Studying the seas from the cloud

Progress of the SeaDataCloud Virtual Research Environment

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

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

[seadatanet.org]
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

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

Some highlights
Marine-ID mapped to B2Acces
Access to messaging a file system

- WEBODV
  Lorem ipsum dolor sit amet, consectetur adipiscing elit sed 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.
Private workspace via B2Drop
Quality control via WebODV

Quality Control Services

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.
Data Extractor

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.
ODV-online

SDN_2015-09_TS_MedSea_QC_done_v2.odv

Depth Range [m] [2.0 - 2285.5]
Bot. Depth (m) 2216
LOCAL_CEN_ID G86200903701000_FS109_261
EDMO_code 261
Instrument Info
Reference
Data set name G86200903701000_FS109
Discipline Administration and dimensions: C
Sample: 923 / 2314
1: Depth (m) 915.2 1
2: ITS-90 water temperature [degrees C] 14.33 1
3: Water body salinity [per mille] 38.961 1
dvdr: Section Latitude 35.68 1

Window 5

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
Water body: north

SeaDataNet
FAN-EUROPEAN INFRASTRUCTURE FOR OCEAN & MARINE DATA MANAGEMENT
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
- Dockerfile [https://github.com/openearth/sdc-visualization](https://github.com/openearth/sdc-visualization)
- Data type. NetCDF / tarfile from B2DROP
- Logging, `import sdc_visualization`
3. Architecture and EUDAT backend
EUDAT services
Technical components

- **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!
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)

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

We are here!
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?