSINGIENSIS

769110

QUADR.

CATAL. 19 <sup>2009</sup> 95

*Orchis curvifolia* Nyl.  
Kon.  
Tivdja, Gub. Novotuz,  
1843. F. N.

(Rchb. fil.) revisio  
E, a. 1895.

*Orchis curvifolia* (Nyl.)

REVISIO DACTYLORCHIDUM

P. Vermeulen - Amsterdam 1939

# How much data is enough data?

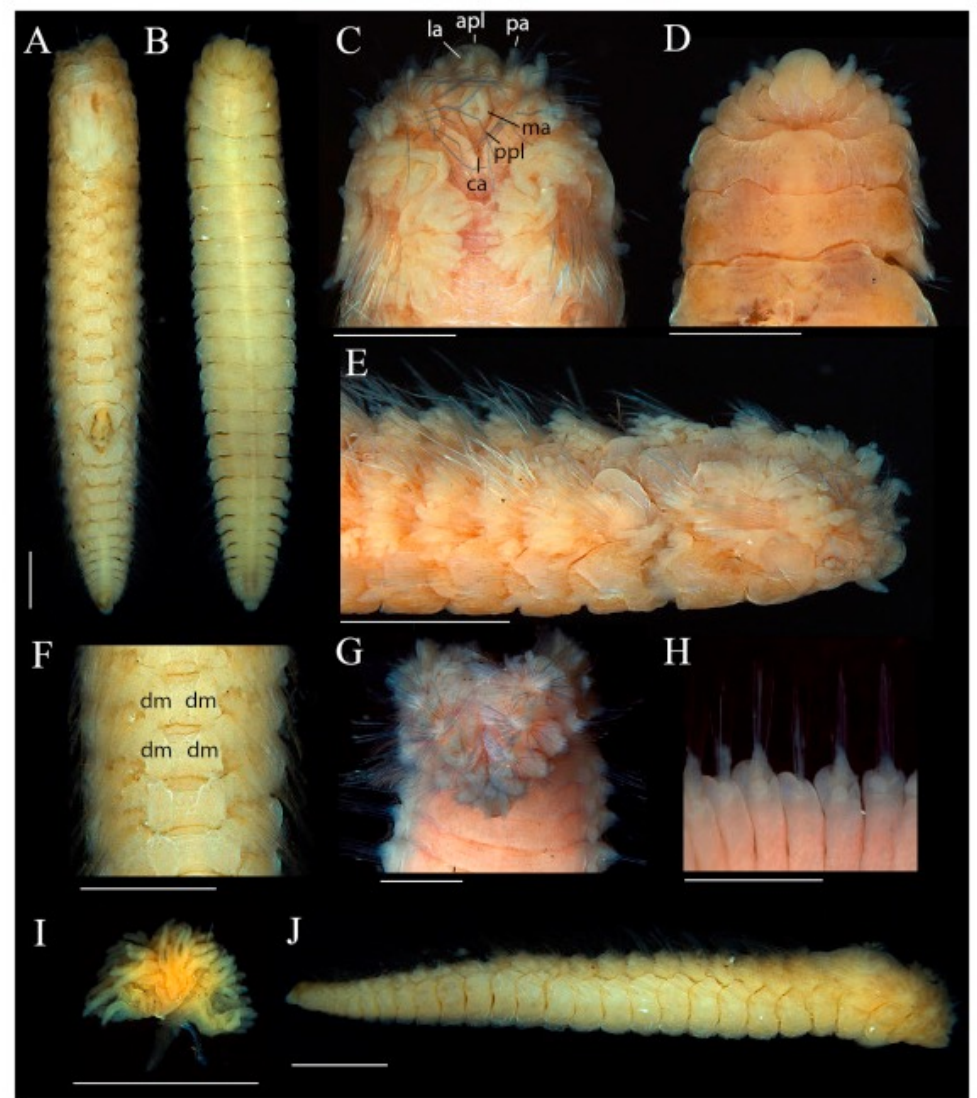
- Minimum Information about a Digital Specimen MIDS

<b>MIDS level</b>	<b>Record extent</b>	<b>Purpose</b>
1	Basic	A basic record of specimen information.
2	Regular	Key information fields that have been agreed over time as essential for most scientific purposes.
3	Extended	Other data present or information known about the specimen, including links to third-party sources.
<i>0</i> <i>(Note)</i>	<i>Bare</i>	<i>A bare or skeletal record making the association between an identifier of a physical specimen and its digital representation, allowing for unambiguous attachment of all other information.</i>

Book-keeping: keeping the digital and the physical  
together

# Keeping track of specimens & digital data

- Which specimen was used for the illustrations in a species description?
- What specimens did Dr. Krivosheina examine in 2004 at Luomus?



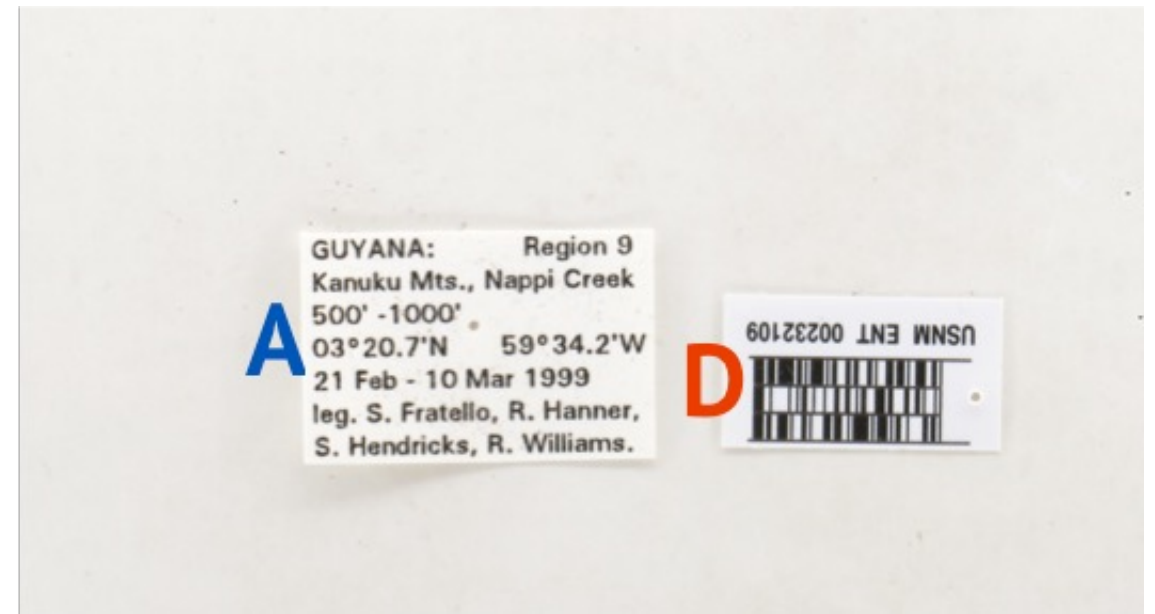
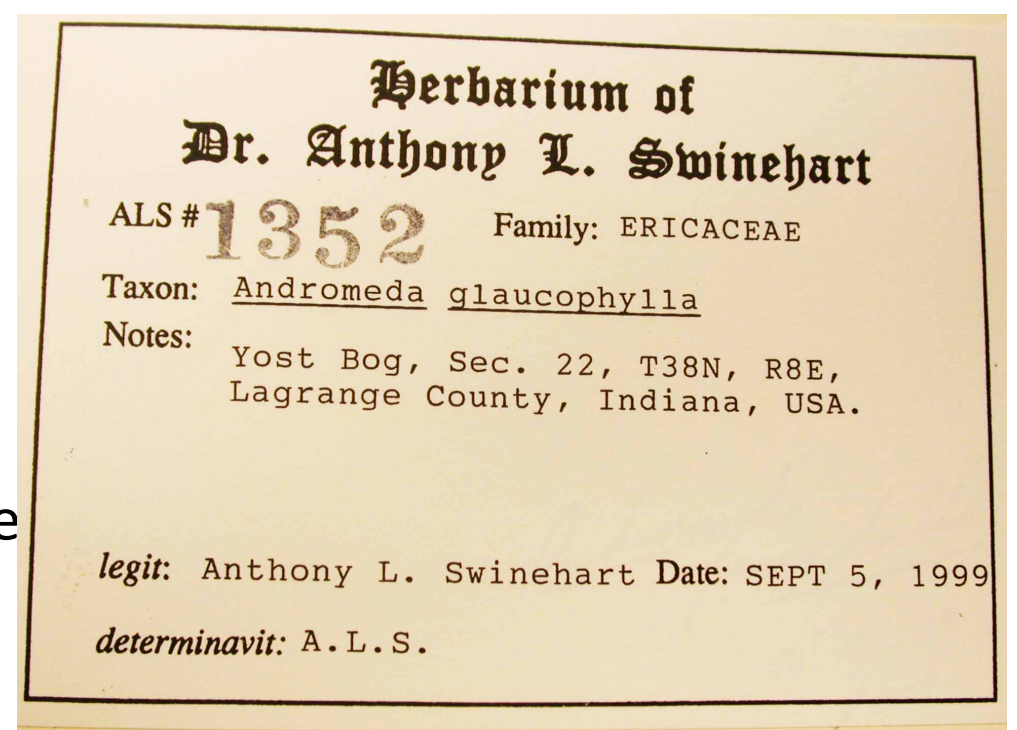
[Download high-res image \(2MB\)](#) [Download full-size image](#)

Fig. 3. *Cryptonome barbada* sp. nov. Stereoscope images. A. entire specimen, dorsal view; B. entire specimen, ventral view; C. anterior region, dorsal view; D. anterior region, ventral view; E. anterior region, lateral view; F. mid-body region, dorsal view; G. anterior region, dorsal view; H. Left side of mid-body region, ventral view; I. Branchia; J. entire specimen, lateral view. Scale bar = 1 mm. apl. anterior prostomial lobe; ca. caruncle; dm. dorsal mound; la. lateral antenna; ma. median antenna; pa. **palp**; ppl. posterior prostomial lobe.



# Unique identifiers

- Solves several problems:
  - Referring to a particular sample in text (e
  - Linking related data
- Stable = unique & understandable



# Stable identifiers

- Still in flux globally
  - See DOI 10.1371/journal.pbio.2001414 : McMurry: "Identifiers for the 21st century: How to design, provision, and reuse persistent identifiers to maximize utility and impact of life science data"
- CETAF Stable Identifiers
  - URIs: Look like web addresses: <http://id.luomus.fi/GV.45118>
- Alternatives
  - LSID - Life Science Identifiers
    - urn:lsid:ncbi.nlm.nih.gov:pubmed:12571434
  - DOI (Digital Object Identifier)
    - 10.1145/2783446.2783605

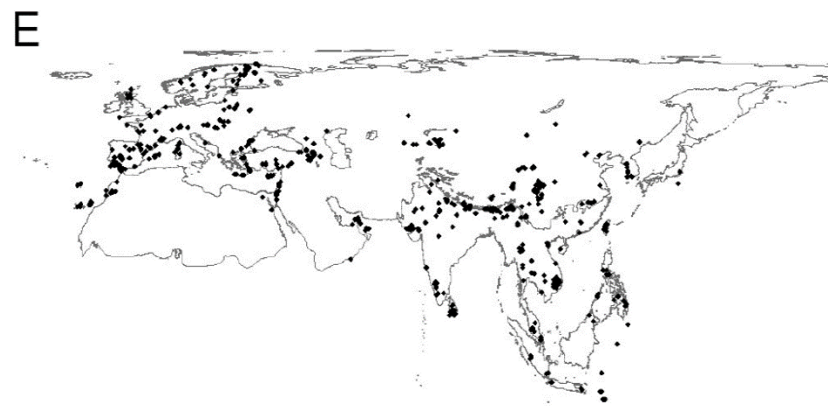
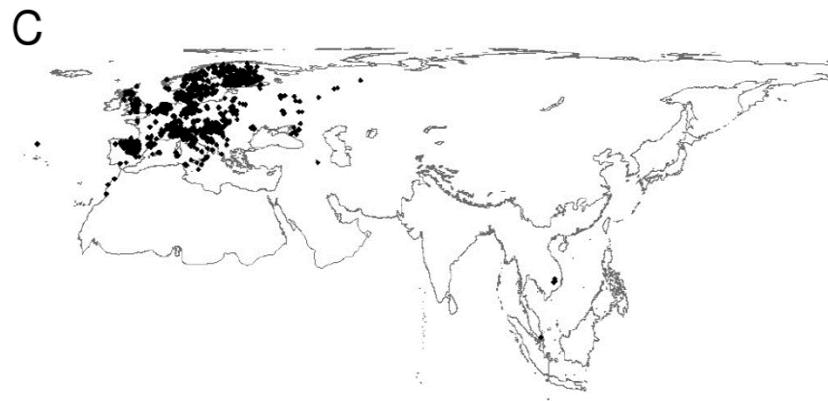
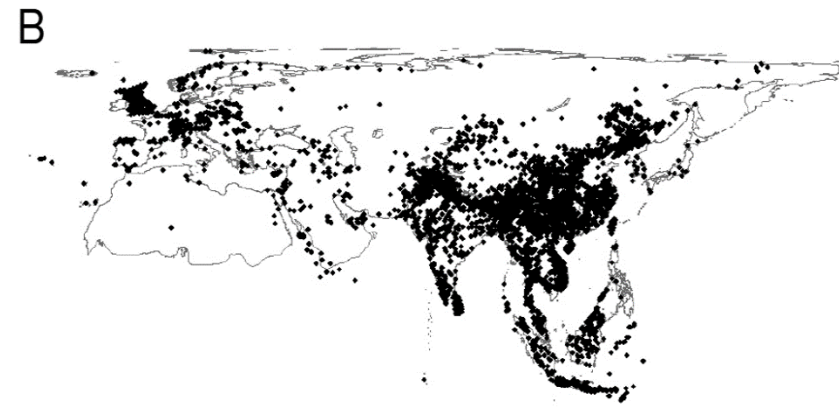
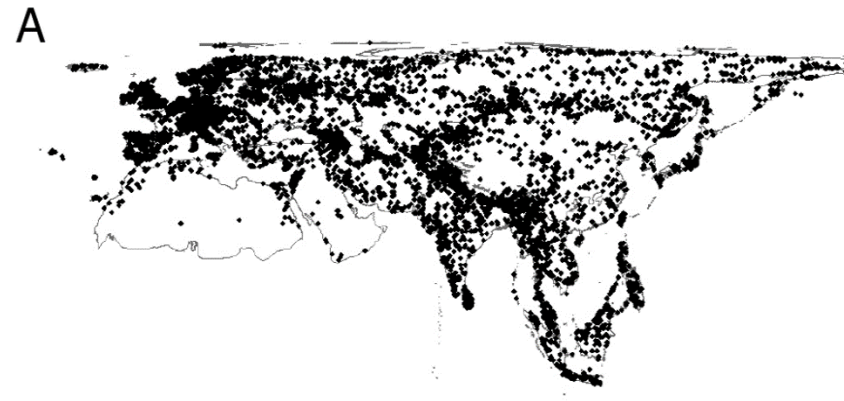
*The one thing to remember from this part!*

Make your work – in collections and  
elsewhere – reproducible:  
use stable identifiers

# Basic plan of the lecture

- Part 2: Finding & acquiring collection specimen data
  - Caveats concerning specimen data
  - GBIF in more detail
  - (Data formats)

The spatial distribution of records from different sources. A) museums, B) literature, C) ringing, D) atlas, and E) website trip reports.



Galliformes

*Boakes et al. 2010* Distorted Views of Biodiversity: Spatial and Temporal Bias in Species Occurrence Data

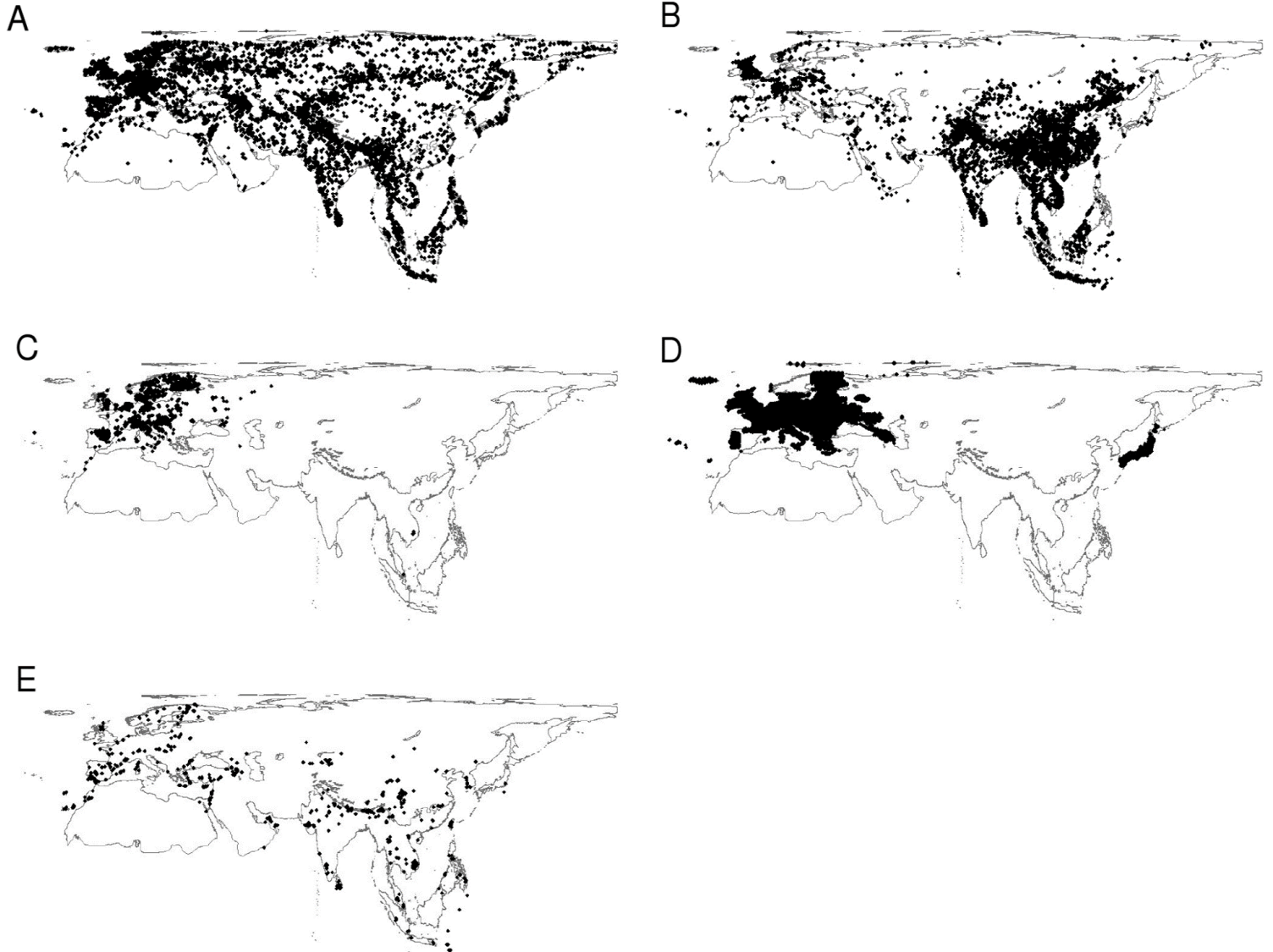
# *Boakes et al. 2010* Distorted Views of Biodiversity: Spatial and Temporal Bias in Species Occurrence Data

- By collating a large historical database of ~170,000 records of species in the avian order **Galliformes, dating back over two centuries and covering Europe and Asia, we investigate patterns of spatial and temporal bias** in five sources of species distribution data: museum collections, scientific literature, ringing records, ornithological atlases, and website reports from “citizen scientists.” **Museum data were found to provide the most comprehensive historical coverage of species' ranges but often proved extremely time-intensive to collect.** Literature records have increased in their number and coverage through time, whereas ringing, atlas, and website data are almost exclusively restricted to the last few decades. Geographically, our data were biased towards Western Europe and Southeast Asia. **Museums were the only data source to provide reasonably even spatial coverage across the entire study region.** In the last three decades, literature data have become increasingly focussed towards threatened species and protected areas, and currently no source is providing reliable baseline information—a role once filled by museum collections.

# Is there bias in this study on bias?

The spatial distribution of records from different sources.

A) museums, B) literature, C) ringing, D) atlas, and E) website trip reports.



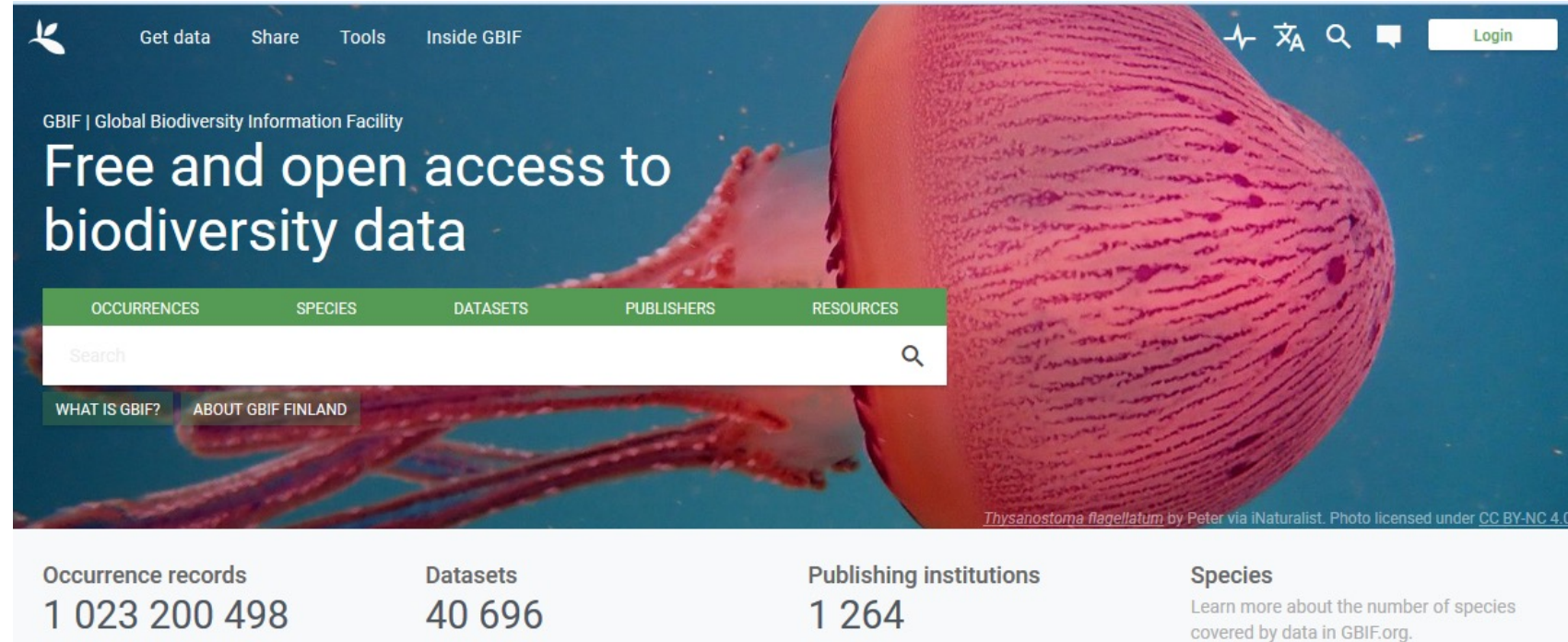
# Morphological/gene data (global)

- DNA
  - BOLD – best traceability back to specimens ([v4.boldsystems.org/](http://v4.boldsystems.org/))
  - Genbank/European Nucleotide Archive – large, but has quality issues ([www.ebi.ac.uk/ena](http://www.ebi.ac.uk/ena))
- Character libraries:
  - Morphobank [morphobank.org/](http://morphobank.org/)
- 2D/3D image/model libraries:
  - very scattered



# Collection sample data sources: global

- GBIF (portal)
  - Original focused mostly on occurrences from observations
  - Species
  - Curated datasets
- Other databases
  - [iobis.org/](http://iobis.org/) (global marine biodiv.)
  - Many others by field or region: CSIRO, India Biodiversity Portal



GBIF | Global Biodiversity Information Facility

## Free and open access to biodiversity data

OCCURRENCES SPECIES DATASETS PUBLISHERS RESOURCES

Search

WHAT IS GBIF? ABOUT GBIF FINLAND

Thysanostoma flagellatum by Peter via iNaturalist. Photo licensed under CC BY-NC 4.0.

Occurrence records	Datasets	Publishing institutions	Species
1 023 200 498	40 696	1 264	Learn more about the number of species covered by data in GBIF.org.

# GBIF

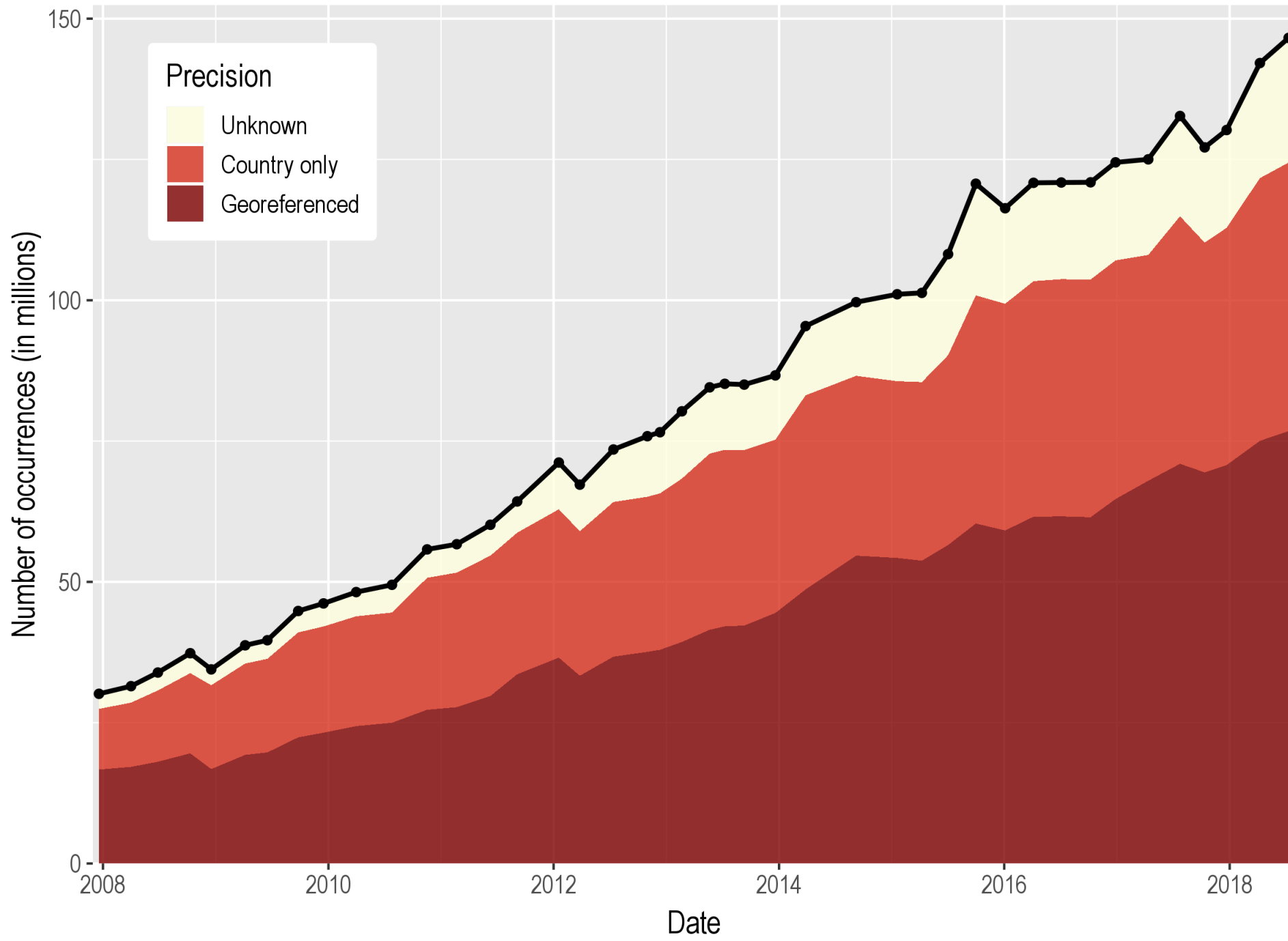
Simple	Advanced
License	▼
Scientific name	▼
Basis of record	▼
Location	▼
Year	▼
Month	▼
Dataset	▼
Country or area	▼
Issues and flags	▼
Media type	▼
Publisher	▼
Institution code	▼
Collection code	▼
Catalog number	▼
Type status	▼



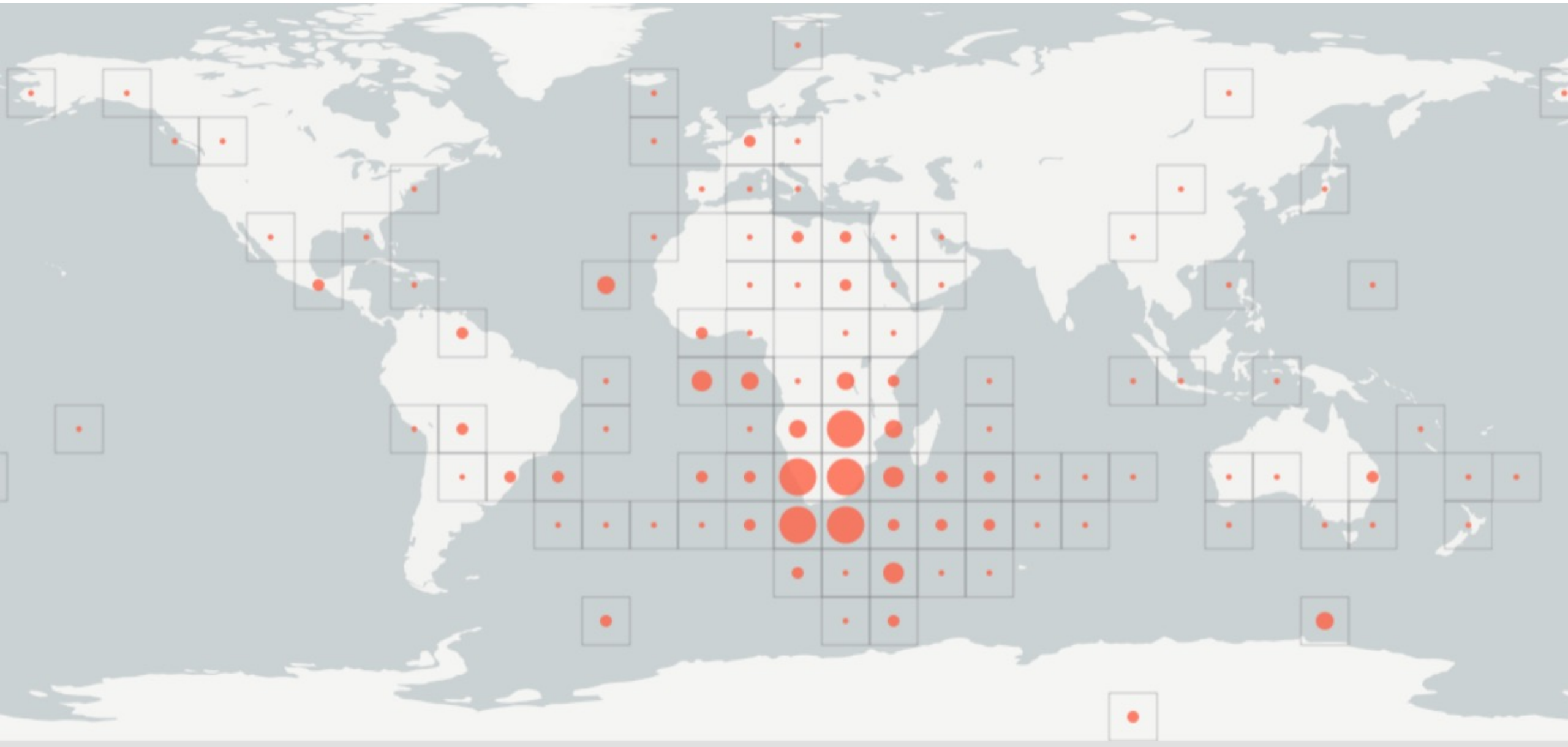
Basis of record		^
<input type="checkbox"/>	Observation	21 737 696
<input type="checkbox"/>	Machine observation	11 004 946
<input type="checkbox"/>	Human observation	805 370 753
<input type="checkbox"/>	Material sample	554 308
<input type="checkbox"/>	Literature	234 405
<input type="checkbox"/>	Preserved specimen	149 053 992
<input type="checkbox"/>	Fossil specimen	10 053 234
<input type="checkbox"/>	Living specimen	1 409 721
<input type="checkbox"/>	Unknown	24 079 355

# Availability of coordinates and country for specimens

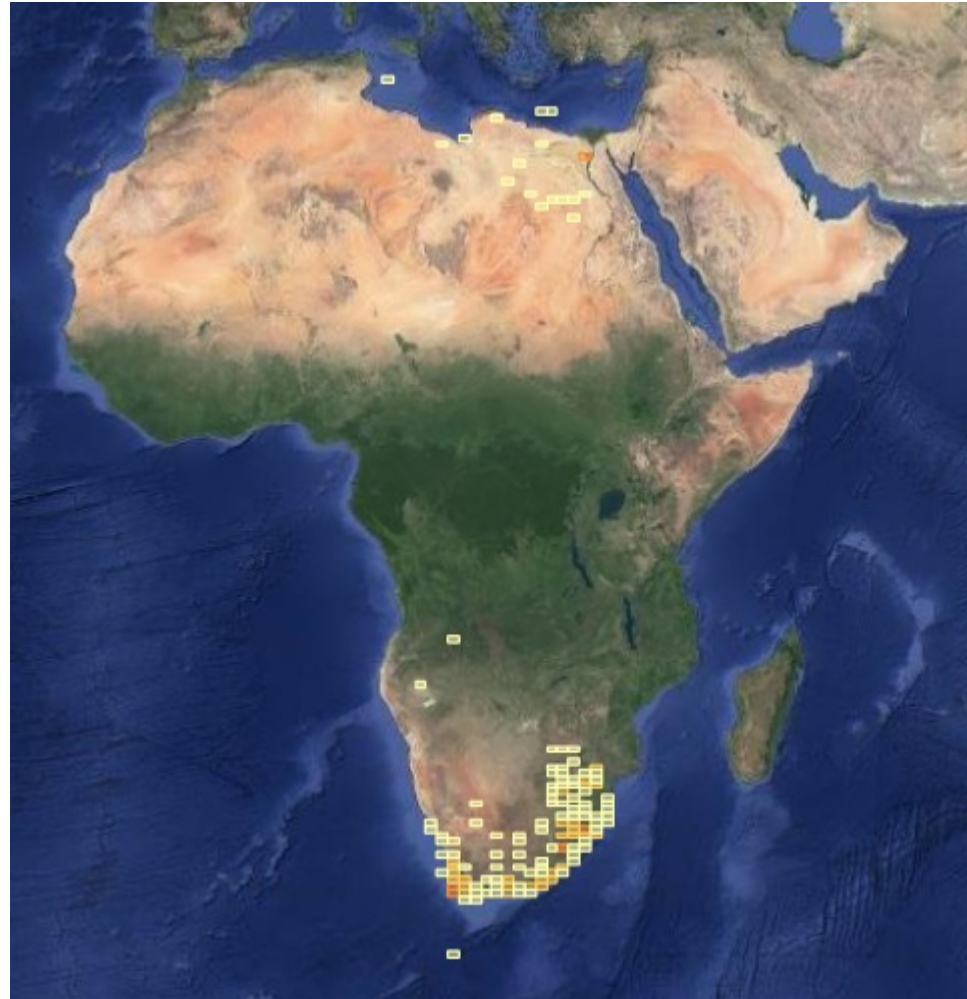
# GBIF



# GBIF "South Africa"



Laji.fi country: "South Africa"



*The one thing to remember from this part!*

All data is biased, some  
data is just wrong



Almost certainly out of time here, so  
everything past this slide will be skipped







for people?  
for institutes?  
for nature?  
for research?

## Discussion

Can we prepare for future research needs today?

Who owns the data (or the specimen)?

# Finding digital collection (etc) data

- choose your requirements **before** searching
- primary & non-primary (aggregator) sources
- collection data is verifiable
- existing vs. digi on demand



# Data formats: occurrence data

- Comma-separated values (CSV)
- Excel files (.xls/.xlsx)
- Darwin Core

# Simple Darwin Core

	A	CF	CG	CH	CI
1	Unit.UnitID	Gathering. Municipality Verbatim	Gathering. BioProvinceVer batim	Gathering. ProvinceV erbatim	Gathering. LocalityVerbatim
46	<a href="http://tun.fi/MY.452398">http://tun.fi/MY.452398</a>			Murmansk	Kuzomen
47	<a href="http://tun.fi/MY.452932">http://tun.fi/MY.452932</a>	Silvaplana		GraubÃ¼nden	
48	<a href="http://tun.fi/MY.460127">http://tun.fi/MY.460127</a>	Parainen	Varsinais-Suomi		Lofsdal
49	<a href="http://tun.fi/MY.460130">http://tun.fi/MY.460130</a>	KolatselkÃ¤		Karelian Republic	
50	<a href="http://tun.fi/MY.460133">http://tun.fi/MY.460133</a>	Kuusamo	Koillismaa		Juuma, PetÃ¤jikkÃ¤puro
51	<a href="http://tun.fi/MY.460136">http://tun.fi/MY.460136</a>	EnontekiÃ¤	EnontekiÃ¤n Lappi		between Vittanki and Mukkav
52	<a href="http://tun.fi/MY.460140">http://tun.fi/MY.460140</a>	Enontekis	EnontekiÃ¤n Lappi		between Naimakka and Vittar
53	<a href="http://tun.fi/MY.460144">http://tun.fi/MY.460144</a>	Muonio	KittilÃ¤n Lappi		on the way to Olostunturi
54	<a href="http://tun.fi/MY.460148">http://tun.fi/MY.460148</a>	Utsjoki	Inarin Lappi		MantojÃ¤rvi
55	<a href="http://tun.fi/MY.460152">http://tun.fi/MY.460152</a>	Enontekis	EnontekiÃ¤n Lappi		KilpisjÃ¤rvitrakten
56	<a href="http://tun.fi/MY.460156">http://tun.fi/MY.460156</a>	Espoo	Uusimaa		
57	<a href="http://tun.fi/MY.460160">http://tun.fi/MY.460160</a>	Vihti	Varsinais-Suomi		PÃ¤ivÃ¤lÃ¤

# Darwin Core

- <http://rs.tdwg.org/dwc/>

```
<dcterms:Location>
  <dwc:locationID>http://guid.mvz.org/sites/arg/127</dwc:locationID>
  <dwc:country>Argentina</dwc:country>
  <dwc:countryCode>AR</dwc:countryCode>
  <dwc:stateProvince>Neuquén</dwc:stateProvince>
  <dwc:locality>Valle Limay, Estancia Rincon Grande, 48 ha area with centroid at this point</dwc:locality>
  <dwc:decimalLatitude>-40.97467</dwc:decimalLatitude>
  <dwc:decimalLongitude>-71.0734</dwc:decimalLongitude>
  <dwc:geodeticDatum>WGS84</dwc:geodeticDatum>
  <dwc:coordinateUncertaintyInMeters>200</dwc:coordinateUncertaintyInMeters>
</dcterms:Location>
```

# Issues with specimen data

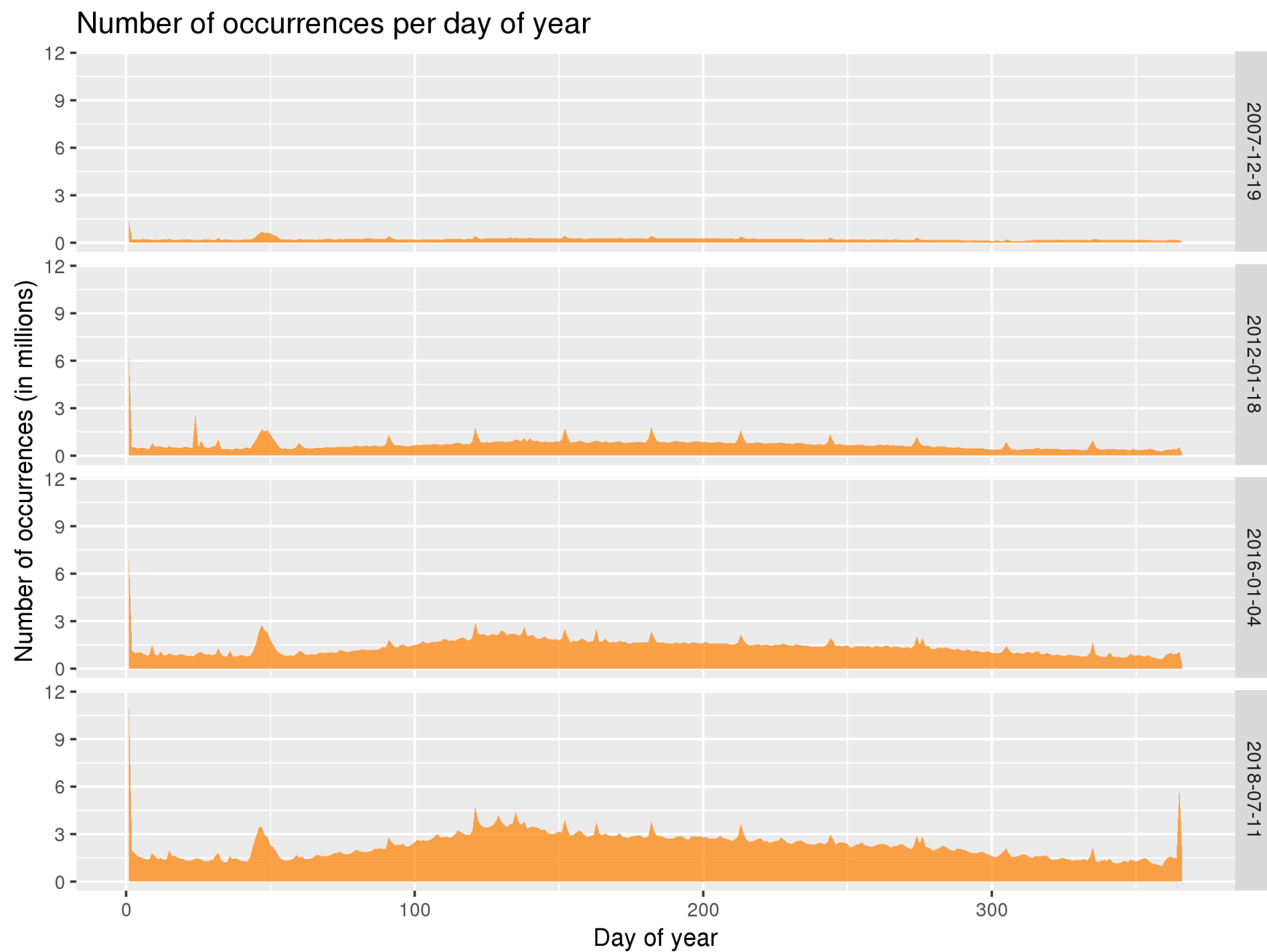
- Always know your original source!
- Occurrence data is extremely biased
  - Work with uniform(cough cough) subsets
  - Normalize as far as possible
- Big data is full of small mistakes

# CETAF Stable identifiers

- Examples
  - <http://id.luomus.fi/GV.45118>
  - <http://mus.utu.fi/ZMAA.TYPE001>
- Museum für Naturkunde Berlin:
  - object at [http://coll.mfn-berlin.org/u/ZMB\\_123](http://coll.mfn-berlin.org/u/ZMB_123)
  - rdf at [http://coll.mfn-berlin.org/u/ZMB\\_123.rdf](http://coll.mfn-berlin.org/u/ZMB_123.rdf)
  - json at [http://coll.mfn-berlin.org/u/ZMB\\_123.json](http://coll.mfn-berlin.org/u/ZMB_123.json)
  - xml at [http://coll.mfn-berlin.org/u/ZMB\\_123.xml](http://coll.mfn-berlin.org/u/ZMB_123.xml)
  - html at [http://coll.mfn-berlin.org/u/ZMB\\_123.html](http://coll.mfn-berlin.org/u/ZMB_123.html)



# GBIF



# IPS-161 BIOLOGICAL COLLECTIONS

---

*Wed 24.viii.*

Lecture 8 - **Jere Kahanpää**/digitization team & **Kari Lahti**

Documentation, databases (including KOTKA etc.) + DIGITALIZATION + OPEN data (FinBIF, GBIF)  
+ biodiversity-informatics  
- citing specimens by unique identifiers

# FinBIF – Finnish Biodiversity Information Facility

## - National Infrastructure of Species Information -

LAJI.FI Lajit Selaa havaintoja Vihko Teemat Foorumi Kari Lahti FI

**LAJI.FI**  
 SUOMEN LAJITIEKESKUS  
 FINLANDS ARTDATACENTER  
 FINNISH BIODIVERSITY INFO FACILITY

30 487 583 havaintoa    31 679 lajia    177 aineistoa

Lajihaku

**Suomen Lajitietokeskus**  
 Suomen Lajitietokeskus kerää ja yhdistää suomalaisen lajitiedon yhtenäiseksi ja avoimeksi kokonaisuudeksi. Laji.fi:ssä voit tutustua lajeihin ja niiden esiintymiseen, selata havaintoja suomalaisista lajitietokannoista sekä pitää kirjaa omista luontohavainnoistasi.

**Lajit**  
 Tutustu lajeihin

**Havainnot**  
 Selaa havaintoja

**Vihko**  
 Ilmoita havaintosi

**Ajankohtaista**

Ajanlaskun ajan rajuin aurinkomyrsky paikannettiin Lapin puista  
 luomus.fi 10.09.2018

Huolto- ja päivityskatko 8.00-10.00 (ohi)  
 tekninen 06.09.2018

Vieraslaji harlekiinileppäpörkki yrittää Suomen valtausta  
 syke.fi 06.09.2018

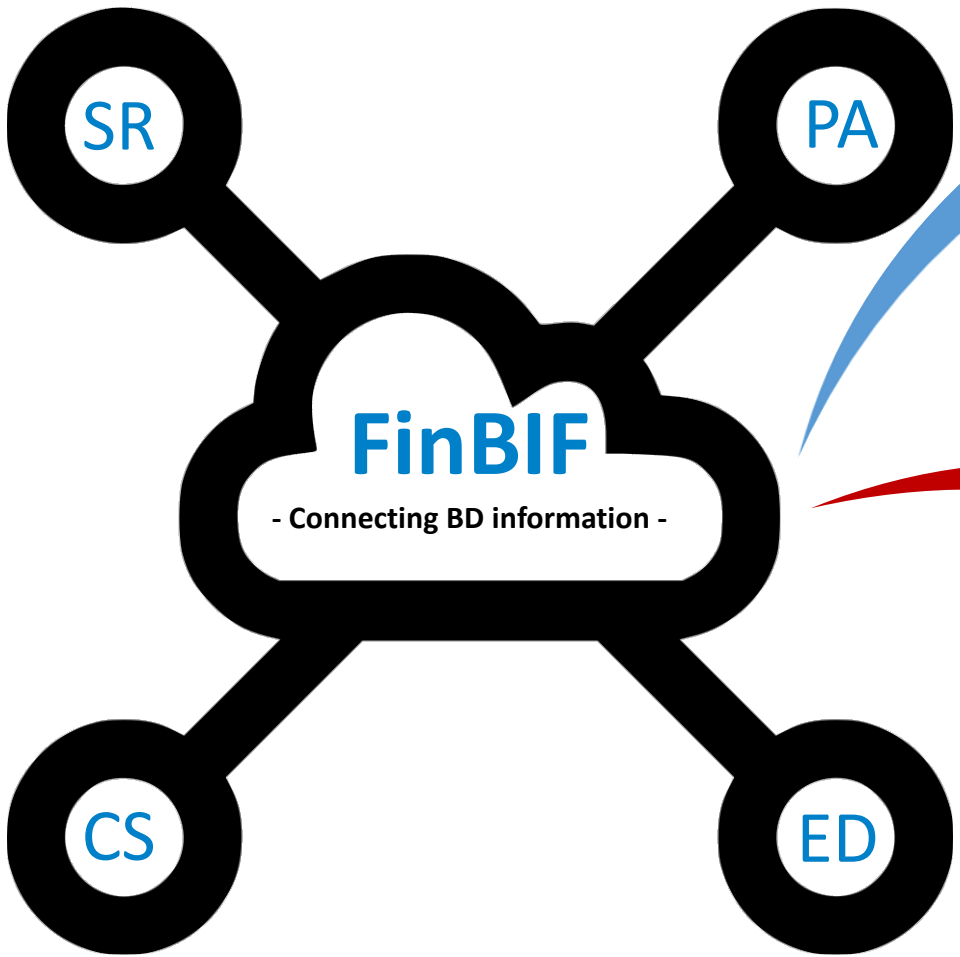
Lyhyt tietoliikennekatko ti 4.9. klo 12:20 - 12:35 (ohi)

Kari LAHTI 24.8.2022

# FinBIF Data Sources and Users

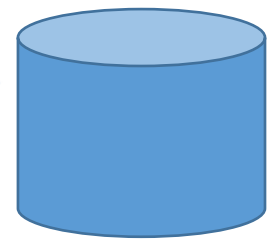
Scientific Research

Public Authority

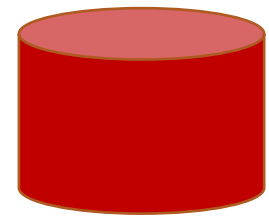
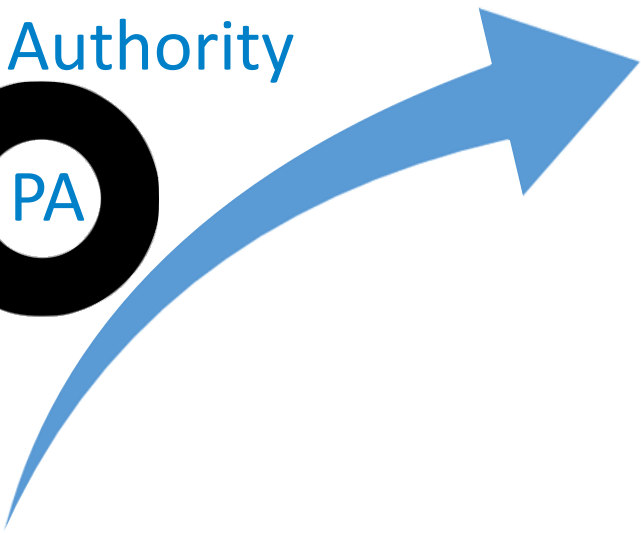


Citizen Science

Education



OPEN DATA

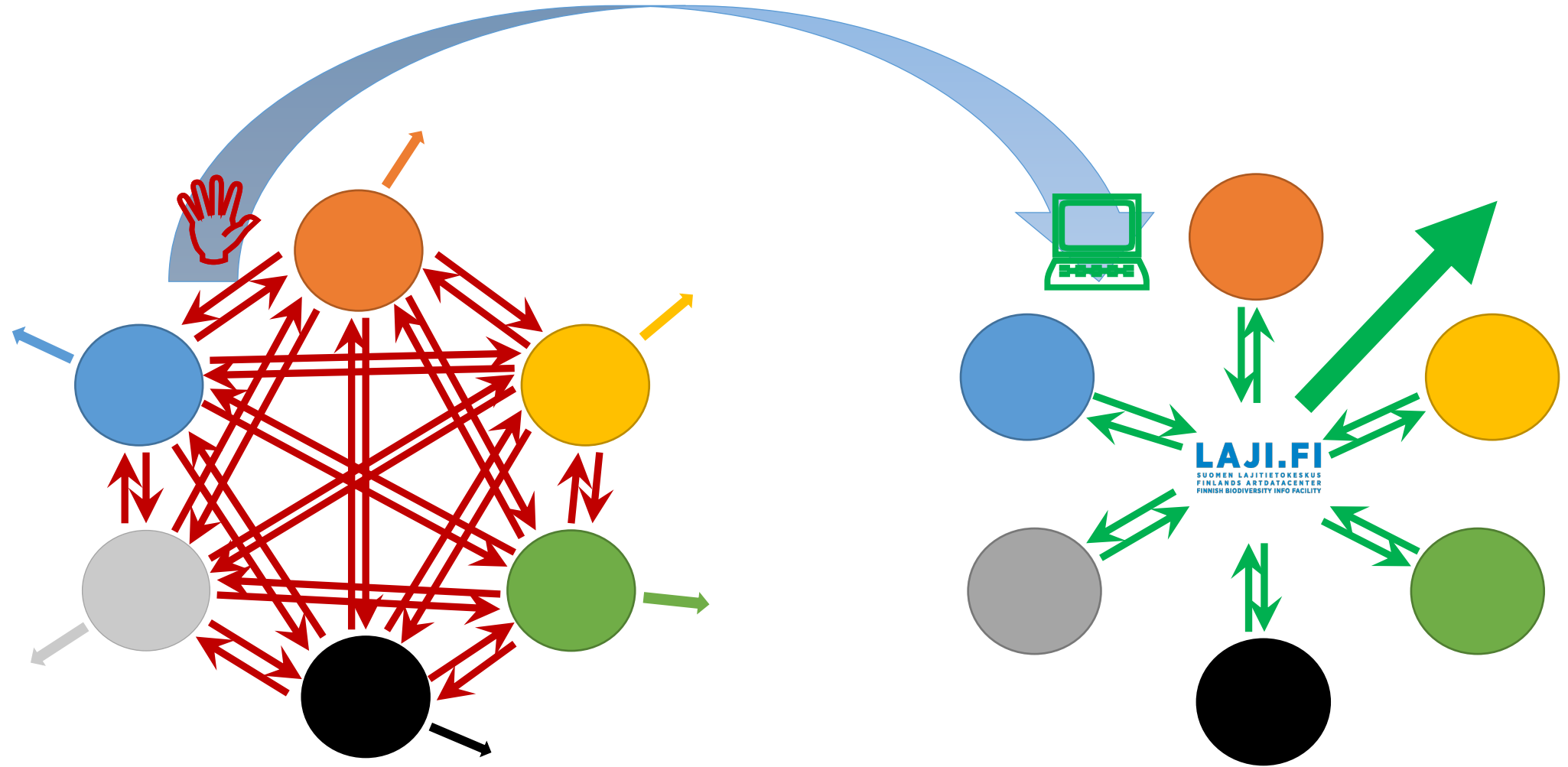


RESTRICTED-USE DATA





# FinBIF – From Chaos towards Harmony



# FinBIF – From Chaos towards Harmony

**Natural Science Collections**



**LUOMUS**  
 LUONNONTIETEELLINEN KESKUSMUSEO

**TURUN YLIOPISTO | Biodiversiteetti**

**LUONNONTIETEELLINEN MUSEO**

**OULUN YLIOPISTO**

**JYVÄSKYLÄN YLIOPISTO UNIVERSITY OF JYVÄSKYLÄ**

**KUOPION LUONNONTIETEELLINEN MUSEO**

**Research Monitoring Mapping**



Amphibian Survey and Monitoring

**SYKE**

**Luke**  
 LUONNONVARAKESKUS

**METSÄHALLITUS**

**LUOMUS**  
 TALVILINTULASKENTA

**TURUN YLIOPISTO**

**Valtakunnallinen päiväperhosseuranta (NAFI)**

**Citizen Science Education sector Associations Enthusiasts**



**Lukiolaiset lepakotutkijat**

**Sieniatlas**

**Pinkka**  
 Species learning environment

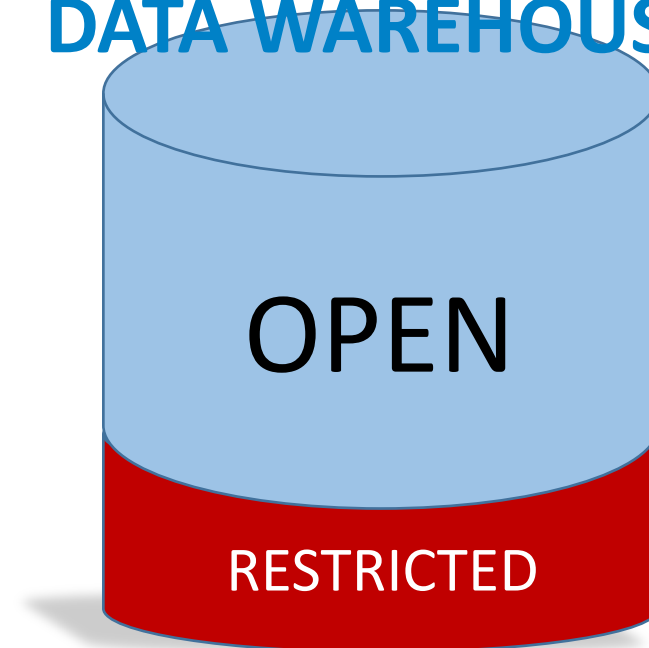
**KOKOELMA-KILPAILU**

**Suomen luonnonsuojeluliitto**



**LAJI.FI**

## FinBIF – Data sharing and usage

LAJI.FI  
DATA WAREHOUSE



publicauthorities.laji.fi

**LAJI.FI**  
**DATA WAREHOUSE**

**LAJI.FI**

34 042 263 osumaa 35 831 lajista

Lataa tiedostona

Kaikki havainnot Suomen havainnot (beta)

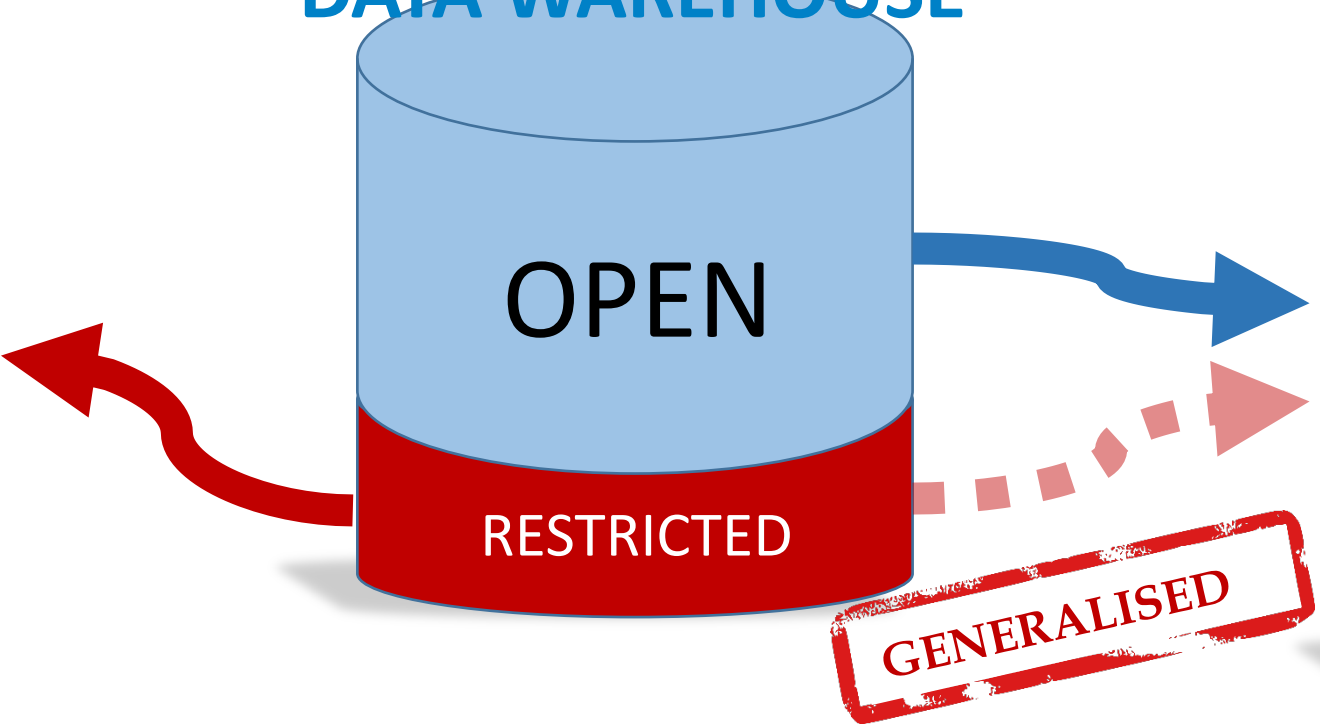
Luettelo Kartta Kuvat Laj

Valitse sarakkeet

Laji	Määr
harmaapäätikka – <i>Picus canus</i>	1
laulujoutsen – <i>Cygnus cygnus</i>	6
korppi – <i>Corvus corax</i>	2
pyy – <i>Tetrastes bonasia</i>	2
tikli – <i>Carduelis carduelis</i>	2
pyrstötäinen – <i>Aegithalos caudatus</i>	8
käpytikka – <i>Dendrocopos major</i>	2
lehtopöllö – <i>Strix aluco</i>	

**RESTRICTED-USE**

**DATA WAREHOUSE LAJI.FI**



**Laji.fi**

LAJI.FI

34 435 915  
havaintoa

35 838  
lajia

406  
ainesta

Lajihaku

**Suomen Lajitietokeskus**

Suomen Lajitietokeskus kerää ja yhdistää suomalaisen lajitiedon yhtenäiseksi ja avoimeksi kokonaisuudeksi. Laji.fi:ssa voit tutustua lajeihin ja niiden esiintymiseen, selata havaintoja suomalaisista lajitietokannoista sekä pitää kirjaa omista luontohavainnoistasi.


**AVOIN**

**TIETOVARASTO LAJI.FI**

# FinBIF – The usage of restricted-use data


publicauthorities.laji.fi

**Public authorities**



- ELY's (Regional Env. Centres)
- Municipalities
- Ministries
- Research institutes

**The Information systems of public authorities**



- LajiGIS
- Hertta
- Aarni
- ArcGIS
- Mapinfo
- ...



**LAJI.FI**

34 042 263 osumaa 35 831 lajista

Lataa tiedostona

Kaikki havainnot Suomen havainnot (beta)

Luettelo Kartta Kuvat Laji

Valitse sarakkeet

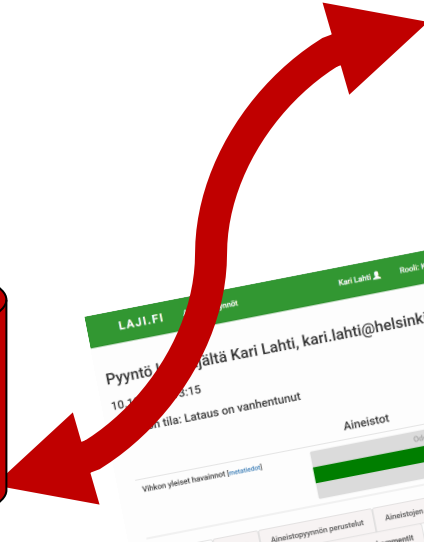
Laji	Määr
harmaapäätikka – <i>Picus canus</i>	1
laulujoutsen – <i>Cygnus cygnus</i>	6
korppi – <i>Corvus corax</i>	2
pyy – <i>Tetrastes bonasia</i>	2
tikki – <i>Carduelis carduelis</i>	2
pyrstötiainen – <i>Aegithalos caedatus</i>	8
käpytikka – <i>Dendrocopos major</i>	2
lehtopöllö – <i>Strix aluco</i>	2

**RESTRICTED-USE**

**Research Assignments Consultation**



- Universities
- Private sector
- Research institutes



*Restricted Data Request Service*

**LAJI.FI**

Pyyntö... Kari Lahti, kari.lahti@helsinki.fi

10... 3:15

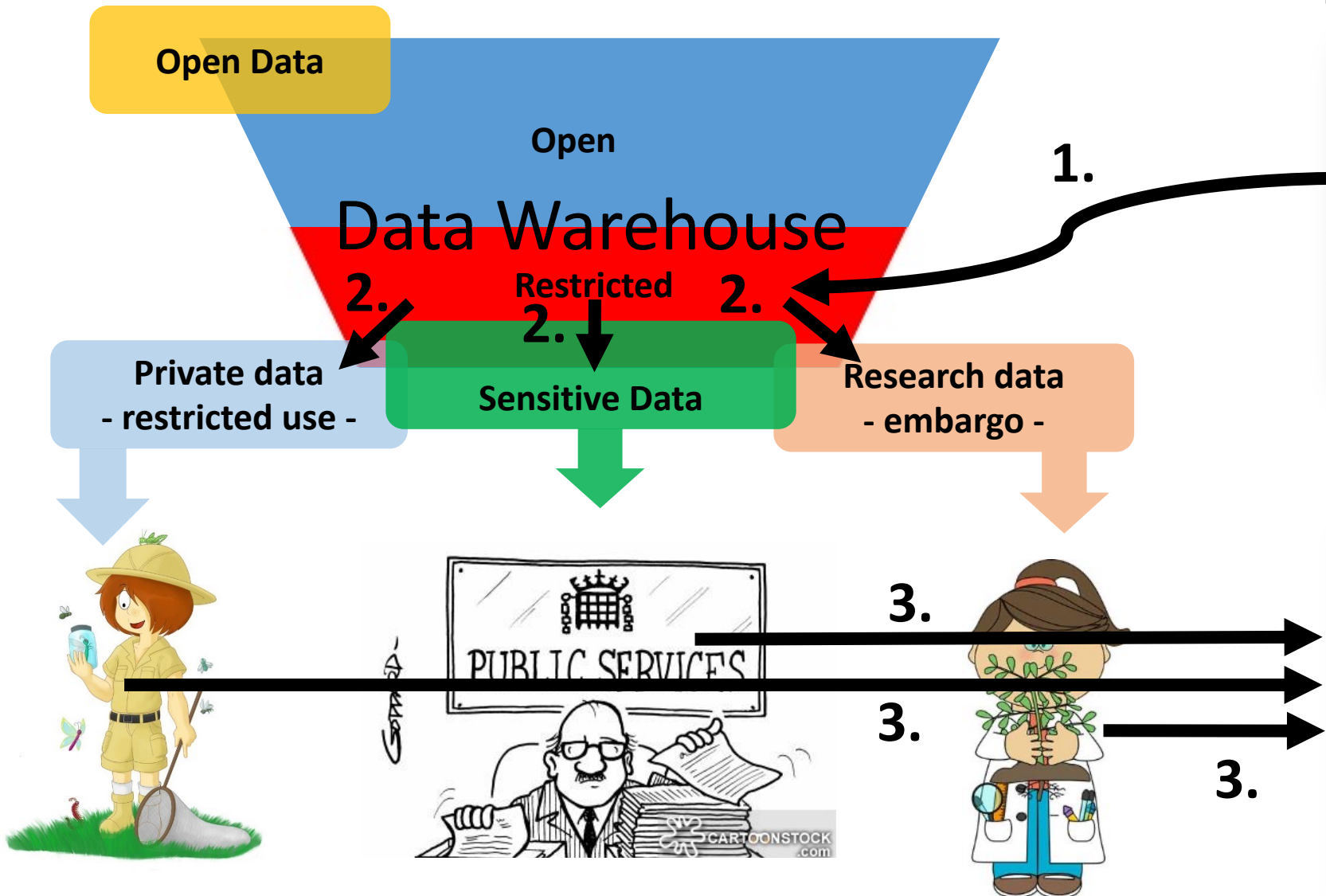
... tila: Lataus on vanhentunut

**Aineistot**

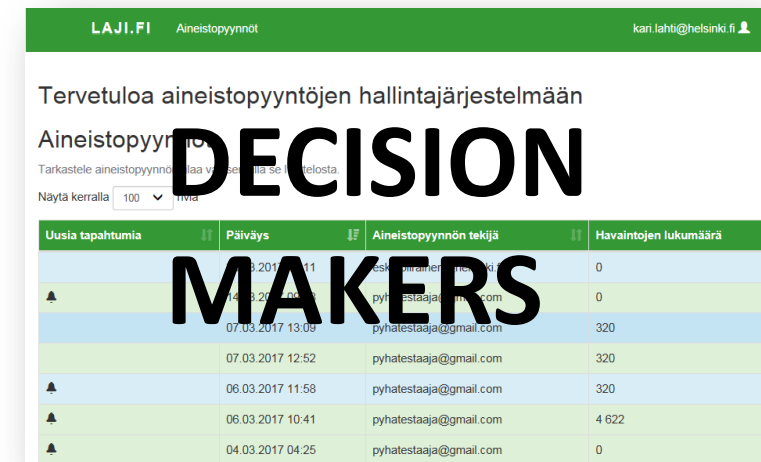
Hyväksytyt

Muutoshistoria	Päiväys	Tekijä	Roolissa
24.02.2020 16:03	Kari Lahti (kari.lahti@helsinki.fi)	Administrattori	Katsottu
20.02.2020 16:57	Mervi Raunisto (mervi.raunisto@helsinki.fi)	Administrattori	Katsottu
20.11.2019 10:06	Kari Lahti (kari.lahti@helsinki.fi)	Administrattori	Katsottu

# FinBIF – The usage of restricted-use data



## RESTRICTED DATA REQUEST SERVICE



## FAIR principles as a “pressure test”

Published 2016\*

Adopted widely

- EC European Open Science Cloud (EOSC) “As Open as Possible, as Closed as Necessary”
- Horizon 2020
  - [Turning FAIR into reality](#) (EUROPA>Publications\_Office of the EU>Publication detail> Turning FAIR into reality)

Aim is to make the data:

- **Findable**
  - **Accessible**
  - **Interoperable**
  - **Re-usable**
1. The elements of the FAIR Principles are related, but independent and separable.
  2. The Principles assist discovery and reuse by third-parties.
  3. The barrier-to-entry is maintained as low as possible.
  4. The Principles function in any combination and incrementally increase degrees of ‘FAIRness’.

\*<https://www.nature.com/articles/sdata201618>, <https://www.go-fair.org/fair-principles/>

## TEST results of FinBIF (self assessment) \*\*\* = max score

### Findable

1. Persistent identifier, PID
2. Rich metadata
3. Registered in a searchable resource
4. PID is specified at the metadata

\*\*\*

### Accessible

1. Data should be retrievable by identifier according to principle "As Open as Possible, as Closed as Necessary"
2. Protocol is open, free, and universally implementable
3. Registration and authorisation supported, where necessary
4. Metadata still available even when the data is no longer available.

\*\*\*

### Interoperable

1. Data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
2. Data use vocabularies that follow FAIR principles.
3. Data include qualified references to other data

\*\*

### Re-usable

1. Meta(data) have a plurality of accurate and relevant attributes.
2. Data are released with a clear and accessible data usage license.
3. Data are associated with their provenance.
4. Data meet domain-relevant community standards

\*\*\*

## Unique PIDs

FinBIF uses a ***persistent HTTP-URI identifier*** for all types of real-life and digital objects (specimens, occurrences, taxa, metadata, persons, organisations, information systems, etc.), as recommended by the World Wide Web Consortium (Best practices for publishing linked data; <https://www.w3.org/TR/Id-bp/>).

The identifier takes the user to an ID redirect service, which redirects the user to a page that shows information about the object in human-readable format. For example, specimen identifiers redirect to information about the specimen and taxon identifiers to a page describing the taxon.

The redirect service can also provide machine-readable data about the object, if the user (client software) requests that using Accept headers.

If partner organisations do not provide HTTP-URI identifiers for their occurrences, FinBIF will use the persistent internal IDs of the data source to generate globally unique URI identifiers.

DOI (Digital Object Identifier) identifiers for data downloads and dataset metadata will be created in the near future. ([https://www.doi.org/driven\\_by\\_doi/DOI\\_Marketing\\_Brochure.pdf](https://www.doi.org/driven_by_doi/DOI_Marketing_Brochure.pdf)).

LAJI.FI

på svenska in English

## Tiedostolataus <http://tun.fi/HBF.5167>

Latauspäivä: 26.8.2020

Osumien lkm: 965

### Rajaukset:

Kohde (laji): Orchidaceae (MX.40029)  
 Eliömaakunta: Ahvenanmaa (A)  
 Lataus tietovarastoon, päivänä tai ennen: 2020-08-26

## Viittausohje

Voit viitata tähän lataukseen seuraavasti:

Suomen Lajitietokeskus/FinBIF. <http://tun.fi/HBF.5167> (haettu 26.8.2020).

Jos käytät vain osaa aineistoista, on suositeltavaa, että viittaat vain niihin aineistoihin. Latauksen osajoukkoon voi viitata seuraavasti (poista käyttämätön aineisto):

Suomen Lajitietokeskus/FinBIF. <http://tun.fi/HBF.5167>, <http://tun.fi/HR.447>, <http://tun.fi/HR.169>, <http://tun.fi/HR.3>

Viitataksesi latauksen yksittäiseen riviin voit käyttää [Document.DocumentID]-kenttää, esimerkiksi:

Suomen Lajitietokeskus/FinBIF. <http://tun.fi/EXMP.1234>, <http://some.org/9876> (haettu 26.8.2020).

## Aineistot

Lataus koostuu seuraavista aineistoista joille on määritelty käyttöoikeuslisenssi:

Kuopio Natural History Museum - KUO Putkilokasvikokoelmat (KUO) - <http://tun.fi/HR.430> [metadata]  
 Creative Commons Nimeä  
 Lisätietoja tämän aineiston käytöstä antaa [outi.vainio@kuopio.fi](mailto:outi.vainio@kuopio.fi)

LajiGIS: Lajin seurantaohjeet - <http://tun.fi/HR.3553> [metadata]  
 Creative Commons Nimeä  
 Lisätietoja tämän aineiston käytöstä antaa [lajigis@metsa.fi](mailto:lajigis@metsa.fi)

Luomus - Hatikka.fi:n havainnot - <http://tun.fi/HR.447> [metadata]  
 Creative Commons Nimeä  
 Lisätietoja tämän aineiston käytöstä antaa [info@laji.fi](mailto:info@laji.fi)

Luomus - Putkilokasvikokoelmat - <http://tun.fi/HR.169> [metadata]  
 Creative Commons Nimeä  
 Lisätietoja tämän aineiston käytöstä antaa [henry.vare@helsinki.fi](mailto:henry.vare@helsinki.fi)

## Lataa tiedosto

Lataamalla tiedoston sitoudut noudattamaan yllä mainittuja käyttöoikeuslisenssejä. Lisenssit löytyvät myös latauksen `readme.txt` tiedostosta.



HBF.5167.zip (0.1 Mt)

Tiedoston vienti Exceliin Tiedoston vienti ArcGIS -paikkatieto-ohjelmaan

## Kotka CMS

- Kotka is one of the two **primary data** management systems of FinBIF
- Kotka applies simple and **pragmatic** approaches. This has helped it grow into a nationally used system.
- The aim is to improve **collection management efficiency** by providing practical tools.
- Kotka **emphasises the quantity** of digitised specimens over completeness of the data. It harmonises practices by bringing all types of collections under one system; the types currently covered include zoological, botanical, mycological and palaeontological museum collections, tissue and DNA samples, and botanic garden and microbial living collections.
- Kotka stores data mostly in a denormalised free text format using a triplestore and a simple hierarchical data model. This allows greater flexibility of use and faster development compared to a normalized relational database.
- Kotka does some data validation, but **quality control** is seen as a continuous process and is **mostly done after the data have been recorded** into the system.
- Kotka is a **web application**. Data can be entered, edited, searched and exported through a browser-based user interface (UI). However, most users prefer to enter new data in customizable MS-Excel templates, which support the hierarchical data model, and upload these to Kotka. Batch updates can also be done using Excel.
- Kotka **stores all revisions** of the data to avoid any data loss due to technical or human error.
- Kotka supports **designing and printing specimen labels** (Heikkinen et al. 2019b), annotations by external users, and handling accessions, loan transactions, and the Nagoya protocol (Kuusijärvi et al. 2019).
- <https://biss.pensoft.net/article/37181/list/19/>



## Decision making

- Sustainable use of Natural Resources
- Land use practices and planning
- Nature Conservation, species protection, Red Data Books
- EU and National Reporting
- Invasive Alien Species; early warning and eradication

## Research

- Species surveys and censuses
- Climate Change indications

## Education

- Schools – Species identification and digital herbaria
- University – Learning environment

**ENVIRONMENT .fi**  
 Joint website of Finland's environmental administration

**Maps and statistics** | Forms, permits and environmental impact assessment

Home > Nature > Species > Species in the EU Habitats and Bird Directives

### Species in the EU Habitats and Bird Directives

**CURRENT**  
 The European Environment Agency (EEA) has published a new report on the state of nature in the EU.

**RELATED TOPICS**  
 Natural habitats referred to in the EU Habitats Directive

**RELATED LINKS**  
 Ministry of the Environment  
 Protection of species

**Linjalaskentomake**

**Pinkka**  
 Species learning environment

**Vieraslajit**  
 Varhaisvaroitussjärjestelmä

Haitallisten vieraslajien (EU-luettelo) havainnoista tulee varhaisvaroitussjärjestelmään aina ilmoitus (kaikista Lajitietokeskukseen liitettujen järjestelmistä) ja siitä lähetetään automaattisesti hälytys asianomaiselle viranomaiselle toimenpiteitä varten. Viranomaiset vastaavat myös tietojen välittämisestä EU:n vaatimusten mukaisesti EASIN-palveluun.

Torjuntatoimet kirjataan Lajitietokeskuksen viranomaispalvelujen Viikko-palvelun vieraslajien torjunta -lomakkeella. Torjuntatoimien tiedot tallennetaan nimellä ja toimia voidaan hakea niiden vaikuttavuuden arvioimiseksi ja raportointia varten.

Suomen Lajitietokeskuksen viranomaispalvelujen osana on maa- ja metsätalousministeriön (MMM) hallinnoiman VISAKE-hankeen (Vieraslajien varhaisvaroituss- ja seuranta järjestelmän kehitys ja tahattomien leviämisvyörien hallinta) tuotoksena rakentunut varhaisvaroitussjärjestelmä, jota EU:n vieraslajiasetus edellyttää jäsenvaltioilta koskien EU:n luettelossa olevia haitallisia vieraslajeja (EU 2016/1141).

Laji	Havainto	Yksilöä	Vanhin	Uusin	Hälytys tehty
karheviuhkalehti ( <i>Cabomba caroliniana</i> )					
keittamajavankaali ( <i>Lysichiton americanus</i> )	21	237	10.05.2005	18.09.2016	
afrikanvesihätä ( <i>Lagarosiphon major</i> )					
kellusvesihyasintti ( <i>Eichhornia crassipes</i> )					
isoärvä ( <i>Myriophyllum aquaticum</i> )					
purpurakudzu ( <i>Pueraria montana var. lobata</i> )					
lautarusolehti ( <i>Ludwigia grandiflora</i> )					
loikonsuolehti ( <i>Ludwigia peploides</i> )					
raastotatar ( <i>Persicaria perfoliata</i> )					
lvisutlatva ( <i>Baccharis halimifolia</i> )					
linahelmikki ( <i>Parthenium hysterophorus</i> )					
persianjättiputki ( <i>Heracleum persicum</i> )	217	98	04.08.1871	17.07.2016	
armenianjättiputki ( <i>Heracleum sosnowskyi</i> )	29	83	01.09.2005	13.08.2016	
sumasammakonputki ( <i>Hydrocotyle ranunculoides</i> )					
sahasabora ( <i>Pseudorasbora parva</i> )					
rohmutokko ( <i>Percocottus glenii</i> )					
härkilemmikko ( <i>Lithobates palustris</i> )					

**Suomen lajien uhanalaisuus 2010**  
 Punainen kirja  
 The 2010 Red List of Finnish Species

## Challenges encountered with database and data

1. Biggest challenge is to convince data owners to **share** their data especially as **Open Data**
2. The diverse use of **different taxonomies**, taxonomic backbones and scientific names in **defining the same taxon concepts** creates a huge challenge, which we are trying to tackle by applying **Linked Data principles with taxon concept URI-identifiers**. Harmonising the used taxonomies to be linked or redefined with the national taxonomy of FinBIF is the ultimate national goal. Nordic-Baltic pilot to link the regional taxonomies is under way through NeIC led project DeepDive.
3. Endless need to **provide tools** to assist users in the **process of sharing** their data **and using** FinBIF data (Excel imports-exports, E-forms, GIS-application support, API interfaces...). Data is stored in such a huge variety of forms – standards enormously needed.
4. **Data flow issues** from a content standpoint are mainly concerning how to deal with the **data quality**, how to handle **data sensitivity**, how to **manage scientific research data** to allow enough time for analysing and publishing and at the same time share the e.g. raw species occurrence data asap for needed use (land use planning and practices, EIA etc.)
5. Data policies are often institutional and quite often protect the institution's internal potential benefits instead of supporting open data. Licencing, sensitive information and use-restrictions are most difficult issues to solve when designing the data policy. To cover the legal aspects is another challenge.

# Summary

## FinBIF

