Bridging the gap between data and sensor information

Roland Koppe, Ana Macario
Challenges

- Increasing **heterogeneity** of instruments, data formats, standards, methods, ...

- Small over large to very large projects, but limited **resources**

- **Scalability** of technical solutions, and important of **curation** and help
A modular virtual research infrastructure designed to support scientific workflows, in particular in the flow from sensor observations to archives.
Objectives

- **Generic** infrastructure for data flows
- **Sustainable** and up-to-date services
- **Interoperable** and support for standards
- Seamless **integration** of information systems
- **User-friendly** and simple services
Data Flow Framework – O2A

SENSOR
Manage platform, sensor metadata

STREAM
Near real-time streaming of large data volume

DASHBOARD
Monitoring of near real-time data

DATA ACQUISITION

NEAR REAL TIME DATA

WORKSPACE
Solutions for data storage, processing and long-term preservation

IMPAIR
Data Flow Framework – O2A

**DASHBOARD**
Monitoring of near real-time data

**ANALYSIS**
Data viewing and analysis solutions;
Map-based visualization services

**PORTAL**
One-stop-shop framework
Interoperability services

**WORKSPACE**
Solutions for data storage,
processing and long-term preservation

**STORAGE ARCHIVE**

**REPOSITORIES**
Data and data products
Publications, presentations,
field reports
<table>
<thead>
<tr>
<th>Info</th>
<th>Device (Short Name)</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>AWIEV Underwater Observatory Svalbard (SVLUWCOBS)</td>
</tr>
<tr>
<td>i</td>
<td>AWIEV Atmosphere Observatory (awipev)</td>
</tr>
<tr>
<td>i</td>
<td>AWIEV Permafrost Observatory Bayelva (Bayelva)</td>
</tr>
<tr>
<td>i</td>
<td>Base Belgrano II (Belgrano II)</td>
</tr>
<tr>
<td>i</td>
<td>Base San Martín (San Martín)</td>
</tr>
<tr>
<td>i</td>
<td>Beagle Channel Underwater Observatory (BEAUWCOBS)</td>
</tr>
<tr>
<td>i</td>
<td>Cuxhaven intercalibration facility (cux_inter)</td>
</tr>
<tr>
<td>i</td>
<td>Dallmann (dallmann)</td>
</tr>
<tr>
<td>i</td>
<td>FINO (fino)</td>
</tr>
<tr>
<td>i</td>
<td>HAUSGARTEN observatory (HAUSGARTEN)</td>
</tr>
<tr>
<td>i</td>
<td>Heigoland Underwater Observatory (HELUWCOBS)</td>
</tr>
<tr>
<td>i</td>
<td>Kohnon (kohnen)</td>
</tr>
<tr>
<td>i</td>
<td>LTO Samoylov station (samoylov)</td>
</tr>
<tr>
<td>i</td>
<td>Neumayer III Acoustic Observatory PALAOA (neumayer_iii_palaoa_obs)</td>
</tr>
<tr>
<td>i</td>
<td>Neumayer III Atmosphere Observatory (neumayer_iii_metobs)</td>
</tr>
<tr>
<td>i</td>
<td>Neumayer III Geomagnetic Observatory (neumayer_iii_mgobs)</td>
</tr>
<tr>
<td>i</td>
<td>Neumayer III Infrasound Observatory (neumayer_iii_infrobs)</td>
</tr>
<tr>
<td>i</td>
<td>Neumayer III Seismology Observatory (neumayer_iii_seisobs)</td>
</tr>
<tr>
<td>i</td>
<td>Neumayer III (neumayer_iii)</td>
</tr>
</tbody>
</table>
AWIPEV Underwater Observatory Svalbard

State: Construction

ID: 220

Parent:

Device URN: station:svluwobs

Short Name: SVLUWOBSS

Long Name: AWIPEV Underwater Observatory Svalbard

Description:
AWIPEV_UNS is designed as an experimental platform project between the HZG and the AWI to host sensors and sensor units in a polar fjord system. The system provides underwater data connection and power as well as a server infrastructure for system and sensor control. Attached is an upward looking ADCP and a CTD plus sensors for turbidity, oxygen, chl-a fluorescence.

Download sensor metadata as: Sensor ML | JSON
Sensor descriptions for provenance and reduced data integration effort

Versioning and citability – use case in RDA PID WG

Interoperability and standards

~1700 descriptions available and counting
## Available sensors

Filter sensors: Neum

<table>
<thead>
<tr>
<th>Sensor code</th>
<th>Sensor description</th>
<th>Last date</th>
<th>Age</th>
<th>Last value</th>
<th>Unit</th>
<th>Platform ID</th>
<th>Sensor ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>station_neumayer</td>
<td>Neumayer III: Ashtec: Latitude</td>
<td>2018-11-02 14:00:00</td>
<td>6 minutes</td>
<td>70.66725</td>
<td></td>
<td>6</td>
<td>681</td>
</tr>
<tr>
<td>station_neumayer</td>
<td>Neumayer III: Ashtec: Longitude</td>
<td>2018-11-02 14:00:00</td>
<td>6 minutes</td>
<td>-5.27887</td>
<td></td>
<td>6</td>
<td>552</td>
</tr>
<tr>
<td>station_neumayer</td>
<td>Neumayer III: GPS: Latitude</td>
<td>2018-11-02 14:00:00</td>
<td>6 minutes</td>
<td>79.69615</td>
<td></td>
<td>6</td>
<td>559</td>
</tr>
<tr>
<td>station_neumayer</td>
<td>Neumayer III: GPS: Longitude</td>
<td>2018-11-02 14:00:00</td>
<td>6 minutes</td>
<td>-5.27907</td>
<td></td>
<td>6</td>
<td>500</td>
</tr>
<tr>
<td>station_neumayer</td>
<td>hPa: Neumayer III:metobs:air_pressure</td>
<td>2018-11-02 14:00:00</td>
<td>6 minutes</td>
<td>961.1 hPa</td>
<td></td>
<td>6</td>
<td>956</td>
</tr>
<tr>
<td>station_neumayer</td>
<td>deg C: Neumayer III:metobs:air_temperature_10m</td>
<td>2018-11-02 14:00:00</td>
<td>6 minutes</td>
<td>-18.6 deg C</td>
<td></td>
<td>6</td>
<td>950</td>
</tr>
<tr>
<td>station_neumayer</td>
<td>deg C: Neumayer III:metobs:air_temperature_3m</td>
<td>2018-11-02 14:00:00</td>
<td>6 minutes</td>
<td>-18.6 deg C</td>
<td></td>
<td>6</td>
<td>951</td>
</tr>
<tr>
<td>station_neumayer</td>
<td>m: Neumayer III:metobs:visibility</td>
<td>2018-11-02 14:00:00</td>
<td>6 minutes</td>
<td>75000.0 m</td>
<td></td>
<td>6</td>
<td>957</td>
</tr>
<tr>
<td>station_neumayer</td>
<td>deg: Neumayer III:metobs:wind_direction_10m</td>
<td>2018-11-02 14:00:00</td>
<td>6 minutes</td>
<td>282.7 deg</td>
<td></td>
<td>6</td>
<td>952</td>
</tr>
<tr>
<td>station_neumayer</td>
<td>deg: Neumayer III:metobs:wind_direction_3m</td>
<td>2018-11-02 14:00:00</td>
<td>6 minutes</td>
<td>283.9 deg</td>
<td></td>
<td>6</td>
<td>953</td>
</tr>
<tr>
<td>station_neumayer</td>
<td>m/s: Neumayer III:metobs:wind_velocity_10m</td>
<td>2018-11-02 14:00:00</td>
<td>6 minutes</td>
<td>2.0 m/s</td>
<td></td>
<td>6</td>
<td>954</td>
</tr>
<tr>
<td>station_neumayer</td>
<td>m/s: Neumayer III:metobs:wind_velocity_3m</td>
<td>2018-11-02 14:00:00</td>
<td>6 minutes</td>
<td>2.4 m/s</td>
<td></td>
<td>6</td>
<td>955</td>
</tr>
<tr>
<td>station_neumayer</td>
<td>Neumayer III: Weather:</td>
<td>2018-11-02 14:00:00</td>
<td>6 minutes</td>
<td></td>
<td></td>
<td>6</td>
<td>974</td>
</tr>
<tr>
<td>station_neumayer</td>
<td>Neumayer III: Weather: Air pressure (43m NN)</td>
<td>2018-02-20 08:16:01</td>
<td>255 days</td>
<td>975.3 hPa</td>
<td></td>
<td>6</td>
<td>42</td>
</tr>
<tr>
<td>station_neumayer</td>
<td>Neumayer III: Weather: Air temperature (2m)</td>
<td>2018-02-20 08:16:01</td>
<td>255 days</td>
<td>-8.8 C</td>
<td></td>
<td>6</td>
<td>39</td>
</tr>
<tr>
<td>station_neumayer</td>
<td>Neumayer III: Weather: Relative humidity (2m)</td>
<td>2018-02-20 08:16:01</td>
<td>255 days</td>
<td>68.0 %</td>
<td></td>
<td>6</td>
<td>43</td>
</tr>
<tr>
<td>station_neumayer</td>
<td>Neumayer III: Weather: Wind direction (10m)</td>
<td>2018-02-20 08:16:01</td>
<td>255 days</td>
<td>96.0 deg</td>
<td></td>
<td>6</td>
<td>41</td>
</tr>
<tr>
<td>station_neumayer</td>
<td>Neumayer III: Weather: Wind velocity (10m)</td>
<td>2018-02-20 08:16:01</td>
<td>255 days</td>
<td>7.6 m/s</td>
<td></td>
<td>6</td>
<td>40</td>
</tr>
</tbody>
</table>

15 / 757 sensor(s) are registered for this data service.

### Request data

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Begin</td>
<td>2018-11-01</td>
</tr>
<tr>
<td>End</td>
<td>2018-11-02</td>
</tr>
<tr>
<td>Format</td>
<td>JSON</td>
</tr>
</tbody>
</table>
Data Web Service 1.0
[ Base URL: /data-xxl/rest ]
https://dashboard.awi.de/data-xxl/rest/swagger/json

The data web service allows accessing and storing near real-time and delayed mode data.

- Ingest bound to sensor descriptions
- Generic driver and QC/QA approach
- Monitoring / alarming functionality
- Access to data, quality flags, aggregation statistics, unit conversion, …
**Water temperatures - NyÅlesund/Svalbard - N 78° 55.200 / E 11° 54.00**

Focus on specific dates or values by horizontal or vertical left-click dragging a zoom window. Double-click for focus reset.

Click on sensor names in the plots to select/deselect data.

---

**Value: sbe38_657:temperature**

The variable "#sbe38_657:temperature" is measured by a temperature probe (SBE38, Company SeaBird) which is mounted at 12 m water depth (+/- tide).

---

**Value: ctd_181:temperature**

The variable "#ctd_181:temperature" is measured by a combined conductivity - temperature - density probe (CTD90, Company Sea&Sun, Sensor Aanderaa optode) which is profiling every day at 12:25 hours between 11 m (+/- tide) and the surface. After this cast, the probe is set to one of the five depth strata 9m, 7m, 5m, 3m or 1m for 24 hours. By this procedure, each depth stratum is sampled for 24 hours once a week.

---

**Value: fb_731101:sbe45_0403:temperature**

The variable "#fb_731101:sbe45_0403:temperature" is measured by a land based FerryBox system (Sensor SBE45, ADMI) getting its water from a pumping station in a depth of 11 m (+/- tide) close to the base of the underwater observatory.
User-customizable, flexible dashboards for data monitoring

- Analysis of near-real time and delayed-mode data
- Based on sensor descriptions and configurations
Data Flow Framework – O2A

SENSOR
Manage platform, sensor metadata

STREAM
Near real-time streaming of large data volume

DASHBOARD
Monitoring of O2A real-time data

ANALYSIS
Data viewing and analysis solutions; Map-based visualization services

PORTAL
One-stop-shop framework Interoperability services

DATA ACQUISITION

NEAR REAL TIME DATA

WORKSPACE
Solutions for data storage, processing and long-term preservation

STORAGE ARCHIVE

REPOSITORIES
Data and data products Publications, presentations, field reports
Workspace

- Science community workspace for data sharing and data analytics within the Helmholtz Data Federation (HDF)
- State-of-the-art storage, replicated between Bremerhaven and Potsdam
- User-friendly “one-click” compute solutions with virtual machines and containers
- Hadoop big data analysis based on Hortonworks data flow and data platform
- Raster data management and analysis with rasdaman, developments
Cloud management with VMware vRealize

Providing tailored short-life containers and virtual machines for scientists

Including support for hybrid clouds, e.g. Amazon WS

Templates for e.g. RStudio and Jupyter
```r
sensors <- dws.sensors("station:svluwobs:svluw2:ctd_181")
sensors
head(data)
library(ggplot2)
library(scales)
data$timestamp <- as.POSIXct(data$datetime, format = "%Y-%m-%dT%H:%M:%S")

gg <- ggplot(data, aes(x = timestamp, y = station.svluwobs.svluw2.ctd_181.pressure..mean...dbar., col = station.svluwobs.svluw2.ctd_181.temperature..mean...C.)) + geom_point(aes(col = station.svluwobs.svluw2.ctd_181.temperature..mean...C.)) + labs(title = "Svalbard temperature at depths", x = "datetime", y = "depth [m]", col = "temperature [°C]"") + scale_color_gradientn(colours = rev(rainbow(5)), name = "temperature [°C]" ) + scale_x_datetime(limits = date_format("%Y-%m-%d", date_breaks = "7 days"))
options(repr.plot.width = 10, repr.plot.height = 5)
plot(gg)
```
Using data web services

We show the use of data web services Python interface for

Load data and metadata

Import the data web service Python interface. The data web
https://github.com/rkoppe/dws. Download the dws.py file an

```
In [1]:
1. from dws import dws
2. import matplotlib.pyplot as plt
3. import matplotlib.dates as mdates
4. import pandas as pd
```

Find some sensors offering temperature at Svalbard. Visit
https://sensorweb.no/ to find useful platforms and further metadata. Platforms, devic

```
In [2]:
1. sensors = dws.sensors('station:svluwobs*temperature*',
2. pd.DataFrame.from_dict(sensors)
```

<table>
<thead>
<tr>
<th>code</th>
<th>id</th>
</tr>
</thead>
<tbody>
<tr>
<td>station:svluwobs:svluw2:ctd_103:temperature_se...</td>
<td>1257</td>
</tr>
<tr>
<td>station:svluwobs:svluw2:adcp_17374:temperature</td>
<td>1250</td>
</tr>
<tr>
<td>station:svluwobs:svluw2:sbe38_657:temperature</td>
<td>427</td>
</tr>
<tr>
<td>station:svluwobs:svluw2:ctd_181:temperature</td>
<td>441</td>
</tr>
<tr>
<td>station:svluwobs:fb_731101:oxygen_sensor_574:...</td>
<td>750</td>
</tr>
<tr>
<td>station:svluwobs:fb_731101:temperature_freshwa...</td>
<td>753</td>
</tr>
<tr>
<td>station:svluwobs:fb_731101:temperature_container</td>
<td>754</td>
</tr>
<tr>
<td>station:svluwobs:fb_731101:temperature_outside</td>
<td>755</td>
</tr>
<tr>
<td>station:svluwobs:fb_731101:sbe45_0403:temperature</td>
<td>746</td>
</tr>
<tr>
<td>station:svluwobs:svluw2:ctd_578:temperature_se...</td>
<td>1024</td>
</tr>
<tr>
<td>station:svluwobs:svluw2:ctd_964:temperature_se...</td>
<td>1312</td>
</tr>
</tbody>
</table>

- Foster sharing of code incl. documentation
- Support native bookkeeping of changes, versioning
- Promote methods for reproducible results
- Notebooks as supplement for publications
Data Flow Framework

SENSOR
Manage platform, sensor metadata

STREAM
Near real-time streaming of large data volume

DASHBOARD
Monitoring of near real-time data

ANALYSIS
Data viewing and analysis solutions; Map-based visualization services

PORTAL
One-stop-shop framework Interoperability services

DATA ACQUISITION

NEAR REAL TIME DATA

WORKSPACE
Solutions for data storage, processing and long-term preservation

STORAGE ARCHIVE

REPOSITORIES
Data and data products Publications, presentations, field reports
- Map visualization and composition of data products
- Maintaining world base maps in different projections based on RTOPO
- Providing standard products, e.g. satellite-based chlorophyll a and sea ice
- Data archiving and publication (DOI) including curation and harmonization
- Linking with articles, platforms, devices, sensors, expeditions, orcid
- Semi-automatic data flows when using O2A

OAI-PMH, ISO19115, RDA WGs / IGs
Portal as central access point to linked and combined data

- Platforms, devices and sensor overviews
- Expeditions
- Data, publications and reports
- Interactive data products
- Services

FRAM
FRONTIERS IN ARCTIC MARINE MONITORING

Layers

Dynamic layers
- Sea-ice coverage
- Chlorophyll a

Product layers
- Fram Strait DTM
- Knipovich Ridge DTM
- Hausgarten - major currents
- FESOM - sea water temperature
- FESOM - sea water salinity

Platform layers
- Platforms
- Buys

→ Demo
The expedition’s planned duration is **350 days** in the ice.

During the expedition, RV Polarstern will be resupplied by **4 icebreakers**: Akademik Fedorov (RUS), Admiral Makarov (RUS), Oden (SWE), Xue Long (CHN).

The ice will drift at an average speed of roughly **7 km per day**.

Around **300 people** will work in the background for the expedition in order to realize it.

For **60-90 days** the research icebreaker Polarstern will be less than **200 km** away from the geographic North Pole.

Throughout the year, a total of **600 experts** will be on board, and be exchanged in phases.

**6,000 t** of fuel will be required by the research icebreaker Polarstern alone.

RV Polarstern will cover a total of some **2,500 km**.
Current Work

- Setup virtual environments for applications like 3D modelling of the ocean floor
- Develop building blocks for e.g. machine learning methods for images and videos
- User-transparent scaling of computing resources, e.g. with Kubernetes
- Repository of drivers for data ingest and quality procedures incl. best practices
- Computing on GPUs between “standard” virtualization and HPC
- ...
Learned and Outlook
Approach

- Illustrate **added value** for individuals, give incentives
- Make **hands on use** as easy as possible
- Provide **fast applicable** tools and best practices with examples
- Perform regular **user sessions** to foster community and user-network, beyond projects
- Collect and give **feedback**
Outlook and Preparations

- Data Management Plans as seamless starting point from project idea, over data ingest and science workspaces to publications
- Legacy systems: refactoring and further integration incl. harmonization of existing vocabularies, e.g. for devices, actions, ...
- New data center on board of Polarstern, remote synchronization of services and data between Arctic and Antarctic over satellites
- Streamline sample management solutions for, e.g. cores and specimen and flow integration
Best Practices Needed

- Continuous measurements, time series
  - how to align tracked sensor information with data
- Combined datasets – PID(s) for events of distinct instruments
  - how to combine it useful
- Agreements on formats and semantics – multi-partner and multi-discipline projects
- DMPs simple and actionable
- Integrating data products, prepare it for the community and beyond
- …
Thank you very much for your attention!

Ana Macario, Angela Schäfer, Antonia Haas, Christian Schäfer-Neth, Hans Pfeiffenberger, Peter Gerchow, Stefanie Schumacher, …

roland.koppe@awi.de