Introduction to Darwin Core Archives

Luke Marsden

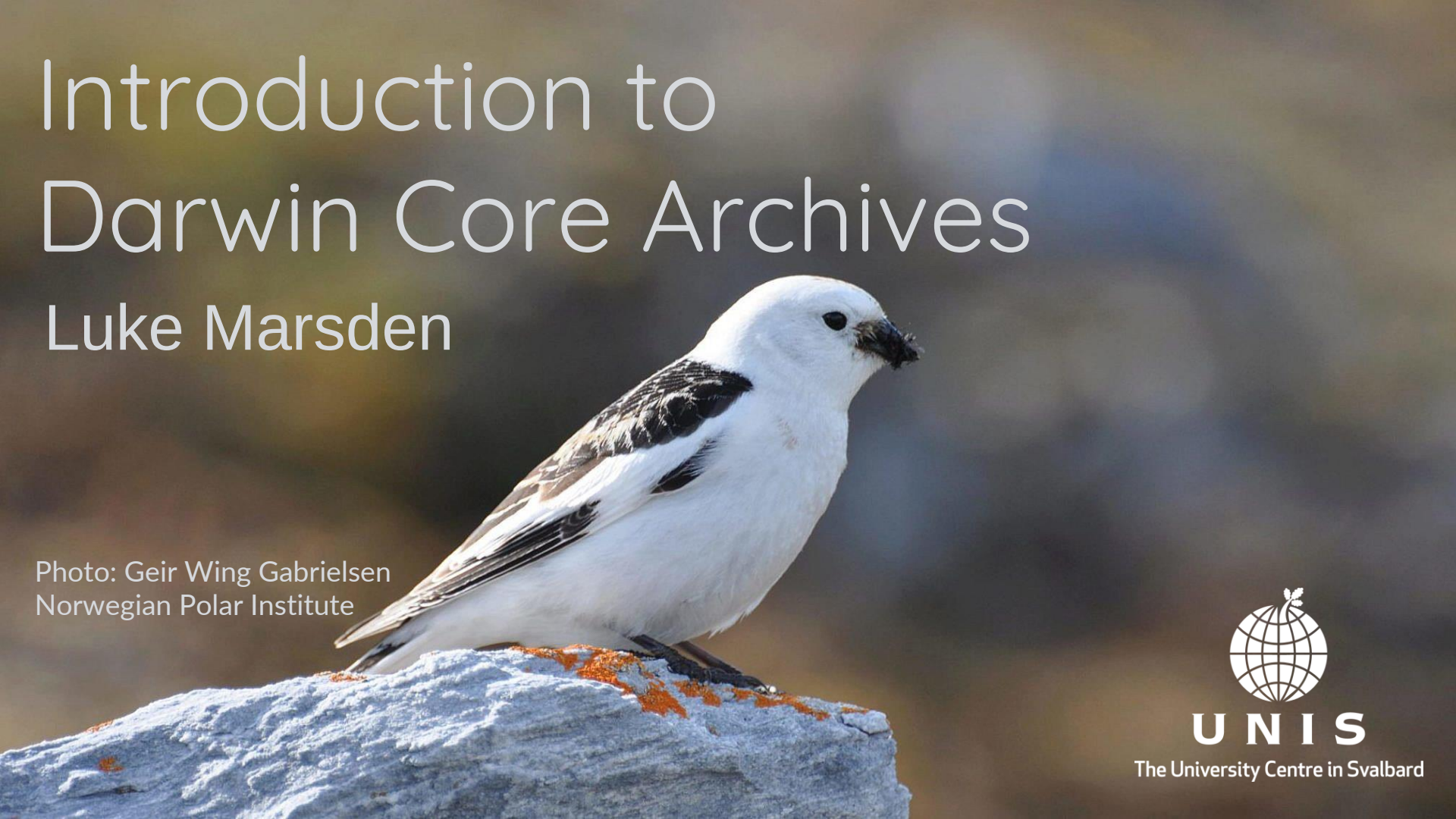
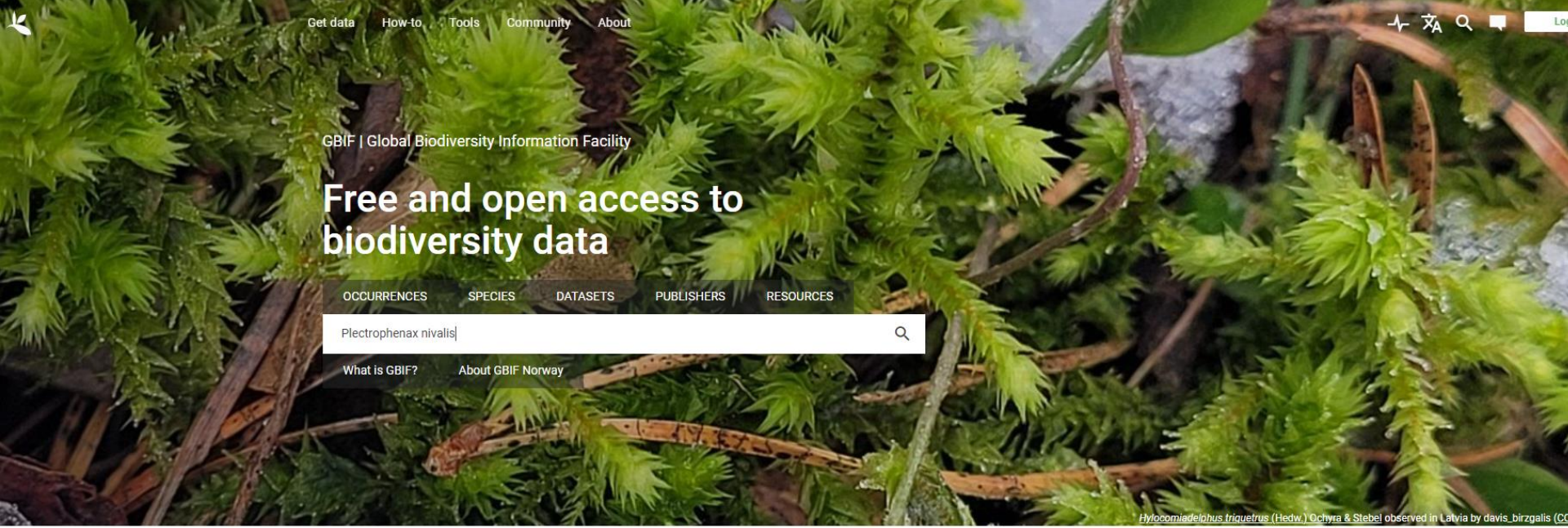


Photo: Geir Wing Gabrielsen
Norwegian Polar Institute

UNIS
The University Centre in Svalbard



2,162,818,707
Occurrence records

68,978
Datasets

1,825
Publishing institutions

7,149
Peer-reviewed papers using data

Occurrence search

gbif.org/occurrence/map?taxon_key=2491719

Get data | How-to | Tools | Community | About | Login

Occurrences

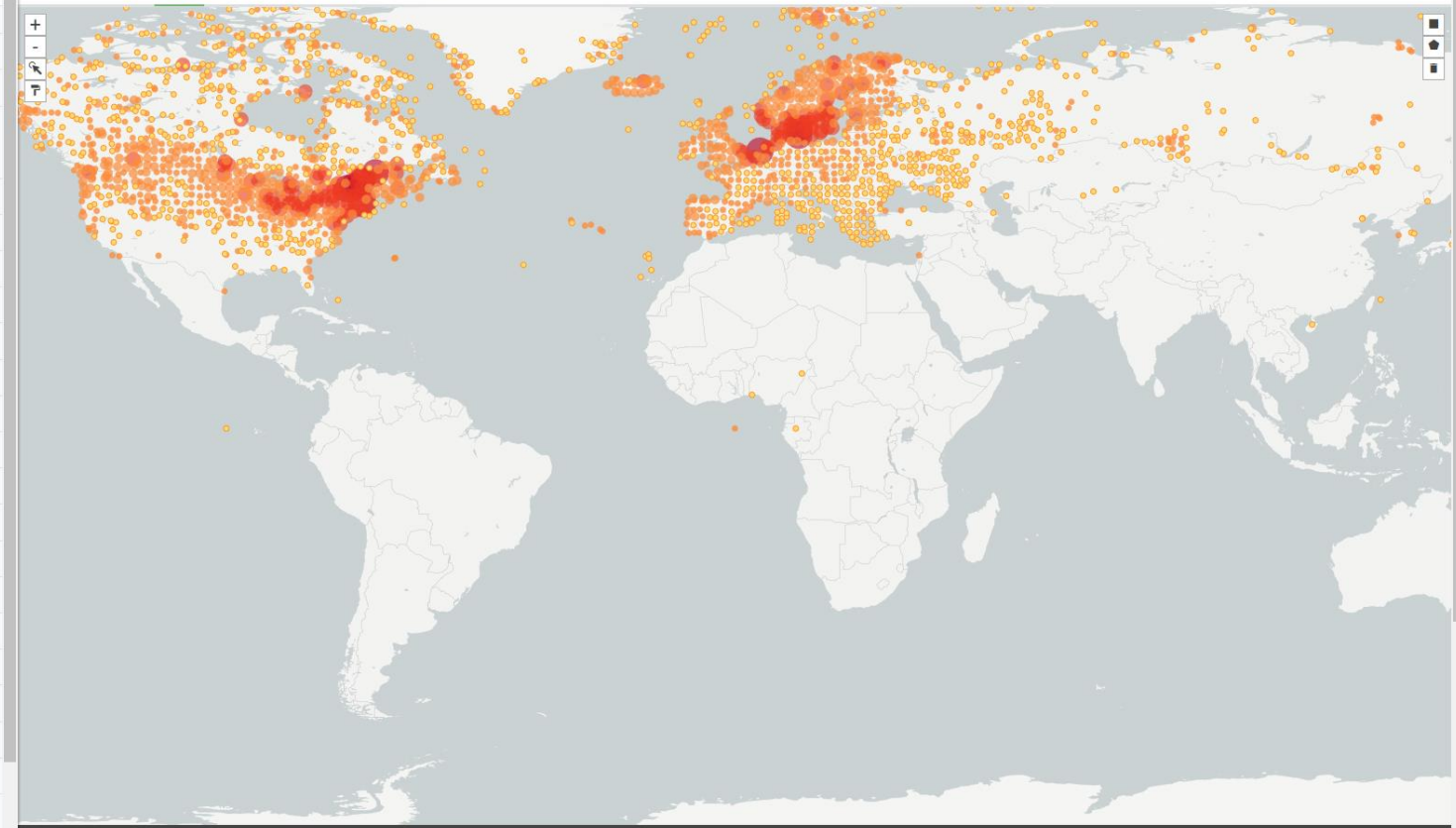
Search all fields

Simple | Advanced

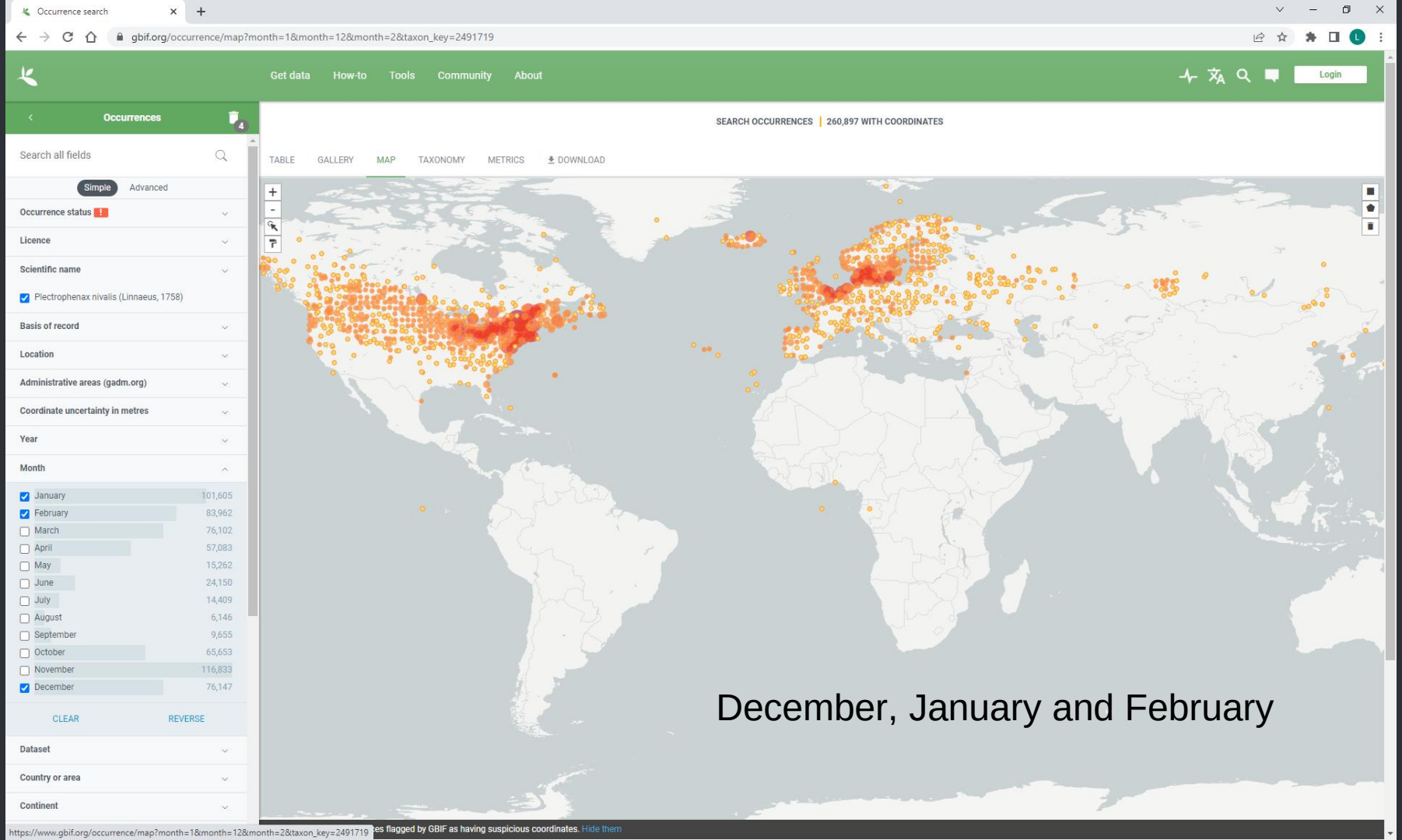
- Occurrence status
- Licence
- Scientific name
  - Plectrophenax nivalis (Linnaeus, 1758)
- Basis of record
- Location
- Administrative areas (gadm.org)
- Coordinate uncertainty in metres
- Year
- Month
- Dataset
- Country or area
- Continent
- Issues and flags
- Media type
- Publisher
- Institution code
- Collection code
- Catalogue number
- Type status
- IUCN Global Red List Category

SEARCH OCCURRENCES | 648,333 WITH COORDINATES

TABLE | GALLERY | MAP | TAXONOMY | METRICS | DOWNLOAD



This map contains occurrences flagged by GBIF as having suspicious coordinates. Hide them


Occurrence search
gbif.org/occurrence/map?month=1&month=12&month=2&taxon_key=2491719
Get data
How-to
Tools
Community
About
Login
Occurrences
Search all fields
Simple
Advanced
Occurrence status
Licence
Scientific name
Plectrophenax nivalis (Linnaeus, 1758)
Basis of record
Location
Administrative areas (gadm.org)
Coordinate uncertainty in metres
Year
Month
January 101,605
February 83,962
March 76,102
April 57,083
May 15,262
June 24,150
July 14,409
August 6,146
September 9,655
October 65,653
November 116,833
December 76,147
CLEAR
REVERSE
Dataset
Country or area
Continent
SEARCH OCCURRENCES | 260,897 WITH COORDINATES
TABLE
GALLERY
MAP
TAXONOMY
METRICS
DOWNLOAD
December, January and February
https://www.gbif.org/occurrence/map?month=1&month=12&month=2&taxon_key=2491719

Occurrence status !
Licence
Scientific name
Plectrophenax nivalis (Linnaeus, 1758)
Verbatim scientific name
Basis of record
Location
Administrative areas (gadm.org)
Coordinate uncertainty in metres
Year
Month
Dataset
Country or area
Continent

Issues and flags
Media type
Publisher
Hosting organization
Institution code
Institution
Collection code
Collection
Catalogue number
Type status
IUCN Global Red List Category
Recorded by
Recorded by ID
Identified by

Elevation
Depth
Locality
Water body
State province
Repatriated
Is in cluster
DWCA extension
Protocol
Event ID
Sampling protocol
Data network
Programme
Project

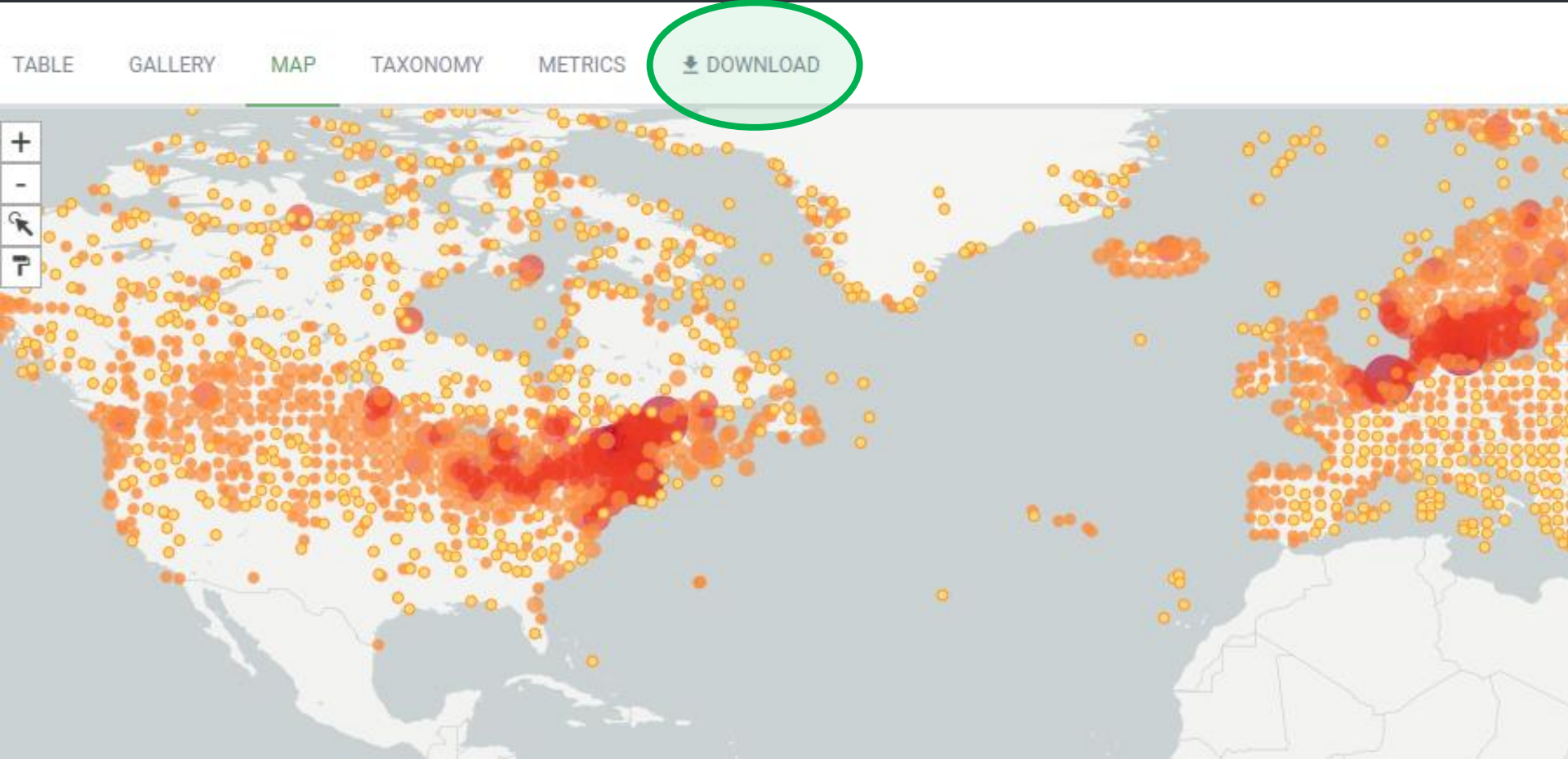


TABLE
GALLERY
MAP
TAXONOMY
METRICS
DOWNLOAD

SEARCH OCCURRENCES | 263,344 RESULTS

TAXONOMY METRICS DOWNLOAD

DOWNLOAD OPTIONS

	Raw data	Interpreted data	Multimedia	Coordinates	Format	Estimated data size
SIMPLE	✗	✓	✗	✓ (if available)	Tab-delimited CSV (for use in Excel, etc.) ?	**142 MB** (31 MB zipped for download)
DARWIN CORE ARCHIVE	✓	✓	✓ (links)	✓ (if available)	Tab-delimited CSV (for use in Excel, etc.) ?	**434 MB** (96 MB zipped for download)
SPECIES LIST	✗	✓	✗	✗	Tab-delimited CSV (for use in Excel, etc.) ?	

DOWNLOAD | 25 MAY 2022

Under processing

DOI 10.15468/dl.9fp7jn

Parent DOI

Preparing CANCEL

PLEASE USE THIS CITATION IN PUBLICATIONS

GBIF.org (25 May 2022) GBIF Occurrence Download https://doi.org/10.15468/dl.9fp7jn

Copy

FILTER APPLIED 25 MAY 2022

RERUN QUERY

The download has been started and is currently being processed.

Please expect up to 3 hours for the download to complete. Most downloads will complete within 15 minutes.

A notification email with a link to download the results will be sent to the following address once ready: lhmarsden@hotmail.com

Licence: Unspecified

Make sure to read the data user agreement and citation guidelines.

API

OCCURRENCE DATASET | REGISTERED AUGUST 6, 2012

Norwegian Species Observation Service

Published by The Norwegian Biodiversity Information Centre (NBIC)

The Norwegian Biodiversity Information Centre . • Hoem S

DATASET METRICS ACTIVITY DOWNLOAD HOME PAGE

27,344,440 OCCURRENCES | 727 CITATIONS

Citizen science species observations from the Species Observation Service in Norway (Artsobservasjoner).

ARTSDATABANKEN

Publication date: May 24, 2022

Metadata last modified: May 24, 2022

Hosted by: The Norwegian Biodiversity Information Centre (NBIC)

Licence: CC BY 4.0

How to cite DOI 10.15468/zjbzel

Child DOI

27,344,440 Occurrences

100% With taxon match

99.9% With coordinates

99.9% With year

27,344,384 GEOREFERENCED RECORDS

Read more about literature, how it's discovered and linked to GBIF-mediated data.

Climatic heterogeneity and low precipitation dictate the global distribution of range-edge hotspots of major tree species

Literature

Lerner, D. Fernandez-Martinez, M. Livne-Luzon, S. Belmaker, J. Penuelas, J. Klein, T. (2022) Research Square
Abstract Understanding the causes of the arrest of species distributions has been a fundamental question in ecology and evolution. For trees, these questions are of particular interest due to their long lifespan and sessile nature. A surge in data-availability evokes for a macro-ecological analysis ...

Range-edge • climate change • global • latitude diversity gradient • species distribution

Preprint Open access

Data referenced in study DOI 10.15468/dl.ajen6k

Species Identification Based on DNA of Selected Mushrooms from Myanmar (2) Geastrum courtecuissei and Phallus merulinuswith Disjunct Distribution?

Literature

Hosaka, K. Nam, K. Linn, W. Aung, M. (2022) Bulletin of the National Museum of Nature and Science. Series B, Botany
To clarify biogeographic patterns of two mushroom species (Phallus merulinus and Geastrum courtecuissei) previously reported from Myanmar, sequence data of the internal transcribed spacer (ITS) region of nuclear ribosomal DNA were retrieved from GenBank. The BLAST search and phylogenetic analyses of...

BLAST • Geastrales • Phallales • Phallomycetidae • South America • Southeast Asia

Journal article Peer-reviewed

Data referenced in study DOI 10.15468/dl.eqhc8y DOI 10.15468/dl.h7yfrm

中国历史时期猕猴桃属（Actinidia）遗存的发现与利用

Literature

杨, 晗. (2022) ()
(no abstract available)

Generic

Data referenced in study DOI 10.15468/dl.9x7seq

Google Scholar

Update on the development of the Agouti platform for collaborative science with camera traps and a tool for wildlife abundance estimation

Literature

Casaer, J. Desmet, P. ENETWILD Consortium, . Jansen, P. Liefting, Y. Rowcliffe, J. (2022) EFSA Supporting Publications
Camera traps (CT) provide an easy and non-invasive way to study wildlife. It is also possible to estimate densities if accurate protocols are followed in a standardised way and additional parameters are estimated. However, the processing and storage of the thousands of images that a typical CT study...

Journal article Open access Peer-reviewed

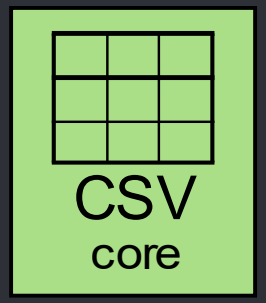
Data referenced in study DOI 10.15468/dl.up2acp

How is this possible?!

- Standardised file structure
  - Darwin Core Archive
- Rich and common metadata
- Controlled vocabularies
  - Darwin Core Terms
  - Ecological Metadata Language

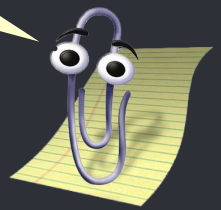
Darwin Core Archive

Darwin core terms

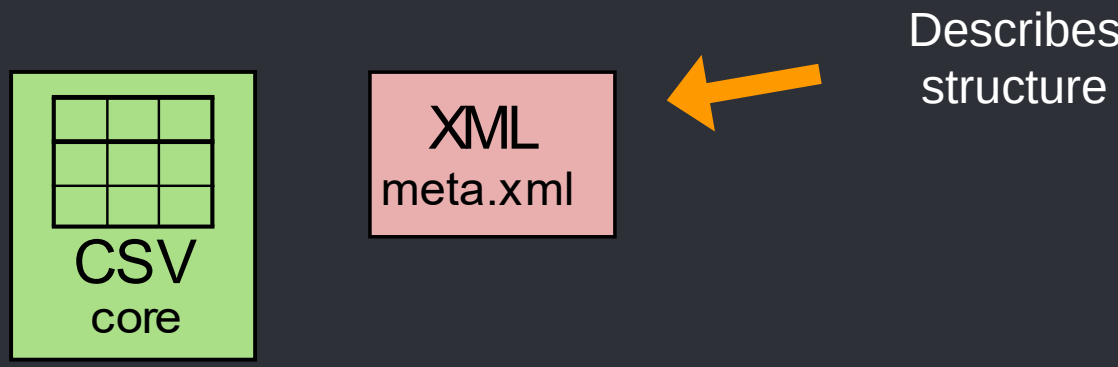


occurrenceID	basisOfRecord	scientificName	eventDate	…
094a5053-6877-46cc-8c0a-0e35f84e402c	Human_observation	Plectrophenax Nivalis (Linnaeus, 1758)	2022-04-29	
a517433f-365e-47da-ba4b-0294db00db6c	Fossil_specimen	Plectrophenax Nivalis (Linnaeus, 1758)	2022-04	
925aa4a9-d8bb-4429-a4ee-94cdd6760b1a	Machine_observation	Plectrophenax Nivalis (Linnaeus, 1758)	2022-04-29T15:19:23Z	

https://dwc.tdwg.org/terms/



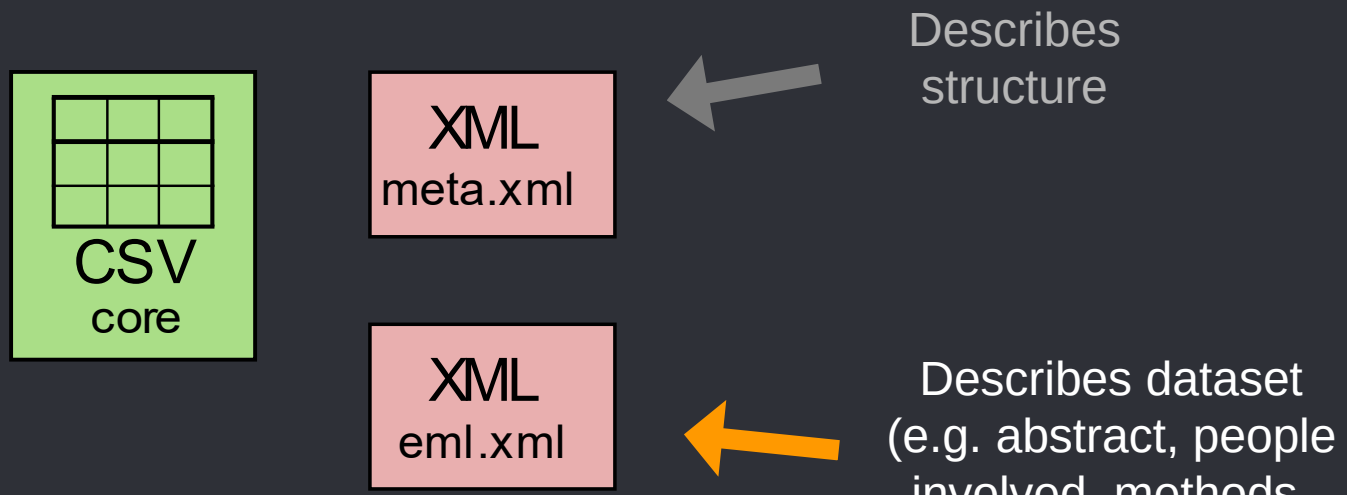
Darwin Core Archive



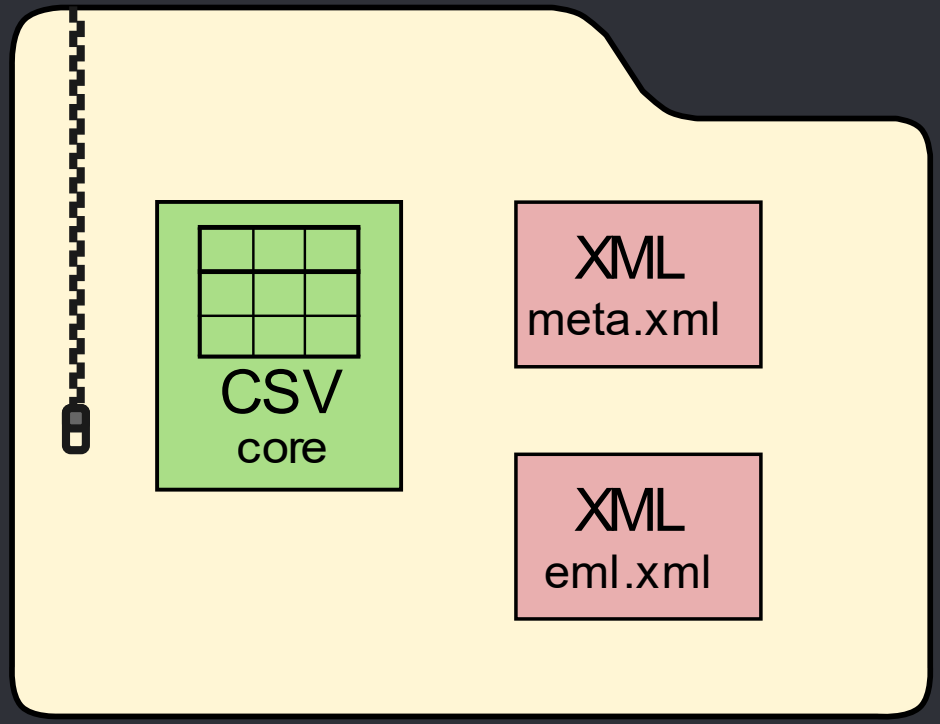
meta.xml

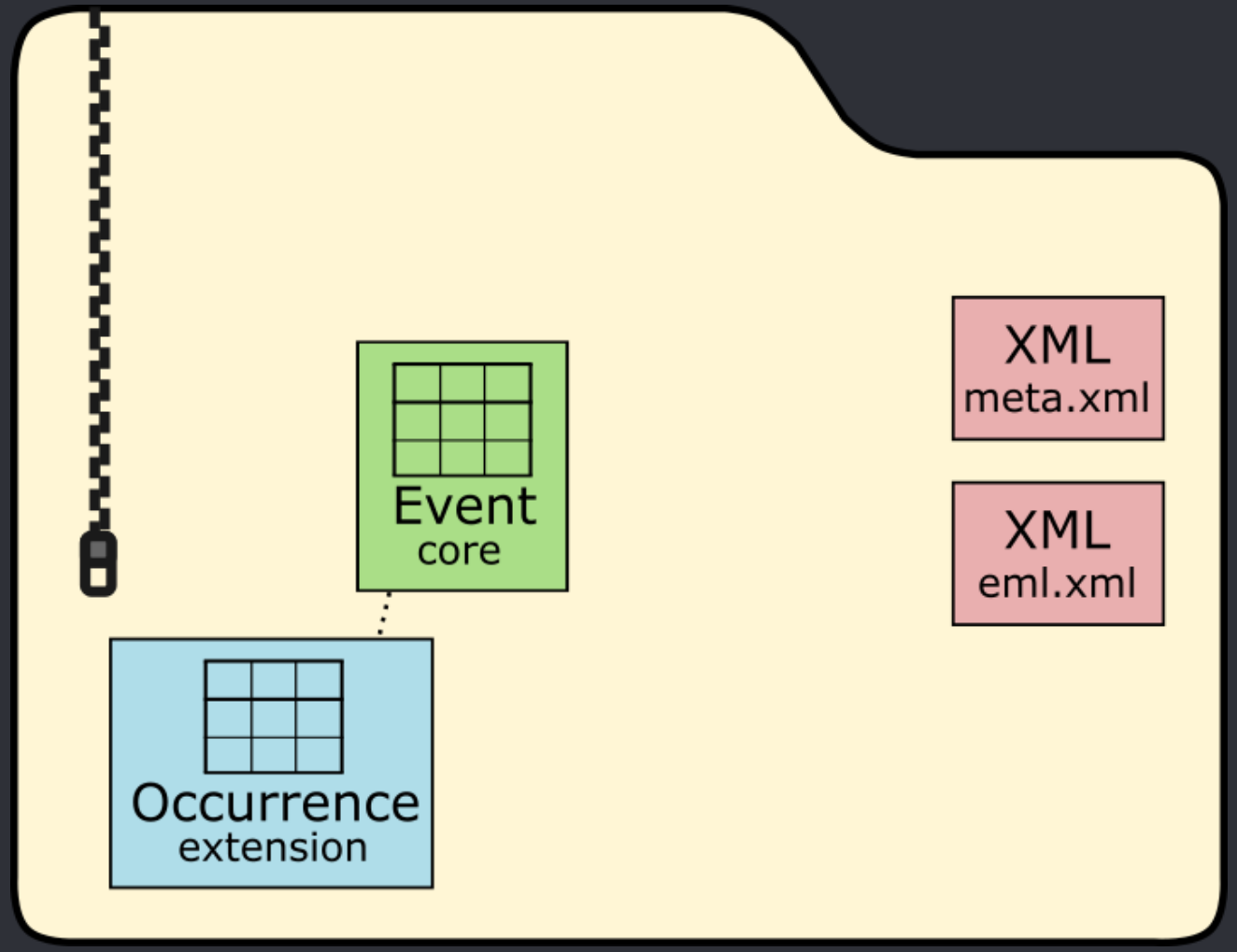
```xml
<archive
    xmlns="http://rs.tdwg.org/dwc/text/" metadata="eml.xml">
    <core encoding="UTF-8" fieldsTerminatedBy="\t" linesTerminatedBy="\n" fieldsEnclosedBy=""
        ignoreHeaderLines="1" rowType="http://rs.tdwg.org/dwc/terms/Occurrence">
        <files>
            <location>occurrence.txt</location>
        </files>
        <id index="0" />
        <field index="1" term="http://rs.tdwg.org/dwc/terms/occurrenceID"/>
        <field index="2" term="http://rs.tdwg.org/dwc/terms/basisOfRecord"/>
        <field index="3" term="http://rs.tdwg.org/dwc/terms/scientificName"/>
        <field index="4" term="http://rs.tdwg.org/dwc/terms/eventDate"/>
        <field index="5" term="http://rs.tdwg.org/dwc/terms/recordedBy"/>
        <field index="6" term="http://rs.tdwg.org/dwc/terms/occurrenceStatus"/>
        <field index="7" term="http://rs.tdwg.org/dwc/terms/samplingProtocol"/>
        <field index="8" term="http://rs.tdwg.org/dwc/terms/decimalLatitude"/>
        <field index="9" term="http://rs.tdwg.org/dwc/terms/decimalLongitude"/>
        <field index="10" term="http://rs.tdwg.org/dwc/terms/geodeticDatum"/>
        <field index="11" term="http://rs.tdwg.org/dwc/terms/coordinateUncertaintyInMeters"/>
        <field index="12" term="http://rs.tdwg.org/dwc/terms/georeferenceProtocol"/>
        <field index="13" term="http://rs.tdwg.org/dwc/terms/scientificNameID"/>
        <field index="14" term="http://rs.tdwg.org/dwc/terms/kingdom"/>
        <field index="15" term="http://rs.tdwg.org/dwc/terms/taxonRank"/>
        <field index="16" term="http://rs.tdwg.org/dwc/terms/vernacularName"/>
    </core>
</archive>
```

Darwin Core Archive



Darwin Core Archive




Event
core
Occurrence
extension
XML
meta.xml
XML
eml.xml

Event Core

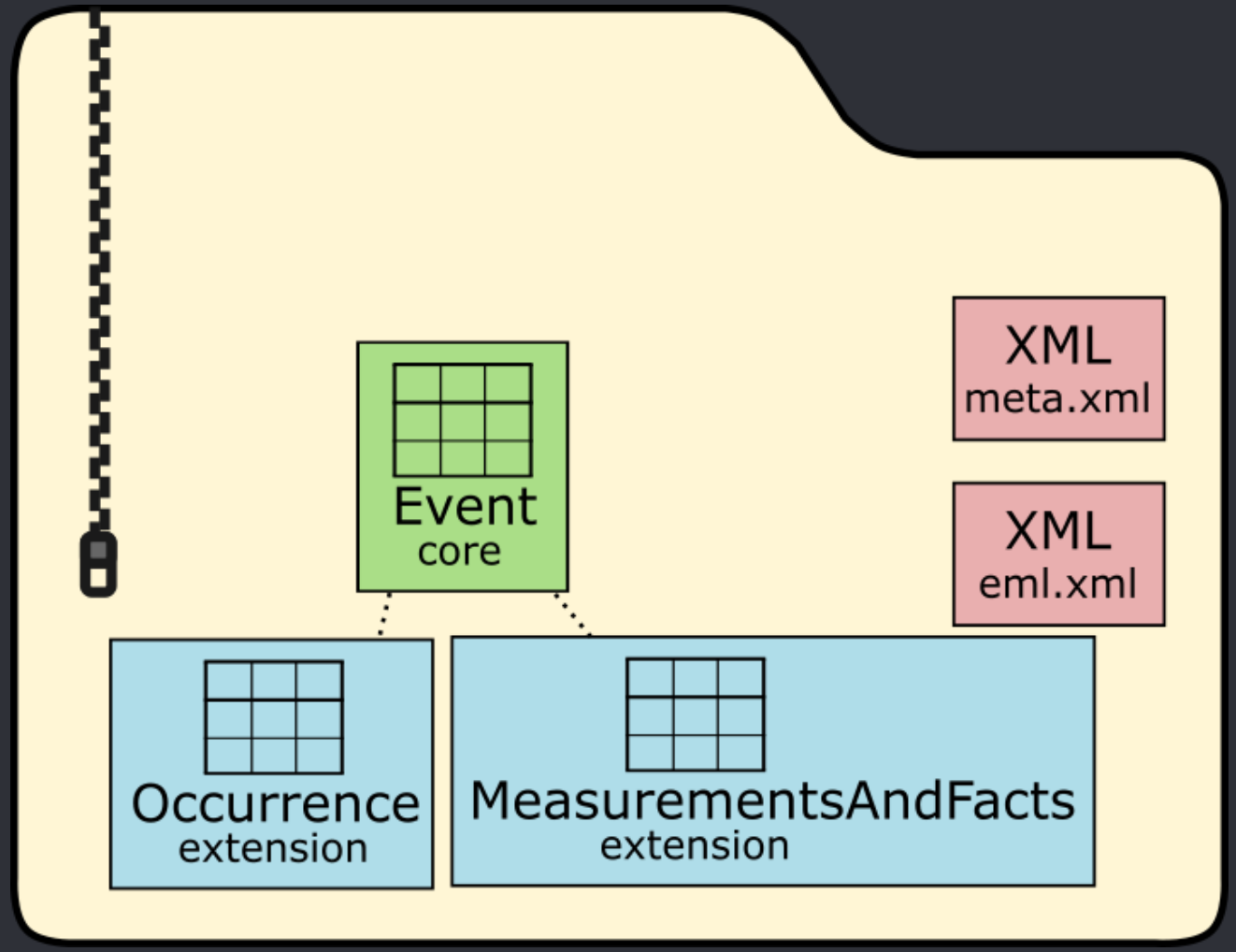
eventID	eventDate	decimalLatitude	samplingProtocol	…
UUID 1	2021-09-24T12:17:41Z	81.2534	Campelen Trawl	
UUID 2	2021-09-25T19:20:58Z	80.1398	Campelen Trawl	

Occurrence extension

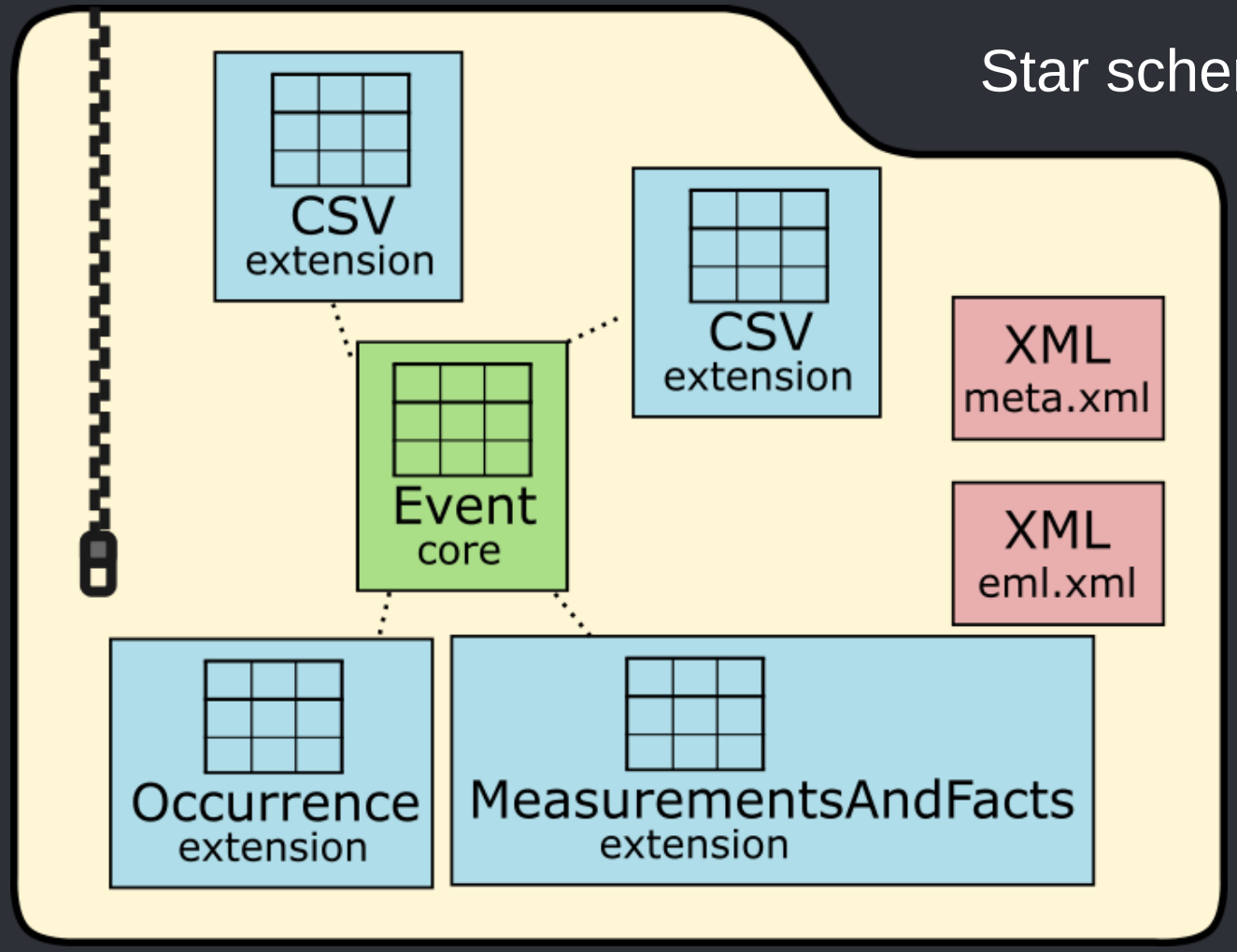
occurrenceID	eventID	scientificName	…
UUID 3	**UUID 1**	Boreogadus saida	
UUID 4	**UUID 1**	Gadus morhua	
UUID 5	**UUID 2**	Melanogrammus aeglefinus	
UUID 6	**UUID 2**	Gadus morhua	

Why?

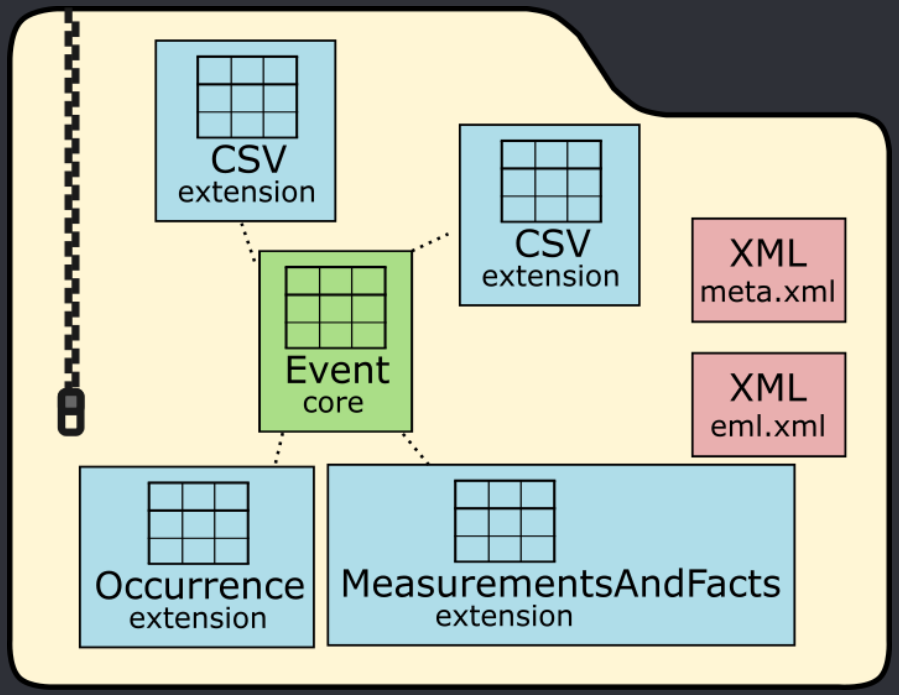
- Many-to-one
- More efficient
- Possible to register environment measurements (temperature, salinity…)


Event
core
XML
meta.xml
XML
eml.xml
Occurrence
extension
MeasurementsAndFacts
extension

All linked by eventID



Darwin Core Archive

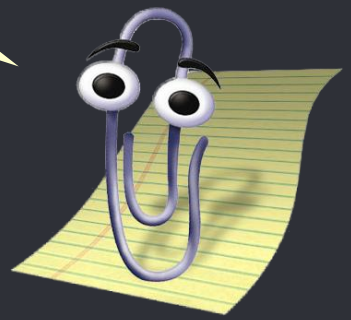


Easy method:

1. Excel templates for CSV files
2. Create XML files and archive using GBIF's Integrated Publishing Toolkit

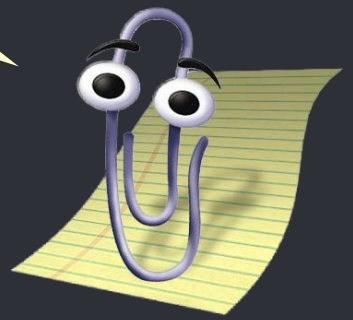
GBIF Darwin Core templates

https://www.gbif.org/news/82852/new-darwin-core-spreadsheet-templates-simplify-data-preparation-and-publishing



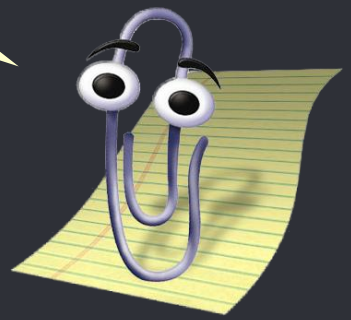
Nansen Legacy templates

https://sios-svalbard.org/cgi-bin/darwinsheet/?setup=aen

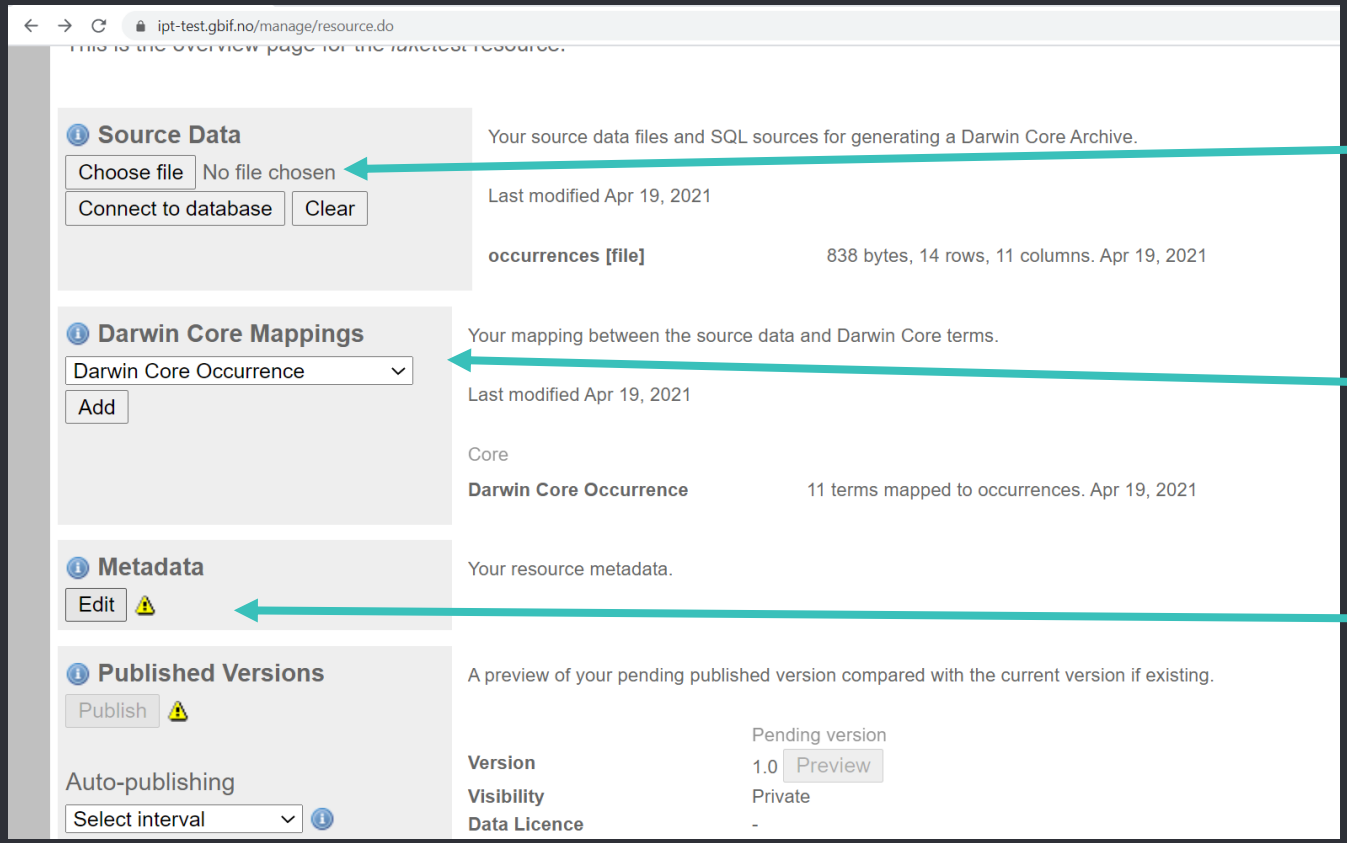


Integrated Publishing Toolkit

https://www.gbif.org/ipt



Integrated Publishing Toolkit



Your Excel files

Mapping columns to Darwin Core terms (semi-automatic)

Fill in form to describe dataset

Useful links

- Darwin Core terms:
  - https://dwc.tdwg.org/terms/
- OBIS Darwin Core manual
  - https://manual.obis.org/
- GBIF Darwin Core templates
  - https://www.gbif.org/news/82852/new-darwin-core-spreadsheet-templates-simplify-data-preparation-and-publishing
- Nansen Legacy template generator
  - https://sios-svalbard.org/cgi-bin/darwinsheet/?setup=aen
- GBIF Integrated Publishing Toolkit
  - https://www.gbif.org/ipt