Nansen Legacy tools for data management

Luke Marsden (lukem@unis.no)
Pål Ellingsen, Øystein Godøy, Tove Gabrielsen
Things we will cover

• Introduce the project

• How are data logged?

• Where are the data?

• Introduce tools that can be used by broader research community
Nansen Legacy Project

• 210 researchers, 10 Norwegian research institutions

• Multidisciplinary

• Northern Barents Sea, multiple cruises

• FAIR data management principles (Wilkinson et al., 2016)
  • Findable, Accessible, Interoperable, Reusable
Overview of how data are logged in the project

• Template generator to log samples/events
  • https://www.sios-svalbard.org/cgi-bin/darwinsheet/?setup=aen

• Samples labelled physically, logged electronically. Data without physical sample also logged.

• Samples/events searchable on SIOS.
  • https://sios-svalbard.org/reports/aen_multi

• Sampling protocols documented and published (link above).
UUIDs

• Universally unique ID

• Used for every sample/event

• 32 digit hex number, can be represented as data matrix for simplicity:

  f69d4072-49be-11eb-b378-0242ac130002

• Chance of collision: $10^{-14}$
UUIDs

• Universally unique ID

• Used for every sample/event

• 32 digit hex number, can be represented as data matrix for simplicity:

  f69d4072-49be-11eb-b378-0242ac130002

• Chance of collision: $10^{-14}$
Hierarchy

Gears/Activities
CTD, Box core, Drone, Dive…
Niskin bottles, L-ADCPs…

Samples
UUID #5
UUID #6

Subsamples
UUID #11
UUID #12

Introduce the project → How are data logged? → Where are the data? → Introduce tools
Hierarchy: Example 1

Introduce the project → How are data logged? → Where are the data? → Introduce tools
Hierarchy: Example 2

Gears/Activities

Dive missions

100 ml water

Photo #1

Photo #2

Drone flight

Introduce the project → How are data logged? → Where are the data? → Introduce tools
Stations

• 25 transect stations, including 7 process stations

• Around 500 stations visited in all
  • Transects, moorings, one off deployments…
Introduce the project → How are data logged? → Where are the data? → Introduce tools

## Metadata

- Searchable online
- Rich metadata
- Contact persons
- Linked to physical samples
Metadata

- Searchable online
- Rich metadata
- Contact persons
- Linked to physical samples
Introduce the project → How are data logged? → Where are the data? → Introduce tools

Metadata

• Searchable online

• Rich metadata

• Contact persons

• Linked to physical samples

Search for all CTD Chlorophyll A data from 2018
<table>
<thead>
<tr>
<th>Event Date</th>
<th>Station Name</th>
<th>Gear Type</th>
<th>Sample Type</th>
<th>Event ID</th>
<th>Parent Event ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018-08-20</td>
<td>SICE3</td>
<td>CTD w/bottles</td>
<td>Chlorophyll a tot</td>
<td>1a5fa299-9e0d-11e8-aecb-8c164557e456</td>
<td>dfdd710a-9c88-11e8-91c9-005055a2b019</td>
</tr>
<tr>
<td>2018-08-20</td>
<td>SICE3</td>
<td>CTD w/bottles</td>
<td>Chlorophyll a tot</td>
<td>1a5fa29b-9e0d-11e8-aecb-8c164557e456</td>
<td>dfdd7104-9c88-11e8-91c9-005055a2b019</td>
</tr>
<tr>
<td>2018-08-20</td>
<td>SICE3</td>
<td>CTD w/bottles</td>
<td>Chlorophyll a tot</td>
<td>1a5fa29a-9e0d-11e8-aecb-8c164557e456</td>
<td>dfdd710c-9c88-11e8-91c9-005055a2b019</td>
</tr>
<tr>
<td>2018-08-20</td>
<td>SICE3</td>
<td>CTD w/bottles</td>
<td>Chlorophyll a tot</td>
<td>1a5fa29c-9e0d-11e8-aecb-8c164557e456</td>
<td>dfdd710e-9c88-11e8-91c9-005055a2b019</td>
</tr>
<tr>
<td>2018-08-20</td>
<td>SICE3</td>
<td>CTD w/bottles</td>
<td>Chlorophyll a tot</td>
<td>1a5fa298-9e0d-11e8-aecb-8c164557e456</td>
<td>dfdd7112-9c88-11e8-91c9-005055a2b019</td>
</tr>
<tr>
<td>2018-08-20</td>
<td>SICE3</td>
<td>CTD w/bottles</td>
<td>Chlorophyll a tot</td>
<td>1a5fa29f-9e0d-11e8-aecb-8c164557e456</td>
<td>dfdd7105-9c88-11e8-91c9-005055a2b019</td>
</tr>
<tr>
<td>2018-08-20</td>
<td>SICE3</td>
<td>CTD w/bottles</td>
<td>Chlorophyll a tot</td>
<td>1a5fa29e-9e0d-11e8-aecb-8c164557e456</td>
<td>dfdd7100-9c88-11e8-91c9-005055a2b019</td>
</tr>
</tbody>
</table>
Introduce the project → How are data logged? → Where are the data? → Introduce tools

Metadata

- Searchable online
- Rich metadata
- Contact persons
- Linked to physical samples
Metadata: Other ways to search

- Sample Search
  - Sample ID Search
  - Multiple ID Search
  - Station Search
  - Station Overview
  - Full sample log as CSV (without metadata)
  - Excel Template Generator

- Nansen Legacy Documents
  - Sampling Protocol (v1)
  - Sampling Protocol (v2)
  - Sampling Protocol (v3)
  - Sampling Protocol (v4.2)
  - Sampling Protocol (v5)
  - Sampling Protocol (v6)
Published data

• All data accessible via SIOS

• Single access point

• Links to data repositories
Published data

For tutorials on how to find data and how to use the results page, click here: https://sios-svalbard.org/tutorials
Published data
Published data

Interact directly with selected products from the map by clicking on the highlighted features. Select products from the table below to
<table>
<thead>
<tr>
<th>Dataset name</th>
<th>Institutions</th>
<th>Abstract</th>
<th>Collection period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chlorophyll A and phaeopigments Nansen Legacy cruise 2019, station P2</td>
<td>University Centre in Svalbard</td>
<td>This dataset is a collection of the acid-corrected chlorophyll A and phaeopigments measurements taken as part of the Nansen Legacy project. The data is from a single depth profile from station P1 taken on 2019-12-12 at 31.219021941666666 and 75.9996045666667%. Samples were collected from different depths of the water column.</td>
<td>2019-12-12T06:00:00Z to 2019-12-12T06:00:00Z</td>
</tr>
<tr>
<td>Chlorophyll A and phaeopigments Nansen Legacy cruise 2019, station P1</td>
<td>University Centre in Svalbard</td>
<td>This dataset is a collection of the acid-corrected chlorophyll A and phaeopigments measurements taken as part of the Nansen Legacy project. The data is from a single depth profile from station P2 taken on 2019-12-10 at 34.06550771666667% and 77.4996566666667%. Samples were collected from different depths of the water column.</td>
<td>2019-12-10T06:00:00Z to 2019-12-10T06:00:00Z</td>
</tr>
<tr>
<td>Chlorophyll A and phaeopigments Nansen Legacy cruise 2019, station P3</td>
<td>University Centre in Svalbard</td>
<td>This dataset is a collection of the acid-corrected chlorophyll A and phaeopigments measurements taken as part of the Nansen Legacy project. The data is from a single depth profile from station P3 taken on 2019-12-09 at 33.99431746666667% and 78.4996358333333%. Samples were collected from different depths of the water column.</td>
<td>2019-12-09T06:00:00Z to 2019-12-09T06:00:00Z</td>
</tr>
<tr>
<td>Chlorophyll A and phaeopigments Nansen Legacy cruise 2019, station P4</td>
<td>University Centre in Svalbard</td>
<td>This dataset is a collection of the acid-corrected chlorophyll A and phaeopigments measurements taken as part of the Nansen Legacy project. The data is from a single depth profile from station P4 taken on 2019-12-06 at 34.094391799999999% and 78.7326178333333%. Samples were collected from different depths of the water column.</td>
<td>2019-12-06T06:00:00Z to 2019-12-06T06:00:00Z</td>
</tr>
</tbody>
</table>
Published data

Seamless access to Norwegian marine data

Abstract
This dataset is a collection of the acid-corrected chlorophyll A and phaeopigments measurements taken as part of the Nansen Legacy project (www.arvenettetmansen.com). The data is from a single depth profile from station P1 taken on 2019-12-12 at 31.21902/16.69000000E and 75.99980/66.66677N. Samples were collected from different depths of the water column.

Scientific keywords:
- EARTH SCIENCE
- OCEANS
- OCEAN CHEMISTRY
- CHLOROPHYLL

Key words:
- Chlorophyll

Learn more about this dataset and how to download it from the NMDC website.
Sampling protocols

- Introduce the project → How are data logged? → Where are the data? → Introduce tools

### Nansen Legacy Tools
- Sample Search
- Sample ID Search
- Multiple ID Search
- Station Search
- Station Overview
- Full sample log as CSV (without metadata)
- Excel Template Generator

### Nansen Legacy Documents
- Sampling Protocol (v1)
- Sampling Protocol (v2)
- Sampling Protocol (v3)
- Sampling Protocol (v4.2)
- Sampling Protocol (v5)
- Sampling Protocol (v6)
Sampling protocols

• Methodological agreement between the involved researchers

• Continuity and comparable data throughout the 5 years sampling period

• An easily accessible overview over parameters sampled

• Easier cruise planning
Template generator

• Developed following discussions with Dag Endresen of GBIF Norway
  • Global Biodiversity Information Facility

• Heavily developed by Pål Ellingsen (my predecessor).

• Based on Darwin Core, some terms from NetCDF-CF, some other terms.

• Generate standardised templates with common vocabulary.

• Continuously developed and added to.
Darwin Core & NetCDF-CF

• Darwin Core:
  • Focuses mainly on primary biodiversity data (e.g. occurrences of a species)
  • Also ecological data (descriptors, traits, etc)
  • The main part of a Darwin Core Archive is a table (csv, excel) that has column names taken from a common vocabulary.

• NetCDF-CF:
  • Initially developed by atmospheric modellers
  • Developed further to include oceanography, also other disciplines (physical data).
  • Standard names searchable online, with descriptions and units.
Darwin Core Archive

- Can be created by template generator
- **Core file**
  - Columns of data
  - One row per record
- **Extension files**
- **Information on records**
- **meta.xml**
  - Describes structure
- **EML.xml**
  - Describes dataset as whole (e.g. abstract, people involved, methods, citations...)
- Many-to-one relationships
Template generator

Introduce the project → How are data logged? → Where are the data? → Introduce tools
Template generator

- Sample Search
- Sample ID Search
- Multiple ID Search
- Station Search
- Station Overview
- Full sample log as CSV (without metadata)

- Sampling Protocol (v1)
- Sampling Protocol (v2)
- Sampling Protocol (v3)
- Sampling Protocol (v4.2)
- Sampling Protocol (v5)
- Sampling Protocol (v6)

- Excel Template Generator
Template generator: Nansen Legacy setup
Template generator: Darwin Core setup

Introduce the project → How are data logged? → Where are the data? → Introduce tools

Excel Template Generator

Check the boxes next to the terms you want to include in your template and click the Create template button.

Darwin Core terms: A quick reference guide

Each box different DwC class
Template generator: Darwin Core setup

<table>
<thead>
<tr>
<th>EVENT</th>
<th>LOCATION</th>
<th>IDENTIFICATION</th>
<th>ORGANISM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day</td>
<td>Continent</td>
<td>Date Identified</td>
<td>Associated Occurrences</td>
</tr>
<tr>
<td>End Day Of Year</td>
<td>Coordinate Precision</td>
<td>Identification ID</td>
<td>Associated Organisms</td>
</tr>
<tr>
<td>Event Date</td>
<td>Coordinate Uncertainty In Meters</td>
<td>Identification Qualifier</td>
<td></td>
</tr>
<tr>
<td>Event ID</td>
<td>Country</td>
<td>Identification References</td>
<td></td>
</tr>
<tr>
<td>Event Remarks</td>
<td>Country Code</td>
<td>Identification Remarks</td>
<td></td>
</tr>
<tr>
<td>Event Time</td>
<td>County</td>
<td>Identification Verification Status</td>
<td></td>
</tr>
<tr>
<td>Field Notes</td>
<td>Decimal Latitude</td>
<td>Identified By</td>
<td></td>
</tr>
<tr>
<td>Field Number</td>
<td>Decimal Longitude</td>
<td>Type Status</td>
<td></td>
</tr>
<tr>
<td>Habitat</td>
<td>Footprint Spatial Fit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Month</td>
<td>Footprint SRS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parent Event ID</td>
<td>Footprint WKT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sampling Size Unit</td>
<td>Geodelic Datum</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sampling Size Value</td>
<td>Georeferenced By</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sampling Effort</td>
<td>Georeferenced Date</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sampling Protocol</td>
<td>Georeference Protocol</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Start Day Of Year</td>
<td>Georeference Remarks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbatim EventDate</td>
<td>Georeference Sources</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year</td>
<td>Georeference Verification Status</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Higher Geography ID [higherGeographyID]**

Higher Geography ID refers to the geographic region within which the location occurred. The recommended best practice is to use a persistent identifier from a controlled vocabulary such as the Getty Thesaurus of Geographic Names.

Example: "TGN: 1002002" for Prov. Tierra del Fuego, Argentina

No validation

*Template generator:* Darwin Core setup

Introduce the project → How are data logged? → Where are the data? → Introduce tools
Template generator: Darwin Core setup

Introduce the project → How are data logged? → Where are the data? → Introduce tools
Template generator: Darwin Core setup

Introduce the project → How are data logged? → Where are the data? → Introduce tools

Language of template. Not of this page

Dataset ID [datasetID]

No validation

Darwin core info (validation takes precedence for formatting):

An identifier for the set of data. May be a global unique identifier or an identifier specific to a collection or institution.
Template generator: Darwin Core setup

Introduce the project → How are data logged? → Where are the data? → Introduce tools

Column header, and corresponding Darwin Core term

Format requirements

Description of field
Introduce the project → How are data logged? → Where are the data? → Introduce tools

Template generator: Darwin Core setup

Excel Template Generator

Check the boxes next to the terms you want to include in your template and click the Create template button.

Darwin Core Terms: A quick reference

Disposition [disposition]

No validation

Darwin core info (validation takes precedence for formatting):

The current state of a specimen with respect to the collection identified in collectionCode or collectionID. Recommended best practice is to use a controlled vocabulary.

Examples: "in collection", "missing", "voucher elsewhere", "duplicates elsewhere"
Introduce the project → How are data logged? → Where are the data? → Introduce tools

Template generator: Darwin Core setup

<table>
<thead>
<tr>
<th>Basis of Record</th>
<th>Occurrence ID</th>
<th>Individual Count</th>
<th>Organism Quantity</th>
<th>Organism Quantity Type</th>
<th>Date</th>
<th>Country Code</th>
<th>Latitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>basisOfRecord</td>
<td>occurrenceID</td>
<td>individualCount</td>
<td>organismQuantity</td>
<td>organismQuantityType</td>
<td>eventDate</td>
<td>countryCode</td>
<td>decim</td>
</tr>
</tbody>
</table>

When pasting only use 'paste special' / 'paste only', selecting numbers and/or text

Darwin core term; row usually hidden for protection

One row per event/occurrence

Note on how to fill in selected cell
How do I create a Darwin Core Archive from this?

Introduce the project → How are data logged? → Where are the data? → Introduce tools

SDMS WG Training #4: How to create Darwin Core Archives for biological data

Matteo De Stefano

GBIF Integrated Publishing Toolkit

https://www.gbif.org/en/ipt
Excel Template Generator

Check the boxes next to the terms you want to include in your template and click the Create template button.

Darwin Core Terms: A quick reference guide

Create template

No validation

Darwin core info (validation takes precedence for formatting):

The age class or life stage of the biological individual(s) at the time the Occurrence was recorded. Recommended best practice is to use a controlled vocabulary.

Examples: "egg", "eft", "juvenile", "adult", "2 adults 4 juveniles"
<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Basis of Record</strong></td>
<td><strong>Occurrence ID</strong></td>
<td><strong>Individual Count</strong></td>
<td><strong>Date</strong></td>
<td><strong>Country Code</strong></td>
<td><strong>Latitude</strong></td>
<td><strong>Longitude</strong></td>
<td><strong>Geodetic Datum</strong></td>
<td><strong>Coordinate Uncertainty In Meters</strong></td>
<td><strong>Scientific Name</strong></td>
<td><strong>Kingdom</strong></td>
<td></td>
</tr>
</tbody>
</table>

When pasting only 'paste special' / 'paste only', selecting numbers and/or text.
<table>
<thead>
<tr>
<th>Basis of Record</th>
<th>Occurrence ID</th>
<th>Individual Count</th>
<th>Date</th>
<th>Country Code</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Geodetic Datum</th>
<th>Coordinate Uncertainty In Meters</th>
<th>Scientific Name</th>
<th>Kingdom</th>
</tr>
</thead>
<tbody>
<tr>
<td>HumanObservation ab10f59-5308-45bd-aeeb-9d77e808e014</td>
<td>1</td>
<td>2021-04-15</td>
<td>SIM</td>
<td>78.2232</td>
<td>15.6267</td>
<td>WGS84</td>
<td>10 Rangifer tarandus platyrhynchus Animalia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HumanObservation 1fd5188-6d99-46ed-8d5b-08070dc4f2f</td>
<td>2</td>
<td>2021-04-16</td>
<td>SIM</td>
<td>78.2232</td>
<td>15.6267</td>
<td>WGS84</td>
<td>10 Rangifer tarandus platyrhynchus Animalia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HumanObservation c515b4a-9c6d-4856-45f6-445c-b4493c603e36</td>
<td>3</td>
<td>2021-04-17</td>
<td>SIM</td>
<td>78.2232</td>
<td>15.6267</td>
<td>WGS84</td>
<td>10 Rangifer tarandus platyrhynchus Animalia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HumanObservation 601c1d1-829e-4c08-8a12-653c177178</td>
<td>1</td>
<td>2021-04-18</td>
<td>SIM</td>
<td>78.2232</td>
<td>15.6267</td>
<td>WGS84</td>
<td>10 Rangifer tarandus platyrhynchus Animalia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>I</td>
<td>J</td>
<td>K</td>
</tr>
<tr>
<td>------------------------</td>
<td>----------------------------------------</td>
<td>-----------------------</td>
<td>---------</td>
<td>------------------------</td>
<td>----------------------</td>
<td>------------------------</td>
<td>------------------------</td>
<td>------------------------</td>
<td>-----------------------------------------</td>
<td>-----------------------------------------</td>
</tr>
<tr>
<td>basisOfRecord</td>
<td>occurrenceId</td>
<td>IndividualCount</td>
<td>eventDate</td>
<td>countryCode</td>
<td>decimalLatitude</td>
<td>decimalLongitude</td>
<td>geoDatum</td>
<td>coordinateUncertaintyInMeters</td>
<td>scientificName</td>
<td>kingdom</td>
</tr>
<tr>
<td>HumanObservation</td>
<td>abD4f52-5308-453d-eeeb-9677e08e0014</td>
<td>1</td>
<td>2021-04-15</td>
<td>SK</td>
<td>78.2232</td>
<td>15.6267</td>
<td>WGS84</td>
<td></td>
<td>Rangifer tarandus platyrhynchus Animalia</td>
<td></td>
</tr>
<tr>
<td>HumanObservation</td>
<td>14d5188-ed59-4967-bd5b-06070db42f</td>
<td>2</td>
<td>2021-04-16</td>
<td>SK</td>
<td>78.2232</td>
<td>15.6267</td>
<td>WGS84</td>
<td></td>
<td>Rangifer tarandus platyrhynchus Animalia</td>
<td></td>
</tr>
<tr>
<td>HumanObservation</td>
<td>c651b4a-9cf4-485d-b45c-b449c6b0836</td>
<td>3</td>
<td>2021-04-17</td>
<td>SK</td>
<td>78.2232</td>
<td>15.6267</td>
<td>WGS84</td>
<td></td>
<td>Rangifer tarandus platyrhynchus Animalia</td>
<td></td>
</tr>
<tr>
<td>HumanObservation</td>
<td>60dc1cd-189f-e4c0-8a12-b53ccf17717c8</td>
<td>4</td>
<td>2021-04-18</td>
<td>SK</td>
<td>78.2232</td>
<td>15.6267</td>
<td>WGS84</td>
<td></td>
<td>Rangifer tarandus platyrhynchus Animalia</td>
<td></td>
</tr>
</tbody>
</table>

**Occurrence ID**
Dawson core sup to info: An identifier for the Occurrence (as opposed to a particular digital record of the occurrence). In the absence of a persistent global unique identifier, construct one from a combination of identifiers in the record that will mos...
Introduce the project → How are data logged? → Where are the data? → Introduce tools

Save as CSV

File name: occurrences
Save as type: CSV (Comma delimited)
Create New Resource

You can create a new blank resource, upload an existing resource saved as a zipped Darwin Core archive, or upload an existing IPT resource using its zipped resource configuration folder. Please refer to the User Manual for more specific instructions. A short name is required.

Shortname

1. luketest

Type

1. Occurrence

Import from an archived resource

Create
Choose the file you have created using the template generator
Introduce the project → How are data logged? → Where are the data? → Introduce tools

Input file name

Created source file

Analysed file (next slide)

Formatting

Readable
Columns 11
File /srv/ipt/resources/luketest/sources/occurrences.txt
Size 838 bytes
Rows 14
Modified 2021-04-19 12:54:14
Source log Download

Analyze

Number of Header Rows
1
Field Quotes "
Character Encoding UTF-8

Field Delimiter ,
Multi-value Delimiter 
Date Format YYYY-MM-DD
Introduce the project → How are data logged? → Where are the data? → Introduce tools

<table>
<thead>
<tr>
<th>basisOfRecord</th>
<th>occurrenceID</th>
<th>individualCount</th>
<th>eventDate</th>
<th>countryCode</th>
<th>decimalLatitude</th>
<th>decimalLongitude</th>
<th>geodeticDatum</th>
<th>coordinateUncertaintyInMeters</th>
</tr>
</thead>
<tbody>
<tr>
<td>HumanObservation</td>
<td>abf04f59-5308-45bd-aeeb-9d77eb08e014</td>
<td>1</td>
<td>2021-04-15</td>
<td>SJM</td>
<td>78.2232</td>
<td>15.6267</td>
<td>WGS84</td>
<td>10</td>
</tr>
<tr>
<td>HumanObservation</td>
<td>14fd5188-ed59-4967-bdb5-0f070dcb4f2f</td>
<td>2</td>
<td>2021-04-16</td>
<td>SJM</td>
<td>78.2232</td>
<td>15.6267</td>
<td>WGS84</td>
<td>10</td>
</tr>
<tr>
<td>HumanObservation</td>
<td>c615ba4a-9cfd-485d-b45c-b4439c6c0836</td>
<td>4</td>
<td>2021-04-17</td>
<td>SJM</td>
<td>78.2232</td>
<td>15.6267</td>
<td>WGS84</td>
<td>10</td>
</tr>
<tr>
<td>HumanObservation</td>
<td>60dc1cd1-89fe-4c08-8a12-b53cf17717c8</td>
<td>1</td>
<td>2021-04-18</td>
<td>SJM</td>
<td>78.2232</td>
<td>15.6267</td>
<td>WGS84</td>
<td>10</td>
</tr>
</tbody>
</table>

Rows: 14
Modified: 2021-04-19 12:54:14
Introduce the project → How are data logged? → Where are the data? → Introduce tools

Overview: *luketest*

This is the overview page for the *luketest* resource.

**Source Data**

- Choose file: No file chosen
- Connect to database
- Clear

Your source data files and SQL sources for generating a Darwin Core Archive.

- Last modified: Apr 19, 2021

**Darwin Core Mappings**

- Darwin Core Occurrence
- Add

Your mapping between the source data and Darwin Core terms.

- Created source file: occurrences [file] 838 bytes, 14 rows, 11 columns. Apr 19, 2021

**Mappings**

Remember 11 columns

**Metadata**

- Edit

Your resource metadata.

**Published Versions**

- Publish

A preview of your pending published version compared with the current version if existing.

- Version
  - 1.0
  - Preview
Darwin Core Occurrence

The category of information pertaining to evidence of an occurrence in nature, in a collection, or in a dataset (specimen, observation, etc.). Replaces version issued 2020-04-15 with a new, limited vocabulary for occurrenceStatus.

Link: http://rs.tdwg.org/dwc/terms/index.htm#Occurrence

Source data

Before you can start mapping concepts, please select a source data file.

Source data

occurrences

Save Cancel
Introduce the project → How are data logged? → Where are the data? → Introduce tools
Introduce the project → How are data logged? → Where are the data? → Introduce tools

<table>
<thead>
<tr>
<th>Field Interval</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>dcterms:accessRights</td>
<td></td>
</tr>
<tr>
<td>dcterms:bibliographicCitation</td>
<td></td>
</tr>
<tr>
<td>dcterms:references</td>
<td></td>
</tr>
<tr>
<td>institutionID</td>
<td></td>
</tr>
<tr>
<td>collectionID</td>
<td></td>
</tr>
<tr>
<td>datasetID</td>
<td></td>
</tr>
<tr>
<td>Use resource DOI</td>
<td></td>
</tr>
<tr>
<td>institutionCode</td>
<td></td>
</tr>
<tr>
<td>collectionCode</td>
<td></td>
</tr>
<tr>
<td>datasetName</td>
<td></td>
</tr>
<tr>
<td>ownerInstitutionCode</td>
<td></td>
</tr>
<tr>
<td>basisOfRecord</td>
<td>basisOfRecord</td>
</tr>
<tr>
<td>Source Sample: HumanObservation</td>
<td>HumanObservation</td>
</tr>
<tr>
<td>Translation: Add</td>
<td></td>
</tr>
</tbody>
</table>
Introduce the project → How are data logged? → Where are the data? → Introduce tools
Introduce the project → How are data logged? → Where are the data? → Introduce tools

**Source Data**

- Choose file: No file chosen
- Connect to database
- Clear

Your source data files and SQL sources for generating a Darwin Core Archive.

Last modified Apr 19, 2021

**occurrences [file]**

838 bytes, 14 rows, 11 columns. Apr 19, 2021

**Darwin Core Mappings**

- Darwin Core Occurrence
- Add

Your mapping between the source data and Darwin Core terms.

Last modified Apr 19, 2021

Core

**Darwin Core Occurrence**

11 terms mapped to occurrences. Apr 19, 2021

**Metadata**

Edit

Your resource metadata.

Last modified Apr 19, 2021

**Published Versions**

Publish

A preview of your pending published version compared with the current version if existing.

**Version**

1.0

**Pending version**

1.0

**Visibility**

Private

**Data Licence**

CC-BY 4.0

**Published on**

-
Darwin Core Archive

- **Core file**: Columns of data, one row per record
- **Information on records**: Can be created by template generator
- **Extension files**: Many-to-one relationships
- **meta.xml**: Describes structure
- **EML.xml**: Describes dataset as whole (e.g. abstract, people involved, methods, citations...)

Archive

- Created by IPT

Introduce the project → How are data logged? → Where are the data? → Introduce tools
NetCDF

• Use AeN setup to create template that includes physical terms

• Tools like Rosetta for converting from CSV to NetCDF.
NetCDF

• Use AeN setup to create template that includes physical terms

• Tools like Rosetta for converting from CSV to NetCDF.
GitHub: Want a similar setup for your project?
That’s all folks!

Exercise?:
- Explore the template
- Looking through metadata catalogue
- Q & A

https://sios-svalbard.org/reports/aen_multi

Image courtesy of Andreas Wolden