SeaDataCloud

Studying the seas from the cloud

Progress of the SeaDataCloud Virtual Research Environment

Merret Buurman, Peter Thijsse and the complete WP10 team
@IMDIS 2018

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

SeaDataCloud

Outline

1. The SeaDataCloud VRE
2. Progress of individual components
3. Architecture and EUDAT backend
4. Outlook

1. The SeaDataCloud VRE

SeaDataCloud

SeaDataClouds VRE main requirements

- Facilitate collaborative and individual research
- Combine data with subsets from other data resources
- Have a high capacity and performance for big data processing and web visualisation services
- Differentiated users, different access to data and data products
- Configure virtual work spaces for individuals or groups, including setting up of dedicated pools of data
- Sharing options, if user decides
- Based and hosted on EUDAT's infrastructure and its B2-… service platforms

Support 5 versatile use cases:

1. **SeaDataNet Temperature and Salinity** water column analysis
2. **EMODNET - Chemistry**, same for bio-geo-chemistry
3. **SeaDataNet Biology Quality Assessment**
4. **EMODNET - HRSM**, DTM processing
5. **Processing and visualising data sets**

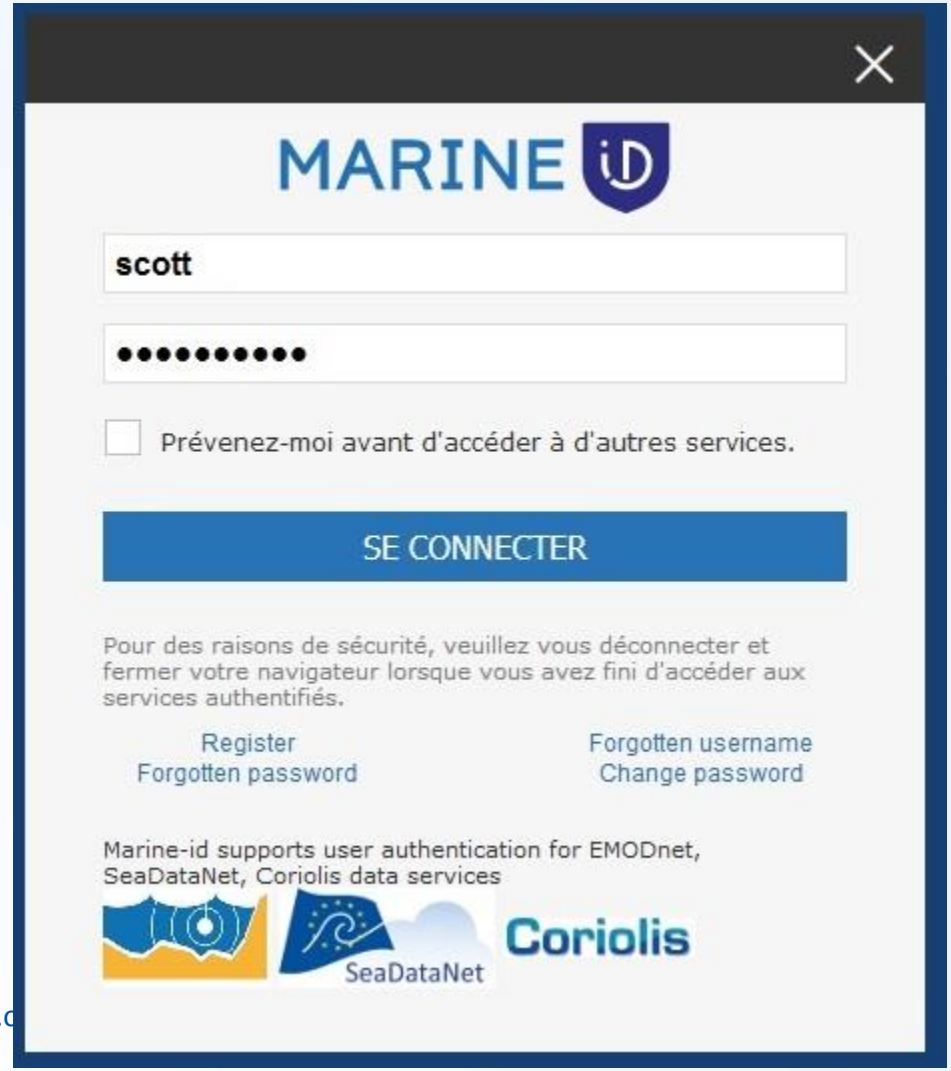
Abstract of T/S use case

log in with single sign on	B2ACCESS + Marine-ID
integration GUI development	Javascript library
apply water column obs quality control with friendly data editor and save result,	webODV
advise data centre of the regional quality control	
be advised of quality control result (email of log of changes/anomalies sorted per DC)	email
configure DIVA interpolation	Jupyter + DIVA library
apply DIVA interpolation, send notification (email) when processing is completed	
visualize interpolation result together with original observations of other observations	
extract and view profiles, time series, hovmuller out of the interpolation result	
publish dataset results (metadata and data), get a DOI	oceanBrowser+sextant-dataCite

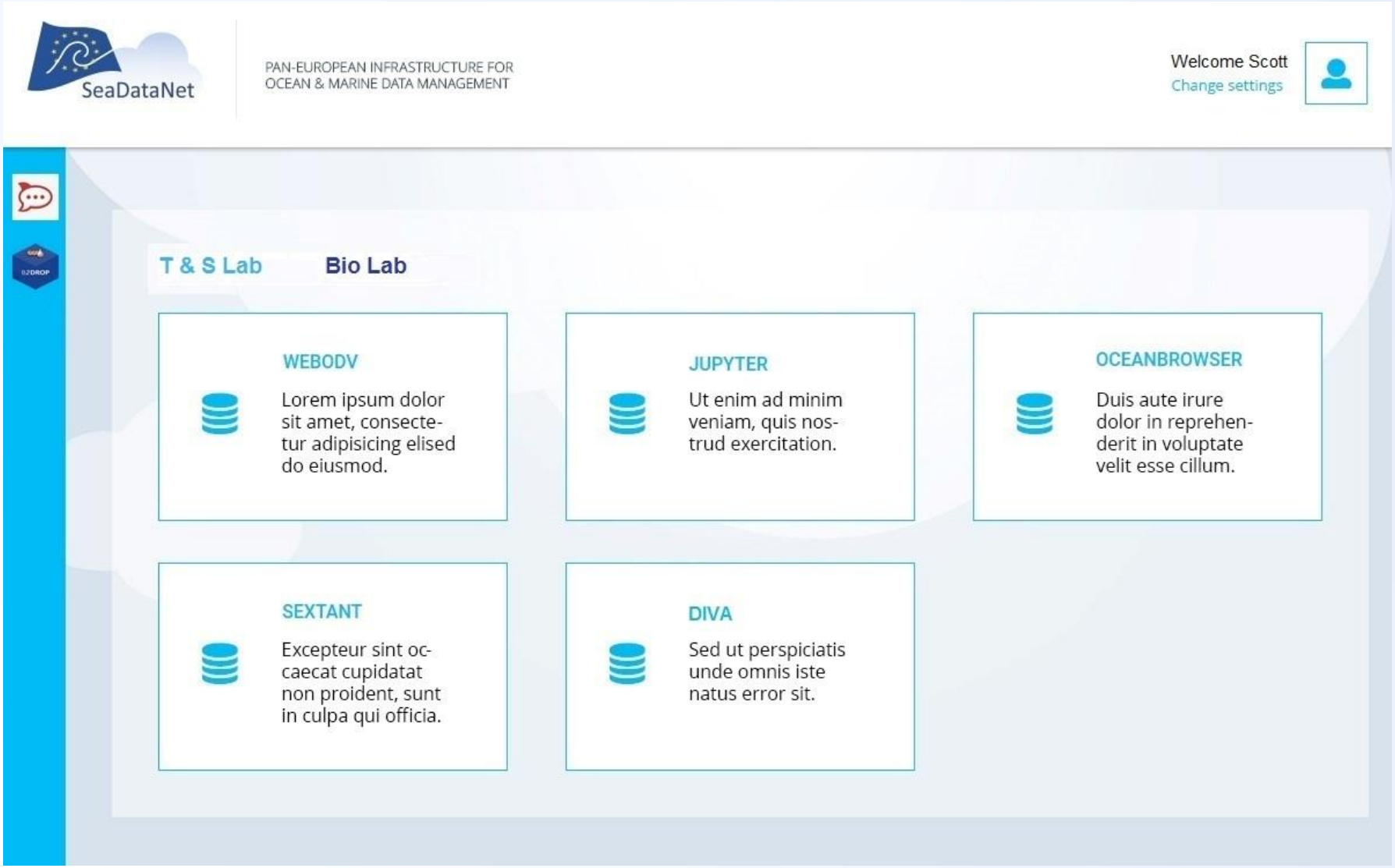
2. Progress per component

Some highlights

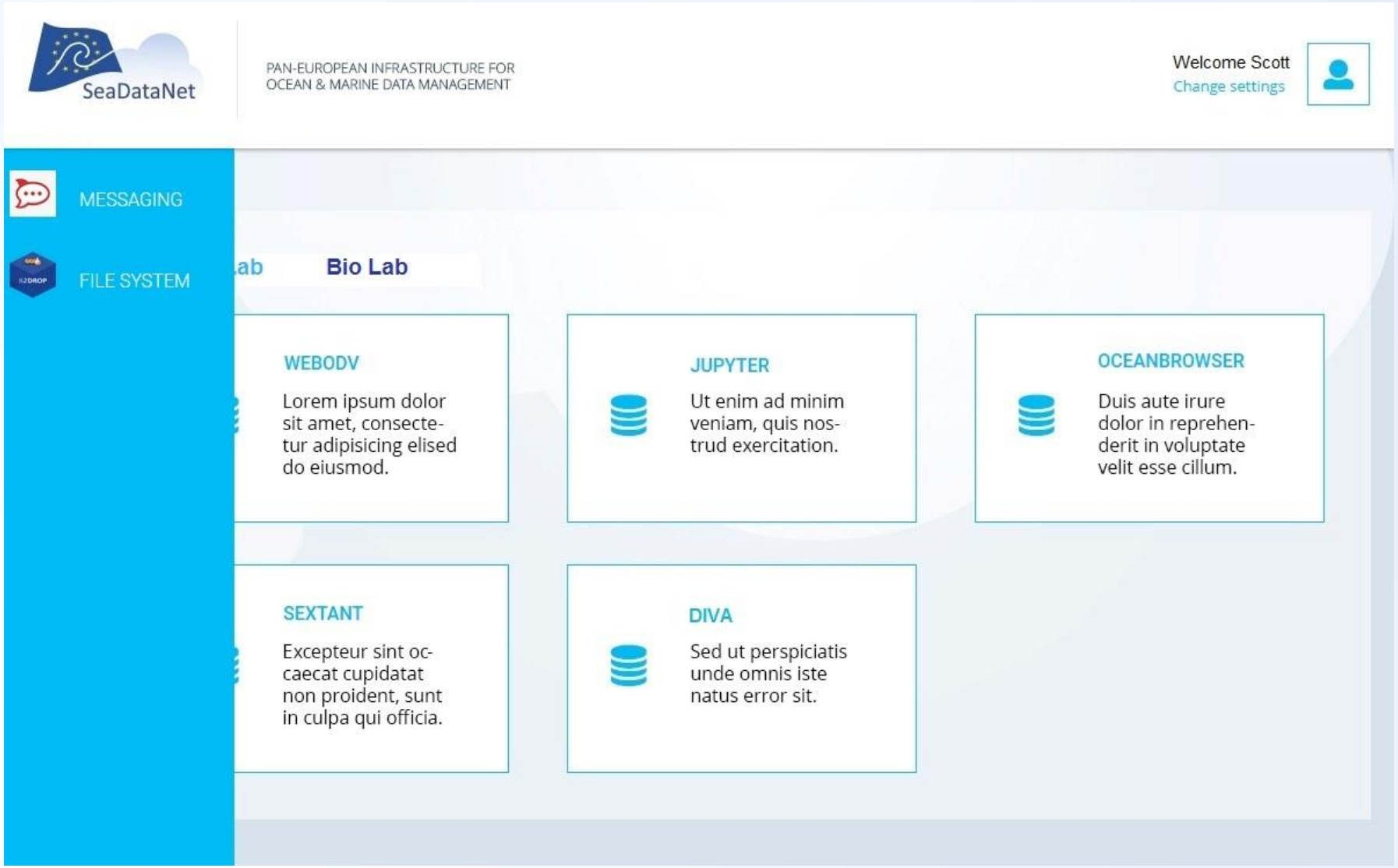
Marine-ID mapped to B2Acces



Dashboard


SeaDataNet
PAN-EUROPEAN INFRASTRUCTURE FOR
OCEAN & MARINE DATA MANAGEMENT
Welcome Scott
Change settings
T & S Lab
Bio Lab
WEBODV
Lorem ipsum dolor sit amet, consectetur adipisicing elised do eiusmod.
JUPYTER
Ut enim ad minim veniam, quis nostrud exercitation.
OCEANBROWSER
Duis aute irure dolor in reprehenderit in voluptate velit esse cillum.
SEXTANT
Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia.
DIVA
Sed ut perspiciatis unde omnis iste natus error sit.

Access to messaging a file system



Private workspace via B2Drop

sdc-vre - Chromium

sdc-vre

seadatacloud/vre

SeaDataNet

PAN-EUROPEAN INFRASTRUCTURE FOR OCEAN & MARINE DATA MANAGEMENT

Welcome Test
Change settings

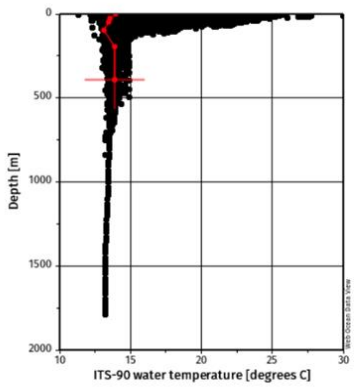
Quality control via WebODV

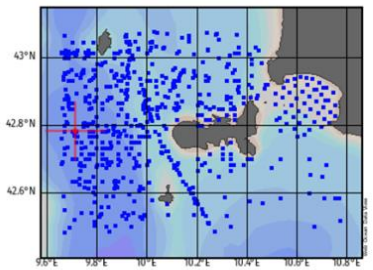
video

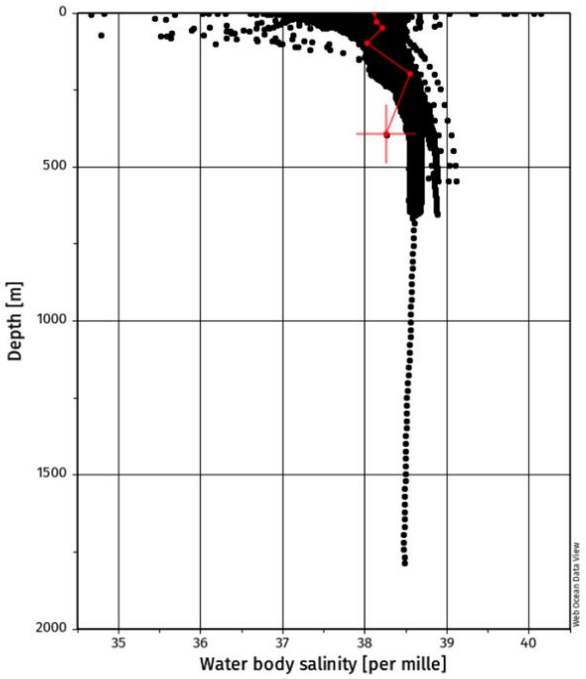
Quality Control Services

← previous | 1. Select data set | 2. QC Editor | 3. Save/Export/Exit | next →

Right click on the data windows or map window to open context menus and choose an options. Apply zoom by double left mouse click or *enter* key if *zoom mode* is active. Cancel *zoom mode* by *ESC* key. To assign a quality flag, right click on the value or flag of the respective variable in the "Sample" table at the bottom right of the page. Use the *arrow buttons* below to navigate from sample to sample within one station. Use double-clicking for larger steps.







Status:

Data set:	VRE/SDN_test_El...
Mouse:	

Station ID: 369

Accession Num...	369
Cruise	EAA2
Station	36950 (B)
Position	9.717° E / 42.783...
Date	25 March 1976
Time	10:05:59
Depth Range [m]	[0.00 - 396.45]
LOCAL_CDI_ID	158829
EDMO_CODE	120

Sample: 6/6

variable	value	flag
1: Depth [...	396.45	1
2: ITS-90 ...	13.88	4
3: Water ...	38.26	4

sdc-vre - Chromium

sdc-vre

seadatacloud/vre

SeaDataNet

PAN-EUROPEAN INFRASTRUCTURE FOR OCEAN & MARINE DATA MANAGEMENT

Welcome Test
Change settings

Data Extractor

← previous | 1. Select stations | 2. Select variables | 3. Download | 4. Exit | next →

Select cruises from the *Cruises menu*. Click *Zoom in* to define a sub-region, *Apply* to select the sub-region, or *Zoom out* to return to global domain. Use the *Required variables* as a station filter.

Selection status

Stations:
Output variables:

Cruises

cruises

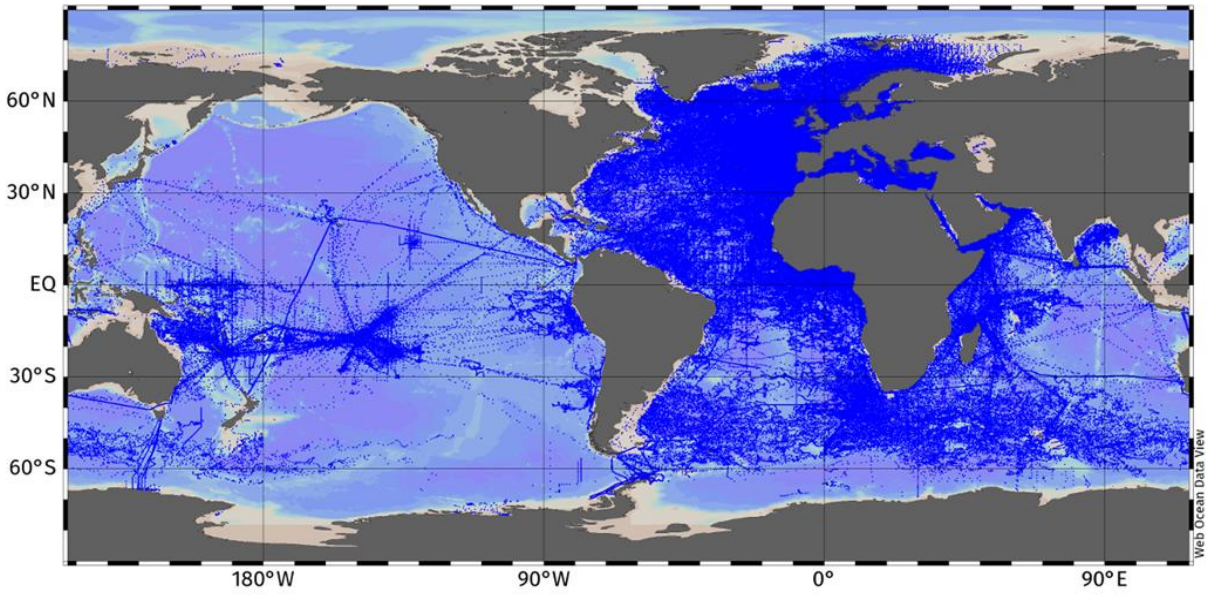
Map domain

Zoom in

Required variables

Nothing selected

Reset



sdc-vre - Chromium

sdc-vre

seadatacloud/vre

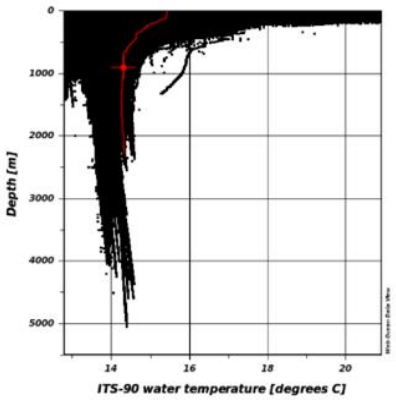
SeaDataNet

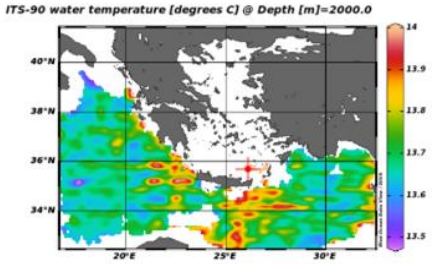
PAN-EUROPEAN INFRASTRUCTURE FOR OCEAN & MARINE DATA MANAGEMENT

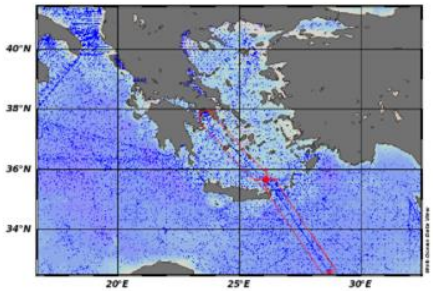
Welcome Test
Change settings

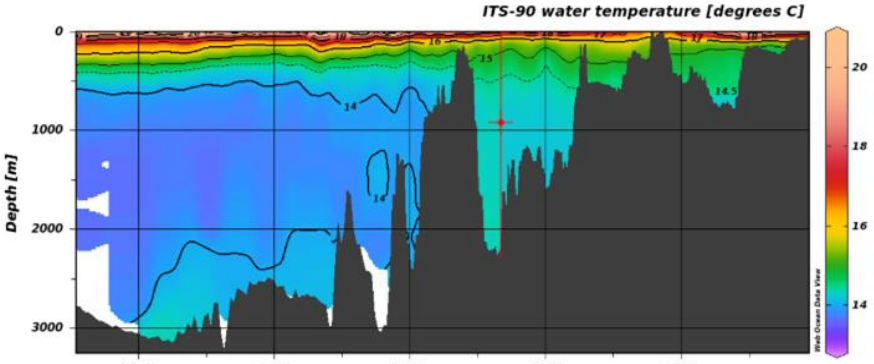
ODV-online

SDN_2015-09_TS_MedSea_QC_done_v2.odv









Station ID: 134757

Depth Range [m]	[2.0 - 2285.5]
Bot. Depth [m]	2316
LOCAL_CDI_ID	GN36200503701000_FS109_269
EDMO_code	269
Instrument Info	
Reference	
Data set name	GN36200503701c.med(FS109)
Discipline	Administration and dimensions;C

Sample: 923 / 2314

1: Depth [m]	915.2	1
2: ITS-90 water temperature [degrees	14.33	1
3: Water body salinity [per mille]	38.961	1
drvd: Section Latitude	35.68	1

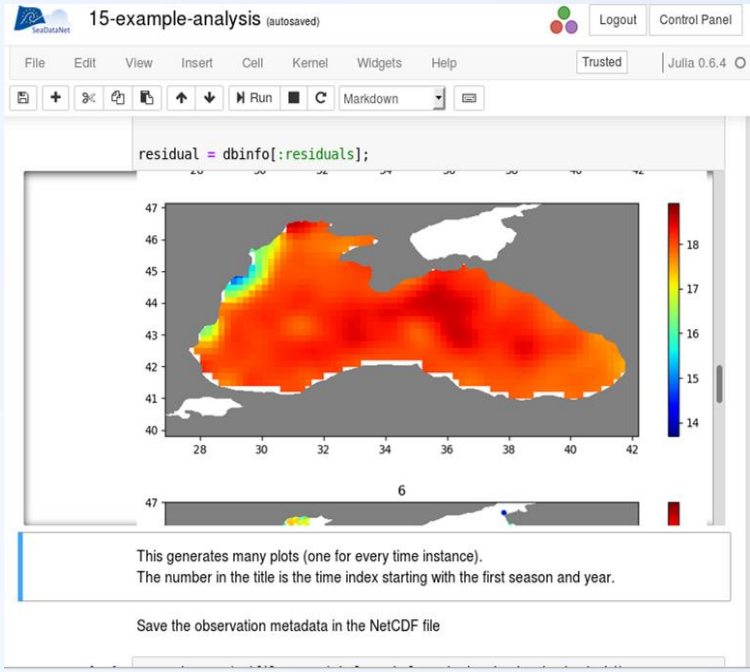
Isosurface Values

Longitude	26.119
Latitude	35.677
Time [yr]	2005.185
Day of Year	68
Depth [m] @ Depth [m]=first	2.0
ITS-90 water temperature [degrees C] @ Depth [	15.43

Window 5

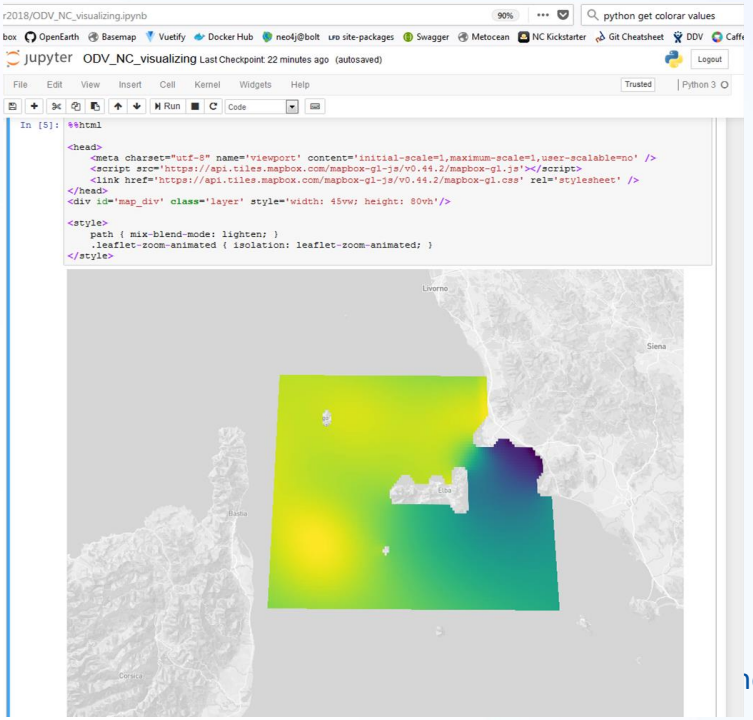
DIVAnd in Jupyter notebooks

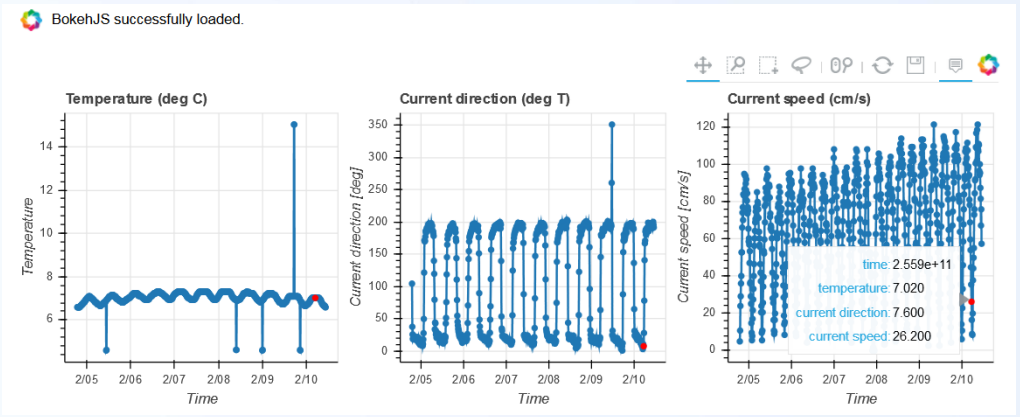
- Jupyter notebooks are integrated web environment
  - Computing
  - Visualization
  - Documentation of code
- DIVAnd extension of DIVA in more than 2 dimensions
- DIVAnd distributed as a Julia package installed into Docker containers



Additional viewing services

- Containers, Notebooks, Package, Visualization
- **Docker**file https://github.com/openearth/sdc-visualization
- Data type. NetCDF / tarfile from **B2DROP**
- Logging, **import sdc_visualization**

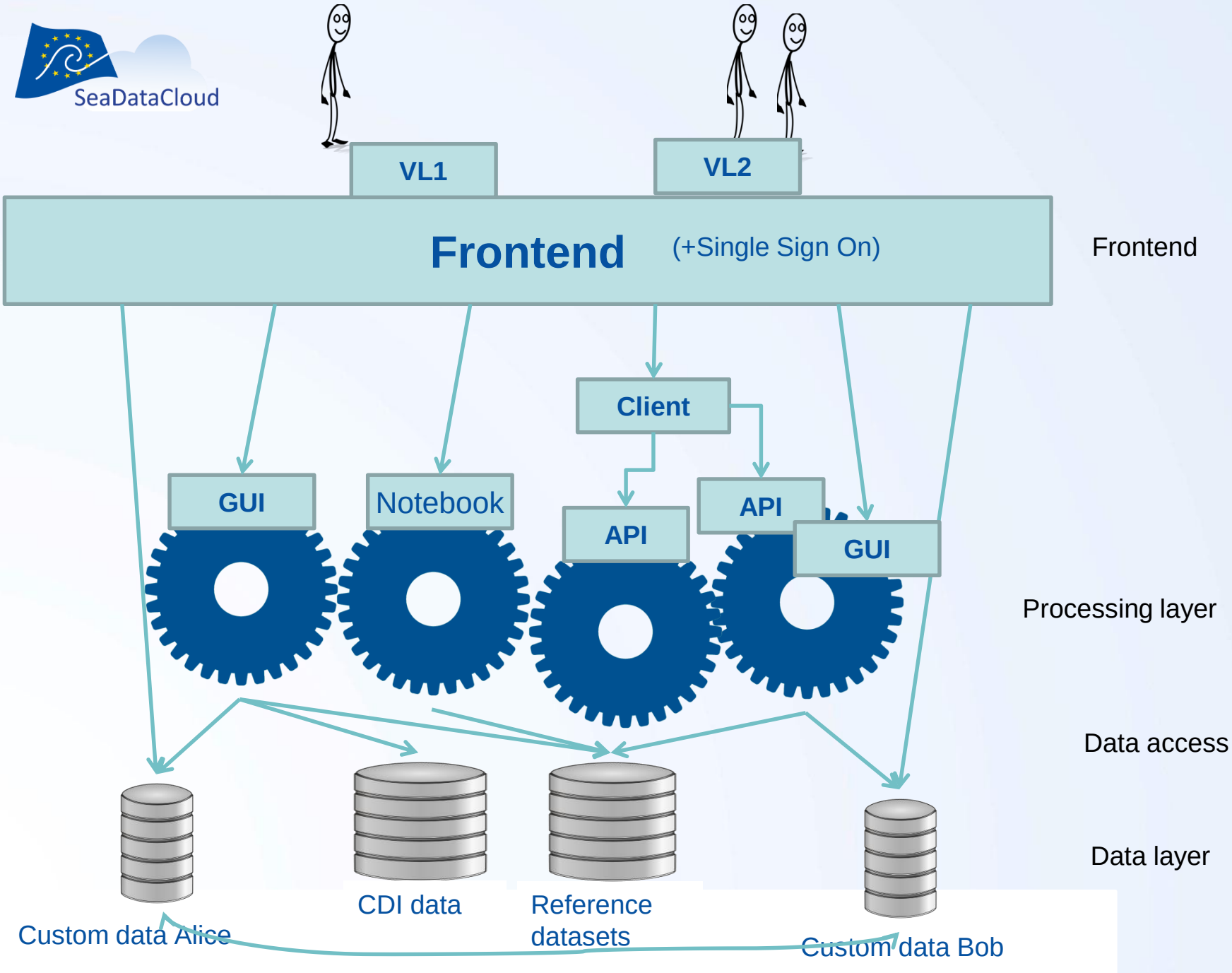




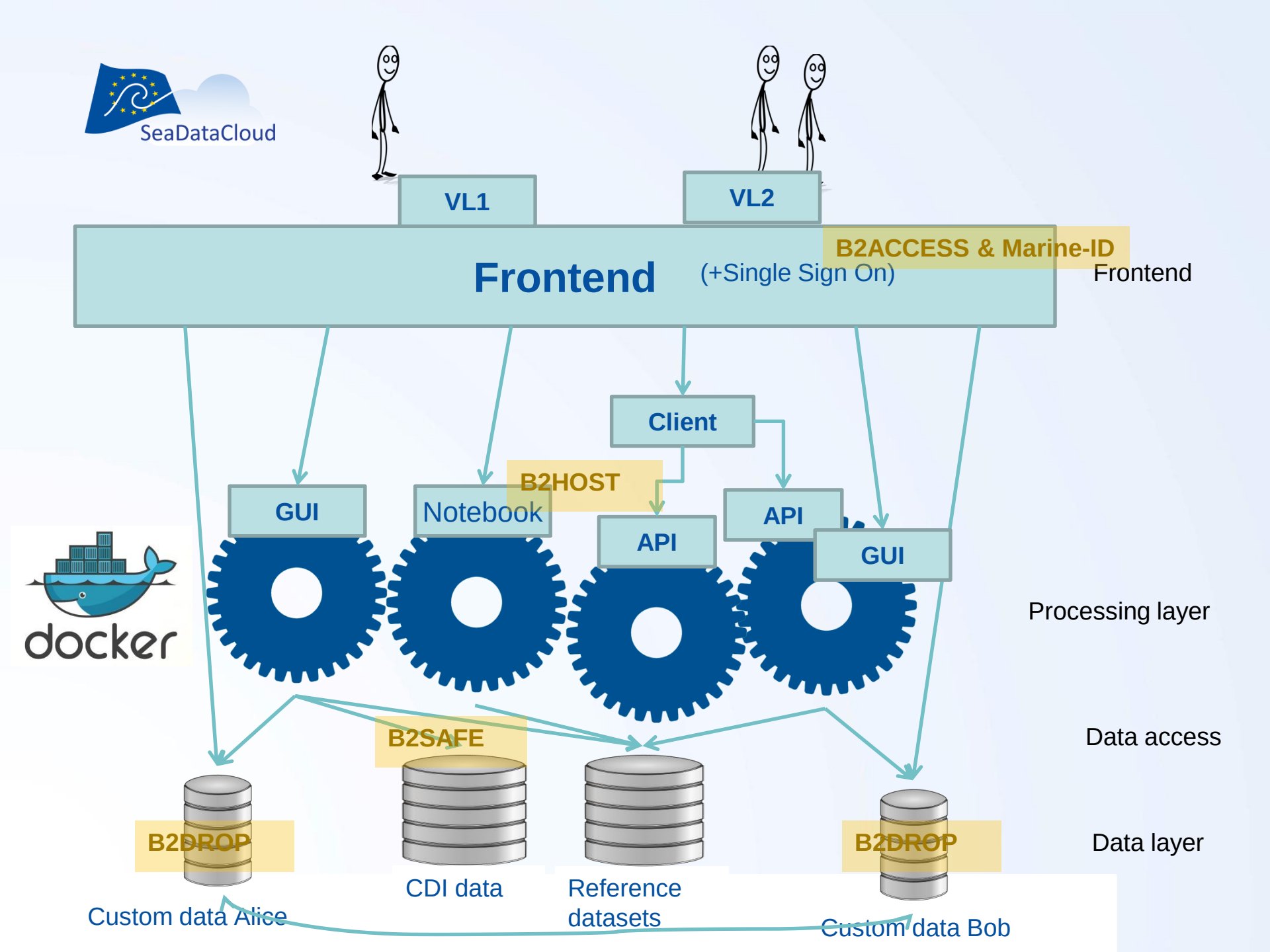
net.org

SeaDataCloud

3. Architecture and EUDAT backend


SeaDataCloud
VL1
VL2
Frontend
(+Single Sign On)
Client
GUI
Notebook
API
API
GUI
CDI data
Reference datasets
Custom data Alice
Custom data Bob
Frontend
Processing layer
Data access
Data layer

EUDAT services


SeaDataCloud
VL1
VL2
B2ACCESS & Marine-ID
Frontend
(+Single Sign On)
Frontend
Client
B2HOST
GUI
Notebook
API
API
GUI
docker
Processing layer
B2SAFE
Data access
B2DROP
B2DROP
Data layer
CDI data
Reference datasets
Custom data Alice
Custom data Bob

SeaDataCloud

Technical components

(slide only for later reference)

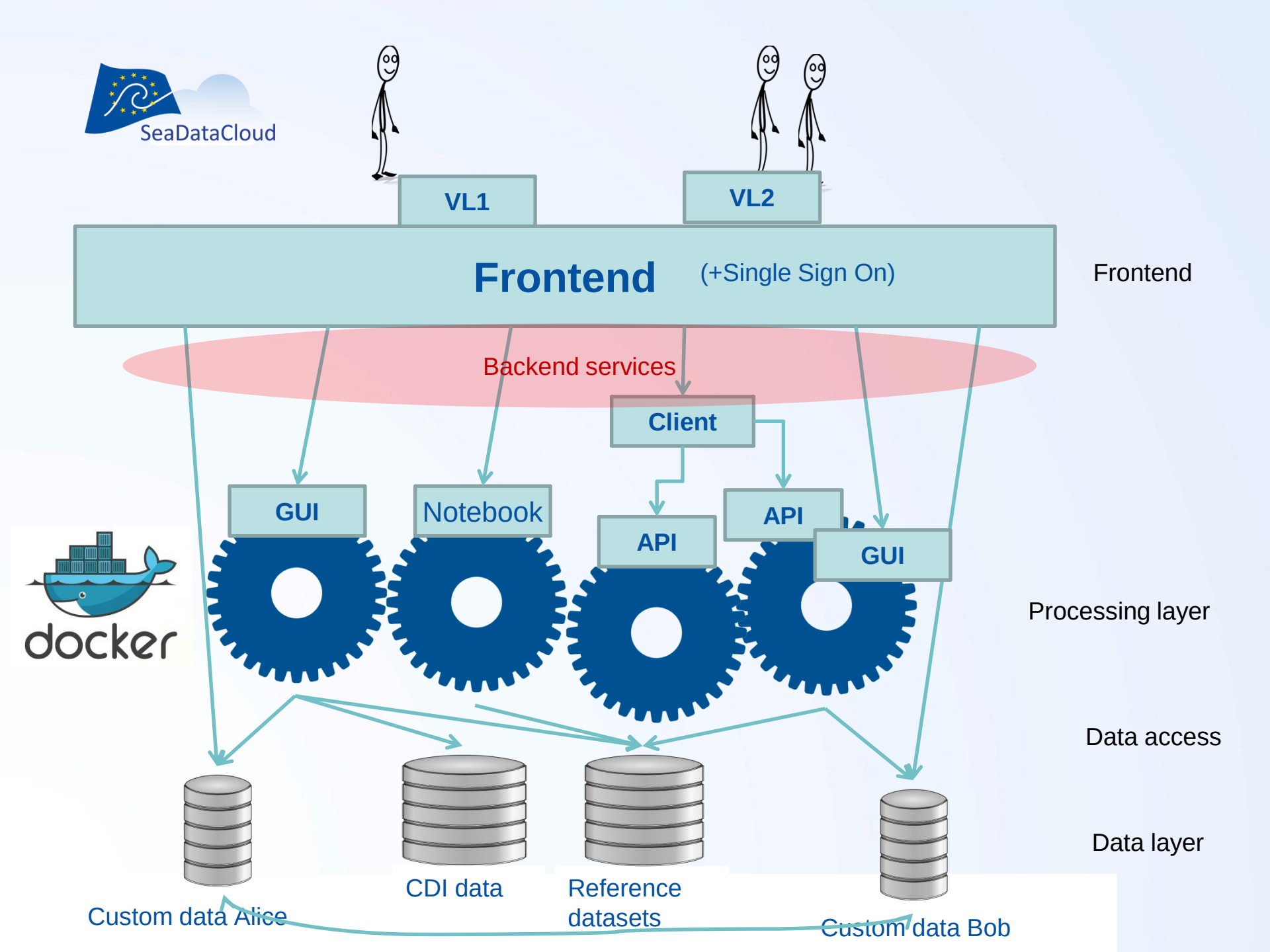
- B2ACCESS: User Authentication and group management
- B2SAFE and B2STAGE: SeaDataNet CDI data and other external datasets
- B2DROP: User data storage and collaboration
- B2HOST: VRE Execution environment
- Jupyter Notebooks: User code execution
- … JavaScript Frontend (GUI) to tie all together!

SeaDataCloud

Execution environment (B2HOST) (slide only for later reference)

Applications are provided as docker containers that run on EUDAT computing centres.

- Processing power
- Efficient access to data
- High availability / redundancy (5 centres)
- Resource accounting
- Service logging (for services run on EUDAT infrastructures)
- Docker registry (SDN-private, incl. security scan)

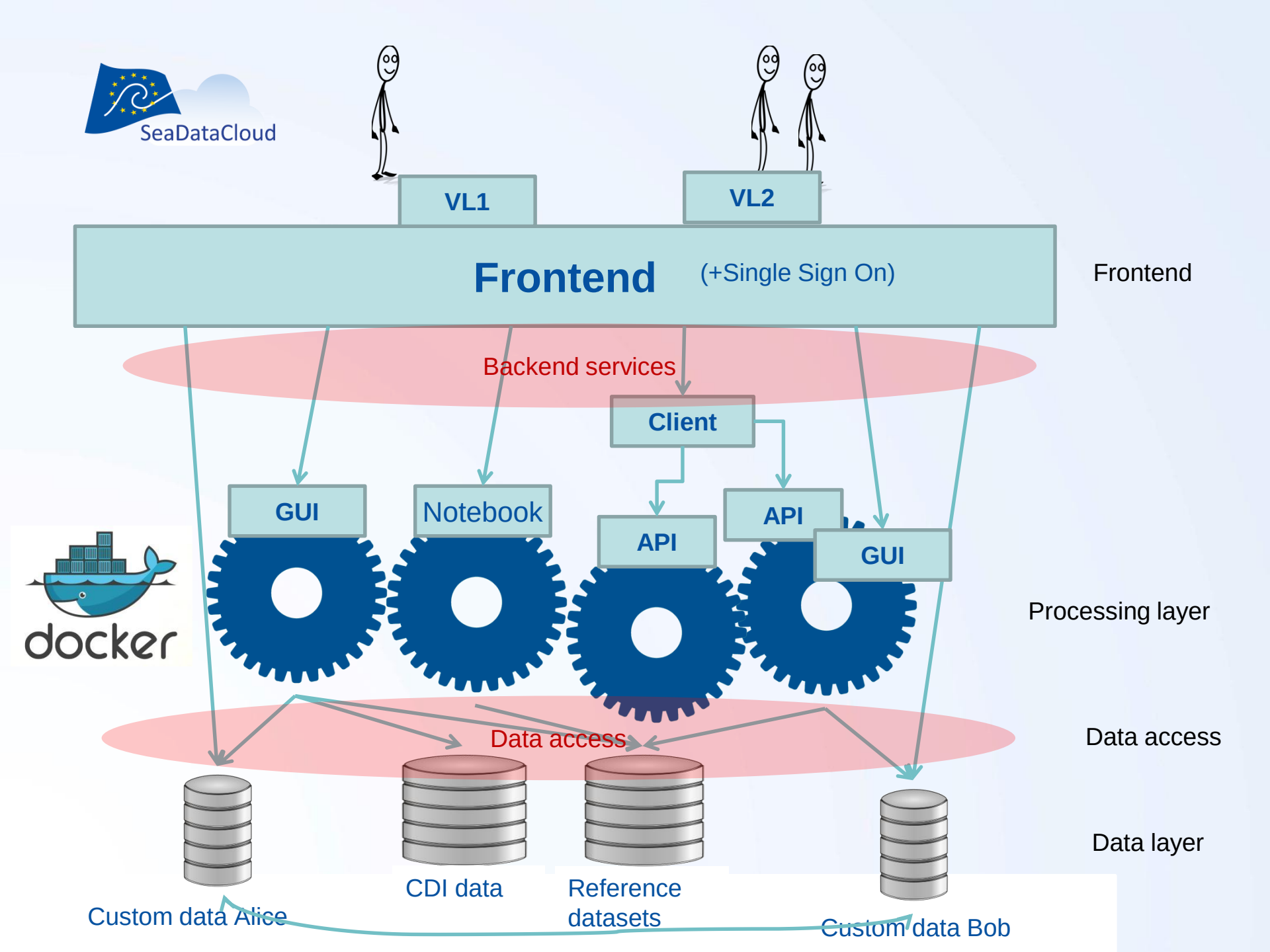

SeaDataCloud
VL1
VL2
Frontend
(+Single Sign On)
Backend services
Client
GUI
Notebook
API
API
GUI
docker
CDI data
Reference datasets
Custom data Alice
Custom data Bob
Frontend
Processing layer
Data access
Data layer

SeaDataCloud

Backend services

The services that the infrastructure provides to the frontend as HTTP requests.

- Starting/stopping/hibernating services (containers)
- Requesting external datasets
  to be moved close to processing
- Requesting B2DROP mount
  (unless directly done by services)


SeaDataCloud
VL1
VL2
Frontend
(+Single Sign On)
Backend services
Client
GUI
Notebook
API
API
GUI
docker
Data access
CDI data
Reference datasets
Custom data Alice
Custom data Bob
Frontend
Processing layer
Data access
Data layer

Data access

- Challenge: Efficient access to data, but no local copies (for accounting reasons).

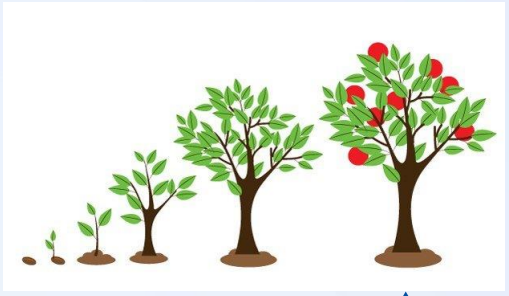
- Custom data + CDI data sits on B2DROP (nextcloud) and is mounted via WebDAV.
- SeaDataNet and EMODnet products sit on Sextant, (to be made) accessible via OpenDAP.
- More external datasets will be included.

4. Outlook

Outlook

SeaDataCloud

Next steps until June 2019:



- Beautiful design to be integrated
- Customization for different groups
- Push CDI data from portal to VRE
- Enable usage of SDN products
- etc.

After June 2019: Advanced features

(workflows, group management, communications, data publications, etc.)

Questions, or suggestions?