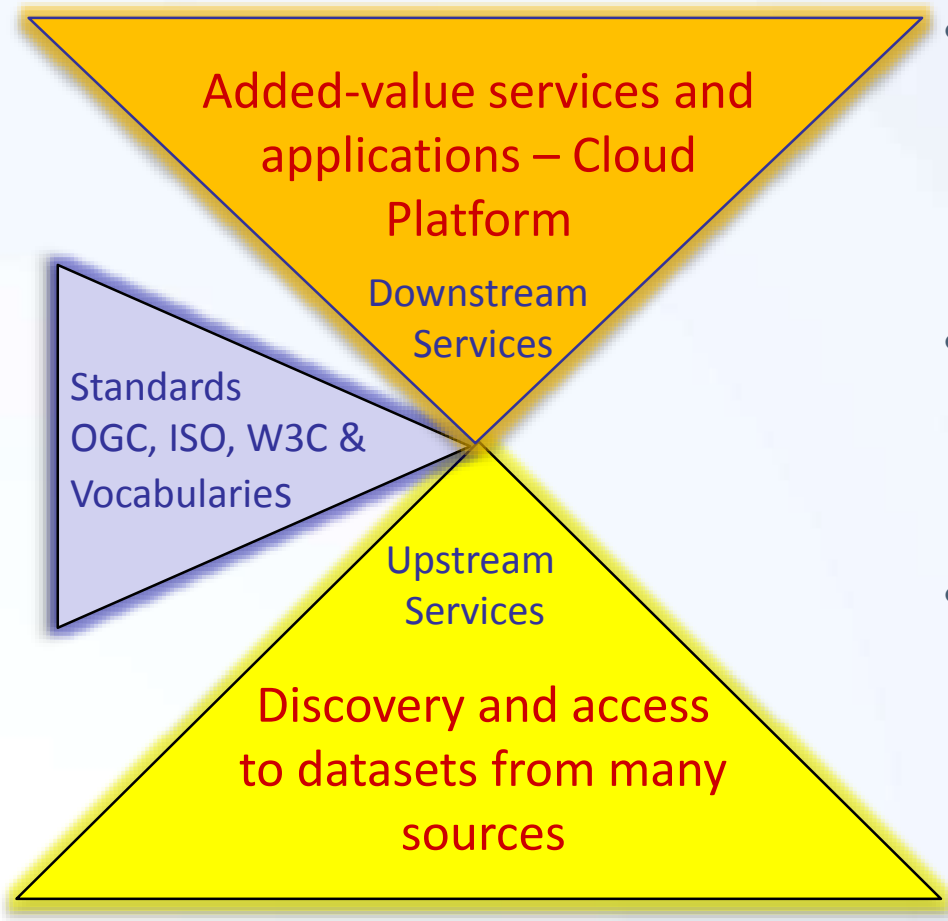
SeaDataCloud

Developing a Virtual Research Environment (VRE)

Dick M.A. Schaap, MARIS, SDC Technical Coordinator

SDC Review, Brussels – Belgium, 6 December 2018
sdn-userdesk@seadatanet.org – www.seadatanet.org

Leading concept for SeaDataCloud



- Providing a cloud platform with common services for data pre-processing, subsetting, analyses, visualizations, publishing, DOIs…
- Applying common standards and interoperability solutions for providing harmonised data and metadata
- Providing harmonised discovery and access to data output from multiple sources, such as European research and monitoring data gathering, but also from other European and international data infrastructures

SDC VRE needs to:

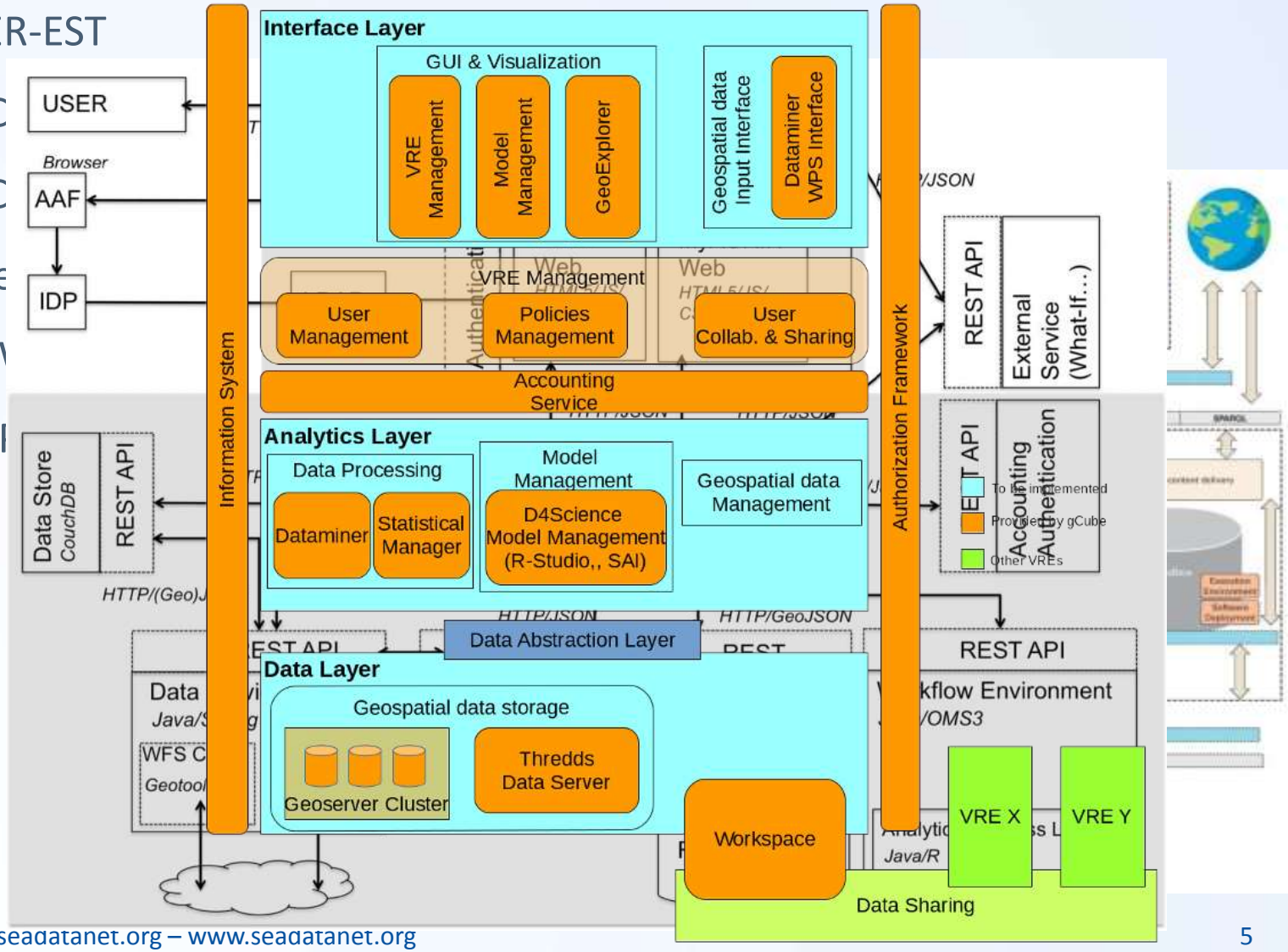
- Facilitate collaborative and individual research:
- Combine data with subsets from other data resources, such as ingested collections
- Have a high capacity and performance for big data processing and state-of-the-art web visualisation services
- Respect privacy of users and differences in data policies.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

SDC VRE needs to:

- Allow flexible expansion with new tools
- Facilitate building different independent workflows or virtual labs
- 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 it B2-… service platforms

Other VRE's analysed

- EVER-EST
- NEC
- NEC
- Blue
- Life
- Ecol



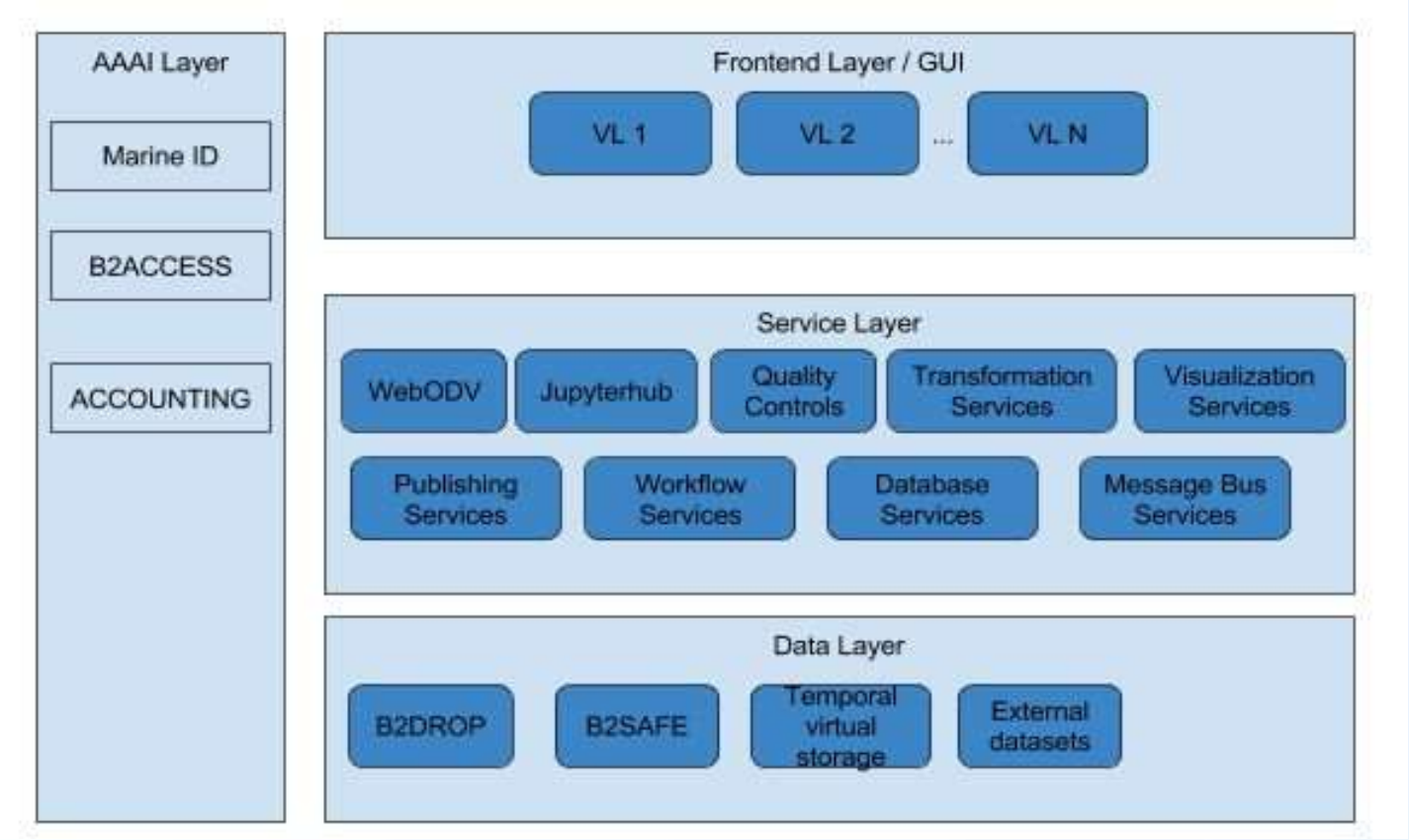
Analysis of existing VREs

- Mostly the same expectations with respect to community building, data sharing, processing and analysis tools
- Authorisation/Authentication layer both in portal layer as well as on top of service layer
- API's for each (processing) service
- Communication and (meta)data standards are key to success
- Front end applications are various: From self created workflows, to VRE virtual labs, to dedicated user interfaces. But all run on same set of services and data.
- Need to distinguish well between typical VRE modules (communication, GUI, etc) and the fundamental architecture

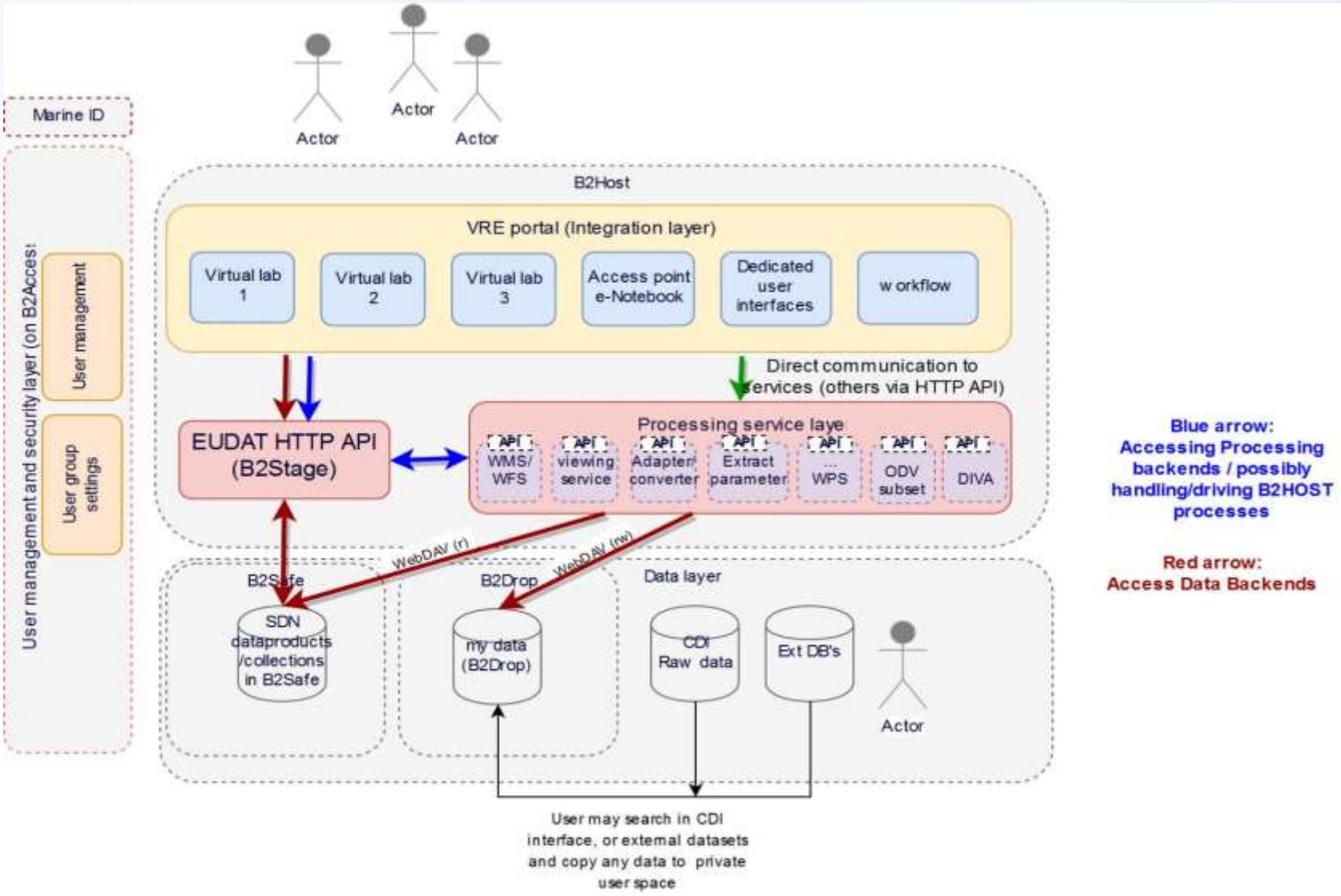
SDC VRE is developed for use cases:

1. **SeaDataNet Temperature and Salinity climatology generation *(Simona)***
2. **EMODnet Chemistry:** comparable use case for bio-geo-chemistry
3. **SeaDataNet: Biology Quality Assessment**
4. **EMODnet Bathymetry:** DTM processing pilot

SDC VRE Architecture – level 1

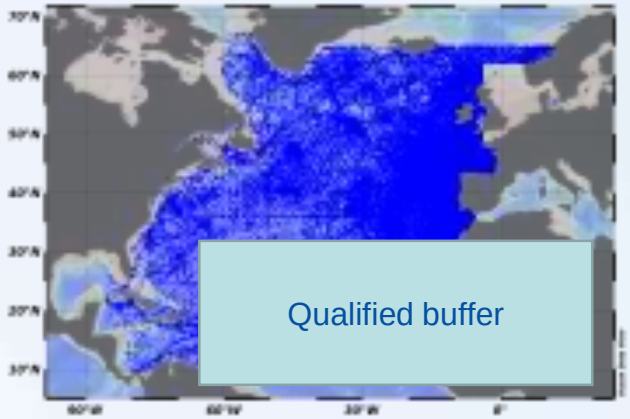


SDC VRE Architecture – level 2

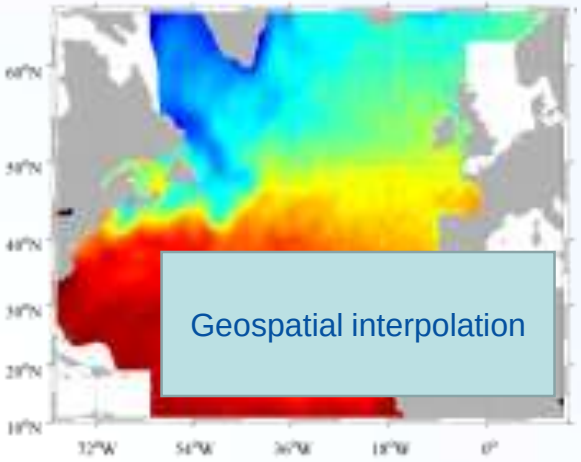


SeaDataNet Temperature and Salinity climatology generation

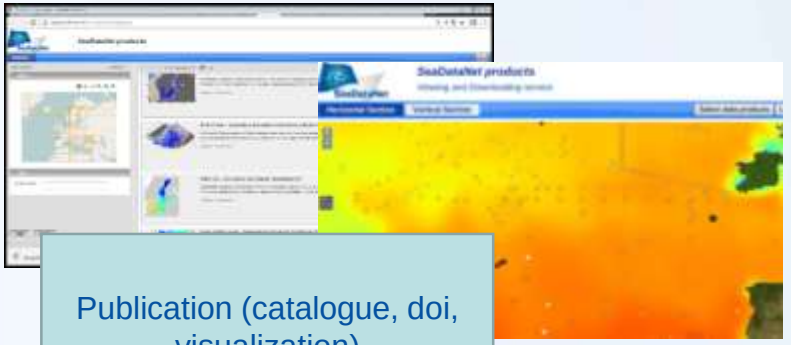
CDI



Qualified buffer



Geospatial interpolation



Publication (catalogue, doi, visualization)

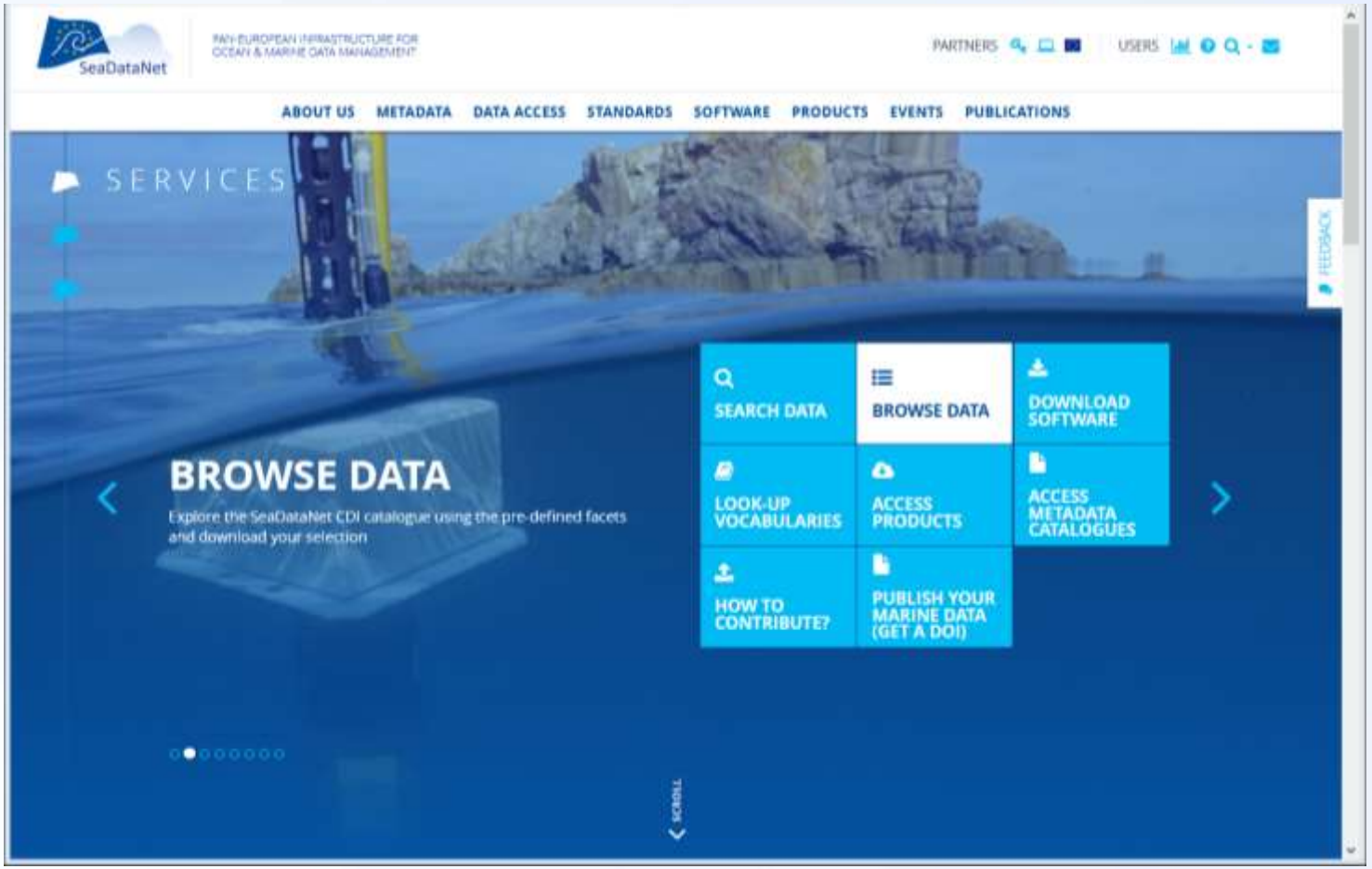
Abstract workflow T/S use case

log in with single sign on	B2ACCESS + Marine-ID
integration GUI development	Javascript library
data pool	CDI service; B2DROP for own data; Brokerage third party portals
subsetting and pre-processing data	ERDDAP; Octopus; …
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+ERDDAP+sextant-dataCite

VRE components

- **Authentication**: exchange has been established between Marine-ID (SDN AAA) and B2ACCESS (EUDAT AAA)
- **Data pool**: exchange with CDI service; B2DROP for user data; brokerage for third party portals
- **VRE base**: B2HOST on CSC servers, JupyterHUB, use of Docker containers
- **Subsetting**: ERDDAP selected as tool and made available in Docker container
- **WebODV**: prototype, REST-API for integration with workflows and Jupyter notebooks ***(Reiner)***
- **DIVAnd**: rewritten to Julia, Jupyter notebook and REST-API ***(Alexander)***
- **Visualisation:** prototypes with OpenEarth and Mapbox on the basis of NetCDF files
- **Biology QA-QC:** specifications for workflow formulated, first test in Docker
- **Accounting:** progress on accounting VRE resources usage KPIs
- **Front end interface:** progress with dashboard design

SeaDataCloud



www.seadatanet.org