Digitization processes in natural history collections: from drawers to data publication
SwissCollNet Annual Workshop 2024

Standardizing Data Standards

Lucas Blattner, Head of Collection Digitization, Natural History Museum Basel
&
Lucia Pollini Paltrinieri, Senior curator invertebrates department, Museo cantonale di storia naturale, Lugano

Bern, 19.01.2024

naturhistorisches museum Basel | archive des lebens

MUSEO CANTONALE DI STORIA NATURALE LUGANO

Spiders (Araneae) of the MCSN: Reconditioning, Revision, Reorganization and Digitization of a Southern Alps collection

Museo cantonale di storia naturale – Lugano
Lucia Pollini Paltrinieri

Project SCN 131-TI

AIMS OF THE PROJECT

- Revise, reorganise, conditioning in high standard and digitise the estimated 9'000 specimens of spiders caught in Ticino and stored at the MCSN

- Material easily available to researchers through his reorganisation and partly re-identification according to current knowledge

- Systematics updating and in line with that used by other Swiss museums

- Data mobilisation through the digitisation on a database structure conforming to Darwin Core and GBIF Swiss node standards

- Training of a young biologist in both spider systematics and museum conservation techniques

- Collaboration between two museums (Lugano & Basel)

INITIAL STATE

- 10+1 projects and ~ 9'000 initial individuals
- Non-homogeneity of conditioning and data

DATA AT THE CURRENT STATE

- Digitisation or homogenisation of data provided
- Updated systematic application
- 815 samples reviewed by specialist
- 564 species in total,
    - of which 3 new species for Switzerland and
    - 1 new for science
    - 4851 vials
- Xxx specimens in collection

araneae
Spinnen Europas

Ambros Hänggi & Christian Kropf

Water Mites in Switzerland Conditioning and Databasing

-

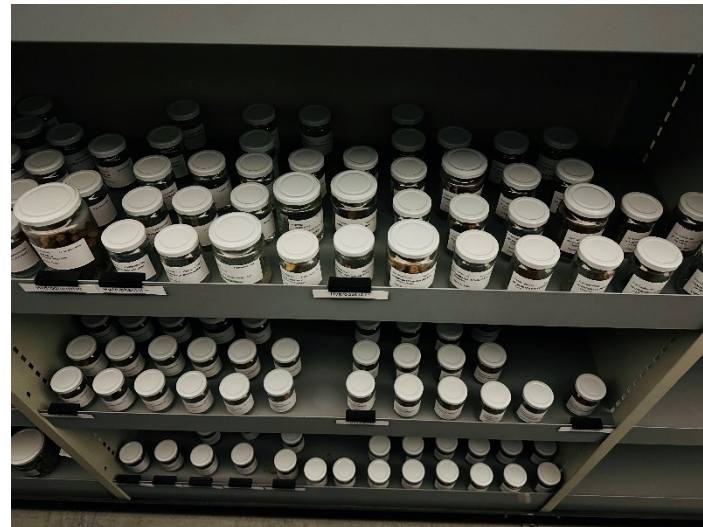Natural History Museum Basel
Holger Frick & Lucas Blattner

Project SCN-123

Aims of the project

- Curating and standardizing app. 7'000 microslides datasets
  - Cross-checking with the objects
  - Standardizing using DarwinCore field descriptors
  - Import into the new CMS Specify @ NMB
- Registering, curating and data standardisation of the liquid collection
  - Newly registering and labelling of app. 3'000 vials
  - Standardizing the data in DwC and importing into Specify
- Identification and registering of potentially important orphant collection objects
- Reorganisation of the collection: from project-specific to alphabetical
- Identification and registering of specimens from the MZL

Initial state

genus	species	n	site (probably irrelevant)	non prep	remark/prep
					all tubes lacking specimens with locality labels, all tubes containing specimens with cancelled locality information
		empty	Älv mossa Mollusc [?] 20/7 14		
					empty tubes: Material transferred to slides?
		empty	Kaukälle mossen 1/8 14		
		empty	Älv mossen 20/7 14		
Pseudosperchon	verrucosus	3	no data	1	1 M/ 1F
		empty	Rättes losch slam		
		empty	Pärekoli 8 mosse		
Megapus	tener	1	no data	0	1 M
Feltria	composita	1	Älv mosse 27/7	0	1 F
Sperchon	squamosus	5	no data	3	1 M / 1 F
		empty	Mossn Myibuk 22/7		
Hygrobates	norvegicus	8	no data	6	1 M / 1 F
Lebertia	stigmatifera	11	no data	9	1 M / 1 F
Sperchon	multiplicatus	3	no data	1	1 M / 1 F
Hygrobates	naicus	2	no data	1	1 F
Sperchon	brevirostris	9	no data	7	1 M / 1 F
Sperchon	lineatus	6	no data	4	1 M / 1 F
		empty	Kövtän mosse mde alger 31/7/14		
Megapus	spinipes	2	Sta [?] sta [?]	1	1 W
Aturus	scaber	1	no data	1	
Lebertia	inaequalis	3	no data	2	1 W
Hygrobates	albinus	16	no data	14	2 M/ 2 F
Oribatida		1	no data	1	

Current state

- All microslide and liquid vial objects have been cross-checked
- Orphant collections are curated and completely registered
- All data is DwC standardized with some custom fields and ready for import into the new CMS
- New, consistent numbering system with QR-codes on the objects
- Collection is reorganized in alphabetic order

Ongoing work:

- QC processes
- Photographing of type specimens
- Identification of new material from MZL

DIGITISATION PROBLEMS

During the data standardisation work, the problems or critical issues we were confronted with were:

- Coodinates
- Sex
- Date
- History of identifications/revisions/change of names of objects in the collection
- dwc:verbatimIdentification
- Clear definition of empty vs. unknown
- Training of staff

DIGITISATION PROBLEMS

Coordinates

- Swiss or international?
  - DarwinCore uses:
    - ➢ dwc:decimalLatitude & dwc:decimalLongitude

  - MCSN uses:
    - ➢ CooX & CooY
    - ➢ Latitudine & Longitudine

  As we work mainly in local projects (using Q Gis and Swiss maps) and in close connection with Info fauna, we decided to use the Swiss coordinate system. To these we have added fields for international coordinates WGS84

- Coordinate uncertainty
  - DarwinCore uses:
    - ➢ meters [dwc:coordinateUncertaintyInMeters]

  - Info fauna uses:
    - ➢ code 0 - 6
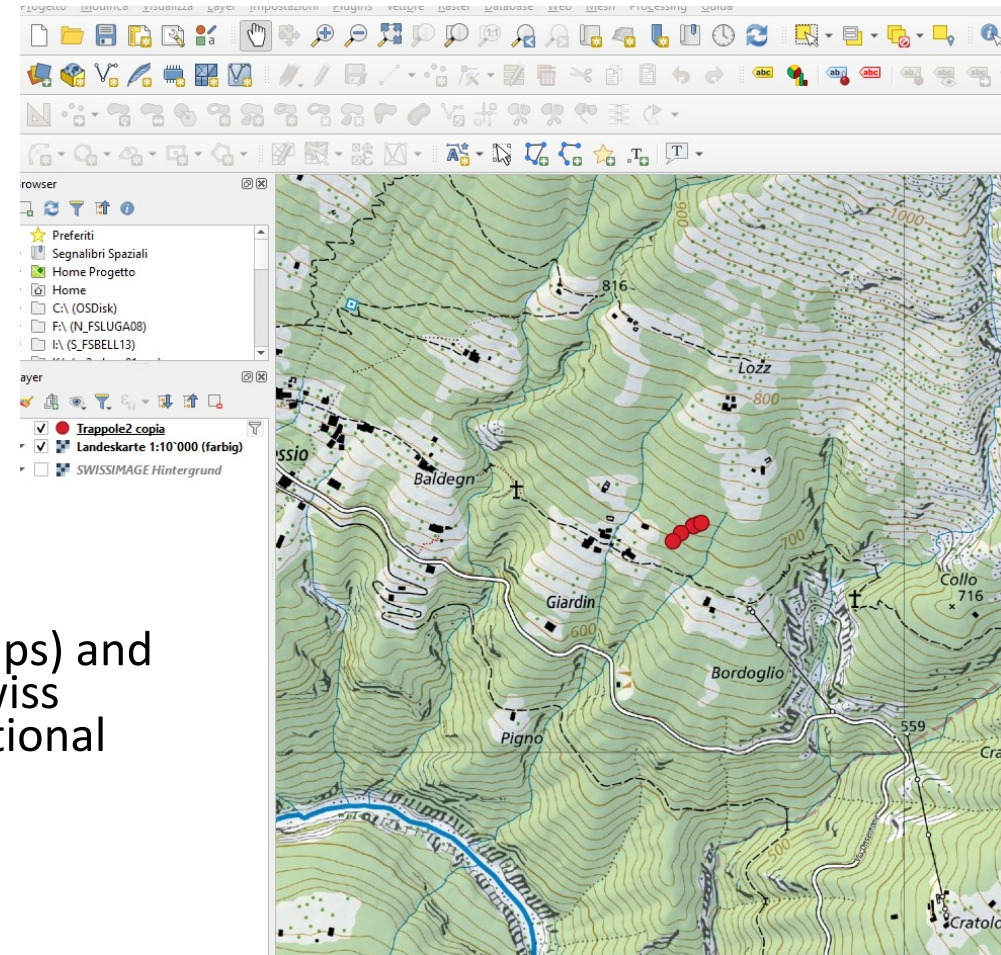
DIGITISATION PROBLEMS

Sex

- DarwinCore has: 1 field [dwc:sex] to be filled with female, male, hermaphrodite

 BUT there may be several individuals of both sexes in a sample (vial)

- MCSN uses:
    - ➢ NoFemmine; NoMaschi; NoIndefinito; NoOperaie

Date

- A field for uncertain dates is lacking, especially for old collections.
    - ➢ p. ex ante 1944
- It is punctual, and enter sampling periods (when using traps) is not user friendly:
    - DarwinCore uses: dwc:startDayOfYear & dwc:endDayOfYear p.ex 1.01-15.02.2019 = ISO 8601-1:2019 & ISO 8601-46:2019
    - MCSN uses: Ginizio; Minizio; Ainizio & Gfine; Mfine; Afine, V
    - Info fauna uses: D, M, YYYY & V
    - GBIF uses: DAYCOLLECTED, MONTHCOLLECTED, YEARCOLLECTED, ENDOFPERIODDAY, ENDOFPERIODMONTH, ENDOFPERIODYEAR

DIGITISATION PROBLEMS

History of identifications

- With the proposed fields it is not possible to keep track of the history of determinations/revisions/change of names of objects in the collection.

- dwc:verbatimIdentification: does it always have to be filled in or only if there is an error or abbreviation?

DIGITISATION PROBLEMS

- Training of staff
- Why data standardisation? It worked until now…
- Clear definition of empty vs. unknown
- Unterstanding DwC – GitHub page is not intuitive
- In-house DwC property explanation and training
- Definition of minimum and maximum DwC property usage to be able to synchronize with CMS
- High degree of self-discipline needed
- QC and QA mechanisms should be implemented

DISCUSSION

- Training of staff

- Why data standardisation? It worked until now…

- Unterstanding DwC – GitHub page is not intuitive

- In-house DwC property explanation and training

- Definition of minimum and maximum DwC property usage to be able to synchronize with CMS

- High degree of self-discipline needed

- QC and QA mechanisms should be implemented

And you?

What problems have you encountered?

Do you propose solutions?…. let's start the discussion!