WP10 – Virtual Research Environment –

Progress of VRE development

WP10 team
Outline

1. SDC VRE – progress summary
2. Demo of individual components
   - Marine-ID and B2Access (Themis, 5 mins)
   - WebODV (Sebastian, 10 mins)
   - DIVAnd and Notebooks (Charles, 10 mins)
   - Biological QC (Simon, 5 mins)
   - Visualisation (Giorgio, 5 mins)
3. VRE deadlines and deliverables
1. VRE – Progress summary
SeaDataClouds VRE needs to:

- Facilitate collaborative and individual research:
- Combine data with subsets from other data resources, such as the ingested collections
- Have a high capacity and performance for big data processing and state-of-the-art web visualisation services
– Respect privacy of users and differences in data policies. Differentiated users, different access to data and data products.

– Be possible to configure virtual work spaces for individuals or groups to work on specific projects, including setting up of dedicated pools of data

– Allow producers to decide whether their outcomes will be shared in the public domain or stay private

– Be based and hosted on EUDAT’s infrastructure based on its B2-… service platforms
Focus on 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
SDC VRE Architecture – level 1

[Diagram showing the SDC VRE Architecture with layers: AAAI Layer, Frontend Layer / GUI, Service Layer, Data Layer, including components like VL 1, VL 2, VL N, WebODV, Jupyterhub, Quality Controls, Transformation Services, Visualization Services, Publishing Services, Workflow Services, Database Services, Message Bus Services, B2DROP, B2SAFE, Temporal virtual storage, External datasets.]
Prototype functions targeted of T/S (narrow abstract of total use case!)

<table>
<thead>
<tr>
<th>Function</th>
<th>Tool/Technology</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
Connecting Marine-ID to B2Access - Themis
Connecting B2ACCESS - Marine-ID

- B2ACCESS is an easy-to-use and secure Authentication and Authorization platform developed by EUDAT. B2ACCESS is versatile and can be integrated with any service.

- Marine-ID is an integrated Registration, Authentication and Accounting infrastructure for marine data users (SeaDataNet, EMODNET, ...).
Connecting B2ACCESS - Marine-ID

Marine-ID has been registered as an Identity Provider in B2ACCESS allowing users to access Services connected to B2ACCESS using their Marine-ID accounts.
Step 1: Demo page

Demo: https://snf-761524.vm.oceanos.grnet.gr/b2access
Step 2: Insert Credentials

![Login Screen](image)
Step 3: Marine-ID Consent Form

Our Identity Provider
(replace this placeholder with your organizational logo / label)

Information to be Provided to Service
- cn
- eduPersonPrincipalName
- givenName
- mail
- sn
- uid

The information above would be shared with the service if you proceed. Do you agree to release this information to the service every time you access it?

Select an information release consent duration:
- Ask me again at next login
  - I agree to send my information this time.
- Ask me again if information to be provided to this service changes
  - I agree that the same information will be sent automatically to this service in the future.
- Do not ask me again
  - I agree that all of my information will be released to any service.

This setting can be revoked at any time with the checkbox on the login page.

Reject Accept
Step 4: B2ACCESS Consent form

SeaDataCloud OAuth2 Authorization Server

A remote client has requested your authorization

SeaDataCloud Test Client
Address: https://snf-761524.vm.okeanos.grnet.gr/b2access/refreshtoken.php

Access to the following information was requested:

- USER_PROFILE
  Provides access to the user’s profile information
  profile
- OpenID Connect user profile scope
- openid
- Enables the OpenID Connect support
  email
- OpenID Connect Email Scope
- GENERATE_USER_CERTIFICATE
Step 5: Yes you have logged in

**B2ACCESS**

**Client ID:** sdc-test-client-id  
**Client Secret:** sdc-test-client-pass

Access Token:  
WCgDDBo0ofRC2TCgW04WHkgZGaD7veJA25FultPN9rU  
NOTE: New access tokens expire in 10 minutes.

Refresh Token:  
QIF15bmc-Lho4BbtPC1217kb5k9AM26Ev8aEYtXpBY  
NOTE: New refresh tokens expire in 1 year.

To generate access tokens from this refresh token use the following curl command:  
curl -X POST -u 'sdc-test-client-id':'sdc-test-client-pass' -d 'client_id=sdc-test-client-id'  
To get the user info from the access token, execute the following curl command:  
curl -H 'Authorization: Bearer WCgDDBo0ofRC2TCgW04WHkgZGaD7veJA25FultPN9rU'
Access to SDC web-based services

1. User clicks on the “login” button in the portal of a SeaDataCloud service and he/she is redirected to B2ACCESS.
   - Then B2ACCESS redirects the user automatically to Marine-ID.
2. The user enters his/her credentials.
3. The user is asked to consent the information that Marine-ID is going to release to B2ACCESS.
   - The user is redirected back to B2ACCESS.
4. If the user logsins through B2ACCESS for the first time, he/she needs to register to the platform. Otherwise, the user will be asked to consent the information that B2ACCESS is going to release to the SeaDataCloud service.
5. The user is redirected back to the SeaDataCloud service.
Register only the first time
More on WebODV integration - Sebastian
Quality control

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.

<table>
<thead>
<tr>
<th>Station ID: 369</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accession Num... 369</td>
</tr>
<tr>
<td>Cruise          EA42</td>
</tr>
<tr>
<td>Station         36950 (8)</td>
</tr>
<tr>
<td>Position        9.717° E / 42.783...</td>
</tr>
<tr>
<td>Date            25 March 1976</td>
</tr>
<tr>
<td>Time            10:05:59</td>
</tr>
<tr>
<td>Depth Range [m] [0.00 - 396.45]</td>
</tr>
<tr>
<td>LOCAL_CDI_ID    158829</td>
</tr>
<tr>
<td>EDMO_CODE       120</td>
</tr>
</tbody>
</table>

Sample: 6/6

<table>
<thead>
<tr>
<th>variable</th>
<th>value</th>
<th>flag</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: Depth...</td>
<td>396.45</td>
<td>1</td>
</tr>
<tr>
<td>2: ITS-90...</td>
<td>13.88</td>
<td>4</td>
</tr>
<tr>
<td>3: Water...</td>
<td>38.26</td>
<td>4</td>
</tr>
</tbody>
</table>
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.

Selection status
Stations:
Output variables:

Cruises
- cruises

Map domain
- Zoom in

Required variables
- Nothing selected

Reset
Select output variables from the Output variables treeview.

- All
  - Depth
  - Temperature
  - Salinity
Data Extractor

Download data in different formats. You will receive a .zip file containing the data in the format of your choice.

Selection status
Stations:
Output variables:

Download
- Spreadsheet
- ODV Collection
- netCDF
- WHP Exchange
REST API -- for developers (us) and advanced users

**cURL Request:**

**JSON Response:**
```
{"station_id":642,"sample_id":1,"sample_count":379,
...
,"datavars_id":[1,2,3],"datavars":["1.98","21.60","38.03"],"flags":["1","1","1"],
...
}
```
More on DIVA integration - Alexander, Charles
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
DIVAnd in Jupyter notebooks

- Loading bathymetry (GEBCO, EMODnet)
- Data formats (ODV spreadsheets, netCDF, mat files etc.)
- Support for databases like: World Ocean Database
  CMEMS INSTAC
- Check for duplicates
- Automated quality checks
- Parameter optimisation
- Error field computation
- Plot results and data
- Generate XML for catalog

Full example notebook

sdn-userdesk@seadatanet.org – www.seadatanet.org
Integrations with SeaDataCloud services

- Querying the vocabulary service
- Querying the EDMO database
- XML files for Sextant
- Extracting data from the ODV REST API
- Embed additional metadata for a given observation based on EDMO code and local CDI id from jupyter notebook
REST API of DIVAnd

- User provides the URL of the data set (netCDF format)
- Provide analysis parameters and metadata
- Download the final analysis
Deployment-Pipeline

- Pipeline of the Docker images for the
  - Jupyter notebook
  - REST API

Source code repository → Code testing → Distribution of Docker images

GitHub → Travis CI → Docker Hub

*Only fully implemented features are merged into the master branch*
Biological QC – Simon
Additional visualisation – Giorgio
- 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`
Docker image updates

- Custom logging format (both for notebooks and python code)
- Updates published to hub.docker.com/openearth/sdc-visualization
- Closer follows docker guidelines
- Some help needed with limiting restrictions needed (--privileged --cap-add SYS_ADMIN --device)
Mapbox application

- Interactive API
- Code is available @ hub.docker.com/openeart/h/sdc-visualization
- Reads NetCDF and subset based on time
- Some help needed with limiting restrictions (--privileged --cap-add SYS_ADMIN –device)
Movie example
Integration layer mockups (UI) - Peter
Private workspace
T & S Lab  |  Bio Lab

WEBODV
Lorem ipsum dolor sit amet, consectetur adipiscing 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.
3. VRE deadlines and deliverables
Where are we now?

- Development kick-off At Deltares
- Virtual code sprint in June 2018
- Physical workshops in Sopot and Helsinki
- Currently in third development cycle of 4 months (Nov 2018 - Feb 2019)
- Two weekly telco’s / trello boards with actions
- Currently in integration phase:
  - Many concepts already tested
  - Basic functions working in the cloud
  - Now development of dashboard for first user group
<table>
<thead>
<tr>
<th>Project month</th>
<th>Component</th>
<th>remark</th>
</tr>
</thead>
<tbody>
<tr>
<td>PM12</td>
<td>VRE specification document</td>
<td>This document</td>
</tr>
<tr>
<td>PM15</td>
<td>Start actual developments VRE</td>
<td>Kickoff to be decided</td>
</tr>
<tr>
<td><strong>PM18</strong></td>
<td><strong>D10.3:</strong> Specification of sub-setting application and development plan</td>
<td>Led by IFREMER</td>
</tr>
<tr>
<td><strong>PM18</strong></td>
<td><strong>D10.5:</strong> Specification of Ocean Data View online and development plan</td>
<td>Led by AWI</td>
</tr>
<tr>
<td><strong>PM18</strong></td>
<td><strong>D10.7:</strong> Specification of DIVA online and development plan</td>
<td>Led by ULiege</td>
</tr>
<tr>
<td><strong>PM18</strong></td>
<td><strong>D10.9:</strong> Specification of Biology Data QC online and development plan</td>
<td>Led by VLIZ</td>
</tr>
<tr>
<td><strong>PM18</strong></td>
<td><strong>D10.13:</strong> Specification of visualisation services and development plan</td>
<td>Led by Deltares</td>
</tr>
<tr>
<td><strong>PM18</strong></td>
<td><strong>D10.19:</strong> Specification of MySeaDataCloud and development plan</td>
<td>Led by EUDAT</td>
</tr>
<tr>
<td>PM24</td>
<td>D10.10: Phase 1 of Biology Data QC online operational</td>
<td>Led by VLIZ</td>
</tr>
<tr>
<td>PM24</td>
<td>D10.17: Specification of SOS viewing services and development plan</td>
<td>Led by 52N</td>
</tr>
<tr>
<td>PM26</td>
<td>D10.15: Specification of Oceanotron services and development plan</td>
<td>Led by IFREMER</td>
</tr>
<tr>
<td><strong>PM30</strong></td>
<td><strong>D10.2</strong> First version Virtual Research Environment (VRE) in the cloud operational</td>
<td>For internal users, related to selected use cases</td>
</tr>
<tr>
<td><strong>PM30</strong></td>
<td><strong>D10.4:</strong> Sub-setting application operational in VRE</td>
<td>Led by IFREMER</td>
</tr>
<tr>
<td><strong>PM30</strong></td>
<td><strong>D10.6:</strong> Ocean Data View online operational in VRE</td>
<td>Led by AWI</td>
</tr>
<tr>
<td><strong>PM30</strong></td>
<td><strong>D10.8:</strong> DIVA online operational in VRE</td>
<td>Led by ULiege</td>
</tr>
<tr>
<td><strong>PM30</strong></td>
<td><strong>D10.14:</strong> Visualisation services operational in VRE</td>
<td>Led by Deltares</td>
</tr>
<tr>
<td>PM30</td>
<td>D10.20: MySeaDataCloud operational</td>
<td>Led by EUDAT</td>
</tr>
<tr>
<td>PM31</td>
<td>D10.18: SOS viewing services for data streams operational</td>
<td>Led by 52N</td>
</tr>
<tr>
<td>PM36</td>
<td>D10.11: Phase 2 of Biology Data QC online operational</td>
<td>Led by VLIZ</td>
</tr>
<tr>
<td>PM40</td>
<td>D10.16: Upgraded Oceanotron services operational for SeaDataNet WP11 data products</td>
<td>Led by IFREMER</td>
</tr>
<tr>
<td>PM42</td>
<td>D10.12: Phase 3 of Biology Data QC online operational</td>
<td>Led by VLIZ</td>
</tr>
<tr>
<td><strong>PM42</strong></td>
<td><strong>VRE operational for internal (and some public?) users</strong></td>
<td>New: No official deliverable</td>
</tr>
</tbody>
</table>
Questions, or suggestions?