SeaDataCloud

The SeaDataCloud VRE - Final release

Sebastian Mieruch (AWI), Merret Buurman (DKRZ), Peter Thijsse (MARIS)
on behalf of the VRE team

SDC Plenary - 29 Oct 2020

sdn-userdesk@seadatanet.org – www.seadatanet.org

1. Introduction + Overview efforts final year

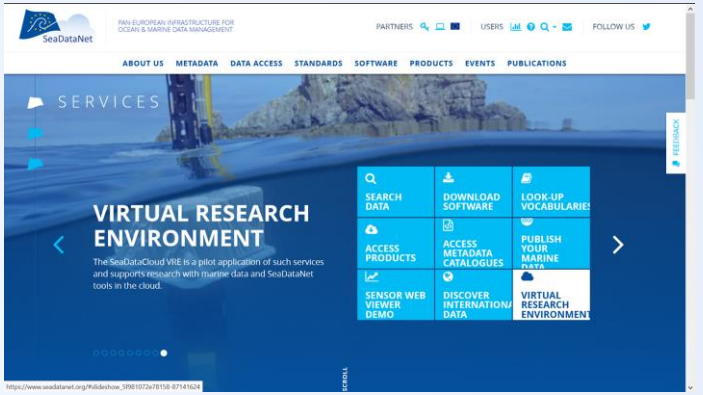
Introduction

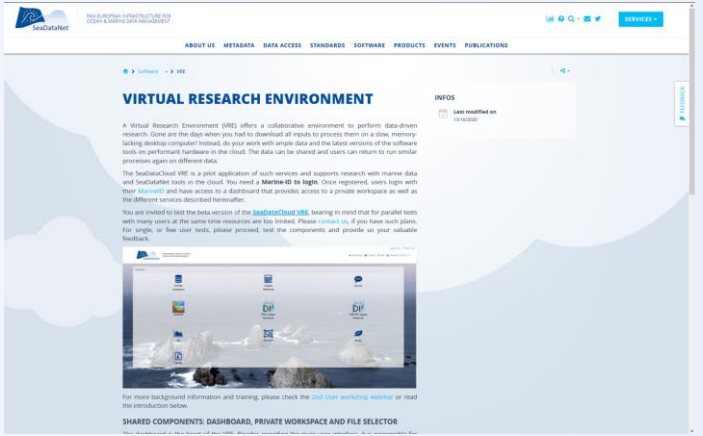
Today's overview of results:

1. Introduction of last year effort
2. Development and production environment
3. System for accounting usage
4. VRE Dashboard
5. WebODV
6. DIVAnd
7. Subsetting via ERDDAP
8. Advanced visualisations
9. Biology QC

Main actions year 4

- During several (virtual) code sprints components have been implemented in the VRE environment and matured
- Solutions for earlier user feedback
- Next SDC and Emodnet Chemistry user consultation meeting (Sept 2020)
- Contribution to SDC webinar (Sept 2020)
- Two versions now: Development version and production version
- Launched and visible on the SDN website now! => https://www.seadatanet.org/Software/VRE





2. Development and production environment

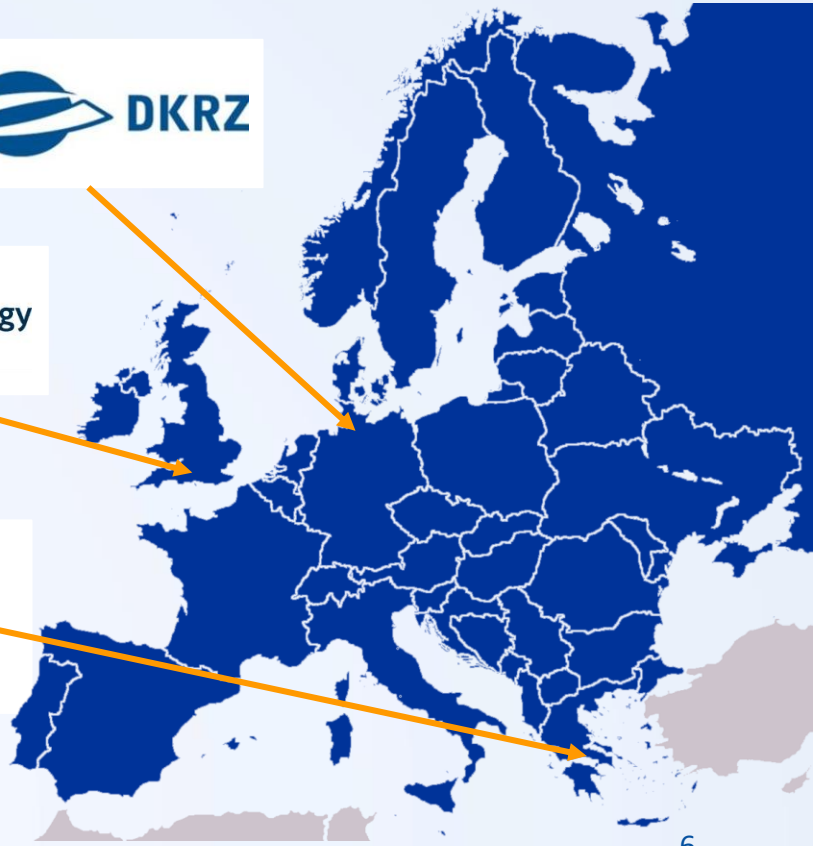
The VRE Production Environment

3 European Data Centers

DKRZ

Science & Technology Facilities Council

grnet



VRE Infrastructure now: Servers

jellyfish.argo.grnet.gr

Processing Services:

- BioQC
- VIZ

vre.seadatanet.org

Central Services:

- Dashboard
- Workspace
- User mgmt.
- + ODV

vre3.argo.grnet.gr

Processing Services:

- ERDDAP

bluewhale.dkrz.de

Processing Services:

- DIVA
- Python/R

VRE Infrastructure now: Data

jellyfish.argo.grnet.gr

Processing Services:

- BioQC
- VIZ

vre.seadatanet.org

Central Services:

- Dashboard
- Workspace
- User mgmt.
- ODV

vre3.argo.grnet.gr

Processing Services:

- ERDDAP

bluewhale.dkrz.de

Processing Services:

- DIVA
- Python/R

Container Cleanup

We now remove user's containers after 14 days without any login.

Many users are not regular. If we keep their containers alive eternally, we waste resources that other users might need!

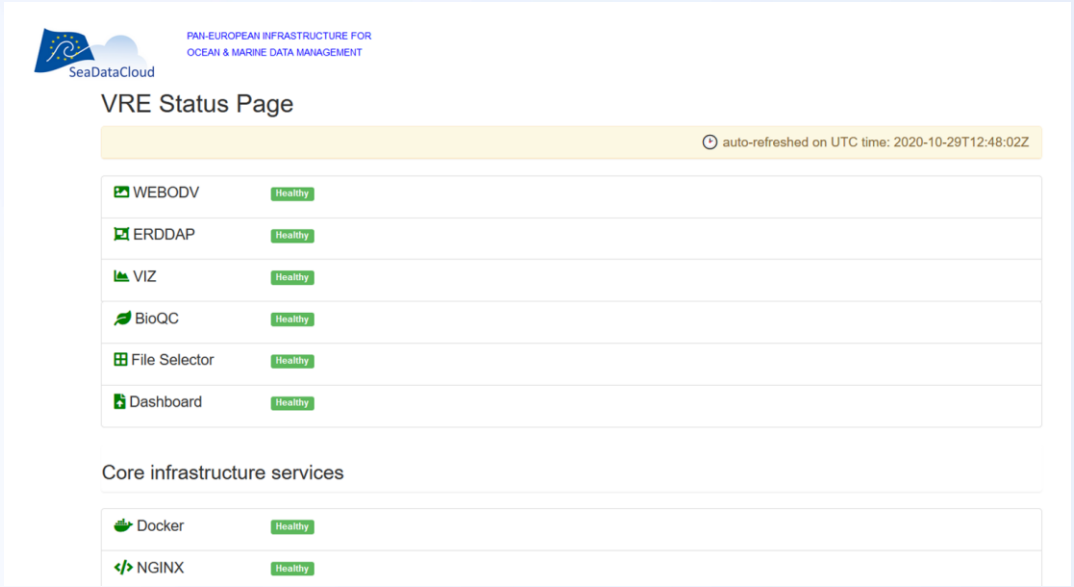
Note: On request, we can "protect" a user's container from deletion!

WIP: Email warning before deletion!

VRE Health Status

A status page informs the users when a service or the core infrastructure has a problem.

Check it out: https://vre.seadatanet.org/status/



VRE statistics

"The ethical alternative"

"Don't damage your reputation with Google Analytics"

"You could lose your customers' trust and risk damaging your reputation if people learn their data is used for Google's "own purposes"."
(Source: https://matomo.org/)

What's new since last plenary?

- Added more RAM and storage
- Improved user data handling
- Introduced health checking
- Introduced container removal
- Improved Firewall settings (no more non-standard ports → less risk of users running into the Firewall; …)

- ...and lots of work on the services!

(Possible) future work

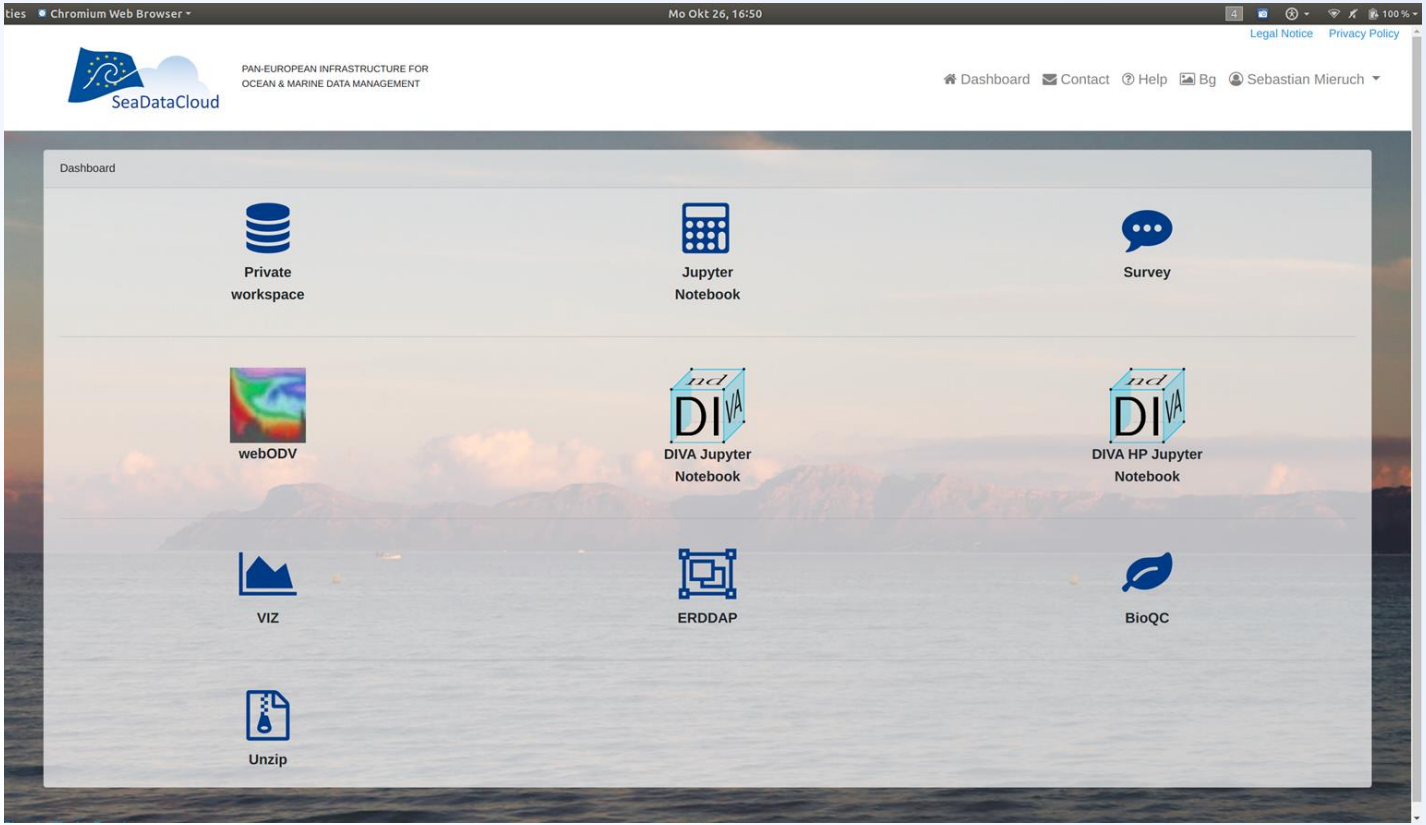
- Email warning before container deletion [WIP]
- CDI Bridge [WIP]
- Better distribution of load (load varies)
- Optimize data distribution (automatic sync, parallel transfers, …)
- (etc.)

Dashboard

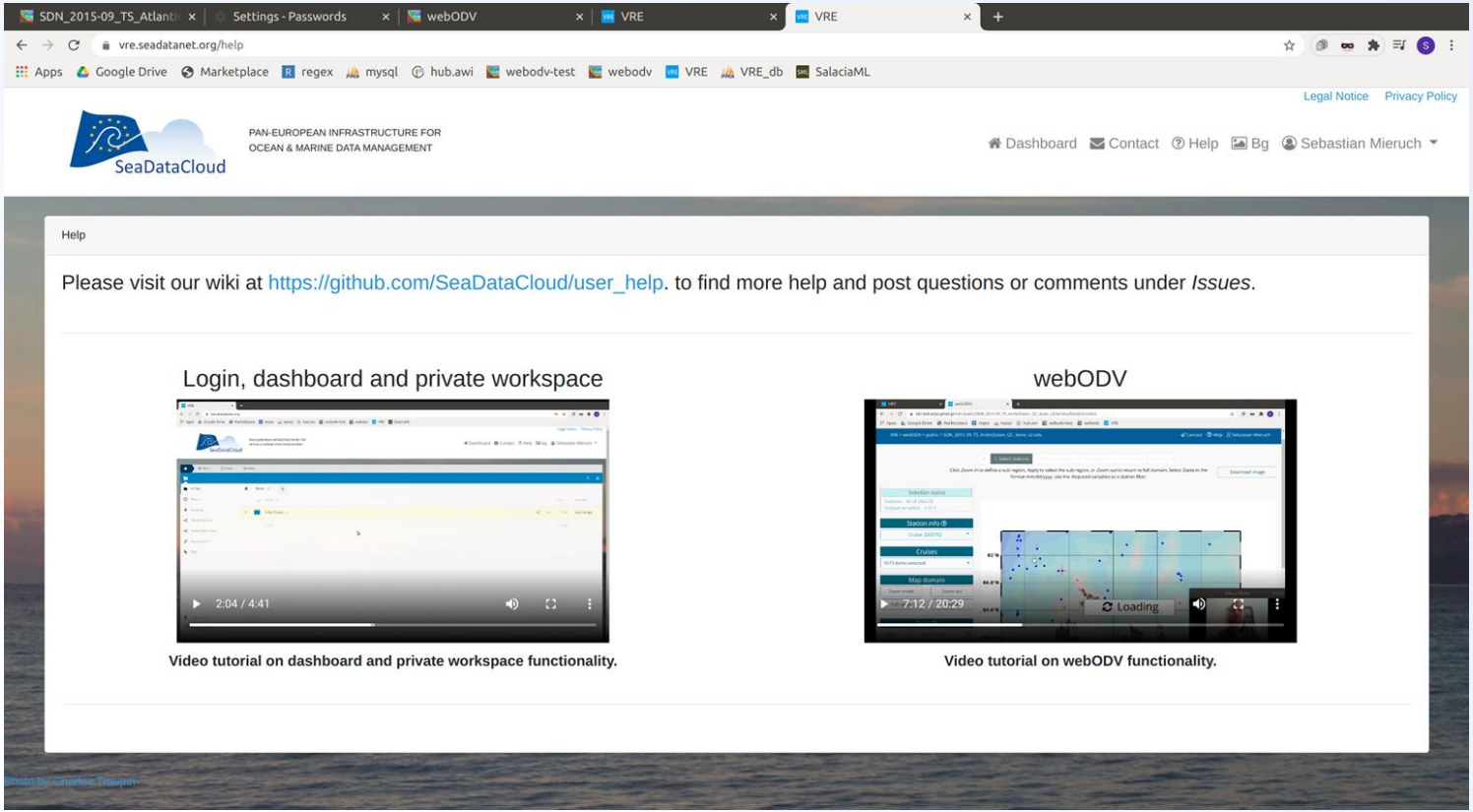
Login


Contact
Legal Notice
Privacy Policy
SeaDataCloud
VIRTUAL RESEARCH ENVIRONMENT
LOGIN WITH MARINE ID
Not registered?
Click on LOGIN and register your MARINE ID.
Photo by Charles Troupin

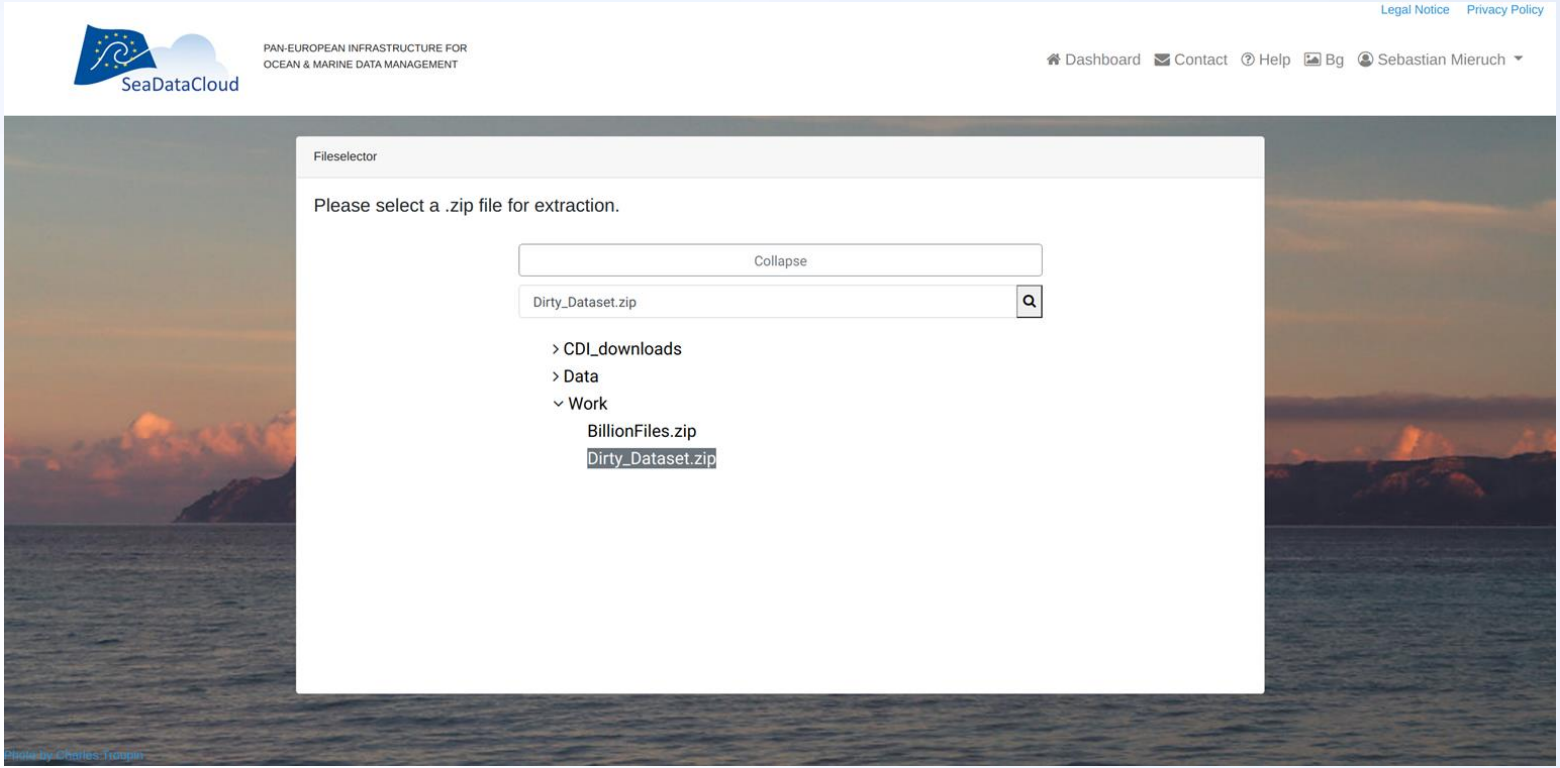
Dashboard


Chromium Web Browser
Mo Okt 26, 16:50
Legal Notice
Privacy Policy
SeaDataCloud
PAN-EUROPEAN INFRASTRUCTURE FOR
OCEAN & MARINE DATA MANAGEMENT
Dashboard
Contact
Help
Bg
Sebastian Mieruch
Dashboard
Private workspace
Jupyter Notebook
Survey
webODV
DIVA Jupyter Notebook
DIVA HP Jupyter Notebook
VIZ
ERDDAP
BioQC
Unzip

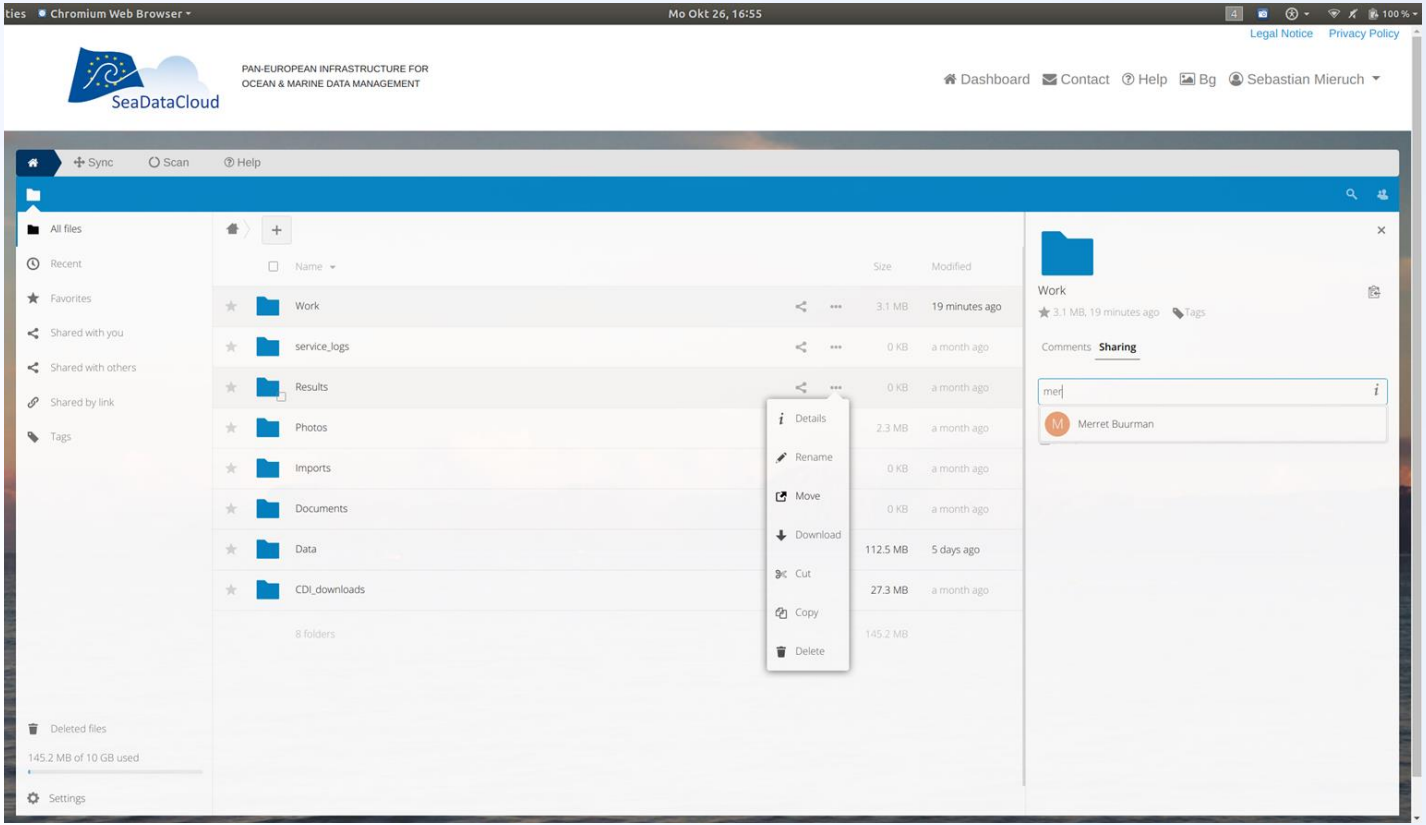
Help - Videos


vre.seadatanet.org/help
Apps
Google Drive
Marketplace
regex
mysql
hub.awi
webodv-test
webodv
VRE
VRE_db
SalaciaML
Legal Notice
Privacy Policy
PAN-EUROPEAN INFRASTRUCTURE FOR
OCEAN & MARINE DATA MANAGEMENT
SeaDataCloud
Dashboard
Contact
Help
Bg
Sebastian Mieruch
Help
Please visit our wiki at https://github.com/SeaDataCloud/user_help. to find more help and post questions or comments under Issues.
Login, dashboard and private workspace
2:04 / 4:41
Video tutorial on dashboard and private workspace functionality.
webODV
7:12 / 20:29
Loading
Video tutorial on webODV functionality.

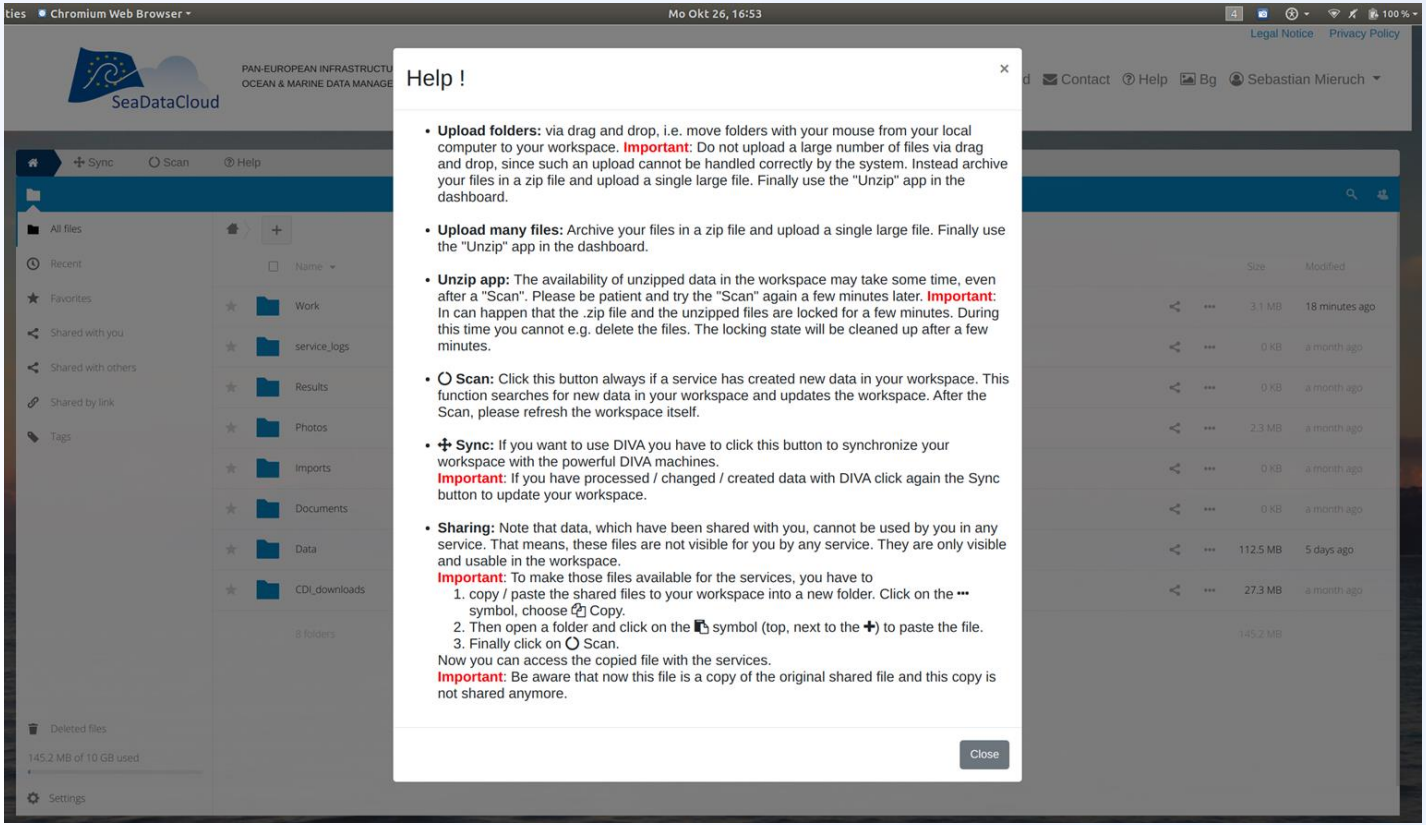
Unzip App


Legal Notice
Privacy Policy
SeaDataCloud
PAN-EUROPEAN INFRASTRUCTURE FOR
OCEAN & MARINE DATA MANAGEMENT
Dashboard
Contact
Help
Bg
Sebastian Mieruch
Fileselector
Please select a .zip file for extraction.
Collapse
Dirty_Dataset.zip
CDI_downloads
Data
Work
BillionFiles.zip
Dirty_Dataset.zip

Workspace

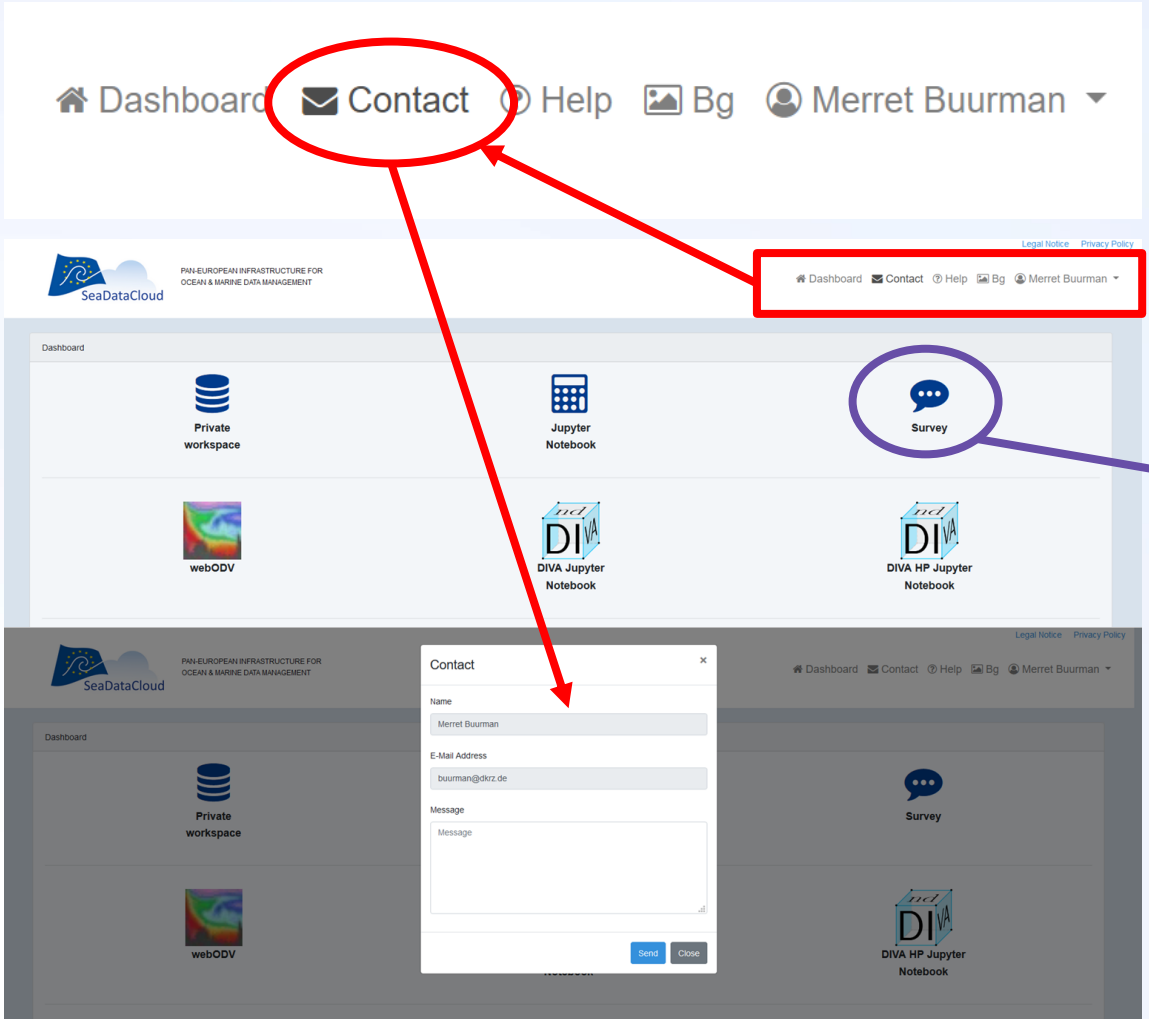

Chromium Web Browser
Mo Okt 26, 16:55
Legal Notice
Privacy Policy
SeaDataCloud
PAN-EUROPEAN INFRASTRUCTURE FOR
OCEAN & MARINE DATA MANAGEMENT
Dashboard
Contact
Help
Bg
Sebastian Mieruch
Sync
Scan
Help
All files
Recent
Favorites
Shared with you
Shared with others
Shared by link
Tags
Deleted files
145.2 MB of 10 GB used
Settings
Name
Size
Modified
Work
3.1 MB
19 minutes ago
service_logs
0 KB
a month ago
Results
0 KB
a month ago
Photos
2.3 MB
a month ago
Imports
0 KB
a month ago
Documents
0 KB
a month ago
Data
112.5 MB
5 days ago
CDI_downloads
27.3 MB
a month ago
8 folders
145.2 MB
Details
Rename
Move
Download
Cut
Copy
Delete
Work
3.1 MB, 19 minutes ago
Tags
Comments
Sharing
mer
Merret Buurman

Workspace - Help



VRE feedback

Please let us know what you think! Thanks a lot! :)



Survey

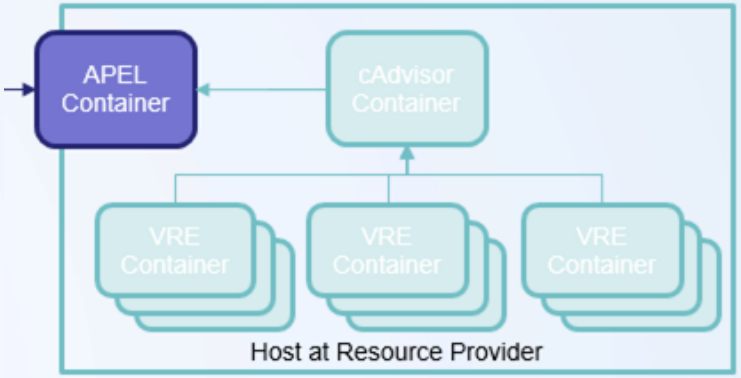
3. Accounting (STFC)

Accounting System

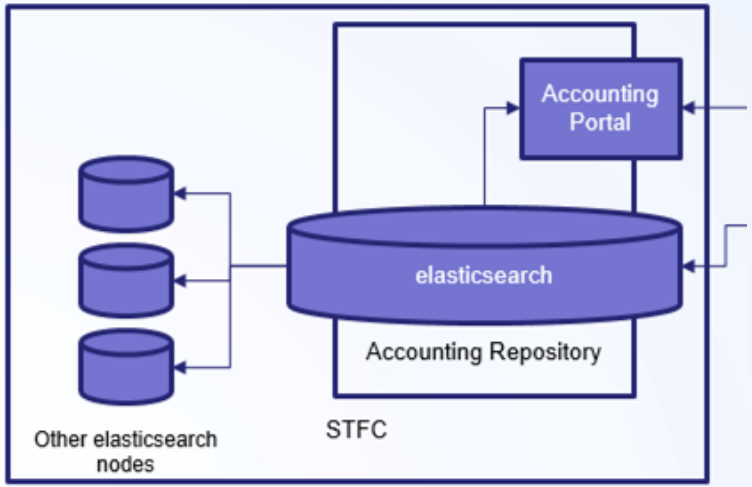
- cAdvisor interacts with Host to monitor container resource usage
  - CPU, Storage and Network usage

- APEL Container periodically polls the cAdvisor API to extract accounting data
  - Frequency of polling can be increased or decreased as needed

- System deployed via docker-compose, just like other containers in the VRE



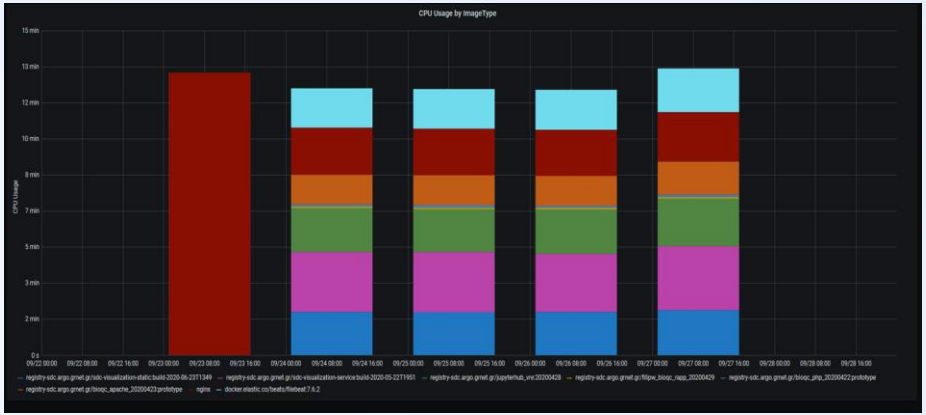
Accounting System

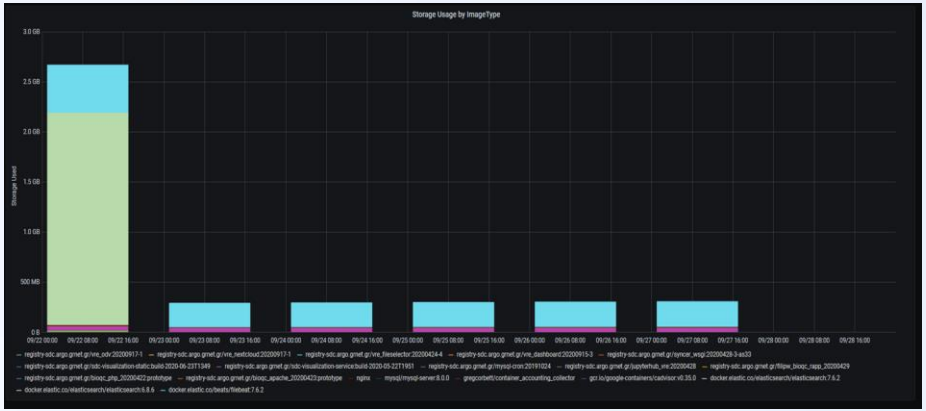


- APEL container compares observed usage with usage already reported
  - This allows it to correctly identify container restarts in lieu of a orchestrator.
- Published data is sent to the accounting portal where it is stored in elasticsearch.
  - Data is then replicated to other nodes in the cluster

Accounting Portal

- Portal graphs usage by container image over time.
- Filters allow the usage of a subset of containers to be graphed.
- Usage can also be filtered by resource provider, i.e. DKRZ or GRNET.

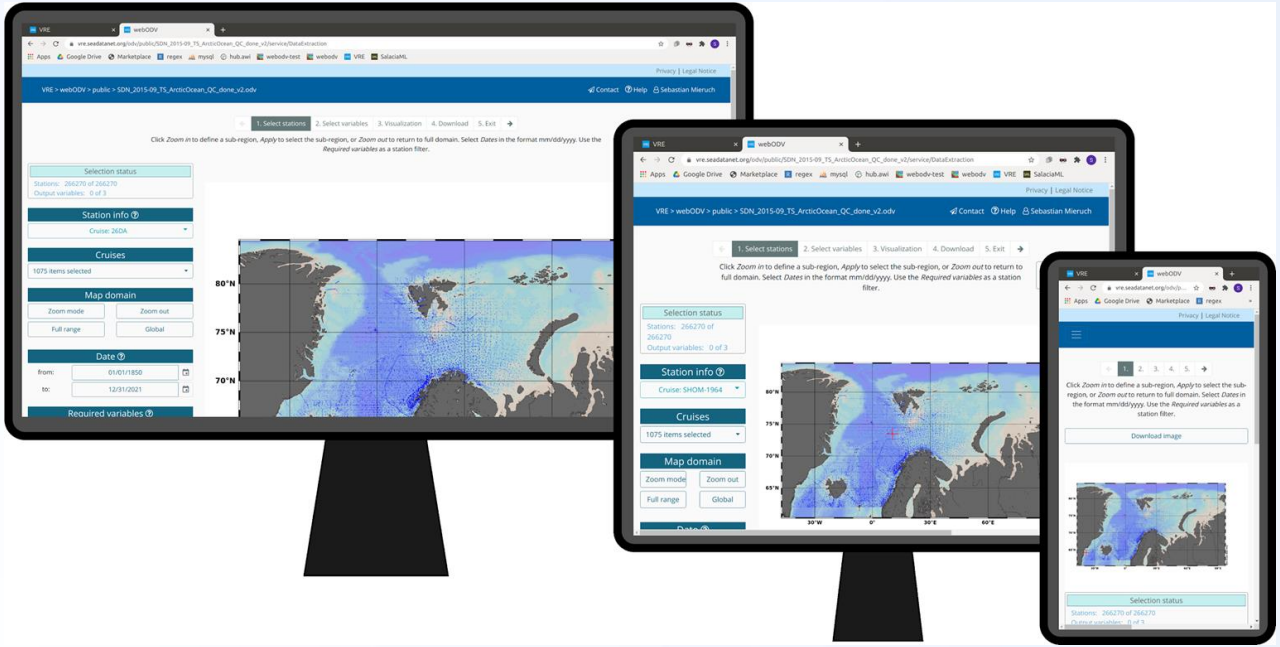




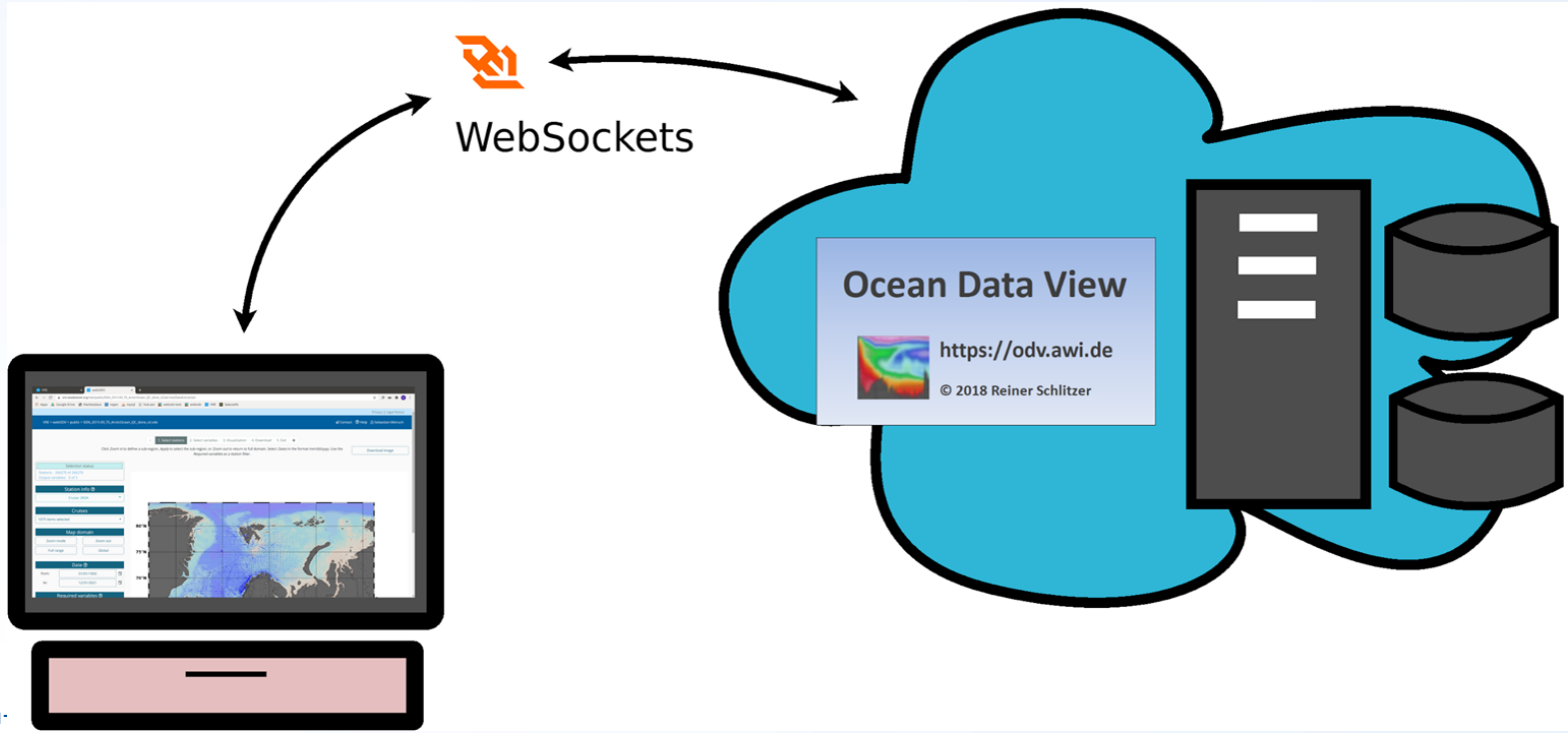
4. webODV

webODV in the VRE

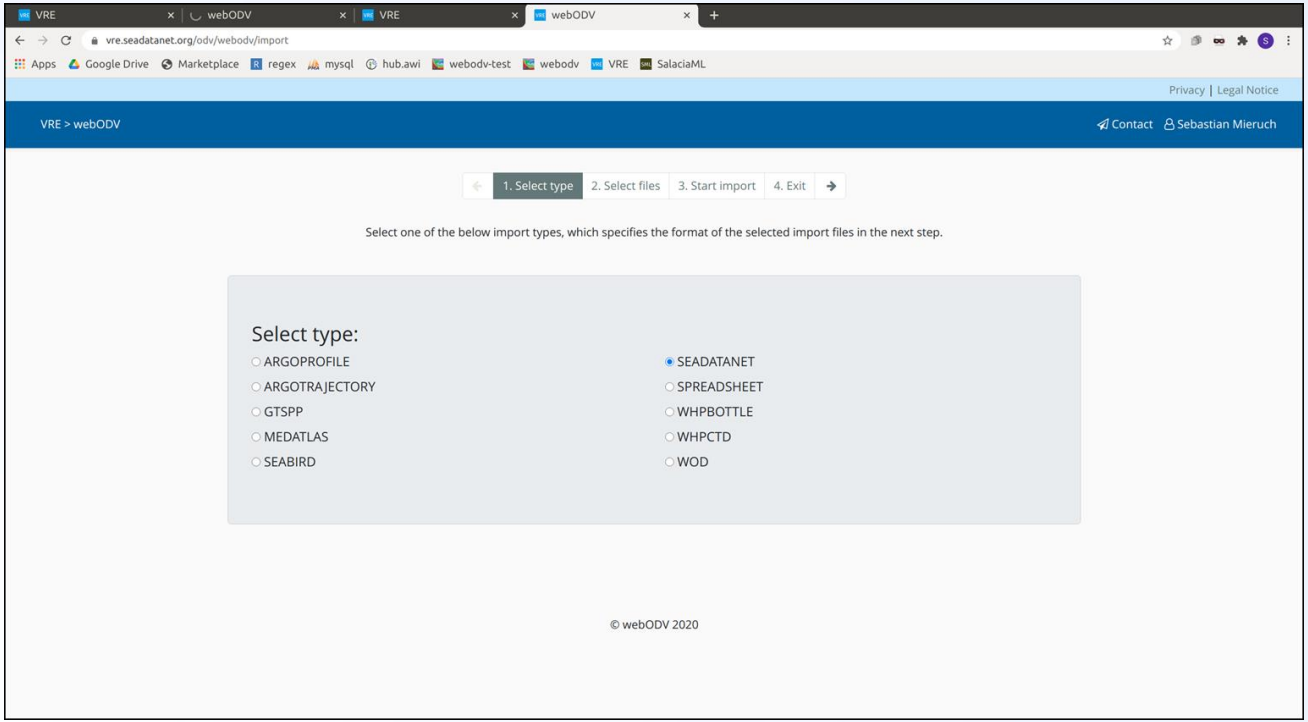
Sebastian Mieruch and Reiner Schlitzer



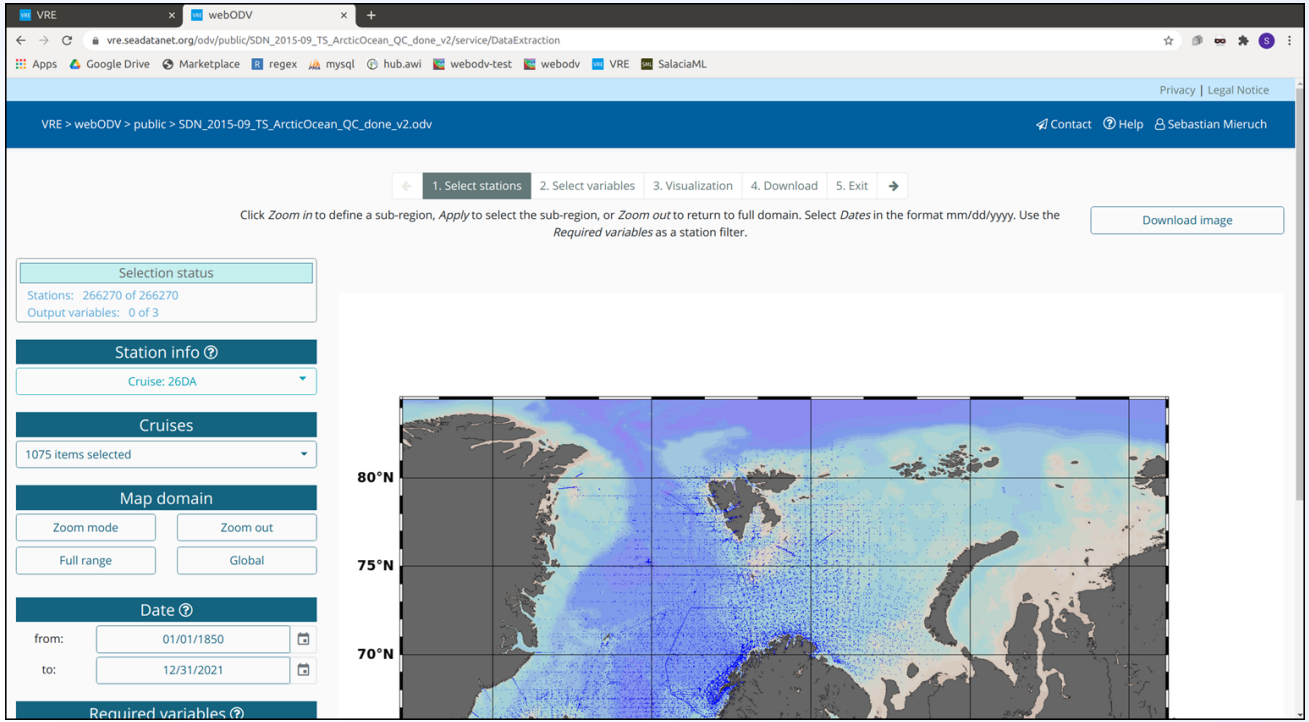
webODV - The concept



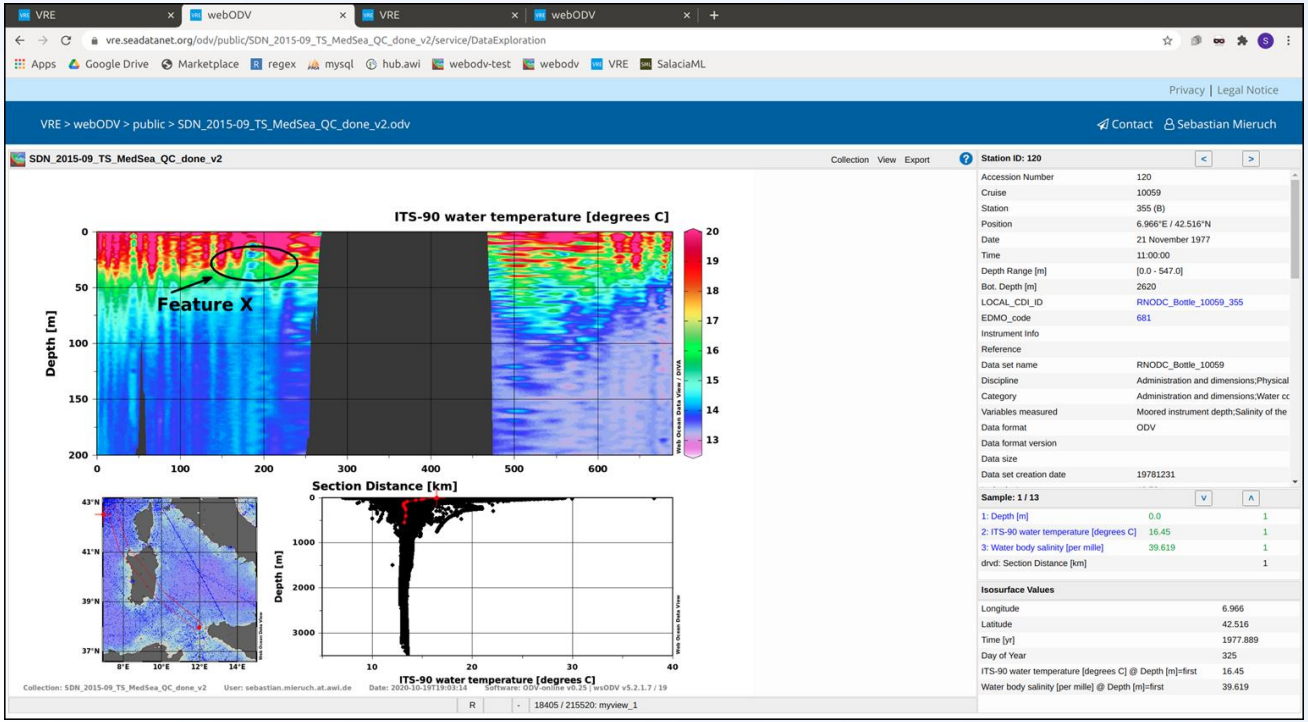
webODV - Import Service



webODV - Extractor Service



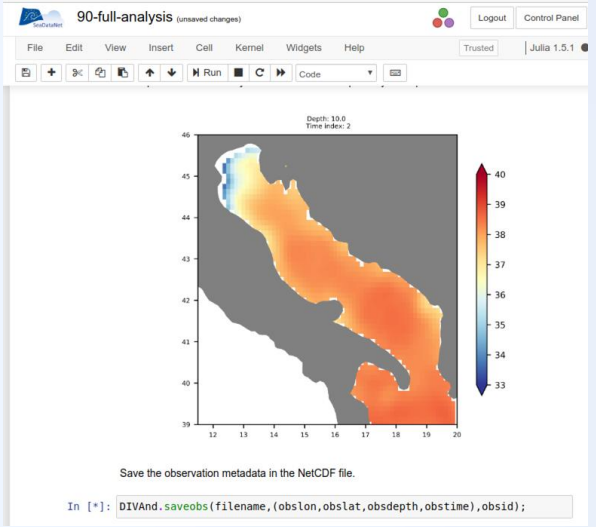
webODV - Explore Service



5. DIVAnd

DIVAnd with jupyter notebooks

- DIVAnd in the VRE is based on jupyter notebooks
- Even more widespread now compared to the start of the project
  - Google colab
  - MyBinder.org
  - WEKEO uses Jupyter notebooks to access ESA satellite data
  - …
- Jupyter, installable on
  - client (user laptop/desktop) → jupyter notebook
  - server (group of users) → jupyterhub
  - Identical user interface
- Jupyterhub reduces the need to download large datasets to the client machine (or upload the final result)



sdn-userdesk@seadatanet.org – www.seadatanet.org

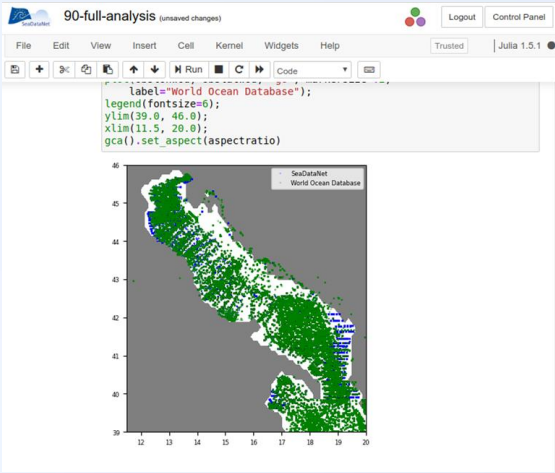
DIVAnd

Most feature of the DIVA tool have been ported to DIVAnd and are available in the VRE, such as:

- **correlation length estimation**
- **XML metadata for Sextant**

New features not in previous DIVA tool:

- Access additional data from **World Ocean Database**
- Detect duplicates based (on distance, time and measurement value)
- Support of new **ODV netCDF format**
- Support for **HF radar data** (manuscript in review at Ocean Dynamics)
  - **New product for Ibiza Channel** with SOCIB (WP11)



DIVA workshop (27 - 30 January 2020)

- 30 registered participants (largest DIVA workshop so far)
  - SeaDataCloud, EMODnet Chemistry, EMODnet Physics EMODnet Biology, …
- Topics covered:
  - Introduction to DIVAnd
  - Background estimation
  - Correlation length
  - Estimation of the expected error
  - ...
- Additional speakers: Nadia Pinardi, Romain Escudier, Simona Simoncelli
- Demo of the VRE during this workshop

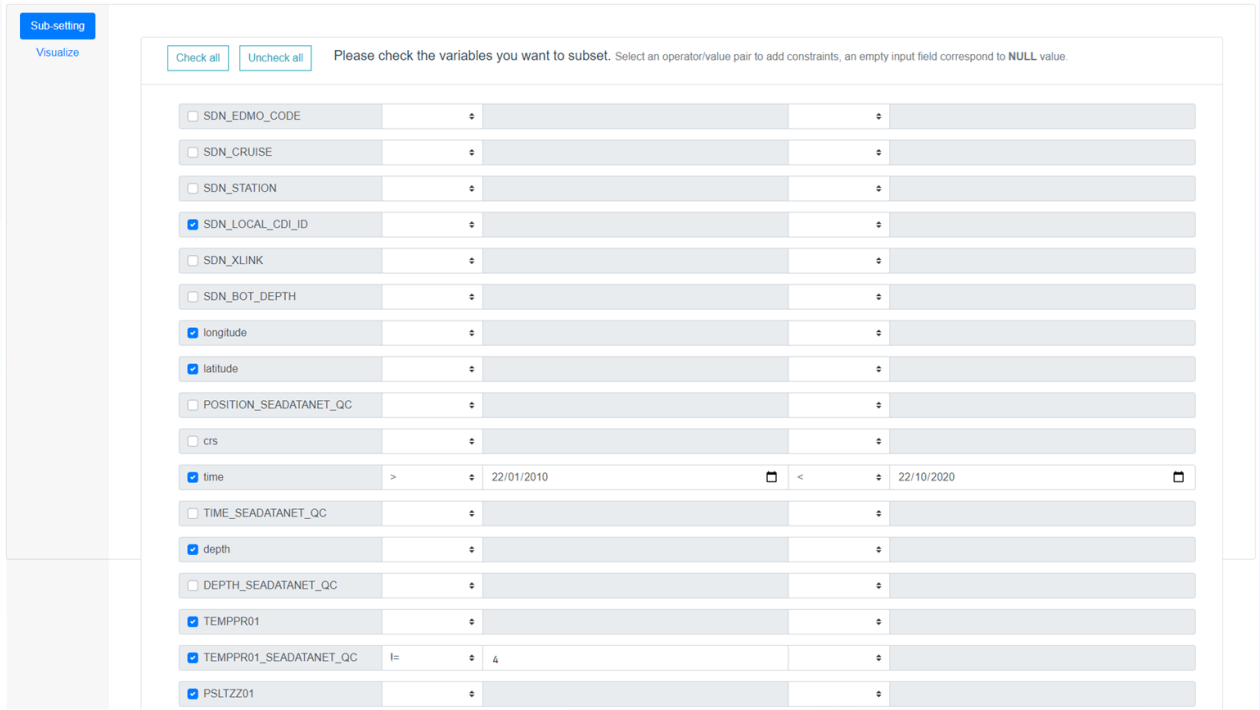
6. Subsetting service using Erddap

Service mechanism

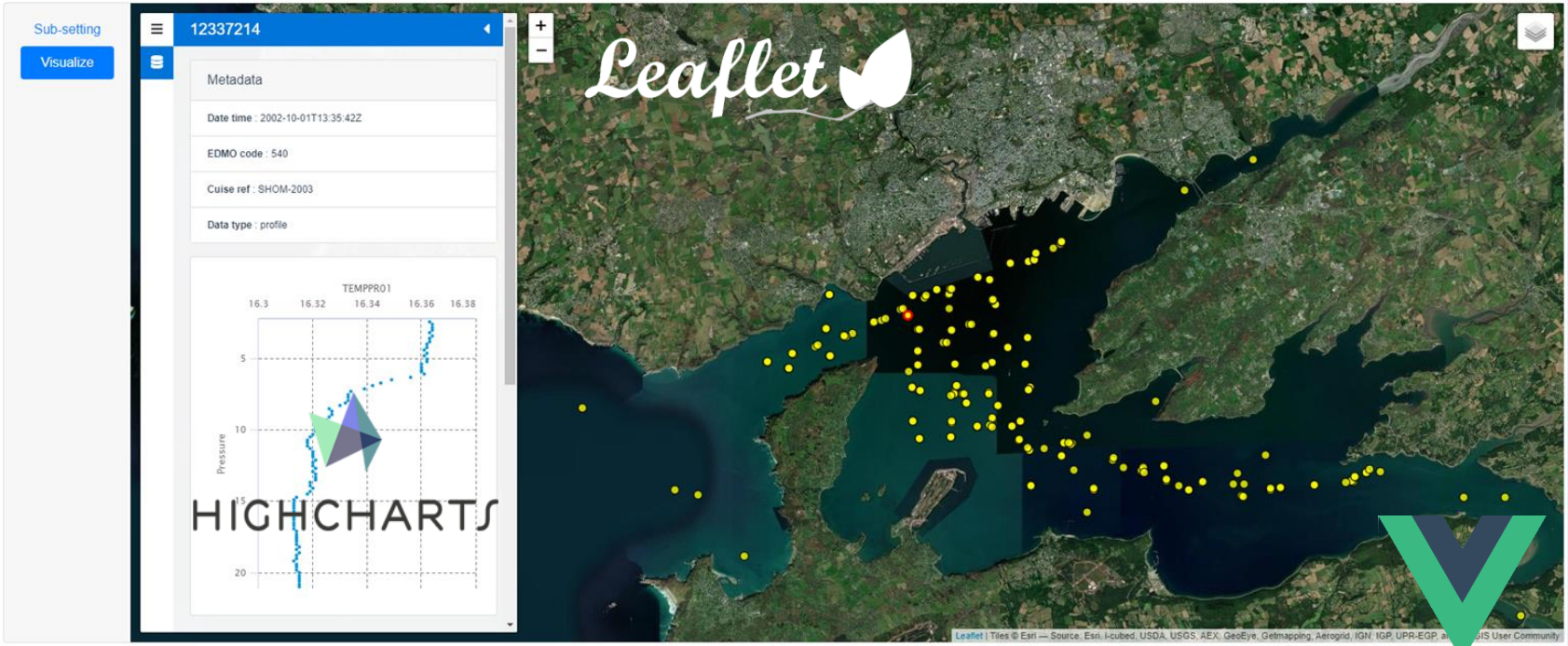
- Generally, Erddap is deployed as a server and proposes a list of several datasets preconfigured.

- Subsetting service needs a most specific targeting user's experience.
- The idea is to let the user choose the dataset he wants to subset and then configure Erddap according to this subset.

Subsetting with Custom data access form



Visualization with modern libraries



Waiting for P02 feature

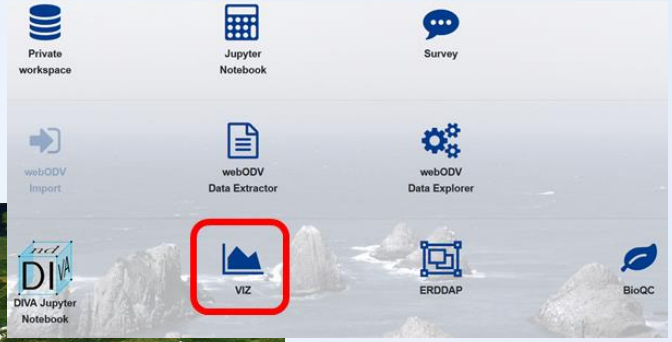
- Waiting for a new version on Erddap including P02 feature to fully finalize the service.
- With this feature we will be able to work with homogeneous datasets thanks to P02 vocabulary.

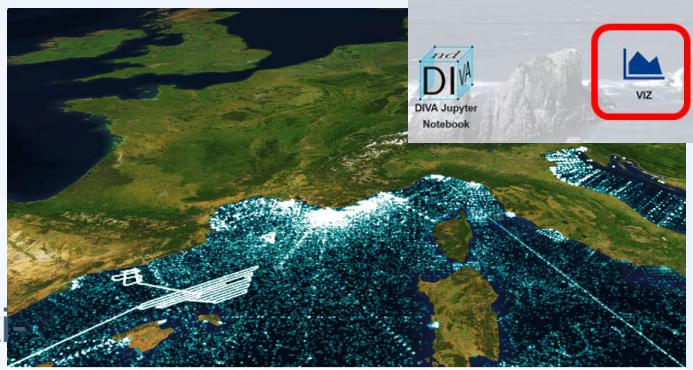
- Version 2.10 of Erddap should be available soon.
- Question is : Does this version will include P02 feature ?

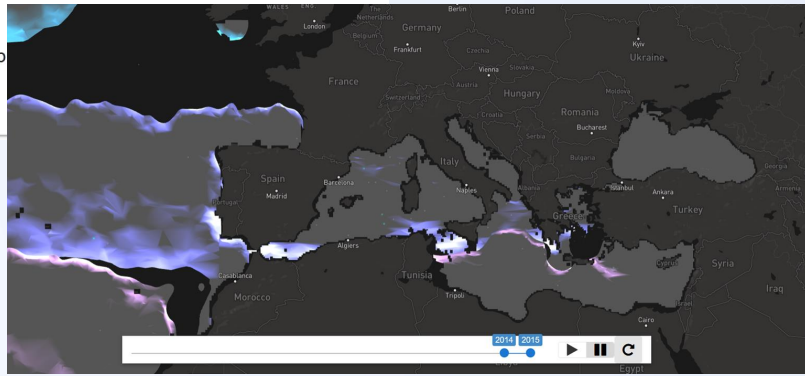
7. Advanced visualisations by Deltares

Giorgio Santinelli, Cindy van de Vries, Fedor Baart

Visualisation services







- Comprehensive visualisation of available data
- Provide a simple way to span through data in space and time
- Framework to inspect large multi dimensional datasets
- Simultaneously visualise observations and climatologies
- https://github.com/openearth/sdc-visualization

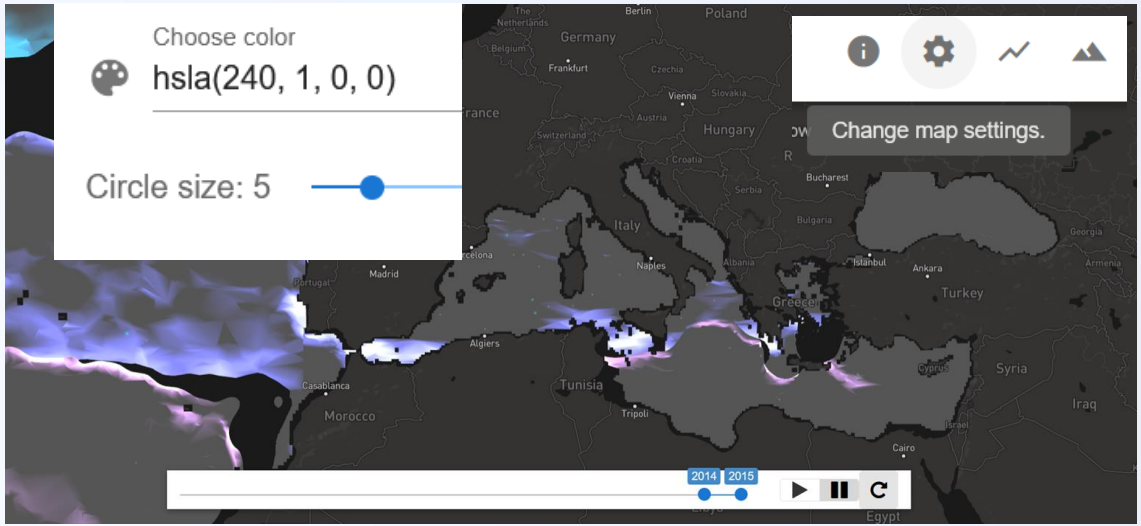
sdn-userdesk@seadatanet.org – www.seadatanet.org

Updates

- Adding Salinity/Temperature layer to map toggle
- Changing point colour/size
- Adding tooltips and github wiki page
- Exporting graphs

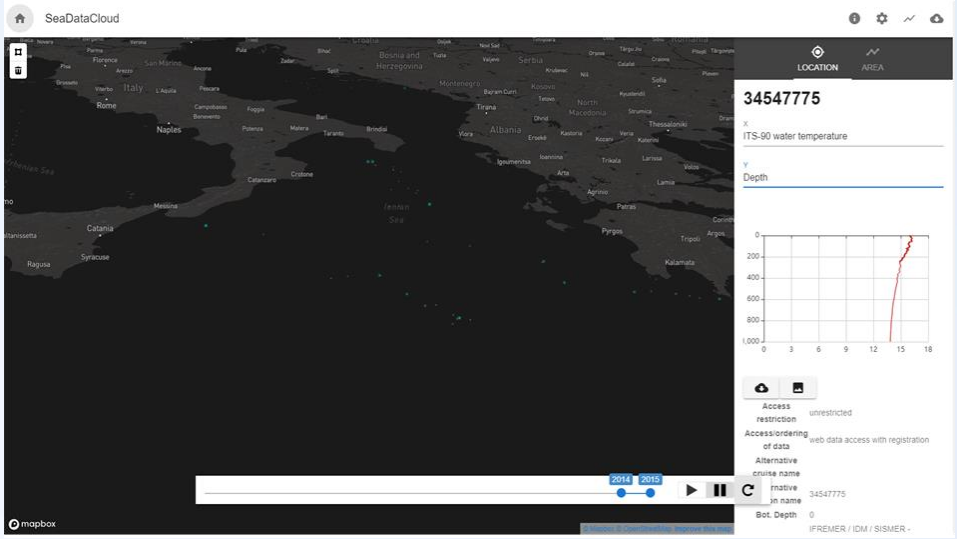
- Components
  - time, map, charts

- Functionalities
  - observations
  - bbox for trajectories
  - rendered gridded files



Visualisation

- Observations
  - Select axes
  - Single profile
  - Metadata
  - Export
  - Trajectories
  - Bbox selection
  - 3d chart
- Improved interface
  - Timeseries
  - Grid



sdn-userdesk@seadatanet.org – www.seadatanet.org

8. Biological QC by VLIZ

Quality Control of Biological datasets in the VRE

Description

Development of a tool to process biological datasets and run some quality control checks on occurrence record level

Analyze the quality and completeness of biology data

Aim

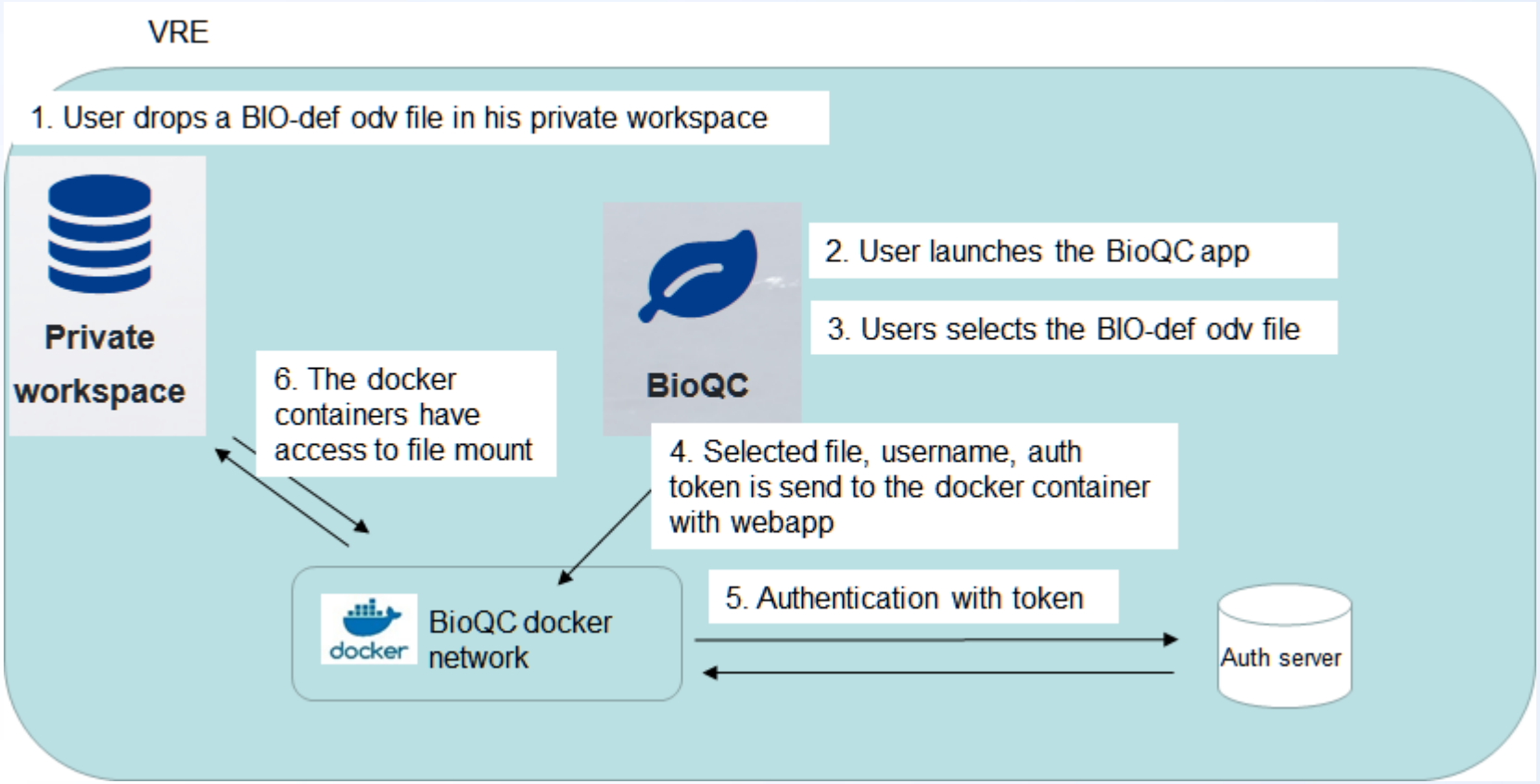
select data that fit for certain analysis

identify possible gaps and errors in datasets

3 types of procedures

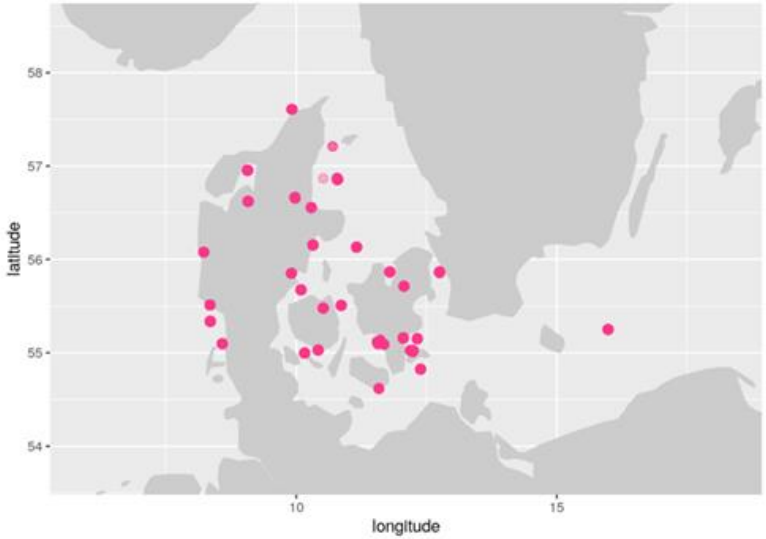
completeness and validity
geographical quality
outlier analysis

Integration into the VRE environment



Data quality report

Map



Issues

General errors and warnings

More than **25** errors and warnings related to this field were found, the first **25** are printed below and shown on a map whenever possible.

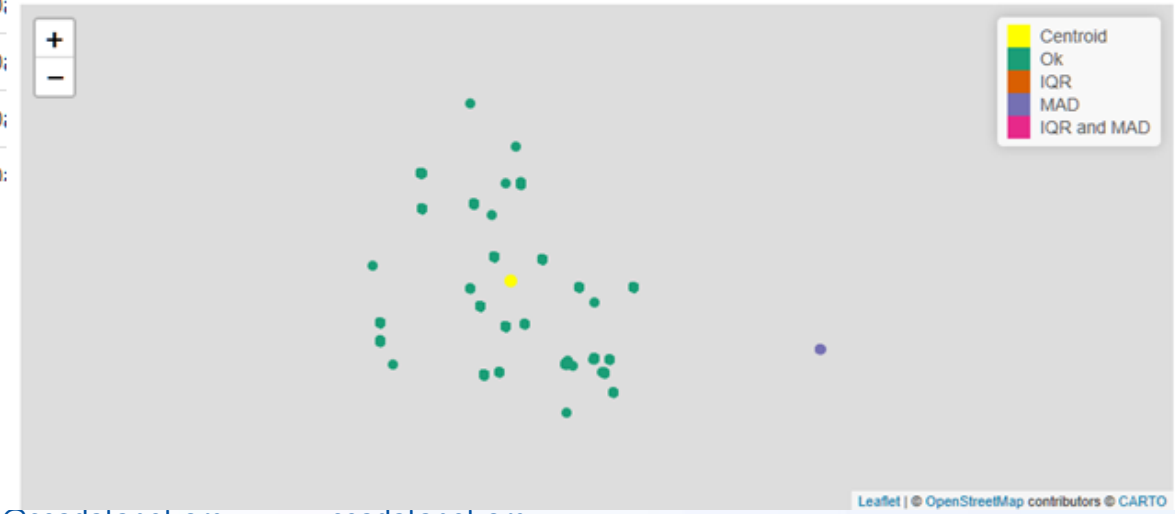
field	row	level	message
NA	18774	warning	Coordinates are located on land
NA	18775	warning	Coordinates are located on land
NA	18776	warning	Coordinates are located on land
NA	18777	warning	Coordinates are located on land

Outliers

Outliers Dataset

More than **25** errors and warnings related to this field were found, the first **25** are printed below and shown on a map whenever possible.

field	row	level	message
Outliers Dataset	3041	warning	spatial [346251.8] is not within MAD limits [-84769.78, 293838.1]
Outliers Dataset	3042	warning	spatial [346251.8] is not within MAD limits [-84769.78, 293838.1]
Outliers Dataset	3043	warning	spatial [346251.8] is not within MAD limits [-84769.78, 293838.1]
Outliers Di			
Outliers Di			
Outliers Di			
Outliers Di			



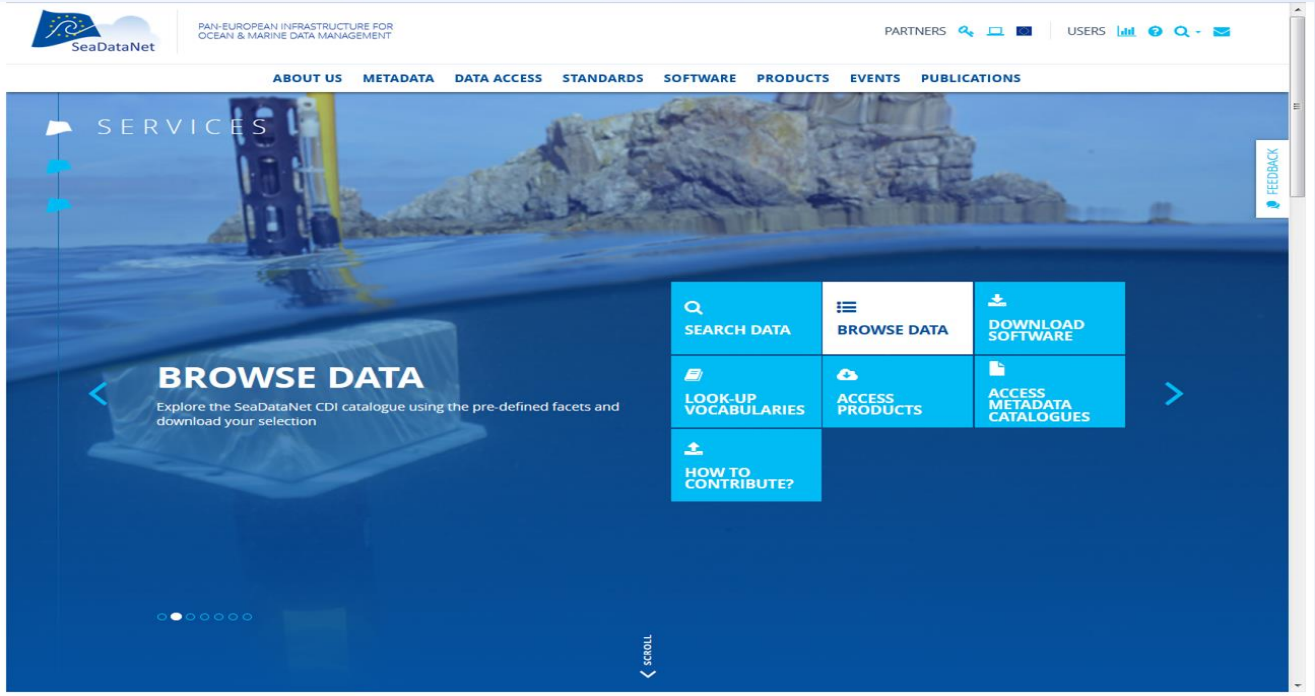
9. Closing remarks and time for questions

Perspective of the VRE

- Now launched as beta-version under SDN website.
- Even demonstrated in test a connection from CDI to the VRE.
- VRE is ready for user feedback, but be careful using it in larger groups/trainings.
- Hope for continuation in SDC2, but VRE as component could be developed further also in separate calls e.g. related to EOSC.

Questions for us??

SeaDataCloud



Keep following us via: VRE.seadatanet.org

sdn-userdesk@seadatanet.org – www.seadatanet.org