

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

An aerial photograph showing a large number of icebergs of various sizes and shapes floating in the dark blue ocean. The icebergs are white and appear to be melting, with some showing a distinct blue-green hue. The water is calm, reflecting the light. The overall scene is a vast, cold, and desolate landscape.

A modular virtual research infrastructure designed to support scientific workflows, in particular in the flow from sensor observations to archives

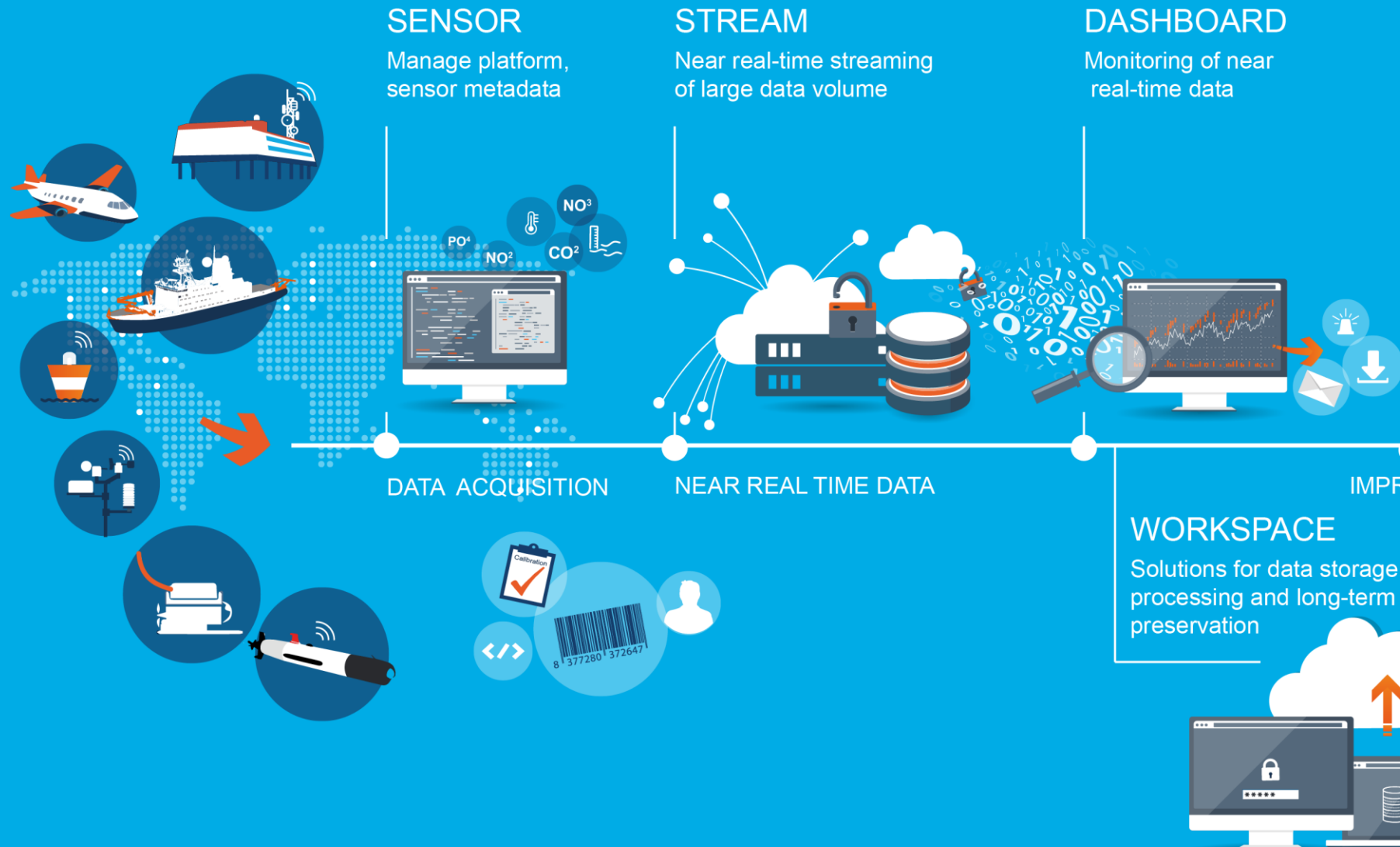
Mission

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



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



IMPROVE DATA → ★★★★★

WORKSPACE

Solutions for data storage,
processing and long-term
preservation

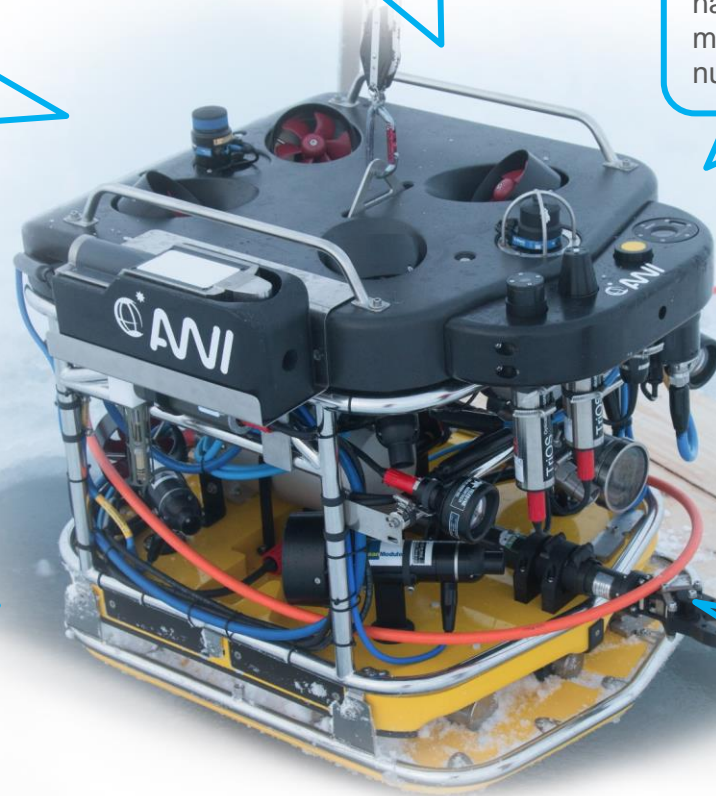
STORAGE ARCHIVE



REPOSITORIES

Data and data products
Publications, presentations,
field reports





Properties
min / max
temperature,
weight,
dimension, ...

Subdevices
multibeam sonar,
fluorometer,
camera, nitrate
sensor, ...

Description
names, manufacturer,
model, description, asset
number, ...

Parameters
backscatter, chl a,
FDOM, depth,
conductivity,
dissolved oxygen,
temperature, ...

Contact
data scientists, editors,
owner, engineer in charge,
principal investigator, ...

Resources
factsheet,
calibration
certificates,
manuals, ...

Action
deployment, mount,
calibration, failure,
information, ...

Platforms -> station

+ Add Item

Show 25 entries

Search:

Info	Device (Short Name)	Tools
	AWIPEV Underwater Observatory Svalbard (SVLUWOBS)	
	AWIPEV Atmosphere Observatory (awipev)	
	AWIPEV Permafrost Observatory Bayelva (Bayelva)	
	Base Belgrano II (Belgrano II)	
	Base San Martín (San Martin)	
	Beagal Channel Underwater Observatory (BEAUWOBS)	
	Cuxhaven intercalibration facility (cux_inter)	
	Dallmann (dallmann)	
	FINO (fino)	
	HAUSGARTEN observatory (HAUSGARTEN)	
	Helgoland Underwater Observatory (HELUWOBS)	
	Kohnen (kohnen)	
	LTO Samoylov station (samoylov)	
	Neumayer III Acoustic Observatory PALAOA (neumayer_iii_palaoa_obs)	
	Neumayer III Atmosphere Observatory (neumayer_iii_metobs)	
	Neumayer III Geomagnetic Observatory (neumayer_iii_magobs)	
	Neumayer III Infrasound Observatory (neumayer_iii_infraobs)	
	Neumayer III Seismology Observatory (neumayer_iii_seisobs)	
	Neumayer III (neumayer_iii)	

AWIPEV Underwater Observatory Svalbard



Overview

Contacts

Actions

Parameters

Resources

Properties

Local Frame

Subdevices

Images

Ingest

Current Version

State:

Construction

Public

Store



ID:

220

Parent:

Device URN:

station:svluwobs

Short Name:

SVLUWOBS

Long Name:

AWIPEV Underwater Observatory Svalbard

Collections:

Description:

AWIPEV_UNNS 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

Serial:

Manufacturer:

Model:

Type:

Station

Asset Number:

Download sensor metadata as: [Sensor ML](#) | [JSON](#)

Close

AWIPEV Underwater Observatory Svalbard

Overview

Contacts

Actions

Parameters

Resources

Properties

Local Frame

Subdevices

Images

Ingest

2018-04-24 14:00:00 Deployment KOL07

SENSOR (2018). Platform metadata for Station AWIPEV Underwater Observatory Svalbard. Configuration from 2018-04-24 16:00:00. Fried Wegener Institute for Polar and Marine Research. <https://hdl.handle.net/10013/sensor.584255b5-7404-4364-9693-88174e88202>

State: **public**

ID: 220

Parent:

Device URN: station:svluwobs

Short Name: SVLUWOBS

Long Name: AWIPEV Underwater Observatory Svalbard

Collections:

Description: AWIPEV_UN 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

Serial:

Manufacturer:

Model:

Type: Station

Asset
Number:

- 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:

	Sensor code		Sensor description	Last date	Age	Last value	Unit	Platform ID	Sensor ID
<input type="checkbox"/>	station:neumayer iii:ashtec:latitude		Neumayer III:Ashtec:Latitude	2018-11-02 14:00:00	6 minutes	-70.66725		6	661
<input type="checkbox"/>	station:neumayer iii:ashtec:longitude		Neumayer III:Ashtec:Longitude	2018-11-02 14:00:00	6 minutes	-8.27988		6	662
<input type="checkbox"/>	station:neumayer iii:gps:latitude		Neumayer III:GPS:Latitude	2018-11-02 14:00:00	6 minutes	-70.66683		6	659
<input type="checkbox"/>	station:neumayer iii:gps:longitude		Neumayer III:GPS:Longitude	2018-11-02 14:00:00	6 minutes	-8.27967		6	660
<input type="checkbox"/>	station:neumayer iii:metobs:air_pressure	hPa	Neumayer III:metobs:air_pressure	2018-11-02 14:00:00	6 minutes	981.1 hPa		6	956
<input type="checkbox"/>	station:neumayer iii:metobs:air_temperature_10m	deg_C	Neumayer III:metobs:air_temperature_10m	2018-11-02 14:00:00	6 minutes	-15.8 deg_C		6	950
<input type="checkbox"/>	station:neumayer iii:metobs:air_temperature_2m	deg_C	Neumayer III:metobs:air_temperature_2m	2018-11-02 14:00:00	6 minutes	-15.8 deg_C		6	951
<input type="checkbox"/>	station:neumayer iii:metobs:visibility	m	Neumayer III:metobs:visibility	2018-11-02 14:00:00	6 minutes	75000.0 m		6	957
<input type="checkbox"/>	station:neumayer iii:metobs:wind_direction_10m	deg	Neumayer III:metobs:wind_direction_10m	2018-11-02 14:00:00	6 minutes	282.7 deg		6	952
<input type="checkbox"/>	station:neumayer iii:metobs:wind_direction_2m	deg	Neumayer III:metobs:wind_direction_2m	2018-11-02 14:00:00	6 minutes	283.9 deg		6	953
<input type="checkbox"/>	station:neumayer iii:metobs:wind_velocity_10m	m/s	Neumayer III:metobs:wind_velocity_10m	2018-11-02 14:00:00	6 minutes	2.8 m/s		6	954
<input type="checkbox"/>	station:neumayer iii:metobs:wind_velocity_2m	m/s	Neumayer III:metobs:wind_velocity_2m	2018-11-02 14:00:00	6 minutes	2.4 m/s		6	955
<input type="checkbox"/>	station:neumayer iii:weather:		Neumayer III:Weather:					6	874
<input type="checkbox"/>	station:neumayer iii:weather:air pressure (43m nn)	hPa	Neumayer III:Weather:Air pressure (43m NN)	2018-02-20 08:16:01	255 days	975.3 hPa		6	42
<input type="checkbox"/>	station:neumayer iii:weather:air temperature (2m)	C	Neumayer III:Weather:Air temperature (2m)	2018-02-20 08:16:01	255 days	-8.8 C		6	39
<input type="checkbox"/>	station:neumayer iii:weather:relative humidity (2m)	%	Neumayer III:Weather:Relative humidity (2m)	2018-02-20 08:16:01	255 days	89.0 %		6	43
<input type="checkbox"/>	station:neumayer iii:weather:wind direction (10m)	deg	Neumayer III:Weather:Wind direction (10m)	2018-02-20 08:16:01	255 days	96.0 deg		6	41
<input type="checkbox"/>	station:neumayer iii:weather:wind velocity (10m)	m/s	Neumayer III:Weather:Wind velocity (10m)	2018-02-20 08:16:01	255 days	7.5 m/s		6	40

18 / 787 sensor(s) are registered for this data service.

Request data

Begin

End

Format

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.

default

GET /data Loads data according to given query parameters.

POST /data Saves data specified in the body

POST /data/bulk Loads data according to the given request in body (see DataRequest object description at the bottom of this site)

DELETE /dataset/{datasetid} Deletes the dataset and the data rows that depends to this dataset

GET /datasets Returns all existing datasets by provided restrictions, the newest is the first

GET /platforms Returns all known platforms

- Ingest bound to sensor descriptions
- Generic driver and QC/QA approach
- Monitoring / alarming functionality
- Access to data, quality flags, aggregation statistics, unit conversion, ...



Research Base

Water temperatures - NyAlesund/Svalbard - N 78° 55.200 / E 11° 54.00

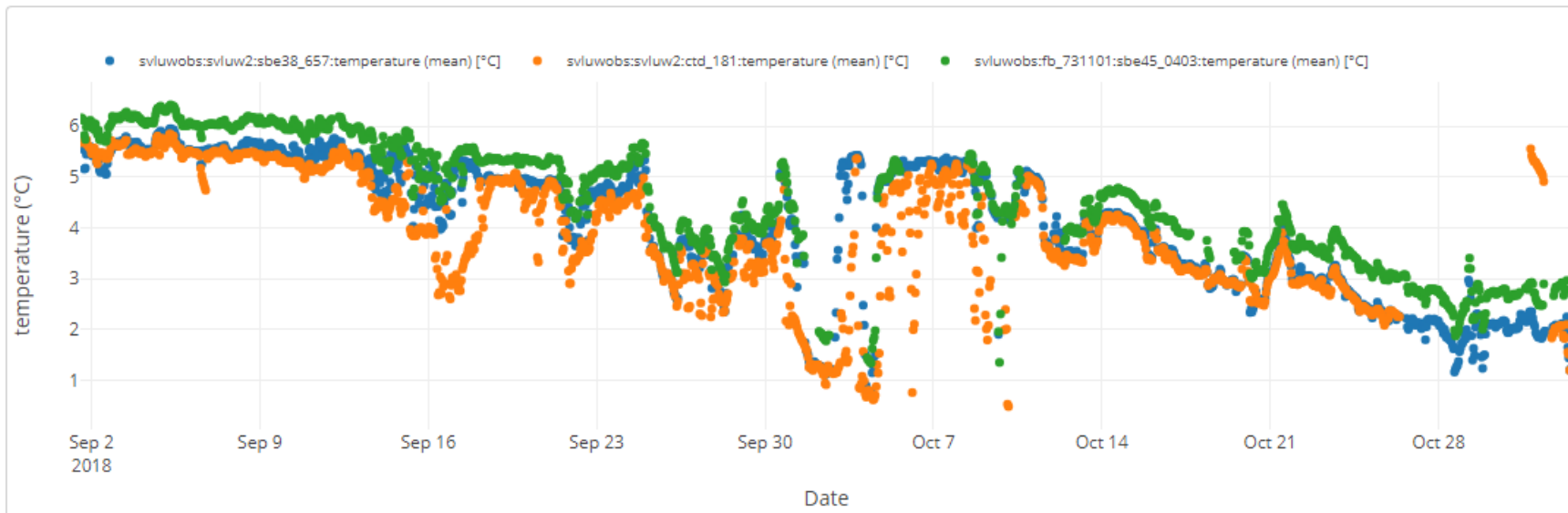


Ny-Alesund 79°N



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 11m (+/- tide) and the surface. After this cast, the probe is set to one of the five depth 9m, 7m, 5m, 3m or 1m for 24hours. 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, ADM) getting its water from a pumping station in a depth of 11m (+/- tide) close to the base of the underwater observatory.



AWI Dashboards Tiles Heatmap 2018-08-01 to now (in days) Hello, rkoppe Logout

Time Range

From: 2018-08-01

To: now

Resolution: day

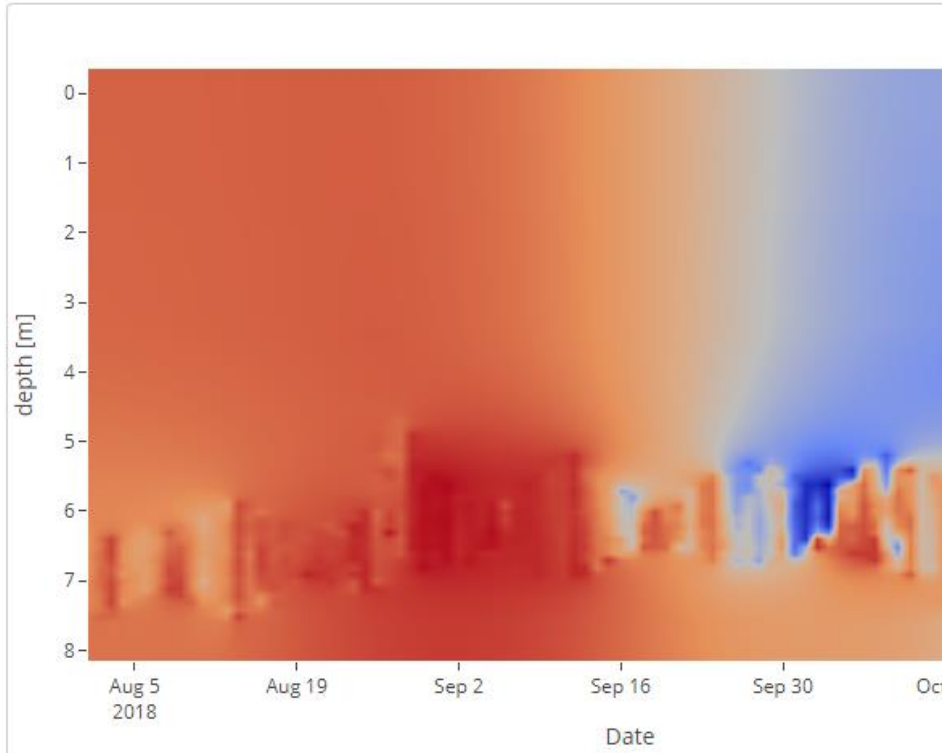
Quick-Ranges

- Last 30 minutes
- Last hour
- Last 3 hours
- Last 6 hours
- Last 12 hours
- Last 24 hours
- Last 2 days
- Last 5 days
- Last week
- Last 2 weeks
- Last month
- Last 2 months

apply

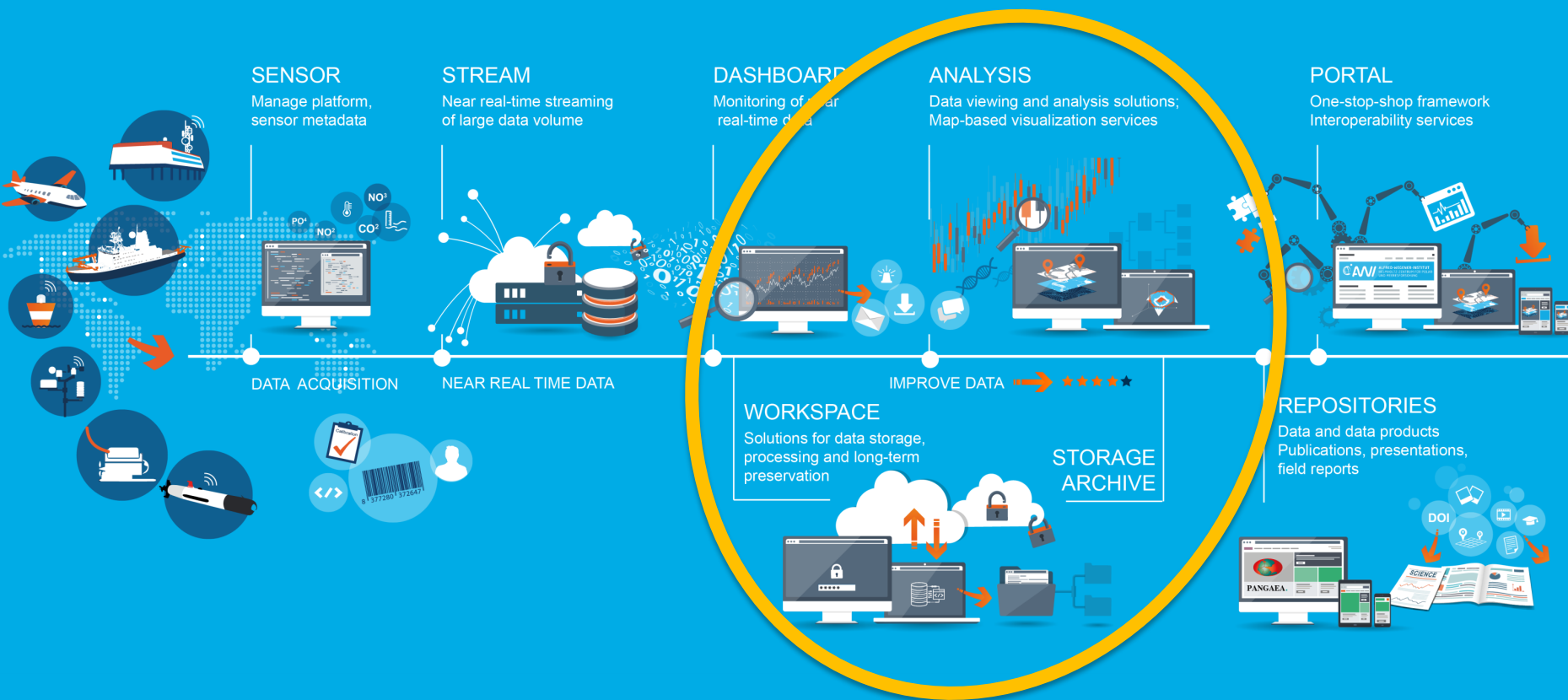
- User-customizable, flexible dashboards for data monitoring

Temperature / depth heatmap for Svalbard.



- Analysis of near-real time and delayed-mode data
- Based on sensor descriptions and configurations

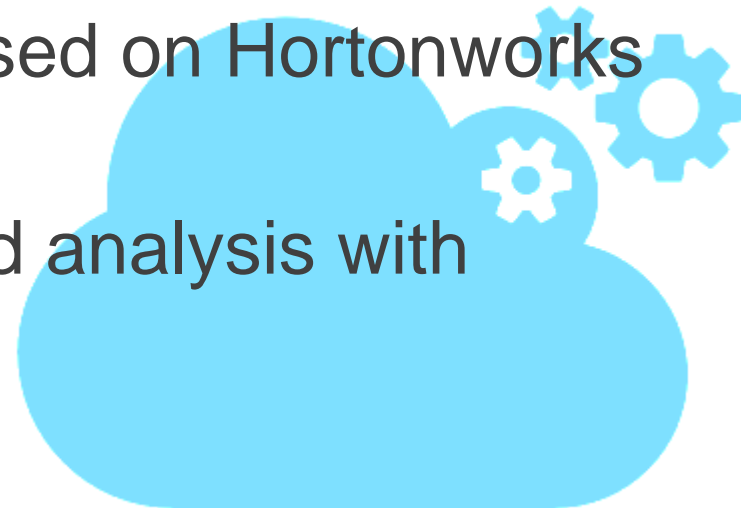
Data Flow Framework – O2A



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




Servicekatalog


Katalog nach von Ihnen benötigten Services durchsuchen.

-  Alle Services i
-  Applications i
-  AWS / EC2 i
-  Container i
-  Container Hosts i
-  JM & AS only i
-  Test & Dev i
-  Virtual Machines i
-  Workshop i


Alle Services (25)




Deep Learning Base AMI...
ami-49655ba2 - Deep Learning Base AMI (Amazon Linux)
Version 8.0 Comes with ...
Anfordern




Jupyterhub Ubuntu18
Jupyterhub auf Ubuntu 18.04
Server mit Anaconda für Python 2.7 und Python 3.4, sowie ...
Anfordern




pangeoscheduler
Container: Pangeo Scheduler
Anfordern




Red Hat Enterprise Linux...
ami-c86c3f23 - Red Hat Enterprise Linux 7.5 (HVM),
SSD Volume Type Red Ha ...
Anfordern




Ubuntu Desktop 16.04
Ubuntu 16.04 Desktop
Anfordern




Jupyterhub
Ubuntu 16.04 mit Jupyter und Anaconda
http://<IP/FQDN> ...
Anfordern




Jupyter Minimal (GMT)
Container: Jupyter Minimal (GMT)
Minimal (GMT) ...
Anfordern



pangeo
Container: ...
Anfordern



RStudio
Ubuntu 16.04 ...
http://<IP/FQDN>:3838 RStudio ...
Anfordern



Ubuntu Server
Ubuntu 16.04 Server
ami-d65fb135 - Ubuntu Serv...
Anfordern

- 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

→ Demo


```
21 sensors <- dws.sensors("station:svluwobs:svluw2:ctd_181*")
22 sensors
23
24
25 data <- dws.get(
26   c("station:svluwobs:svluw2:ctd_181:temperature", "station:svluwobs:svluw2:ctd_181:pressure")
27   begin="2018-10-01",
28   end="2018-11-05",
29   aggregate = "hour")
30 head(data)
31
32 library(ggplot2)
33 library(scales)
34
35 data$timestamp <- as.POSIXct(data$datetime, format = "%Y-%m-%d %H:%M:%S")
36 gg <- ggplot(data, aes(x = timestamp, y = -station.svluwobs.svluw2.ctd_181.pressure.mean.dbar))
37   geom_point(aes(col = station.svluwobs.svluw2.ctd_181.temperature.mean.C))
38   labs(title = "Svalbard temperature at depths", x = "datetime", y = "~ depth [m]", "ok")
39   scale_color_gradientn(colors = rev(rainbow(5)), name = "temperature [°C]")
40   scale_x_datetime(labels = date_format("%Y-%m-%d"), date_breaks = "7 days")
41
42 options(repr.plot.width = 10, repr.plot.height = 5)
43 plot(gg)
44
45
```

```
~/dws/ 6.3/604/
>
> 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)) +
+   geom_point(aes(col = station.svluwobs.svluw2.ctd_181.temperature.mean.C)) +
+   labs(title = "Svalbard temperature at depths", x = "datetime", y = "~ depth [m]", "ok") +
+   scale_color_gradientn(colors = rev(rainbow(5)), name = "temperature [°C]") +
+   scale_x_datetime(labels = date_format("%Y-%m-%d"), date_breaks = "7 days")
>
> options(repr.plot.width = 10, repr.plot.height = 5)
> plot(gg)
>
```

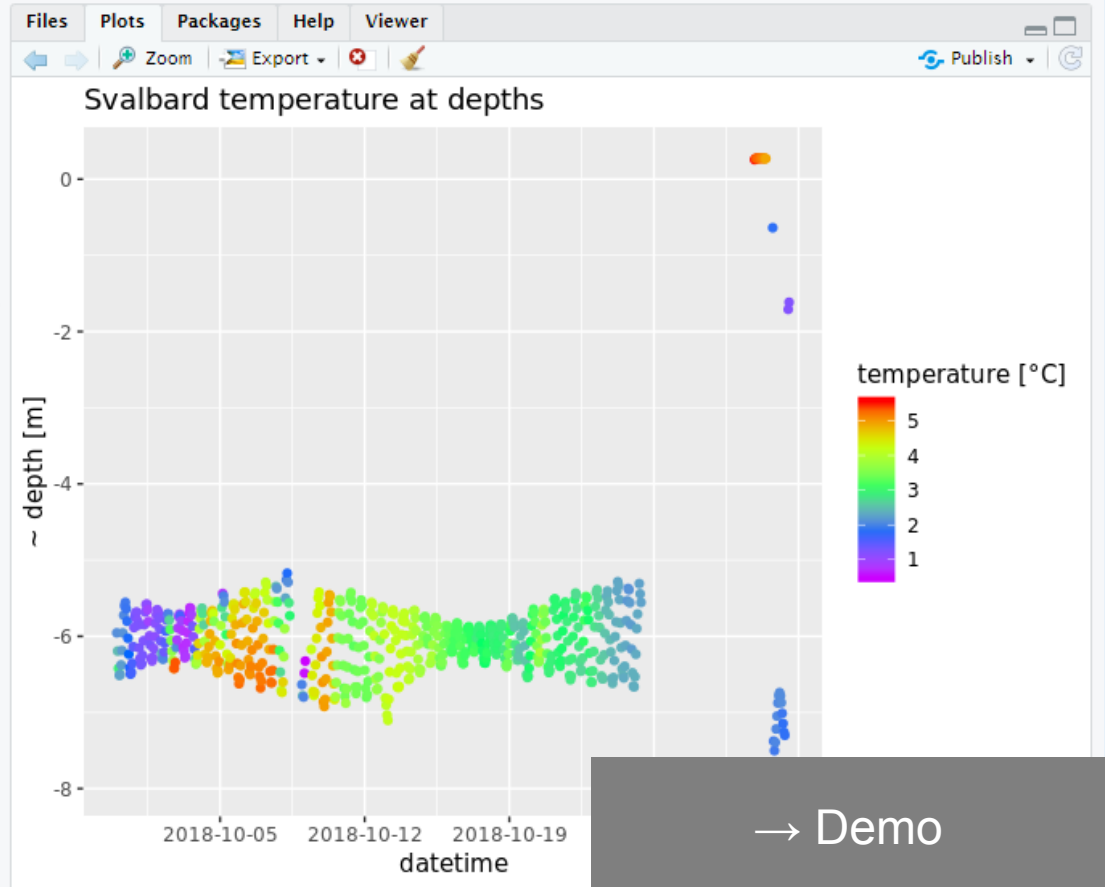
Environment | History | Connections | Git

Global Environment

Data	
data	621 obs. of 4 variables
gg	List of 9
meta	Large list (8 elements, 964.3 Kb)
platform	List of 6
range	List of 4
sensor	List of 7
sensors	9 obs. of 6 variables

Values

dws.DATA_BASE_URL	"https://dashboard.awi.de/data-xxl/rest"
-------------------	--



Using data web services

We show the use of data web services Python interface for an use case: temperature data from Svalbard

Load data and metadata

Import the data web service Python interface. The data web service interface can be found here: <https://github.com/rkoppe/dws>. Download the dws.py file and place it in your working directory.

```
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://senso> and find useful platforms and further metadata. Platforms, devices and sensors have codes (aka um) identifying measurements.

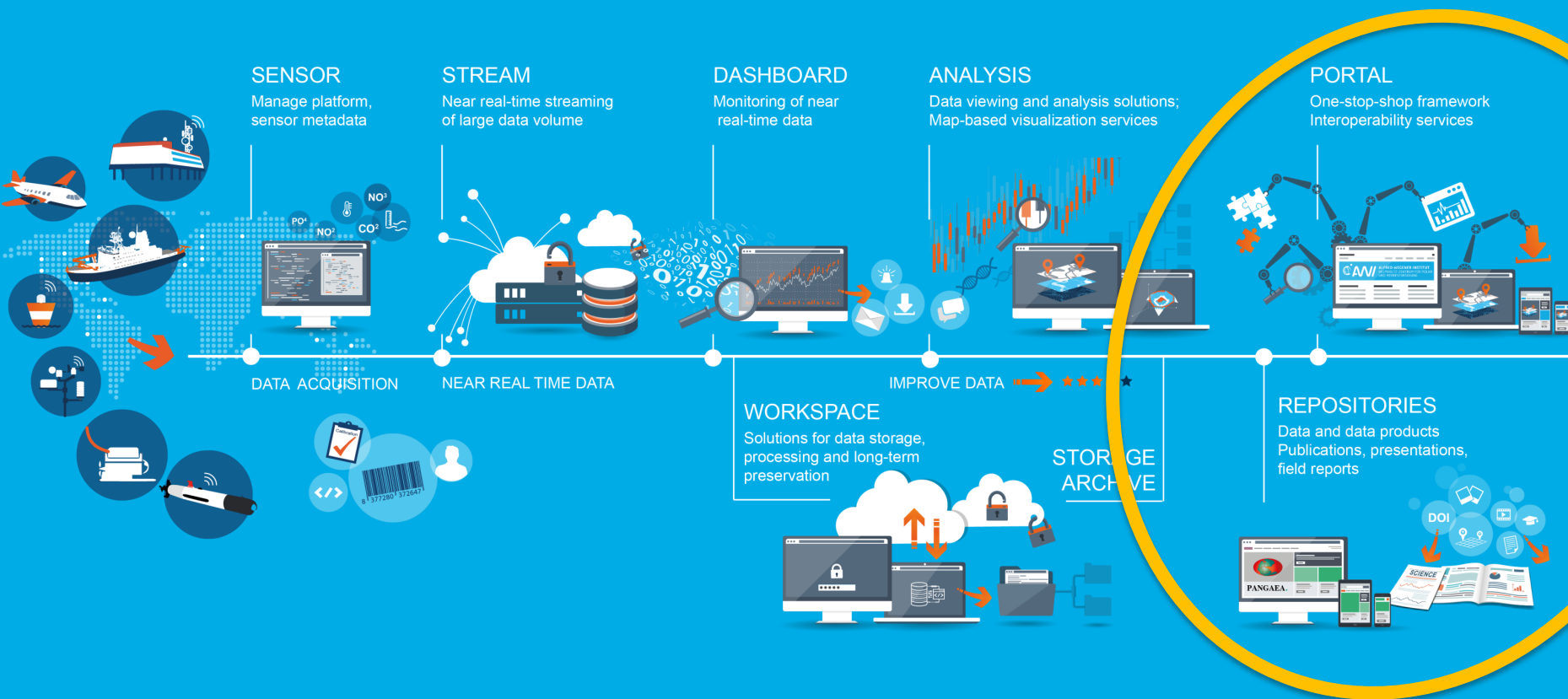
```
In [2]: 1 sensors = dws.sensors('station:svluwobs*temperature*')
        2 pd.DataFrame.from_dict(sensors)
```

Out[2]:

	code	id
0	station:svluwobs:svluw2:ctd_103:temperature_se...	1257
1	station:svluwobs:svluw2:adcp_17374:temperature	1260
2	station:svluwobs:svluw2:sbe38_657:temperature	427
3	station:svluwobs:svluw2:ctd_181:temperature	441
4	station:svluwobs:fb_731101:oxygen_sensor_574:t...	750
5	station:svluwobs:fb_731101:temperature_freshwa...	753
6	station:svluwobs:fb_731101:temperature_contain...	754
7	station:svluwobs:fb_731101:temperature_outside...	755
8	station:svluwobs:fb_731101:sbe45_0403:temperature	746
9	station:svluwobs:svluw2:ctd_578:temperature_se...	1024
10	station:svluwobs:svluw2:ctd_964:temperature_se...	1312
11	station:svluwobs:svluw2:ctd_183:temperature_se...	1321

- Foster sharing of code incl. documentation
- Support native book-keeping of changes, versioning
- Promote methods for reproducible results
- Notebooks as supplement for publications

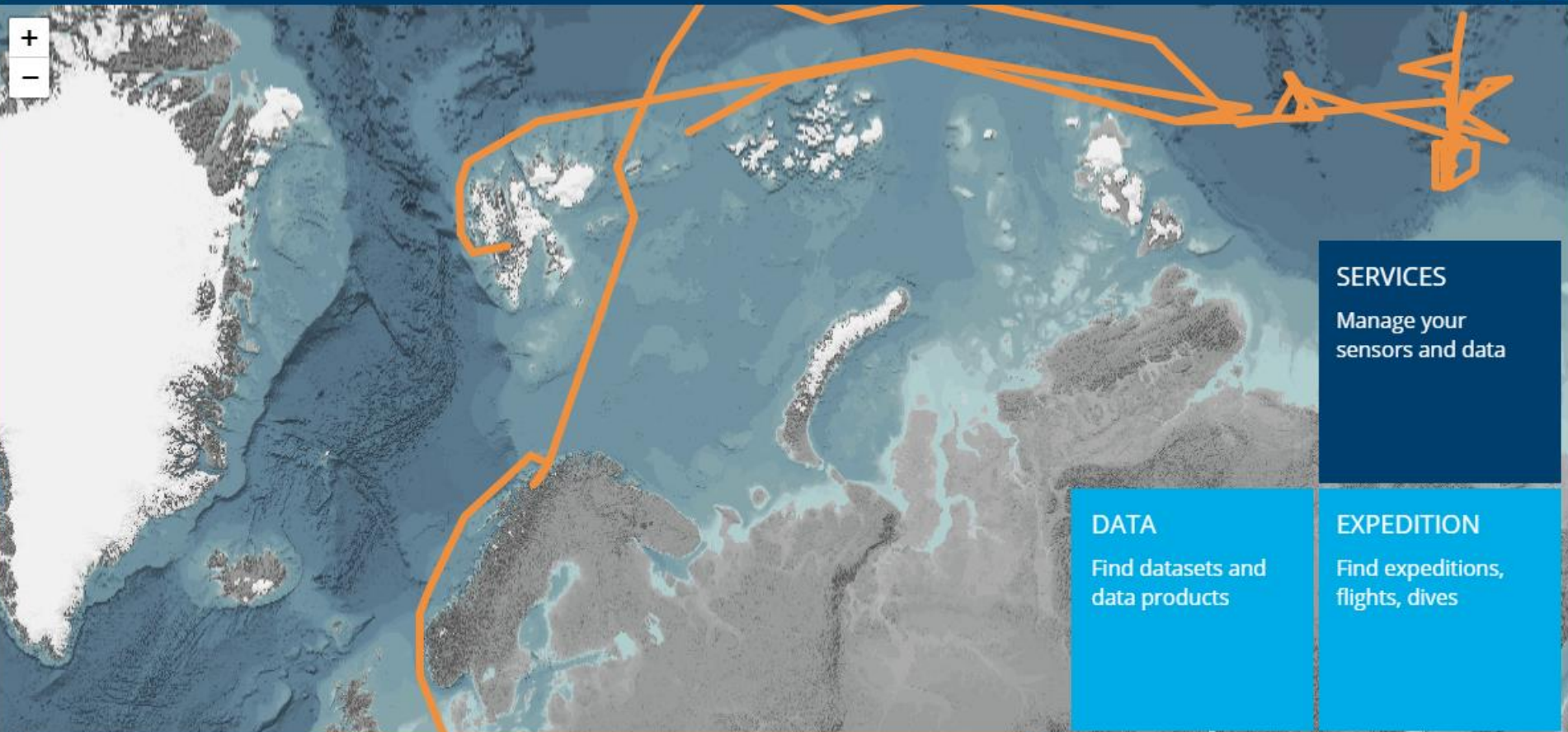
Data Flow Framework



The screenshot displays a web-based map interface. On the left, a 'Map Content' panel is open, showing a 'Data Layers' section with various layers like 'Moorings', 'Drifting Traps', 'Bottom Landers', 'Planned Moorings', 'Buoys', 'HAUSGARTEN Sites', 'Sea-ice Minimum', 'Sea-ice Maximum', 'Maritime Boundaries', and 'Depth Contours [m]'. The 'Buoys' layer is expanded, showing a list of buoy identifiers with corresponding colored dots: 2014T14 (green), 2015S16 (dark green), 2016S44 (red), 2016M1 (pink), 2016S36 (light purple), 2016S45 (purple), 2016S46 (blue), 2016S50 (yellow), and 2016ITP93 (orange). Below the layers is a 'Filter by Year' section with a range from 2013 to 2016. The main map area shows a satellite-style view of a coastal region with a yellow line indicating a path or boundary. A vertical toolbar on the right side of the map contains icons for zooming in (+), zooming out (-), full screen, and other navigation tools. At the bottom of the map, a coordinate box shows 'Latitude: 82.59 Longitude: -34.19' and a scale bar for 200 km.

- 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

Search for author, expedition, project, ...



SERVICES
Manage your sensors and data

DATA
Find datasets and data products

EXPEDITION
Find expeditions, flights, dives

Latest expeditions:
PS115/2 with Polarstern 2018-09-05 - Longyearbyen
HE521 with Heincke 2018-10-30 - Bremerhaven
2018-10-16 - Bremerhaven
2018-11-03 - Bremerhaven

Latest news:
Measuring ice and snow thickness : Poke it with a stick | (Last blogpost from PS115 2 expedition to the Arctic after

COLLECTIONS
Explore data and products thematically grouped

NEAR REAL TIME DATA
Near real time data is presented in hourly averages and no quality control is applied.



FRAM

FRontiers in Arctic Marine Monitoring



Layers

Dynamic layers

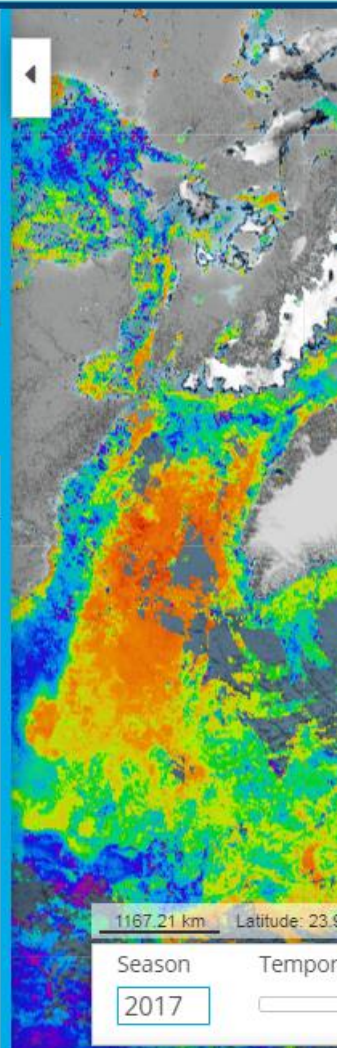
- Sea-ice coverage
- Chlorophyll a

Product layers

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

Platform layers

- Platforms
- Buoys



- Portal as central access point to linked and combined data
- Platforms, devices and sensor overviews
- Expeditions
- Data, publications and reports
- Interactive data products
- Services

→ Demo



In Action

**Frontiers in
Arctic Marine Monitoring**

**DIGITAL
EARTH**

The logo for Digital Earth, with the word 'DIGITAL' in a spaced-out blue font above the word 'EARTH' in a larger, bold blue font. To the right of the text is a cluster of blue and green squares of varying sizes, some overlapping, representing a digital grid or data points.

MOSAIC

The logo for MOSAIC, featuring the word 'MOSAIC' in a large, bold blue font. Below it is a stylized blue ship silhouette. To the left of the ship is a 3D map of the Arctic region, composed of blue rectangular blocks representing ice floes. The text 'International Arctic Drift Expedition' is written in a smaller blue font to the left of the map.

MOSES

The logo for MOSES, with the word 'MOSES' in a large, bold blue font. The letter 'O' is replaced by a green circle containing a white pulse line, resembling a heart rate monitor. Below the word is a blue horizontal bar with the text 'Modular Observation Solutions for Earth Systems' in white.

**HELMHOLTZ
Data Federation | HDF**





The expedition's planned duration is **350 days** in the ice.



The following **17 nations** will participate in the expedition:



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.



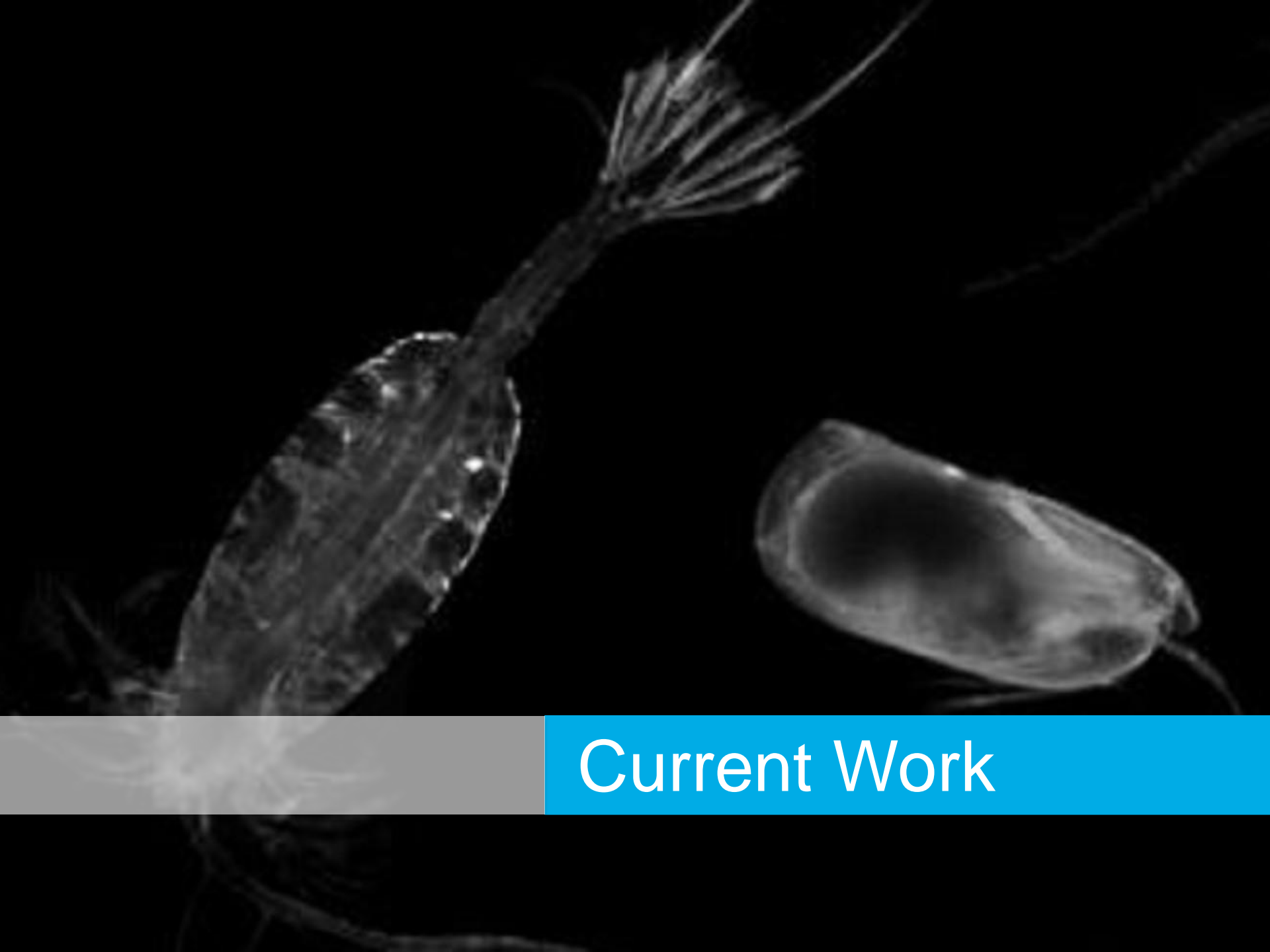
Throughout the year, a total of **600 experts** will be on board, and be exchanged in phases.



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

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 – PIDs 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!

roland.koppe@awi.de

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