



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

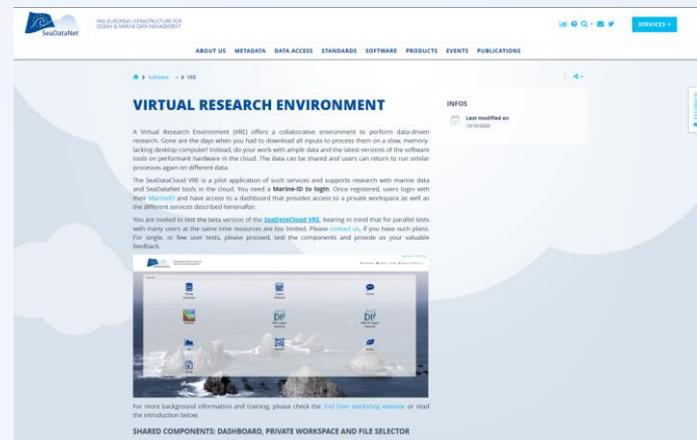
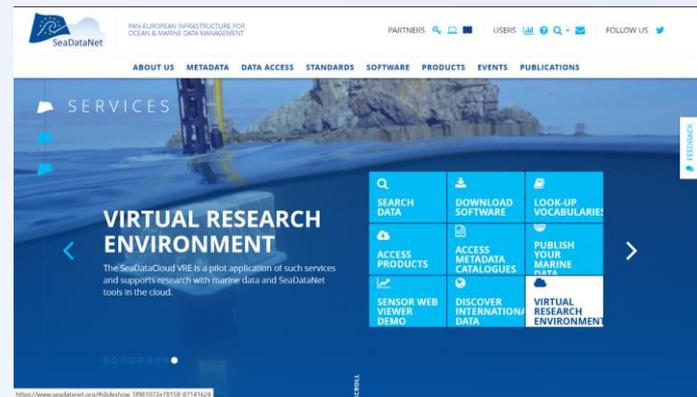
Today's overview of results:

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

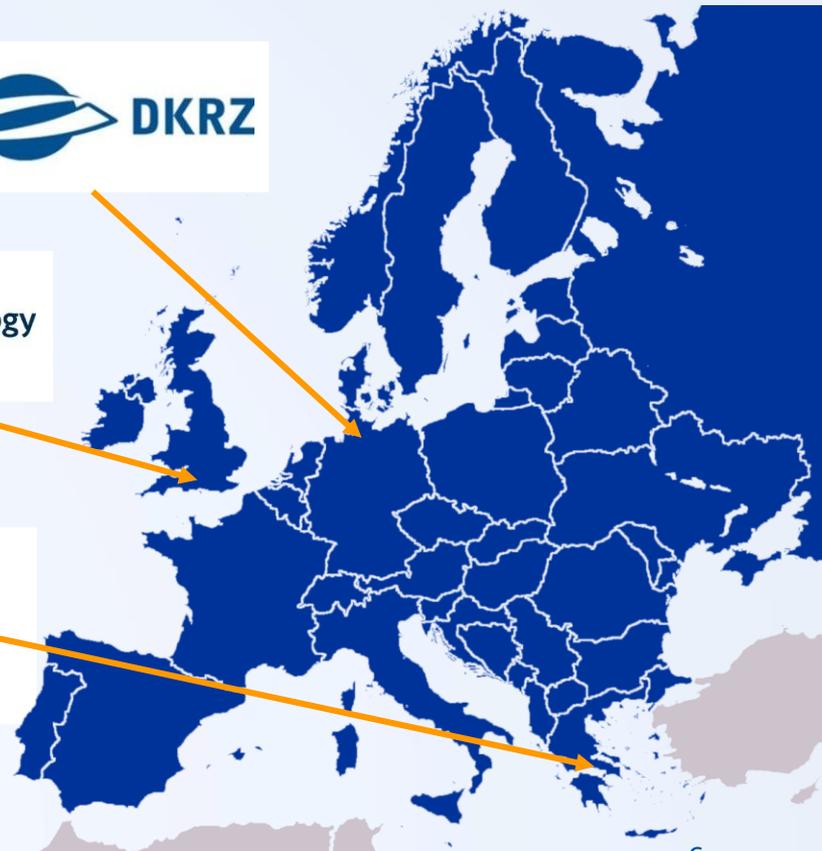
- During several (virtual) code sprints components have been implemented in the VRE environment and matured
- Solutions for earlier user feedback
- Next SDC and Emodnet Chemistry user consultation meeting (Sept 2020)
- Contribution to SDC webinar (Sept 2020)
- Two versions now: Development version and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

<https://www.seadatanet.org/Software/VRE>



2. Development and production environment

3 European Data Centers



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



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

NFS

NFS

NFS

direct

Sync

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!

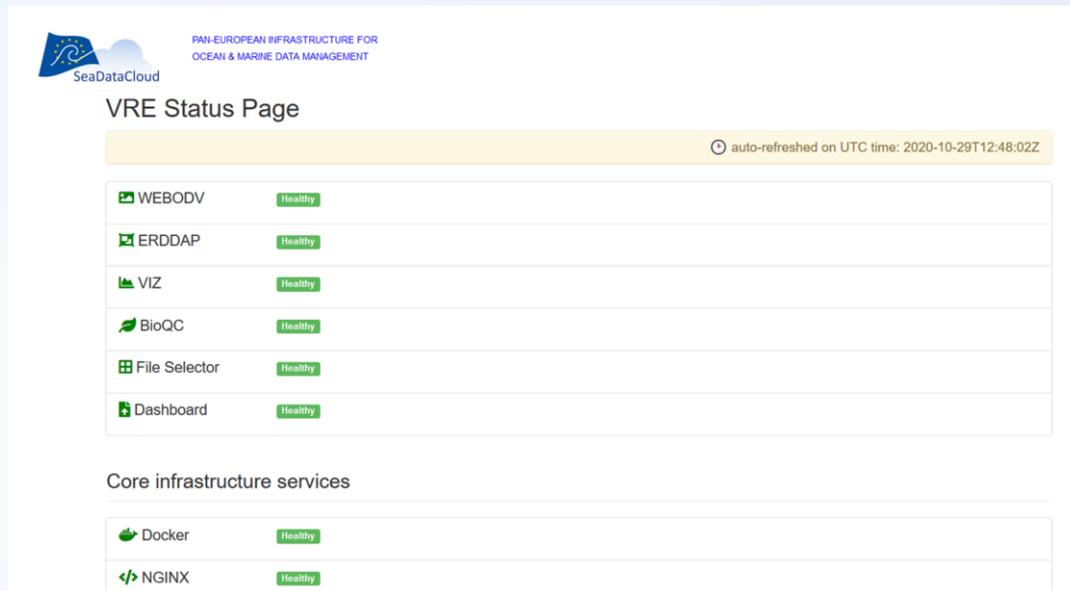
Note: On request, we can “protect” a user's container from deletion!

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/>



The screenshot shows the VRE Status Page from SeaDataCloud. At the top, it features the SeaDataCloud logo and the text "PAN-EUROPEAN INFRASTRUCTURE FOR OCEAN & MARINE DATA MANAGEMENT". The page title is "VRE Status Page". A yellow banner indicates the page is "auto-refreshed on UTC time: 2020-10-29T12:48:02Z". Below this, there are two sections: "VRE Services" and "Core infrastructure services". Each service is listed with a small icon, the service name, and a green "Healthy" status indicator.

Service	Status
WEBODV	Healthy
ERDDAP	Healthy
VIZ	Healthy
BioQC	Healthy
File Selector	Healthy
Dashboard	Healthy

Service	Status
Docker	Healthy
NGINX	Healthy

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



The screenshot shows the SeaDataCloud login interface. At the top right, there are links for 'Contact', 'Legal Notice', and 'Privacy Policy'. The main content area features the SeaDataCloud logo, the text 'VIRTUAL RESEARCH ENVIRONMENT', a blue button labeled 'LOGIN WITH MARINE ID', and a link for 'Not registered?' with a question mark icon, followed by the instruction 'Click on LOGIN and register your MARINE ID.' The background is a scenic view of a coastal town and mountains.

Contact Legal Notice Privacy Policy



SeaDataCloud

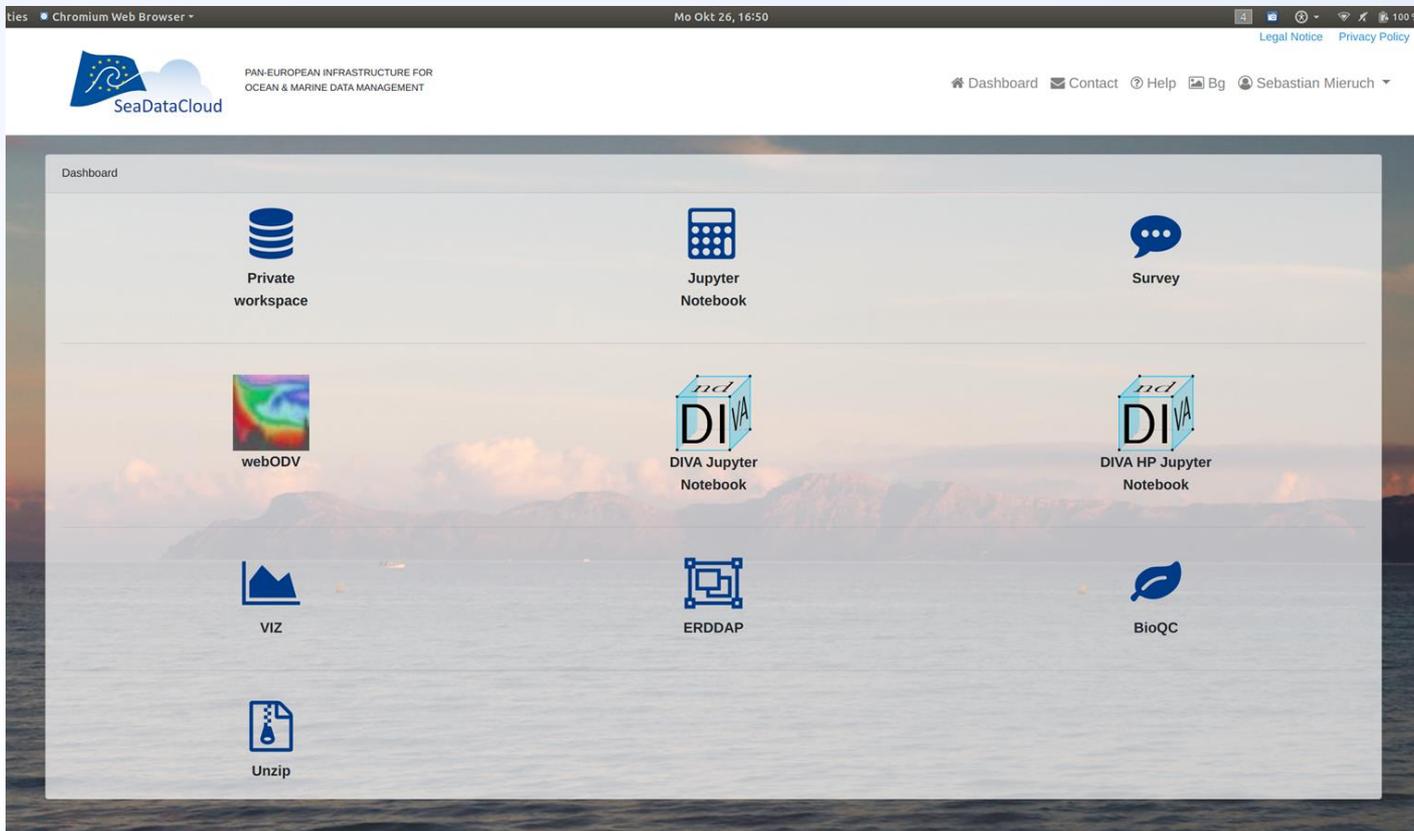
VIRTUAL RESEARCH ENVIRONMENT

LOGIN WITH MARINE ID

Not registered? 

Click on LOGIN and register your MARINE ID.

Photo by Charles Troupin



Chromium Web Browser

Mo Okt 26, 16:50

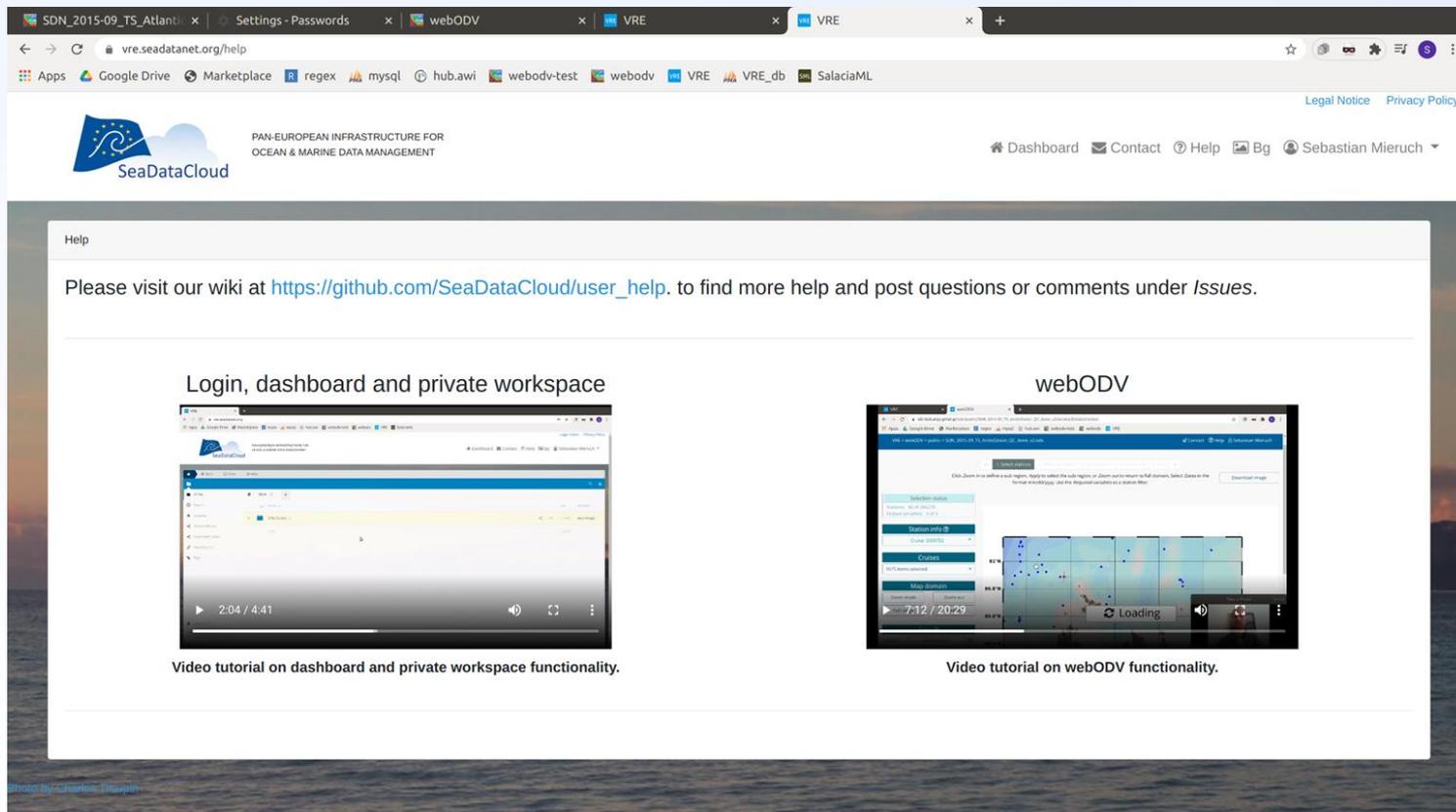
Legal Notice Privacy Policy

SeaDataCloud PAN-EUROPEAN INFRASTRUCTURE FOR OCEAN & MARINE DATA MANAGEMENT

Dashboard Contact Help Bg Sebastian Mieruch

Dashboard

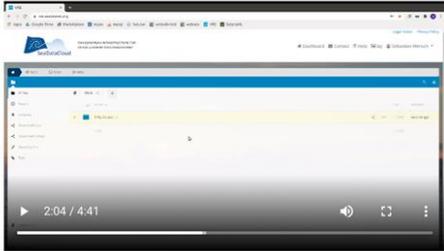
- Private workspace
- Jupyter Notebook
- Survey
- webODV
- DIVA Jupyter Notebook
- DIVA HP Jupyter Notebook
- VIZ
- ERDDAP
- BioQC
- Unzip



Help

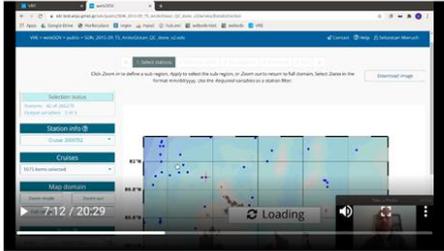
Please visit our wiki at https://github.com/SeaDataCloud/user_help. to find more help and post questions or comments under *Issues*.

Login, dashboard and private workspace



Video tutorial on dashboard and private workspace functionality.

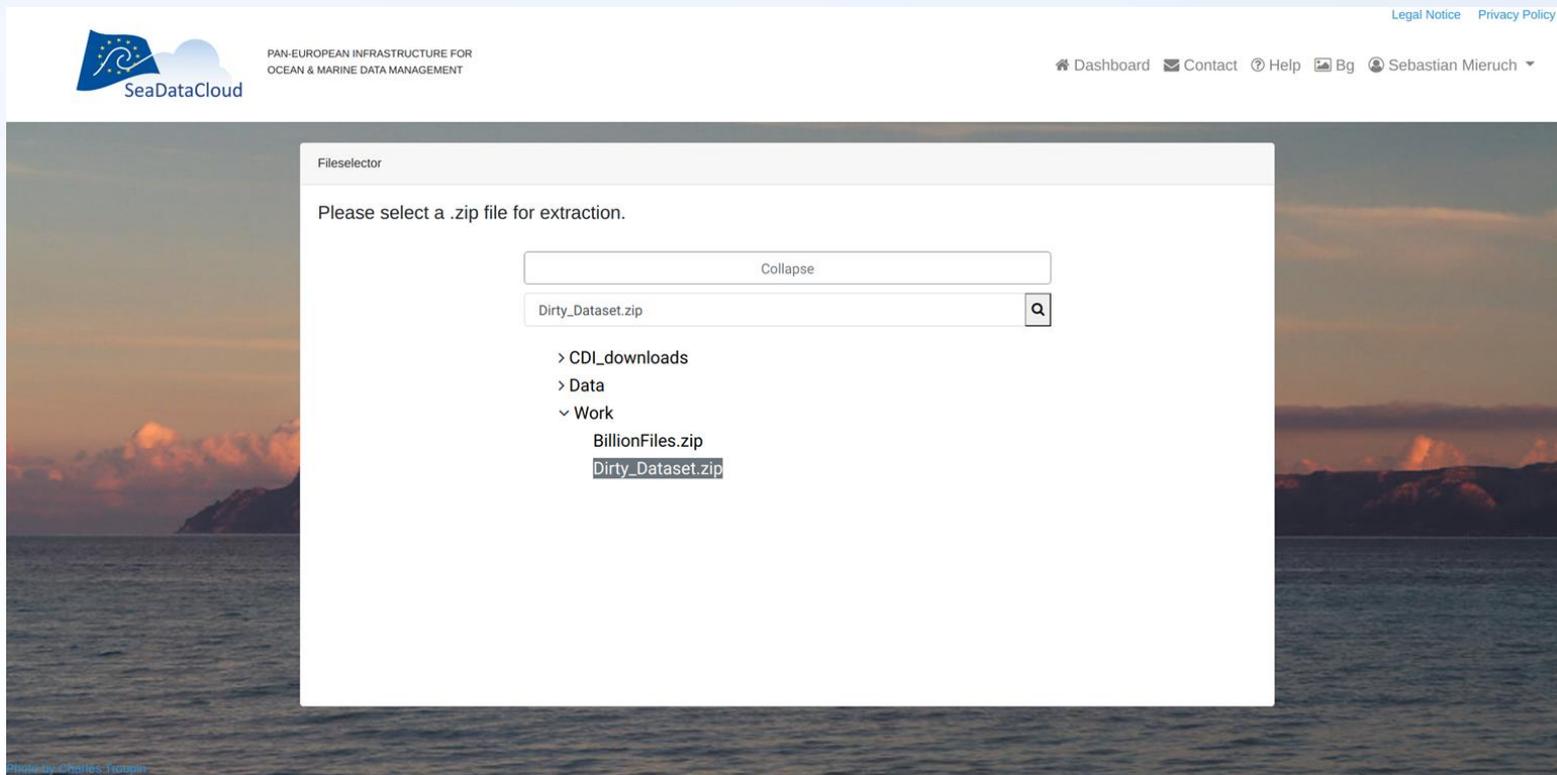
webODV



Video tutorial on webODV functionality.

Photo by Creative Commons

Unzip App

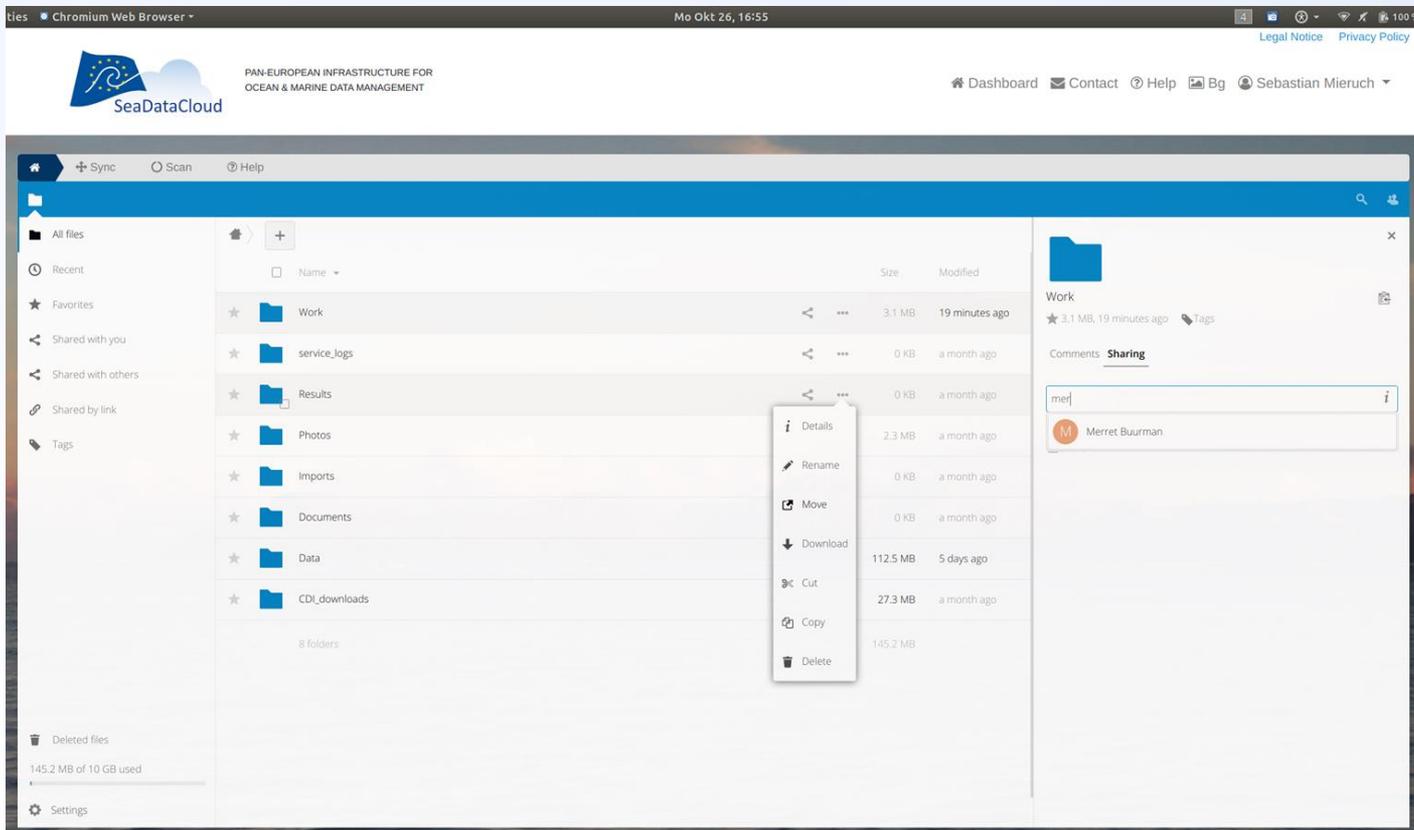


The screenshot shows the SeaDataCloud website header with the logo and navigation links: Legal Notice, Privacy Policy, Dashboard, Contact, Help, Bg, and Sebastian Mieruch. The main content area features a 'Fileselector' dialog box with the instruction 'Please select a .zip file for extraction.' Below this, there is a search bar containing 'Dirty_Dataset.zip' and a search icon. A file tree is displayed with the following structure:

- > CDI_downloads
- > Data
- < Work
 - BillionFiles.zip
 - Dirty_Dataset.zip**

The background of the page is a scenic image of a sunset over the ocean.

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

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		

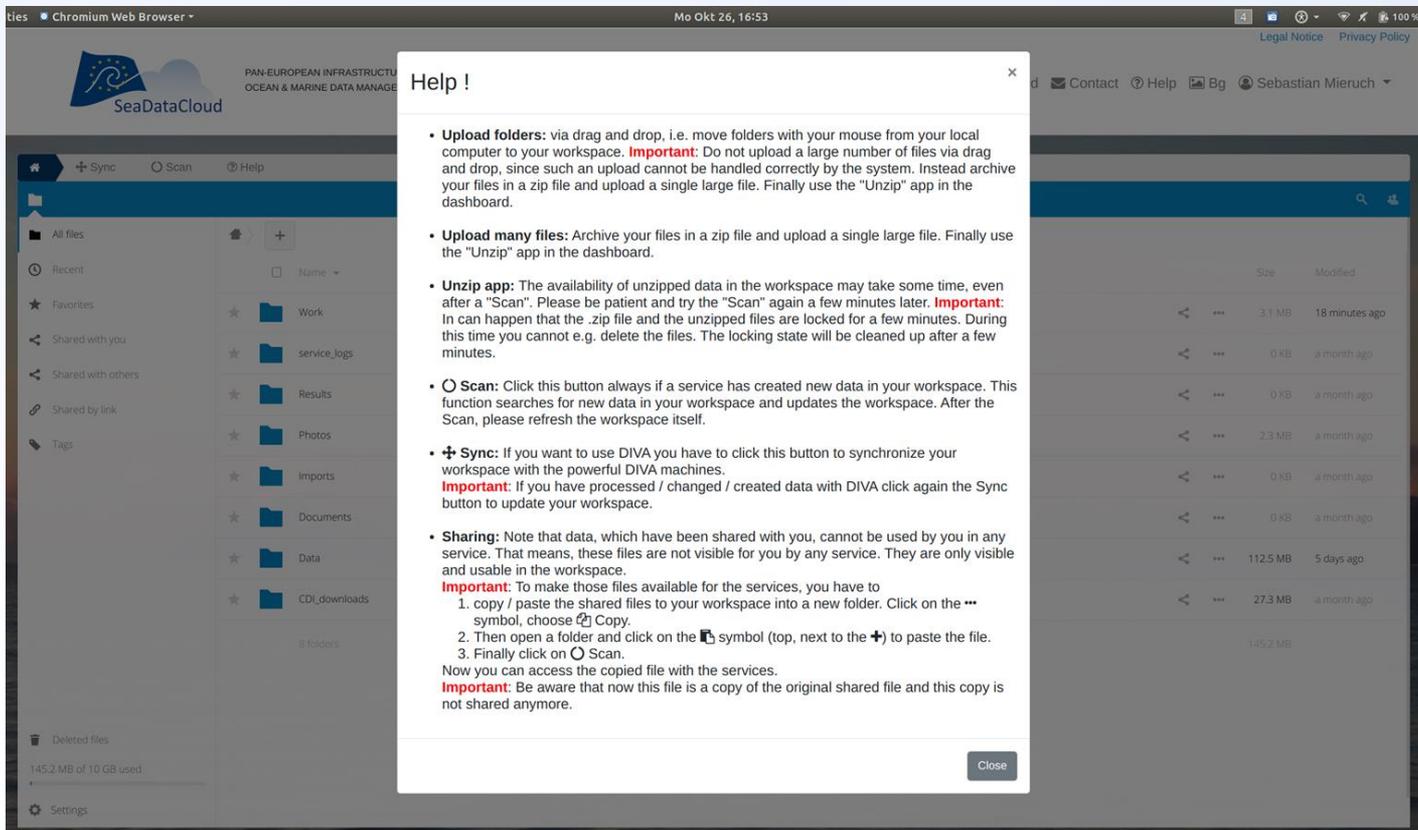
Deleted files | 145.2 MB of 10 GB Used | Settings

Work | 3.1 MB, 19 minutes ago | Tags

Comments | Sharing

mer | Merret Burman

Workspace - Help



The screenshot shows the SeaDataCloud workspace interface in a Chromium Web Browser. A help dialog box titled "Help !" is centered on the screen, providing instructions for various actions. The background interface includes a navigation sidebar with folders like "Work", "service_logs", "Results", "Photos", "Imports", "Documents", "Data", and "CDI_downloads". A table of files is visible on the right side of the workspace.

Help !

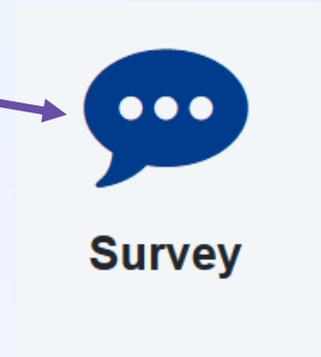
- **Upload folders:** via drag and drop, i.e. move folders with your mouse from your local computer to your workspace. **Important:** Do not upload a large number of files via drag and drop, since such an upload cannot be handled correctly by the system. Instead archive your files in a zip file and upload a single large file. Finally use the "Unzip" app in the dashboard.
- **Upload many files:** Archive your files in a zip file and upload a single large file. Finally use the "Unzip" app in the dashboard.
- **Unzip app:** The availability of unzipped data in the workspace may take some time, even after a "Scan". Please be patient and try the "Scan" again a few minutes later. **Important:** It can happen that the .zip file and the unzipped files are locked for a few minutes. During this time you cannot e.g. delete the files. The locking state will be cleaned up after a few minutes.
- **Scan:** Click this button always if a service has created new data in your workspace. This function searches for new data in your workspace and updates the workspace. After the Scan, please refresh the workspace itself.
- **Sync:** If you want to use DIVA you have to click this button to synchronize your workspace with the powerful DIVA machines. **Important:** If you have processed / changed / created data with DIVA click again the Sync button to update your workspace.
- **Sharing:** Note that data, which have been shared with you, cannot be used by you in any service. That means, these files are not visible for you by any service. They are only visible and usable in the workspace. **Important:** To make those files available for the services, you have to
 1. copy / paste the shared files to your workspace into a new folder. Click on the ... symbol, choose  Copy.
 2. Then open a folder and click on the  symbol (top, next to the +) to paste the file.
 3. Finally click on **Scan**.
 Now you can access the copied file with the services. **Important:** Be aware that now this file is a copy of the original shared file and this copy is not shared anymore.

Close

VRE feedback

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

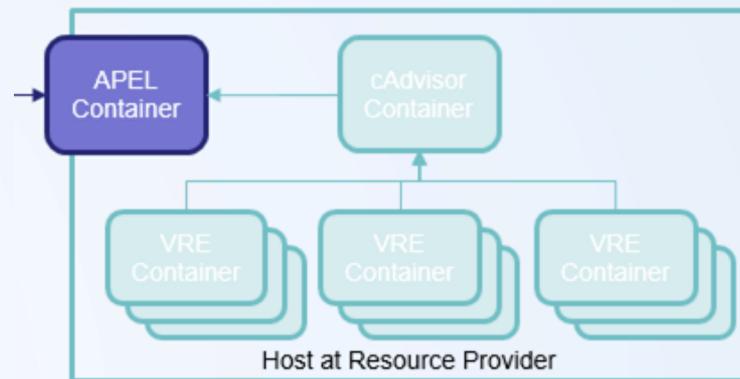
The image shows a screenshot of the SeaDataCloud dashboard. At the top, the navigation bar includes 'Dashboard', 'Contact', 'Help', 'Bg', and 'Merret Buurman'. The 'Contact' link is circled in red. Below the navigation bar, the dashboard features several service tiles: 'Private workspace', 'Jupyter Notebook', 'webODV', 'DIVA Jupyter Notebook', and 'DIVA HP Jupyter Notebook'. A 'Survey' icon (a blue speech bubble) is circled in purple. A red arrow points from the 'Contact' link to a 'Contact' modal window. Another red arrow points from the 'Survey' icon to a larger 'Survey' icon on the right. The 'Contact' modal window contains the following fields: Name (Merret Buurman), E-Mail Address (buurman@dkrz.de), and a Message field. There are 'Send' and 'Close' buttons at the bottom of the modal.



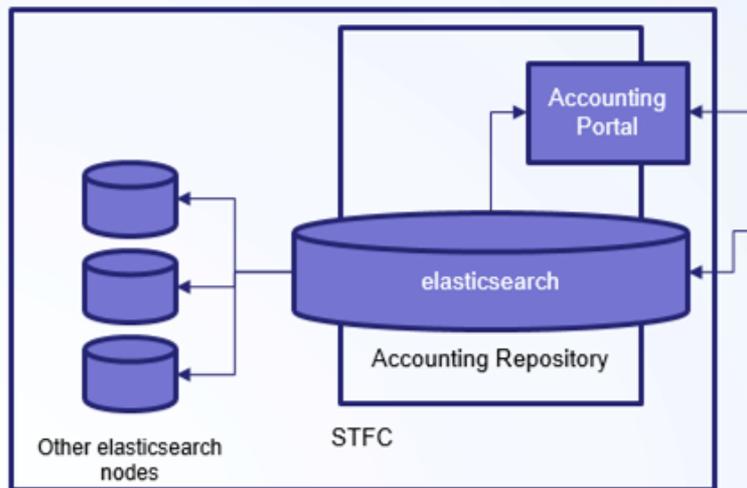
3. Accounting (STFC)

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



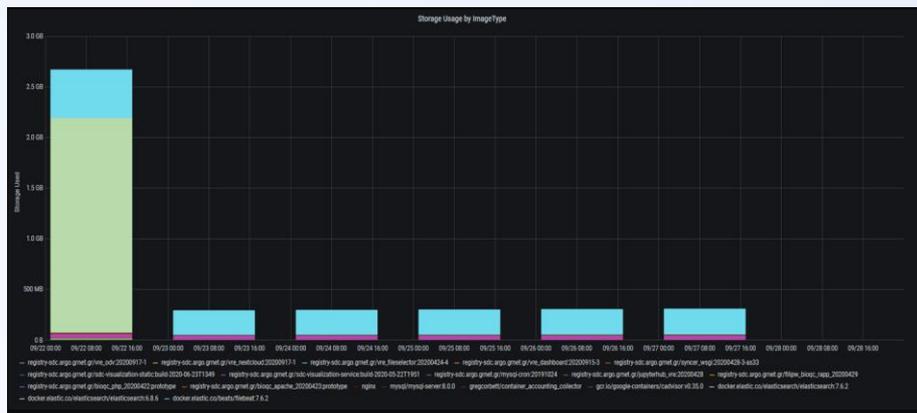
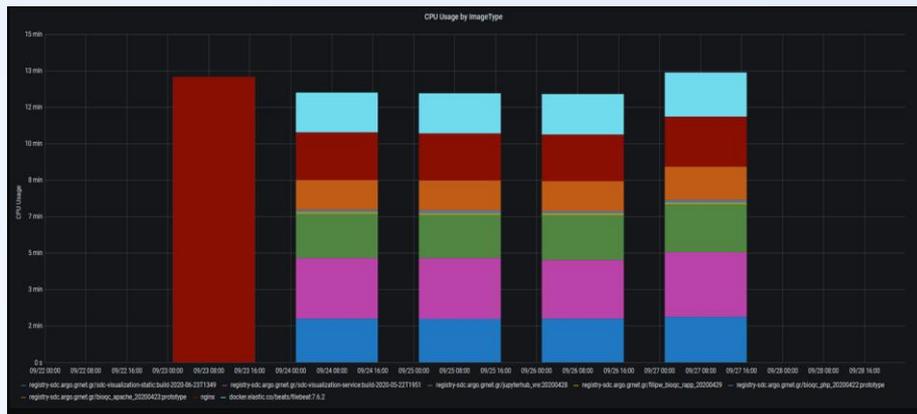
Accounting System



- 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

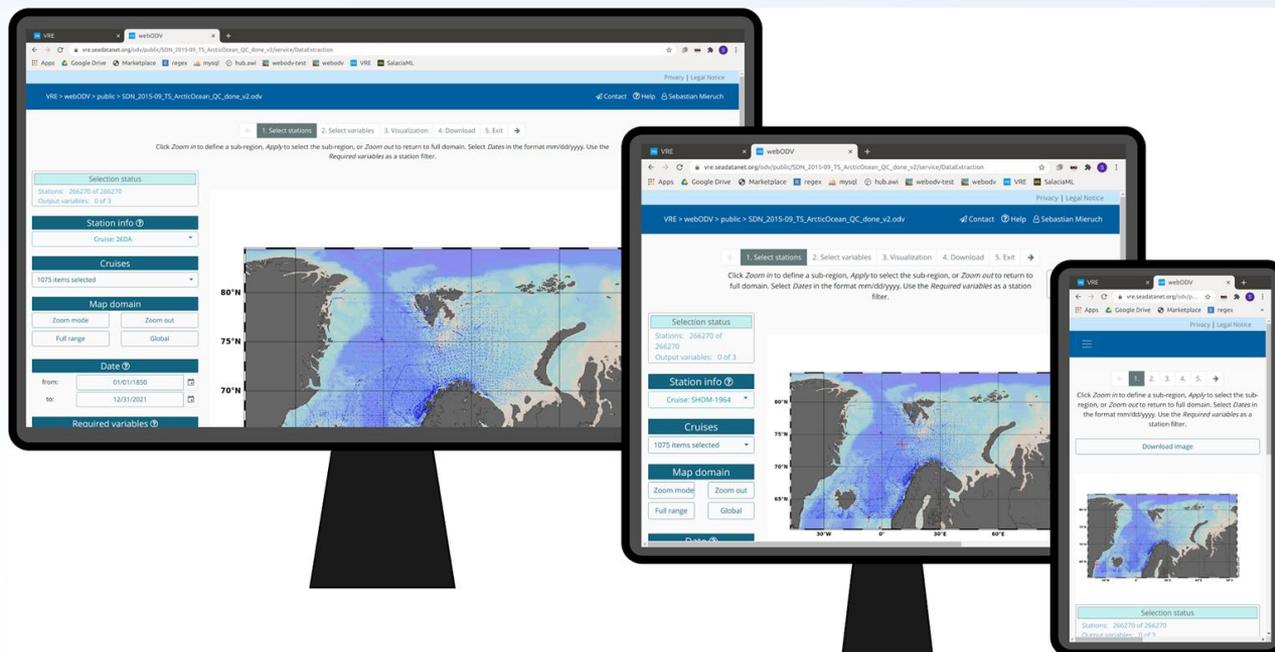
- Portal graphs usage by container image over time.
- Filters allow the usage of a subset of containers to be graphed.
- Usage can also be filtered by resource provider, i.e. DKRZ or GRNET.



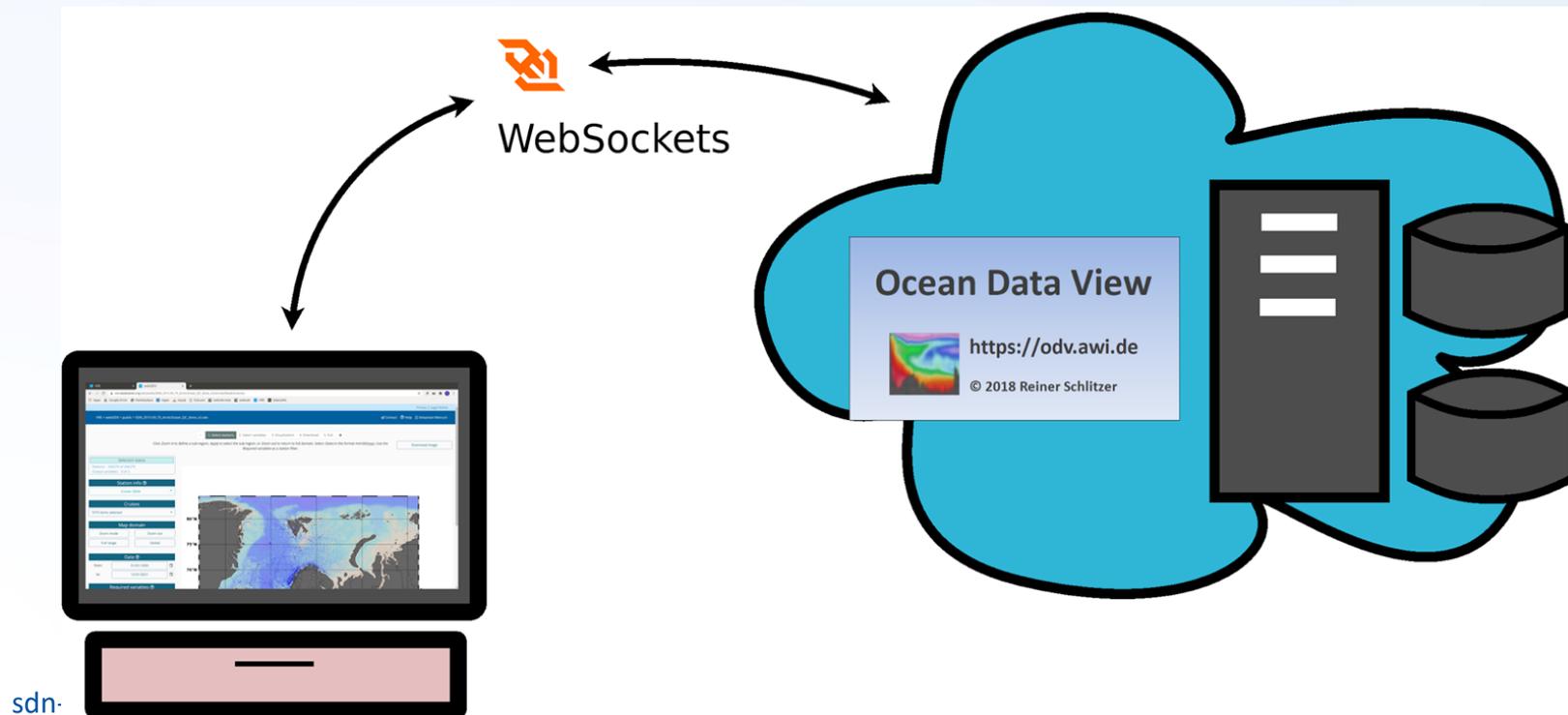
4. webODV

webODV in the VRE

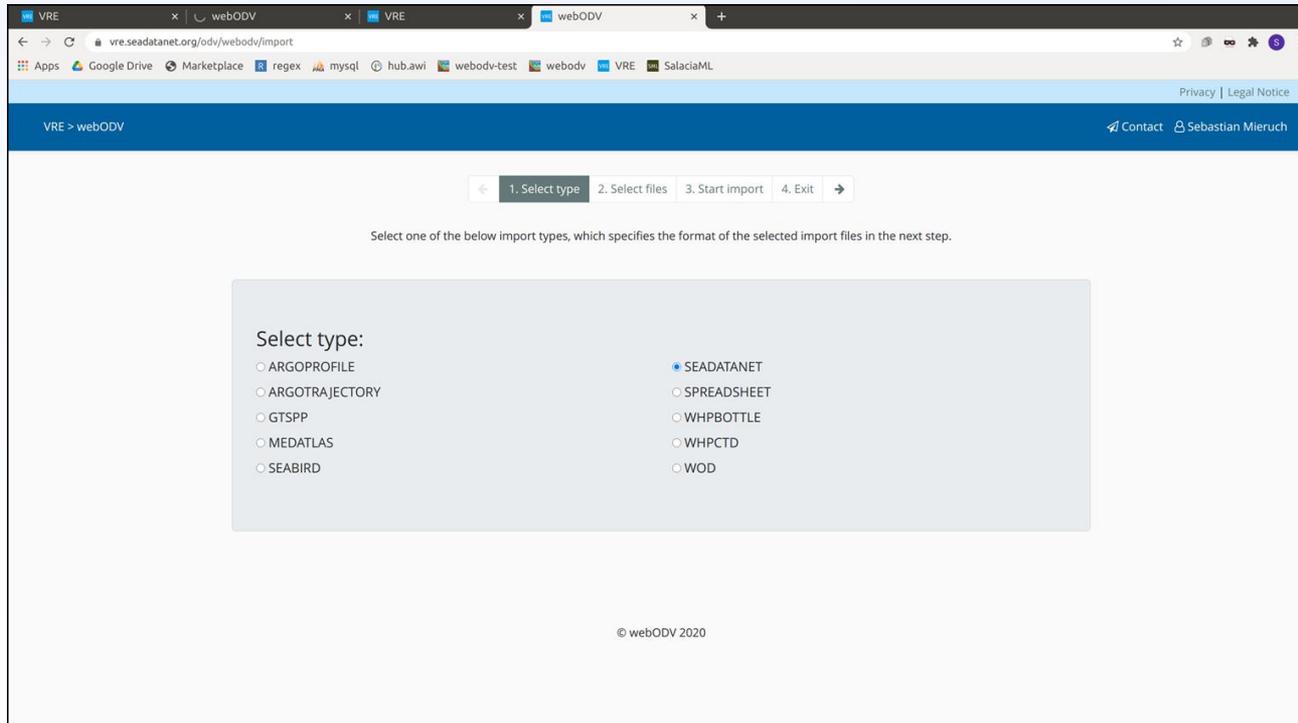
Sebastian Mieruch and Reiner Schlitzer



webODV - The concept

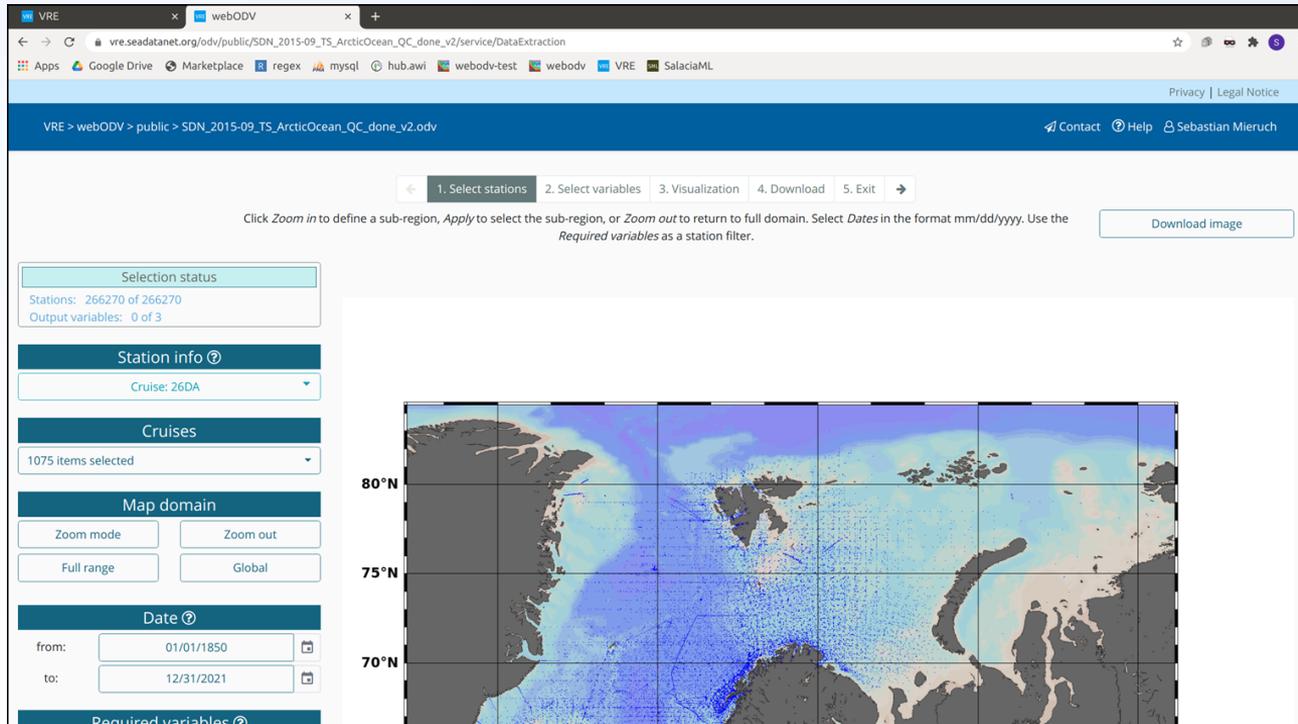


webODV - Import Service



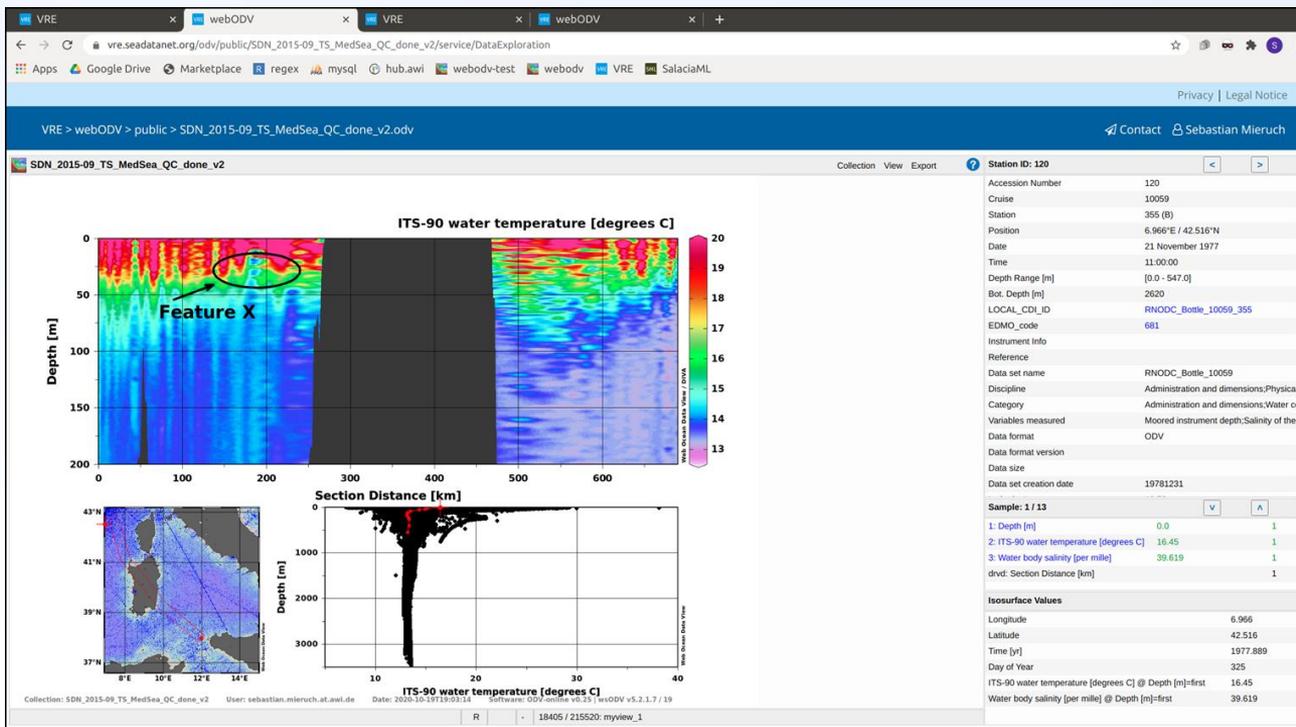
The screenshot shows a web browser window with the URL `vre.seadatanet.org/odv/webodv/import`. The browser's address bar and tabs are visible at the top. Below the browser window, a navigation bar contains the text "VRE > webODV" on the left and "Contact" and "Sebastian Mieruch" on the right. The main content area features a progress indicator with four steps: "1. Select type", "2. Select files", "3. Start import", and "4. Exit". The first step, "1. Select type", is currently active. Below the progress indicator, a text prompt reads: "Select one of the below import types, which specifies the format of the selected import files in the next step." A large light gray box contains the heading "Select type:" followed by two columns of radio button options. The first column lists: ARGOPROFILE, ARGOTRAJECTORY, GTSPP, MEDATLAS, and SEABIRD. The second column lists: SEADATANET (which is selected), SPREADSHEET, WHPBOTTLE, WHPCTD, and WOD. At the bottom center of the page, the copyright notice "© webODV 2020" is displayed.

webODV - Extractor Service



The screenshot displays the webODV interface in a browser window. The address bar shows the URL: `vre.seadatanet.org/odv/public/SDN_2015-09_TS_ArcticOcean_QC_done_v2/service/DataExtraction`. The breadcrumb navigation is: `VRE > webODV > public > SDN_2015-09_TS_ArcticOcean_QC_done_v2.odv`. The interface includes a progress bar with steps: 1. Select stations, 2. Select variables, 3. Visualization, 4. Download, and 5. Exit. A text instruction reads: "Click Zoom in to define a sub-region, Apply to select the sub-region, or Zoom out to return to full domain. Select Dates in the format mm/dd/yyyy. Use the Required variables as a station filter." A "Download image" button is present. On the left, there are several control panels: "Selection status" (Stations: 266270 of 266270, Output variables: 0 of 3), "Station info" (Cruise: 26DA), "Cruises" (1075 items selected), "Map domain" (Zoom mode, Zoom out, Full range, Global), "Date" (from: 01/01/1850, to: 12/31/2021), and "Required variables". The main area features a map of the Arctic region with a grid overlay, showing data points in blue and purple.

webODV - Explore Service



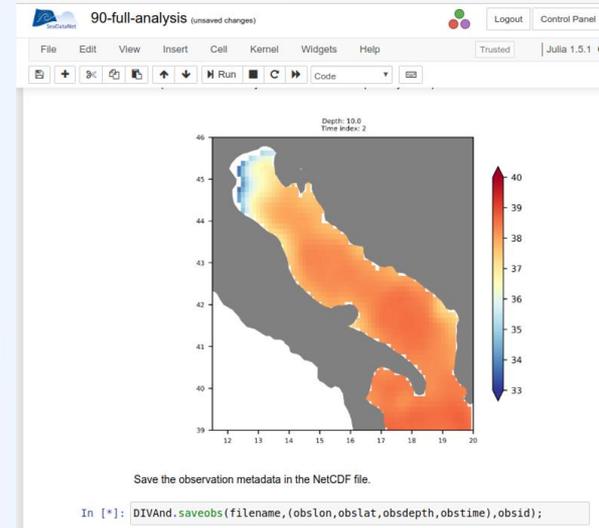
5. DIVAnd

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

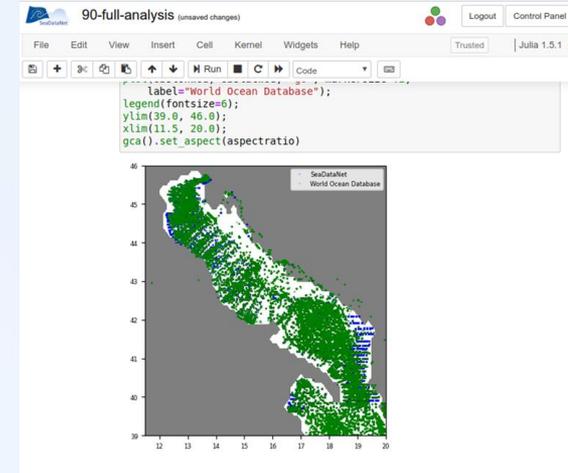


Most feature of the DIVA tool have been ported to DIVAnd and are available in the VRE, such as:

- **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



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

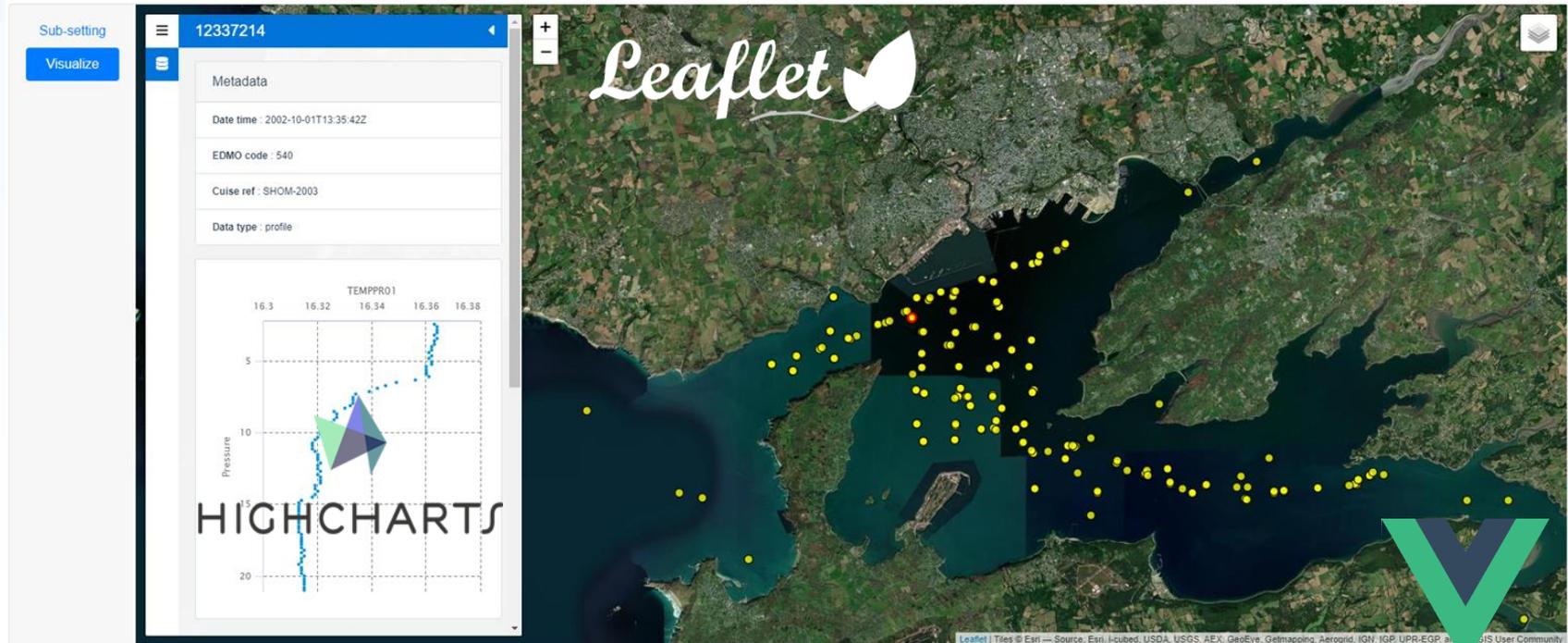
Sub-setting
Visualize

Please check the variables you want to subset. Select an operator/value pair to add constraints, an empty input field correspond to NULL value.

<input type="checkbox"/>	SDN_EDMO_CODE		*						
<input type="checkbox"/>	SDN_CRUISE		*						
<input type="checkbox"/>	SDN_STATION		*						
<input checked="" type="checkbox"/>	SDN_LOCAL_CDI_ID		*						
<input type="checkbox"/>	SDN_XLINK		*						
<input type="checkbox"/>	SDN_BOT_DEPTH		*						
<input checked="" type="checkbox"/>	longitude		*						
<input checked="" type="checkbox"/>	latitude		*						
<input type="checkbox"/>	POSITION_SEADATANET_QC		*						
<input type="checkbox"/>	crs		*						
<input checked="" type="checkbox"/>	time	>	*	22/01/2010	<input type="checkbox"/>	<	*	22/10/2020	<input type="checkbox"/>
<input type="checkbox"/>	TIME_SEADATANET_QC		*						
<input checked="" type="checkbox"/>	depth		*						
<input type="checkbox"/>	DEPTH_SEADATANET_QC		*						
<input checked="" type="checkbox"/>	TEMPPR01		*						
<input checked="" type="checkbox"/>	TEMPPR01_SEADATANET_QC	!=	*	4					
<input checked="" type="checkbox"/>	PSLTZZ01		*						



Visualization with modern libraries



Waiting for P02 feature

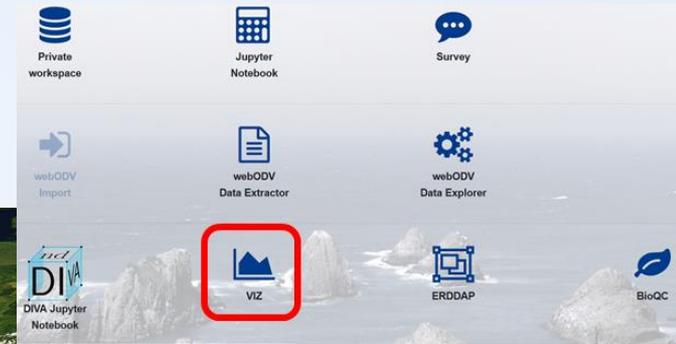
- Waiting for a new version on Erddap including P02 feature to fully finalize the service.
- With this feature we will be able to work with homogeneous datasets thanks to P02 vocabulary.
- Version 2.10 of Erddap should be available soon.
- Question is : Does this version will include P02 feature ?

7. Advanced visualisations by Deltares

Giorgio Santinelli, Cindy van de Vries, Fedor Baart

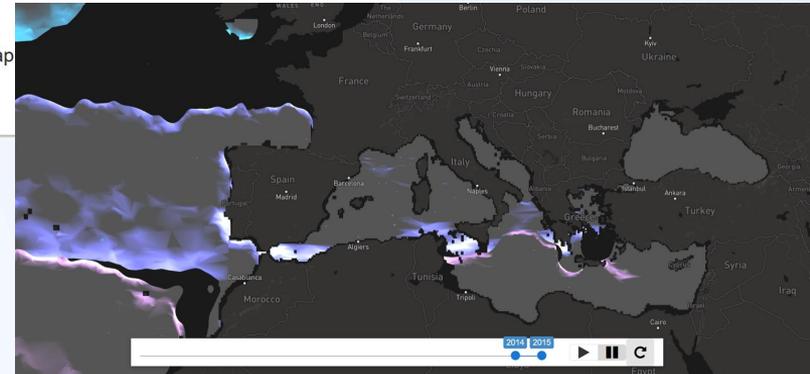
Visualisation services

- 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>



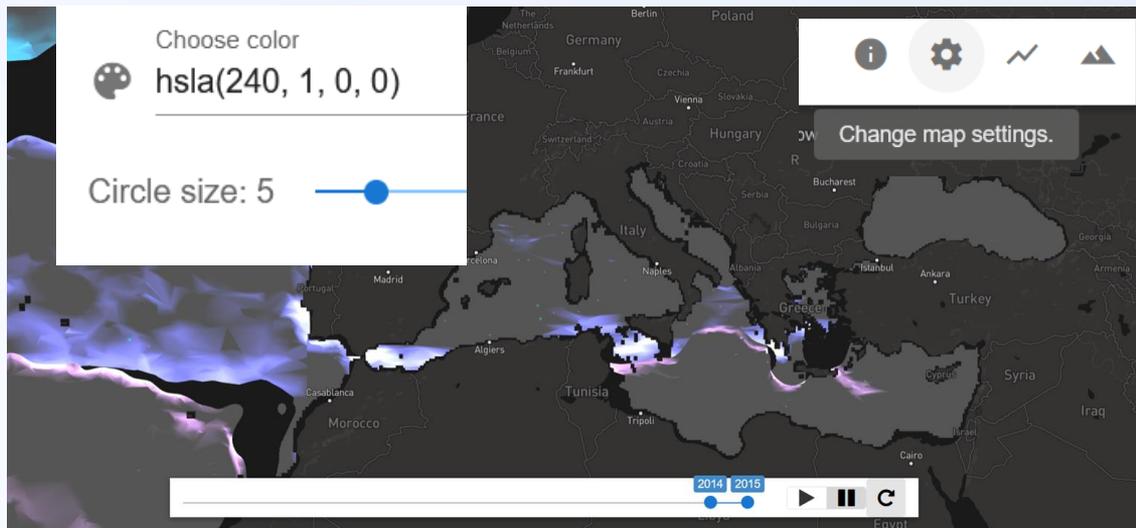
Visualization
Create visualizations on the map

GO



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



- Observations

- Select axes
- Single profile
- Metadata
- Export
- Trajectories
- Bbox selection
- 3d chart

- Improved interface

- Timeseries
- Grid



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

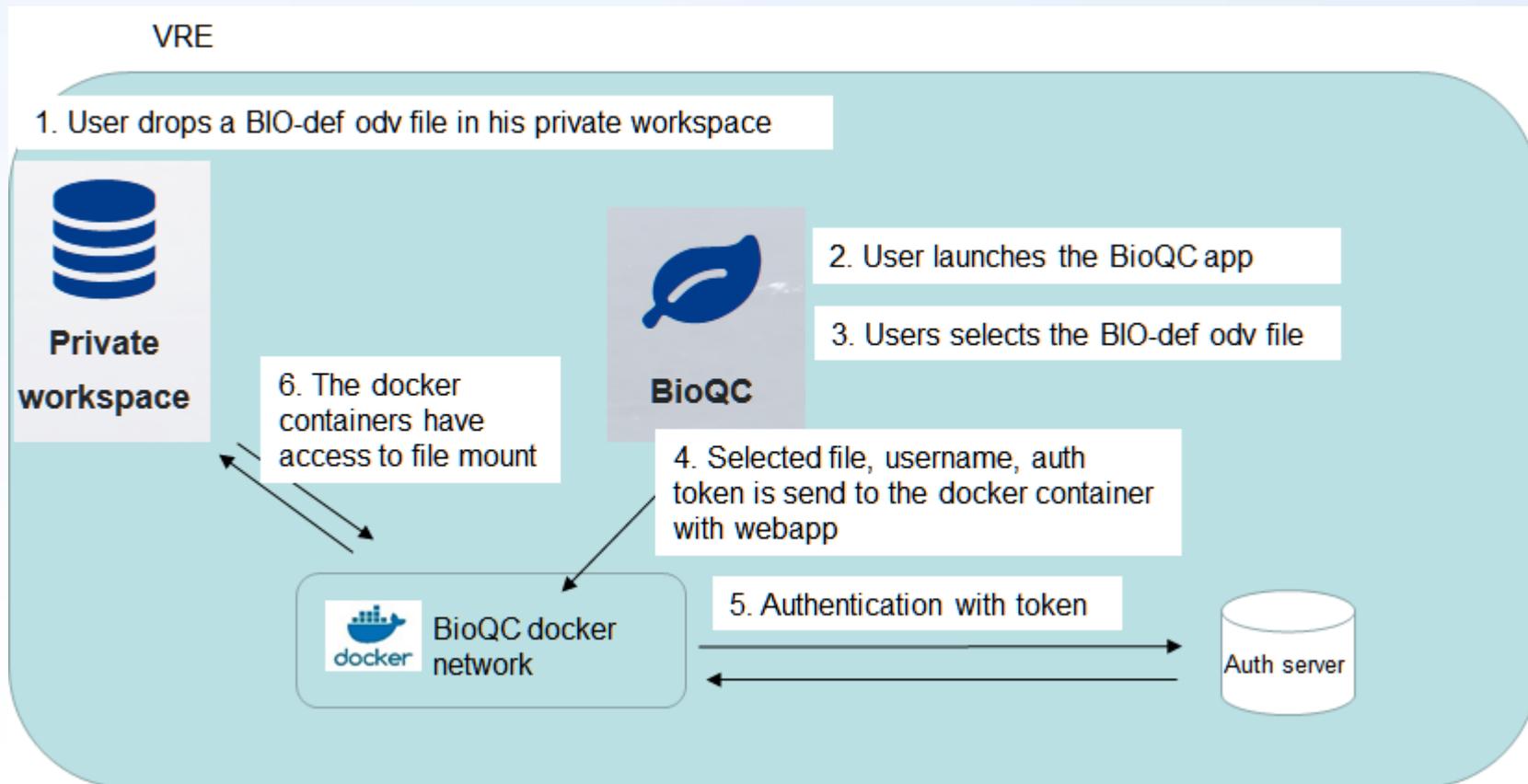
3 types of procedures

completeness and validity

geographical quality

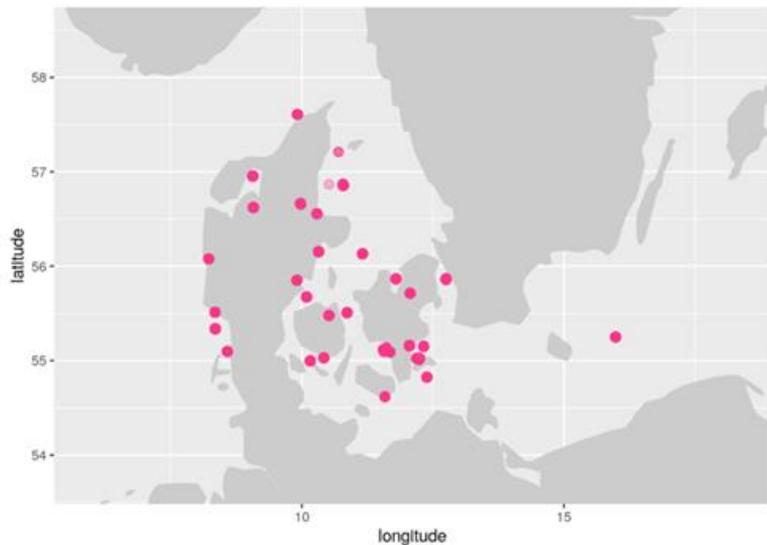
outlier analysis

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.

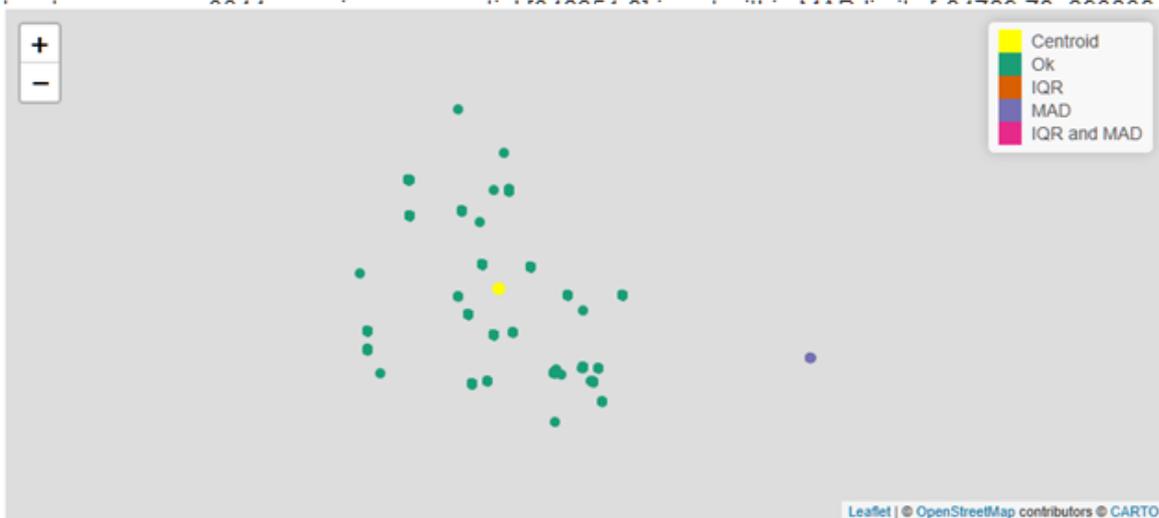
field	row	level	message
NA	18774	warning	Coordinates are located on land
NA	18775	warning	Coordinates are located on land
NA	18776	warning	Coordinates are located on land
NA	18777	warning	Coordinates are located on land

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 Dataset	3044	warning	spatial [346251.8] is not within MAD limits [-84769.78, 293838.1]
Outliers Dataset	3045	warning	spatial [346251.8] is not within MAD limits [-84769.78, 293838.1]
Outliers Dataset	3046	warning	spatial [346251.8] is not within MAD limits [-84769.78, 293838.1]
Outliers Dataset	3047	warning	spatial [346251.8] is not within MAD limits [-84769.78, 293838.1]
Outliers Dataset	3048	warning	spatial [346251.8] is not within MAD limits [-84769.78, 293838.1]
Outliers Dataset	3049	warning	spatial [346251.8] is not within MAD limits [-84769.78, 293838.1]
Outliers Dataset	3050	warning	spatial [346251.8] is not within MAD limits [-84769.78, 293838.1]
Outliers Dataset	3051	warning	spatial [346251.8] is not within MAD limits [-84769.78, 293838.1]
Outliers Dataset	3052	warning	spatial [346251.8] is not within MAD limits [-84769.78, 293838.1]
Outliers Dataset	3053	warning	spatial [346251.8] is not within MAD limits [-84769.78, 293838.1]
Outliers Dataset	3054	warning	spatial [346251.8] is not within MAD limits [-84769.78, 293838.1]
Outliers Dataset	3055	warning	spatial [346251.8] is not within MAD limits [-84769.78, 293838.1]
Outliers Dataset	3056	warning	spatial [346251.8] is not within MAD limits [-84769.78, 293838.1]
Outliers Dataset	3057	warning	spatial [346251.8] is not within MAD limits [-84769.78, 293838.1]
Outliers Dataset	3058	warning	spatial [346251.8] is not within MAD limits [-84769.78, 293838.1]
Outliers Dataset	3059	warning	spatial [346251.8] is not within MAD limits [-84769.78, 293838.1]
Outliers Dataset	3060	warning	spatial [346251.8] is not within MAD limits [-84769.78, 293838.1]
Outliers Dataset	3061	warning	spatial [346251.8] is not within MAD limits [-84769.78, 293838.1]
Outliers Dataset	3062	warning	spatial [346251.8] is not within MAD limits [-84769.78, 293838.1]
Outliers Dataset	3063	warning	spatial [346251.8] is not within MAD limits [-84769.78, 293838.1]
Outliers Dataset	3064	warning	spatial [346251.8] is not within MAD limits [-84769.78, 293838.1]
Outliers Dataset	3065	warning	spatial [346251.8] is not within MAD limits [-84769.78, 293838.1]

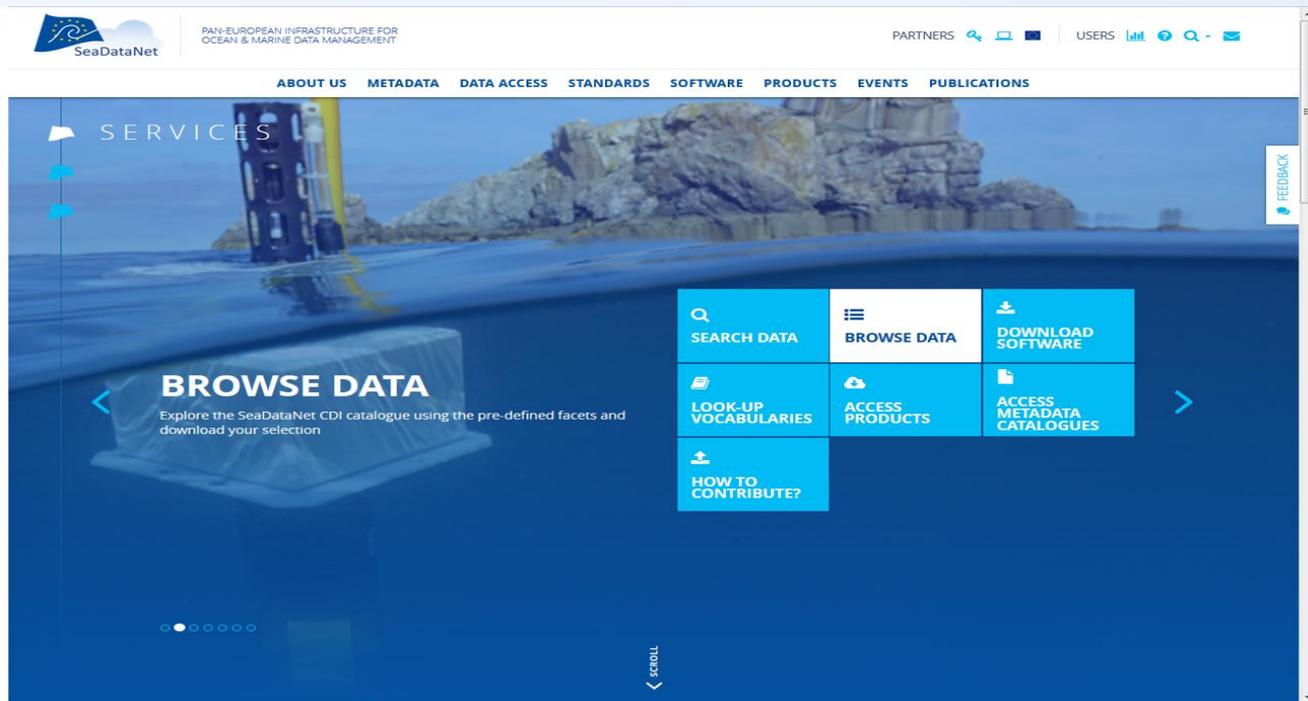


9. Closing remarks and time for questions

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??



The screenshot shows the SeaDataNet website homepage. At the top left is the SeaDataNet logo and the text "PAN-EUROPEAN INFRASTRUCTURE FOR OCEAN & MARINE DATA MANAGEMENT". To the right are navigation links for "PARTNERS", "USERS", and social media icons. A horizontal menu below the header lists: "ABOUT US", "METADATA", "DATA ACCESS", "STANDARDS", "SOFTWARE", "PRODUCTS", "EVENTS", and "PUBLICATIONS". The main content area features a large blue background image of a ship's deck with a yellow crane. The word "SERVICES" is written in white at the top left of this area. A "BROWSE DATA" section is highlighted with a white box and contains the text: "Explore the SeaDataNet CDI catalogue using the pre-defined facets and download your selection". To the right of this section is a grid of service buttons: "SEARCH DATA", "BROWSE DATA", "DOWNLOAD SOFTWARE", "LOOK-UP VOCABULARIES", "ACCESS PRODUCTS", "ACCESS METADATA CATALOGUES", and "HOW TO CONTRIBUTE?". A "FEEDBACK" button is on the right edge, and a "SCROLL" indicator is at the bottom center.

Keep following us via: VRE.seadatanet.org

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