# Compiling FAIR sea ice core data: a reuser perspective

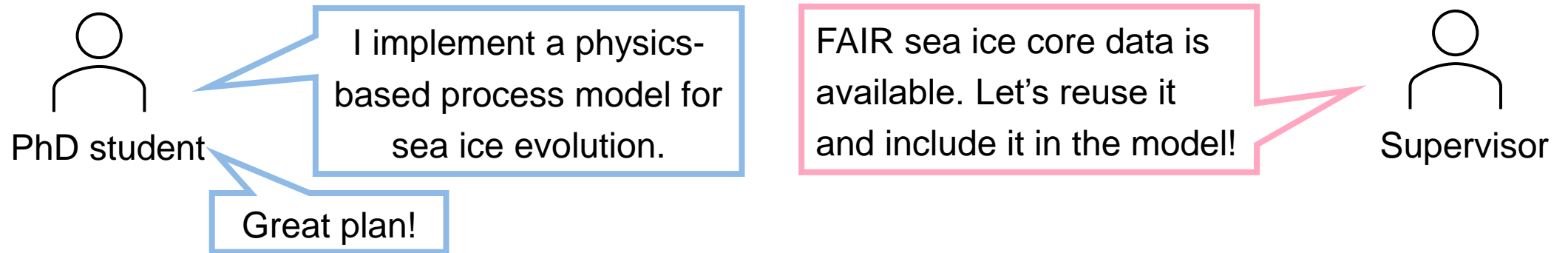**Anna Simson**[1], Anil Yildiz[1], Sebastian Mieruch[2] and Julia Kowalski[1]

[1] Methods for Model-based Development in Computational Engineering, RWTH Aachen University, Germany
[2] Department of Marine Geology, Alfred-Wegener-Institut Helmholtz Zentrum für Polar- und Meeresforschung
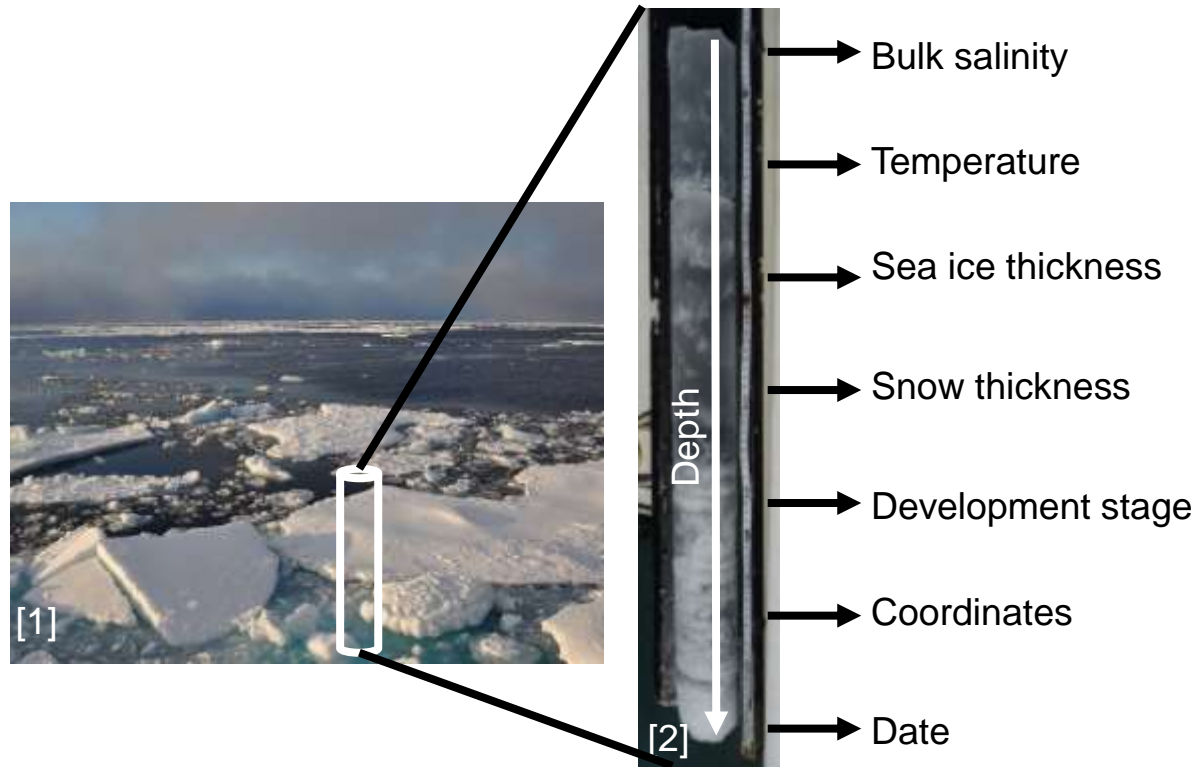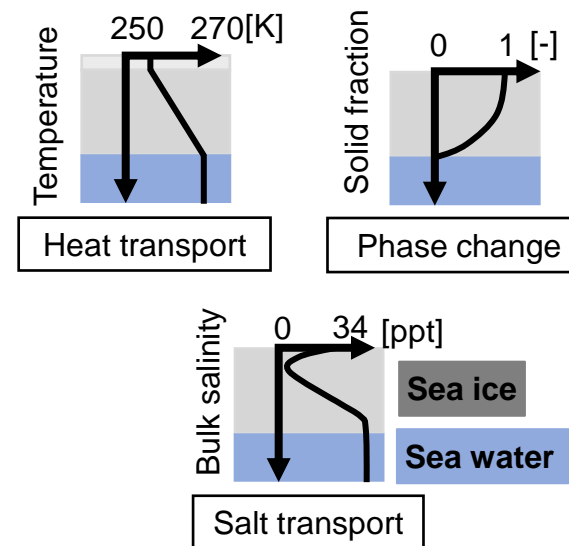
IMDIS 2024

# The reuse perspective: Sea ice core data for reuse scenarios

## Sea ice core data



- Bulk salinity
- Temperature
- Sea ice thickness
- Snow thickness
- Development stage
- Coordinates
- Date

## Reuse scenario A
### Validation data for sea ice evolution model



Heat transport

Phase change

Salt transport

## Reuse scenario B
### Training data for classification

**Development stages**

Old ice

First-year ice

MBD Chair of Methods for Model-based Development in Computational Engineering | RWTH AACHEN UNIVERSITY
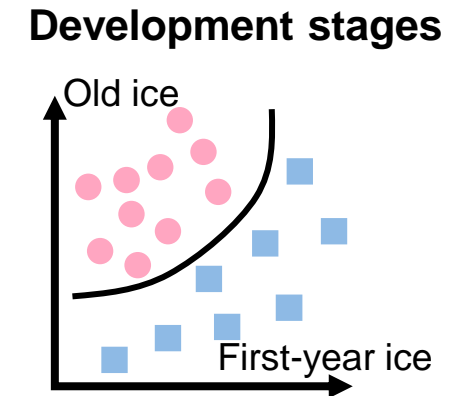
# Formalize reuse scenarios by defining the reuse scope

## Scenario A
### Process model validation

Temperature [K]

250    270

Depth [m]
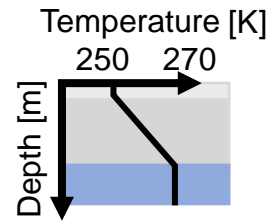
| Inputs | Constraints |
|--------|-------------|
| Salinity sea ice (profile) | Ratio: ppt<br>Spatial resolution: <0.2 m |
| Temperature sea ice (profile) | Unit: K<br>Measurement error: < 5% |
| Solid volume fraction sea ice (profile) | Spatial resolution: <0.2 m<br>Measurement error: < 5% |
| Salinity sea water | Ratio: ppt |
| Thickness sea ice | Unit: m<br>Value: <1m |

## Scenario B
### Classification algorithm

Old ice

First-year ice
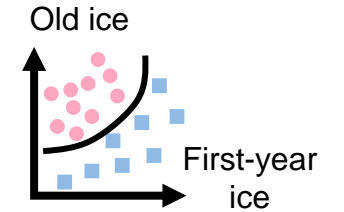
| Inputs | Constraints |
|--------|-------------|
| Date | Format: YYYY-MM-DD |
| Coordinates | Format: decimal degrees |
| Water body | Controlled vocabulary: SeaVoX[1] |
| Mean salinity | Ratio: ppt |
| Development stage | Controlled vocabulary: Sea Ice Nomenclature (SIN)[2] |

> We need a **database** of these elements for various sea ice cores.
> Let's **compile** existing **data sets**!

Anna Simson - simson@mbd.rwth-aachen.de

[1] British Oceanographic Data Centre (2023). DOI: 10.14284/590
[2] World Meteorological Organization (2014).
, URL: https://library.wmo.int/idurl/4/41953

MBD Chair of Methods for Model-based Development in Computational Engineering | RWTH AACHEN UNIVERSITY

# Intuitive approach to search and compile data



**Search for data**

**Technical combination**

| A | B | C | D | E | F |
|---|---|---|---|---|---|
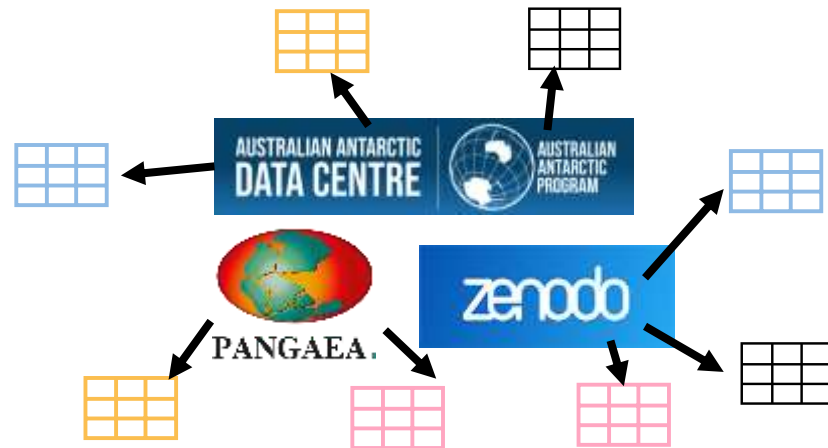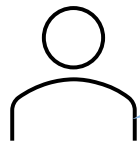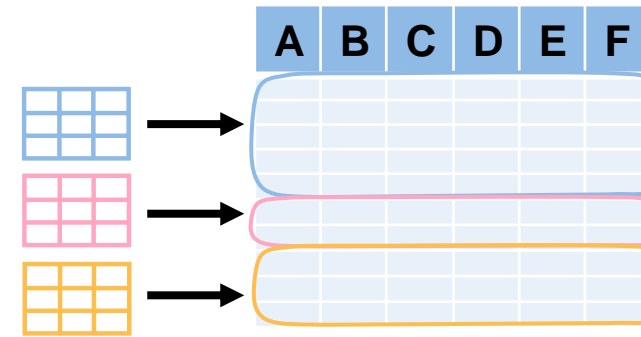
Data sets are **difficult** to reuse in a compiled form. They **vary** in content and structure.

Let's come up with a strategy to reuse data sets in a compiled form.
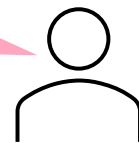
PhD student

Supervisor

- Followed **reusability-targeted** approach
- Created **enriched** sea ice core database **RESICE**

Anna Simson - simson@mbd.rwth-aachen.de

Chair of Methods for Model-based Development in Computational Engineering

RWTH AACHEN UNIVERSITY

# Challenges of searching for reuse scope elements

**Search for data**

Reuse scope elements are not available as requested.

Reuse scope elements are distributed across data sets, articles and reports.
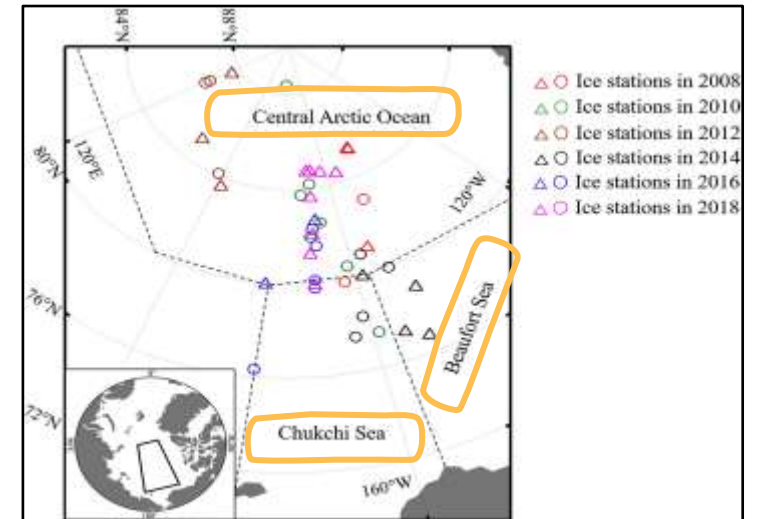
**Data set[1]** provides several reuse scope elements

| ID | Date | Location | Ice thickness (cm) | Snow thickness (cm) | Freeboard (cm) | Bulk ice tempaerature (°C) | Bulk salinity (PSU) | Bulk density (kg/m$^3$) | Air temperature (°C) | Ice type |
|---|---|---|---|---|---|---|---|---|---|---|
| 08-S1 | 27.08.08 | 85.1°N, 147.1°W | 160.7±9.0 | / | / | −0.8±0.4 | 1.1±0.8 | 723±29 | -0,5 | Multiyear ice |
| 08-S2 | 28.08.08 | 85.2°N, 147.3°W | 160.0±0 | / | / | −1.7±0.2 | 0.9±0.7 | 769±48 | -2,8 | Multiyear ice |

**Article[2]** provides further reuse scope elements: Measurement device accuracy and water body

sections into sealed plastic containers. The salinity samples and a salinometer with an accuracy of 0.1 practical salinity unit e salinity at a room temperature.

Anna Simson - simson@mbd.rwth-aachen.de

[1] Wang et al. (2020). DOI: 10.5281/ZENODO.3779867
[2] Wang et al. (2020). DOI: 10.1029/2020JC016371

Chair of Methods for Model-based Development in Computational Engineering

RWTH AACHEN UNIVERSITY

# Method: Element availability matching for each sea ice core



Anna Simson - simson@mbd.rwth-aachen.de

# Plausibility checks of the original data sets

Search for data → **Plausibility checks**

Anna Simson - simson@mbd.rwth-aachen.de

# Plausibility checks of the original data sets

Search for data → **Plausibility checks**

Data sets are redundant on different data repositories.

AADC data set[1]

| site | section | (DD/MM/YYYY) | Long-S | Lat-E | depth (cm) | Temp (°C) | Salinity |
|------|---------|--------------|--------|-------|-----------|-----------|----------|
| 1 | SNOW | | | | 20 | | 0,7 |
| 1 | 1 | 11.09.07 | 64°13.773' | 127°57.132' | 5,0 | -7,9 | 12,3 |
| 1 | 2 | | | | 15,0 | -6,8 | 8,7 |
| 1 | 3 | | | | 43,0 | -4,3 | 8,3 |
| 1 | 4 | | | | 94,0 | -2,0 | 4,8 |

PANGAEA data set[2]

| 1 Event | 2 Date/Time | 3 Latitude | 4 Longitude | 5 Depth ice/snow [m] | 6 t [°C] | 7 Depth ice/snow [m] (Lower limit of Salinity depth...) | 8 Depth ice/snow [m] (Upper limit of Salinity depth...) | 9 Sal |
|---------|-------------|------------|-------------|---------------------|----------|--------------------------------|--------------------------------|-------|
| SIPEX_01 | 2007-09-11 | -64.2295 | 127.9522 | 0.100 | -7.30 | 0.100 | 0.000 | 12.300 |
| SIPEX_01 | 2007-09-11 | -64.2295 | 127.9522 | 0.200 | -6.40 | 0.200 | 0.100 | 8.700 |
| SIPEX_01 | 2007-09-11 | -64.2295 | 127.9522 | 0.300 | -5.40 | 0.460 | 0.400 | 8.300 |
| SIPEX_01 | 2007-09-11 | -64.2295 | 127.9522 | 0.400 | -4.40 | 0.860 | 0.800 | 5.400 |
| SIPEX_01 | 2007-09-11 | -64.2295 | 127.9522 | 0.500 | -4.10 | 0.970 | 0.910 | 4.800 |

→ Detect redundancy and incorporate only one data set

 Anna Simson - simson@mbd.rwth-aachen.de

[1] Trull et al. (2011). AADC. DOI: 10.4225/15/514A9368882AA
[2] Lannuzel (2016). PANGAEA. DOI: 10.1594/PANGAEA.865031

MBD Chair of Methods for Model-based Development in Computational Engineering | RWTH AACHEN UNIVERSITY

# Plausibility checks of the original data sets

Search for data → **Plausibility checks**

Data sets are redundant on different data repositories.

Accuracy of measurement data is exaggeratedly high.

**Data set[1]** provides temperature with three decimals

| t [°C] (Thermometer) |
| --- |
| -6.140 |
| -6.025 |
| -5.912 |

**Instrument manual[2]** (Testo 720): Accuracy $< \pm 0.2$ ℃

➡ Adjust accuracy

Anna Simson - simson@mbd.rwth-aachen.de

[1] Lange et al. (2015). DOI: 10.1594/PANGAEA.842371
[2] Testo (2024). URL: https://static.testo.com/image/upload/Instruction-manual-and-Software/Instruction-manuals/testo-720-instruction-manual-7808.pdf

Chair of Methods for Model-based Development in Computational Engineering

RWTH AACHEN UNIVERSITY

# Plausibility checks of the original data sets

Search for data  →  **Plausibility checks**

## Data set[1] snapshots

SIPEX_05 🔍 ⊞ * *Latitude:* -65.523330 * *Longitude:* 125.752020 * *Date/Time:*
2007-09-17T00:00:00 * *Location:* Indian Ocean 🔍 * *Campaign:* SIPEX 🔍 * *Basis:* Aurora
Australis 🔍 * *Method/Device:* Electroplished stainless steel corer, Lichtert Industrie
(ESSC) 🔍 * *Comment:* Ice thickness: 0.43 m; snow thickness: 0.2 m; pack ice; (flooded/
slushy area), Vandermerwe et al. 2011 DSR

Data sets are redundant on different data repositories.

Accuracy of measurement data is exaggeratedly high.

Data sets are inconsistent.

| 1 ⓘ Event | 2 ⓘ Date/Time | 3 ⓘ Latitude | 4 ⓘ Longitude | 5 ⓘ Depth ice/snow [m] | 6 ⓘ 🖼 t [°C] |
|---|---|---|---|---|---|
| SIPEX_05 | 2007-09-17 | -65.5233 | 125.7520 | 0.650 | -5.80 |
| SIPEX_05 | 2007-09-17 | -65.5233 | 125.7520 | 0.750 | -4.70 |
| SIPEX_05 | 2007-09-17 | -65.5233 | 125.7520 | 0.850 | -4.30 |

Sea ice depth of the measurements exceeds thickness in the comment.

⇒ Find and remove inconsistencies

 Anna Simson - simson@mbd.rwth-aachen.de

Chair of Methods for Model-based Development in Computational Engineering

MBD | RWTH AACHEN UNIVERSITY

# Technical combination of data and metadata per sea ice core
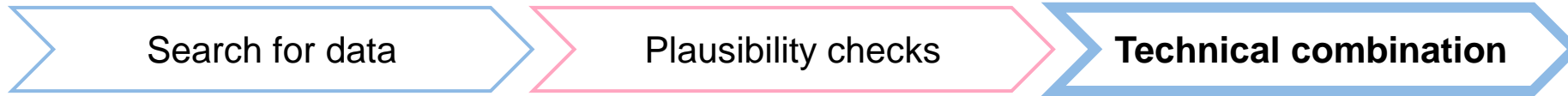
Flexibly combine different sources per sea ice core

Traceability of original sources, corrections and adjustments

**One YAML file per sea ice core:**

Harmonization

```
thickness_sea-ice:
type: scalar
value: 1.43
comment: from Fig. 4.22
doi: 10.2312/BzPM_0586_2009

measurement-device-accuracy-salinity_sea-ice:
type: scalar
value: 0.2
comment: (WTW microprocessor conductivity meter LF 196, accuracy 0.2).
doi: 10.1016/j.dsr2.2010.10.029

temperature_sea-ice:
type: tabulated
value: {0.03: 268.7,0.08: 268.8,[...]}
comment: from C to K and accuracy correction
doi: 10.1594/PANGAEA.865031
```
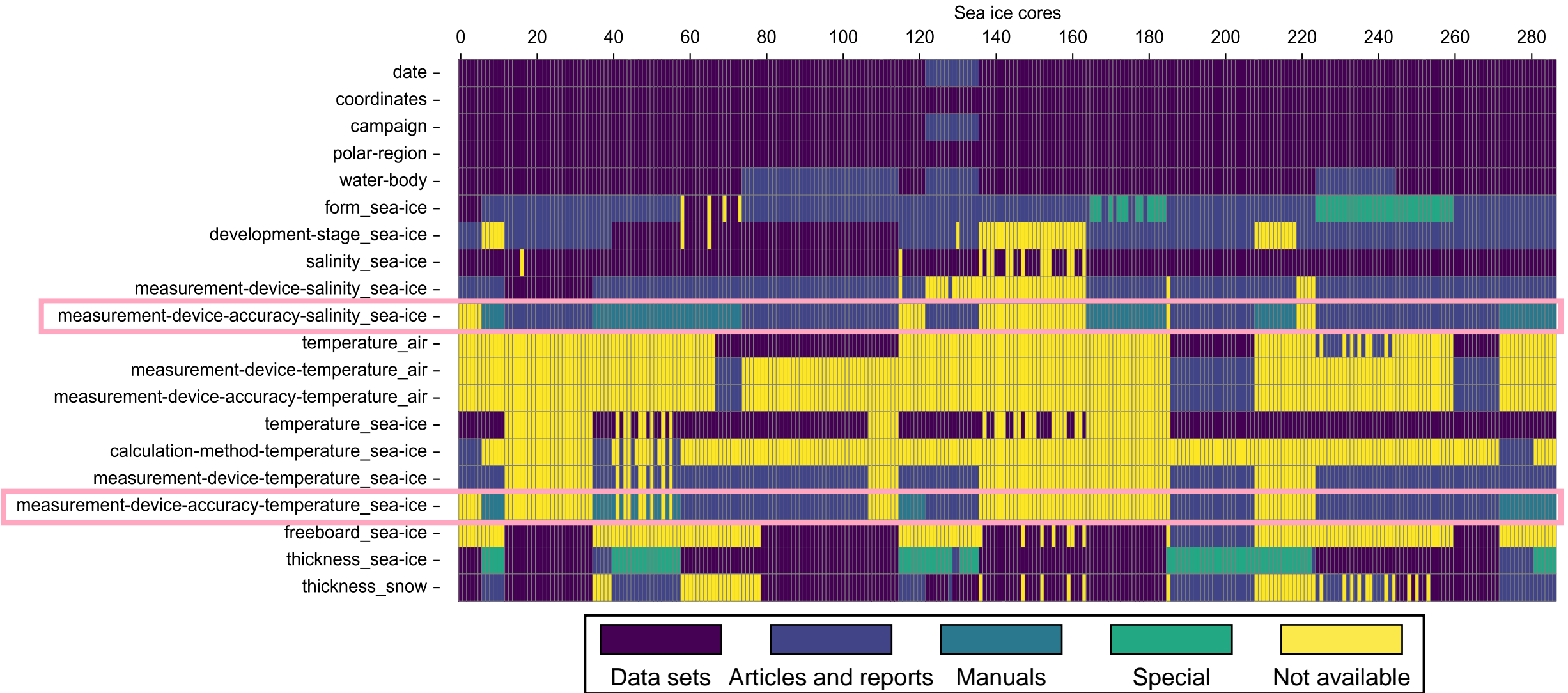
Figure

Excerpt

Plausibility

Anna Simson - simson@mbd.rwth-aachen.de

Anna Simson - simson@mbd.rwth-aachen.de

**Each node is clickable and guides you to the respective doi or url!**

Anna Simson - simson@mbd.rwth-aachen.de

# Automatic metadata enrichment to fulfill reuse scope

| Search for data | Plausibility checks | Technical combination | **Element enrichment** |
|---|---|---|---|

Average spatial resolution and mean of measurements ⟹ Calculate from measurement data

Sea ice development stage following Sea Ice Nomenclature[1] ⟹ Derive with sea ice thickness and from Sea Ice Nomenclature

Names of water bodies following SeaVoX[2] ⟹ Obtain Ocean and Sea name from coordinates using SeaVoX Polygon attributes REGION and SUB_REGION

MBD Chair of Methods for Model-based Development in Computational Engineering | RWTH AACHEN UNIVERSITY

Salinity sea ice (all data)
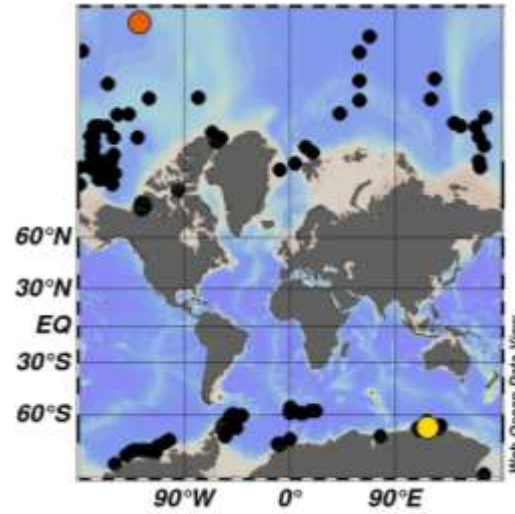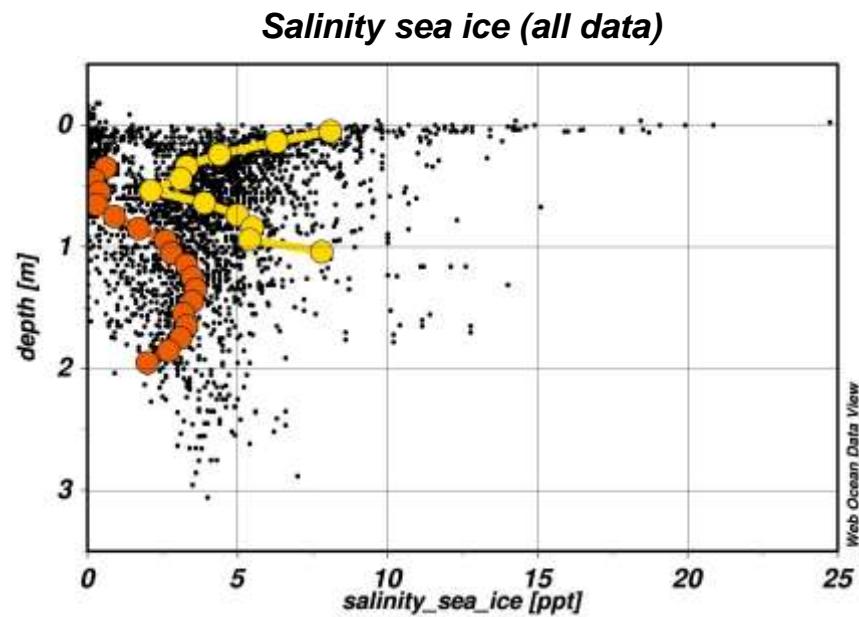
Development stage SIN level 1 (automatically enriched)

# Summary



Anna Simson - simson@mbd.rwth-aachen.de

[1] World Meteorological Organization (2014). URL: https://library.wmo.int/idurl/4/41953
[2] British Oceanographic Data Centre (2023). DOI: 10.14284/590
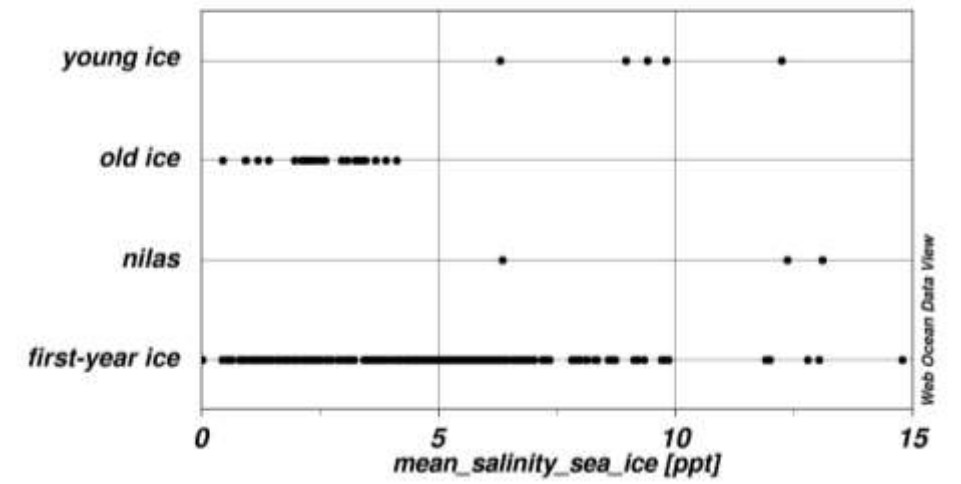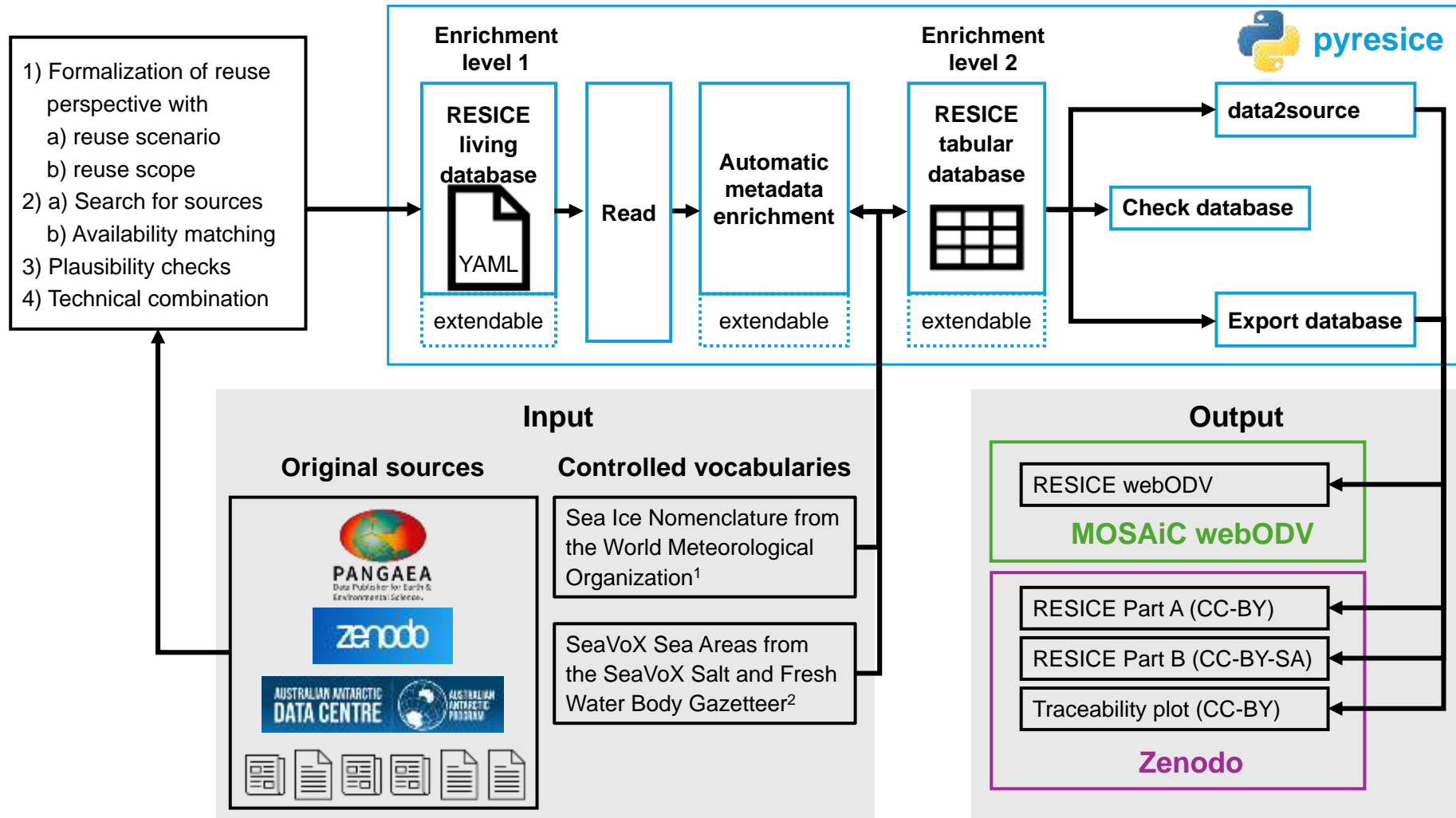
# RESICE resources

RESICE in Zenodo



RESICE in MOSAiC webODV
Go to: Extra>Sea_Ice>RESICE



Interactive data2source traceability



Pyresice Python package



Anna Simson - simson@mbd.rwth-aachen.de

Chair of Methods for Model-
based Development in
Computational Engineering

RWTH AACHEN UNIVERSITY

# Challenges of searching for reuse scope elements

**Data set[1] provides salinity**

| 1 ⓘ Event | 2 ⓘ Date/Time | 3 ⓘ Latitude | 4 ⓘ Longitude | 5 ⓘ Sample label | 6 ⓘ Depth ice/snow [m] | Sal |
|---|---|---|---|---|---|---|
| PS78/203-1 ⓘ | 2011-08-14T20:40 | 85.9710 | 59.3485 | 110814SAL | 0.310 | 0.4 |
| PS78/203-1 | 2011-08-14T20:40 | 85.9710 | 59.3485 | 110814SAL | 0.410 | 0.0 |
| PS78/203-1 | 2011-08-14T20:40 | 85.9710 | 59.3485 | 110814SAL | 0.510 | 0.2 |

Reuse scope element are not available as requested.

**Data set[2] provides temperature**

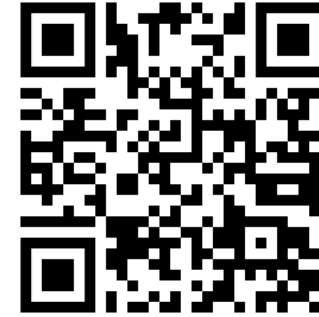| 1 ⓘ Event | 2 ⓘ Date/Time | 3 ⓘ Latitude | 4 ⓘ Longitude | 5 ⓘ Sample label | 6 ⓘ Depth ice/snow [m] | t [°C] |
|---|---|---|---|---|---|---|
| PS78/203-1 ⓘ | 2011-08-14T20:40 | 85.9710 | 59.3485 | 110814ARC | 0.0 | 0.0 |
| PS78/203-1 | 2011-08-14T20:40 | 85.9710 | 59.3485 | 110814ARC | 0.1 | -1.0 |
| PS78/203-1 | 2011-08-14T20:40 | 85.9710 | 59.3485 | 110814ARC | 0.2 | 0.1 |

**Expedition report[3] provides instrument name**

The salinity core was sawed in slices of 10 cm and packed into boxes. After melting of the samples, the salinity was measured onboard with a salinometer of type „WTW Cond 3151". Afterwards small samples were filled up for biological studies.

Reuse scope elements are distributed across data sets, articles and reports.

**Instrument manual[4] provides accuracy**

| SAL | Bereich 0,0 ... 42,0 | |
|---|---|---|
| | Genauigkeit | Meßguttemperatur |
| | ± 0,1 | 5 °C ... 25 °C |

Anna Simson - simson@mbd.rwth-aachen.de

[1] Nicolaus et al. (2012). DOI: 10.1594/PANGAEA.773276
[2] Nicolaus et al. (2012). DOI: 10.1594/PANGAEA.773277
[3] Hendricks et at. (2012). DOI: 10.2312/BzPM_0649_2012
[4] WTW (2004). URL: https://www.labworld.at/wp-content/uploads/2014/10/Cond_315i.pdf

# Challenges of search for reuse scope elements

> **Search for data**

Data set content and reuse scope are not equivalent.

Snapshots of the data set[1]

PS69/542-2 (WS-1) 🔍 * *Latitude Start:* -60.466000 * *Longitude Start:* -40.816300 * *Latitude End:* -60.445500 *
*Longitude End:* -40.784800 * *Date/Time Start:* 2006-09-08T18:52:00 * *Date/Time End:* 2006-09-08T21:25:00 *
*Elevation Start:* -759.0 m * *Elevation End:* -737.0 m * *Location:* Scotia Sea, southwest Atlantic 🔍 * *Campaign:*
ANT-XXIII/7 (PS69 WWOS) 🔍 * *Basis:* Polarstern 🔍 * *Method/Device:* Ice station (ICE) 🔍

| 1 ℹ️ | 2 ℹ️ 🖼️ | 3 ℹ️ 🖼️ | 4 ℹ️ 🖼️ | 5 ℹ️ |
|---|---|---|---|---|
| **Depth ice/snow** [m] | **t** [°C] | **TTT** [°C] | **Sal brine** | **Comment** |
| -1.000 | | -4.5 | | air |
| 0.000 | -5.1 | | | snow |
| 0.020 | -5.0 | | 84.6 | |
| 0.050 | -4.7 | | 79.9 | |
| 0.100 | -4.6 | | 78.4 | |

- Limited availability of reuse scope elements
- Water body does not follow controlled vocabulary SeaVoX

Anna Simson - simson@mbd.rwth-aachen.de

[1] Kramer et al. (2010). DOI:10.1594/PANGAEA.734395.

MBD — Chair of Methods for Model-based Development in Computational Engineering | RWTH AACHEN UNIVERSITY

# Challenges of searching for reuse scope elements

**Search for data**

Reuse scope elements are not available as requested.

**Elements unavailable:**

- Solid volume fraction of sea ice

- Spatial resolution of measurements

**Elements available with adjustment:**

- Measurement error is not available ⇨ replaced with instrument accuracy

- Water body names and development stages do not follow controlled vocabularies

Anna Simson - simson@mbd.rwth-aachen.de

# Technical combination of data and metadata per sea ice core

Search for data → Plausibility checks → **Technical combination**

Harmonize original sources

- Units

- Label names

- Categorical names

- Coordinates

- Date

- Reference depth

Anna Simson - simson@mbd.rwth-aachen.de

Chair of Methods for Model-based Development in Computational Engineering

RWTH AACHEN UNIVERSITY