Ocean Data Interoperability Platform (ODIP)

SeaDataCloud VRE – Analysis of potential architecture options

8h ODIP Workshop – Galway - Ireland– 2 October 2017

European Union's HORIZON 2020 Framework Programme for Research and Innovation
Grant agreement no 654310

Peter Thijsse / Thomas Loubrieu
On behalf of: T Loubrieu, G Leblan, A. Barth, Charles Troupin,Christine Coatanoan, Merret Buurman, Marco rorro, Nicolas Liampotis, Harsha Vathsavay, Claudio Cacciari, Sebastien Mieruch, Reiner Schlitzer, Giorgio Santinelli, Simon Claus, Mickaël Treguer, Dick M.A. Schaap

SeaDataClouds VRE needs to:

- facilitate collaborative and individual research: Using, handling, analysing and processing ocean and marine data into value-added data products which can be integrated, visualised and published using high level visualisation services.
- combine 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
- Based on EUDAT's infrastructure based on it B2-… service platforms

This is an ambitious challenge

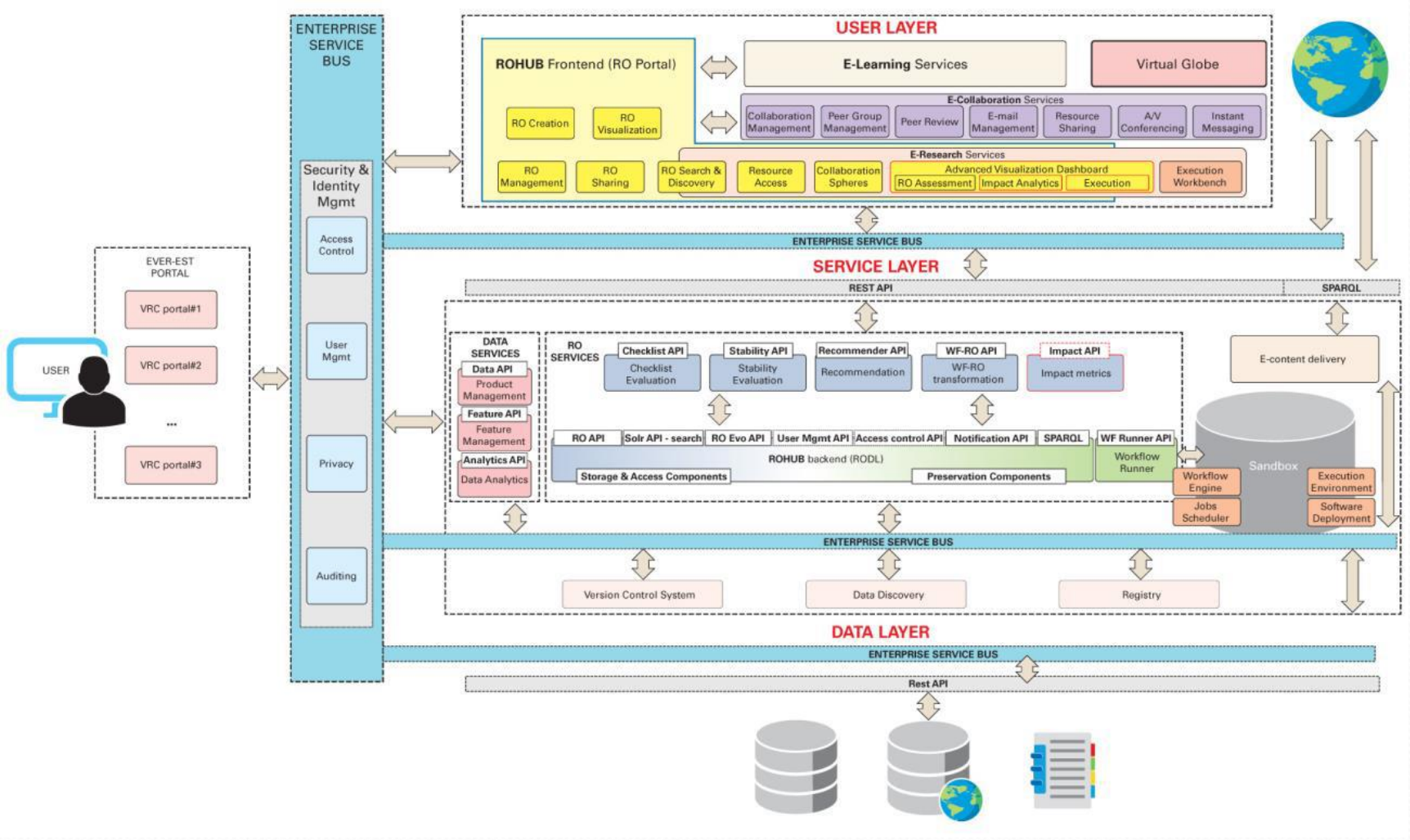
- It needs a solid architecture, ready to be expanded over time
- But: The EUDAT platform is strong and already partly operational
- But: We can learn from existing architectures from other projects

Architecture analysis

EVER-EST

- Discover, assess and process both existing and new heterogeneous Earth Science datasets
- Share data, models, algorithms, scientific results within a community or across communities
- Capture, annotate and store the workflows, processes and results from their research activities
- Work together in a real-time environment that facilitates the sharing of expertise, information and data resources
- Ensure the long-term sustainability and preservation of data, models, workflows, tools and services developed by existing communities of practice

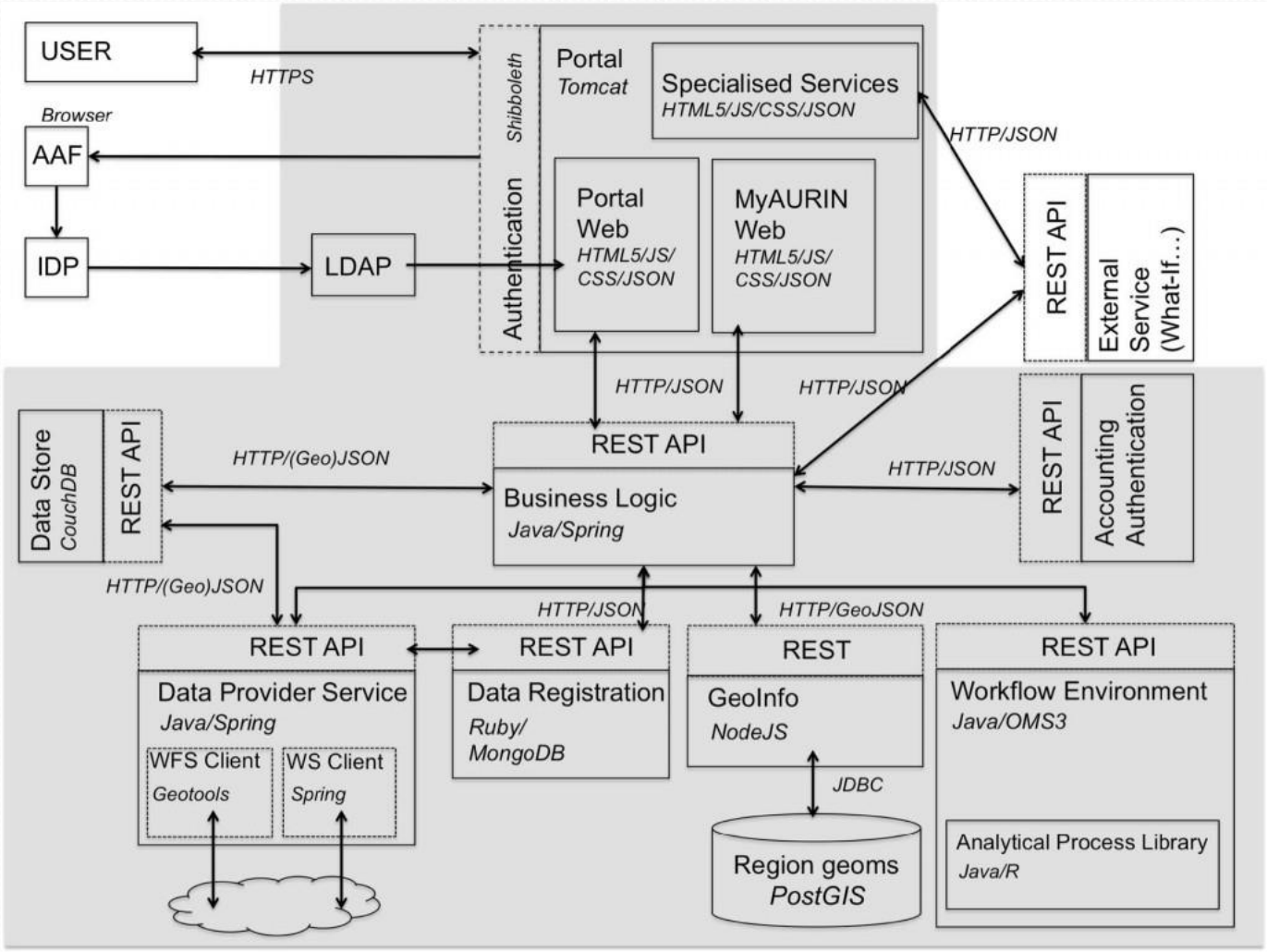
EVER-EST


ENTERPRISE SERVICE BUS
Security & Identity Mgmt
Access Control
User Mgmt
Privacy
Auditing
USER LAYER
ROHUB Frontend (RO Portal)
RO Creation
RO Visualization
RO Management
RO Sharing
RO Search & Discovery
Resource Access
Collaboration Spheres
E-Learning Services
Virtual Globe
E-Collaboration Services
Collaboration Management
Peer Group Management
Peer Review
E-mail Management
Resource Sharing
A/V Conferencing
Instant Messaging
E-Research Services
Advanced Visualization Dashboard
RO Assessment
Impact Analytics
Execution
Execution Workbench
ENTERPRISE SERVICE BUS
SERVICE LAYER
REST API
SPARQL
EVER-EST PORTAL
VRC portal#1
VRC portal#2
VRC portal#3
USER
DATA SERVICES
Data API
Product Management
Feature API
Feature Management
Analytics API
Data Analytics
RO SERVICES
Checklist API
Checklist Evaluation
Stability API
Stability Evaluation
Recommender API
Recommendation
WF-RO API
WF-RO transformation
Impact API
Impact metrics
RO API
Solr API - search
RO Evo API
User Mgmt API
Access control API
Notification API
SPARQL
WF Runner API
ROHUB backend (RODL)
Storage & Access Components
Preservation Components
Workflow Runner
E-content delivery
Sandbox
Workflow Engine
Jobs Scheduler
Execution Environment
Software Deployment
ENTERPRISE SERVICE BUS
Version Control System
Data Discovery
Registry
DATA LAYER
ENTERPRISE SERVICE BUS
Rest API

NECTAR/MARVL

- NecTAR (https://nectar.org.au/) is an Australian cloud infrastructure development with on top of that several VRE projects with a specific theme, e.g. MarVL (Marine Virtual Lab):
  - Configure a range of different community coastal/ ocean and wave models, through a user-friendly web application.
  - Discover and assemble ocean observations from IMOS and AODN in a format that is suitable for model evaluation or data assimilation.
  - Helps marine scientists make better use of ocean observations to improve forecasting and planning for marine and coastal environments.”

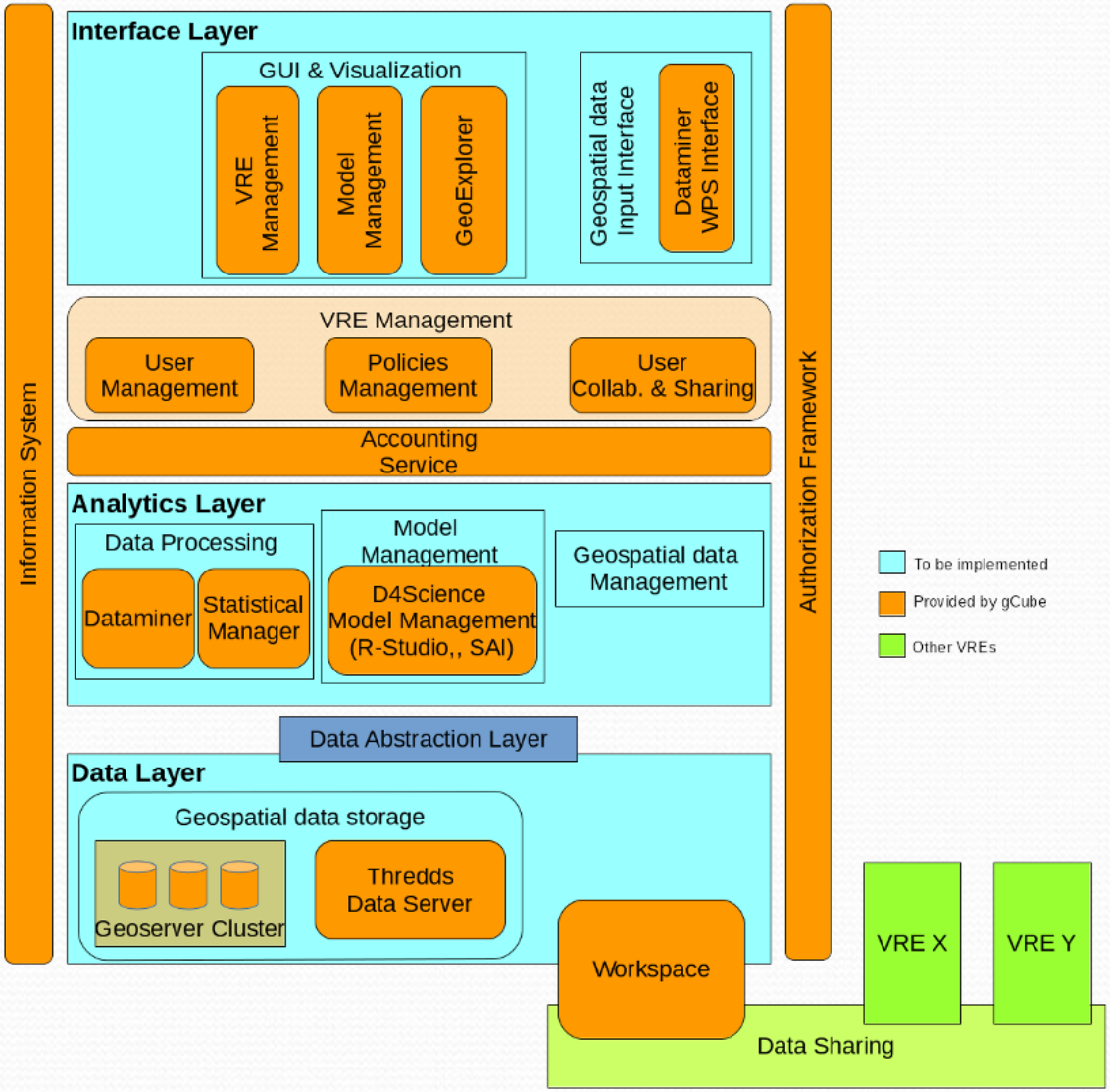
NECTAR / AURIN Architecture



BlueBridge

- BlueBridge (http://www.bluebridge-vres.eu/) is a service provider offering access to VREs. Key features for a Blue Bridge VRE are:
  - Web-based working environment
  - Tailored to serve the needs of a Community of Practice
  - Expected to provide a Community of Practice with the whole array of commodities needed to accomplish the community's goal(s)
  - Open and flexible
  - Promotes fine-grained controlled sharing of both intermediate and final research results
  - Example tools for working on the lab: R Studio to work on R, Data Miner, Species Data

BlueBridge


Information System
Interface Layer
GUI & Visualization
VRE Management
Model Management
GeoExplorer
Geospatial data Input Interface
Dataminer WPS Interface
VRE Management
User Management
Policies Management
User Collab. & Sharing
Accounting Service
Analytics Layer
Data Processing
Dataminer
Statistical Manager
Model Management
D4Science Model Management (R-Studio,, SAI)
Geospatial data Management
Data Abstraction Layer
Data Layer
Geospatial data storage
Geoserver Cluster
Thredds Data Server
Workspace
Data Sharing
VRE X
VRE Y
Authorization Framework
To be implemented
Provided by gCube
Other VREs

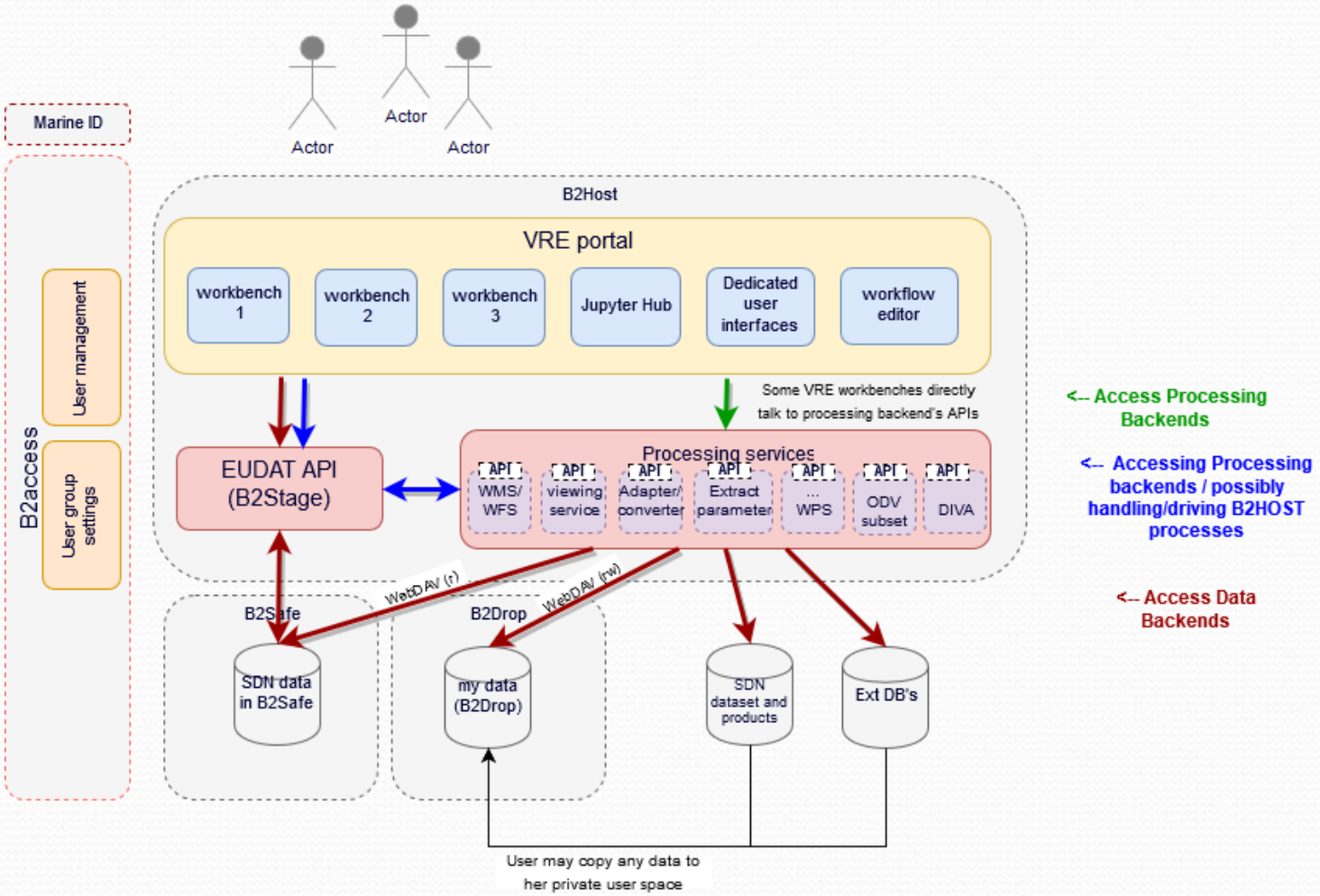
Other VRE's observed

- LifeWatch Marine VRE (using Taverna as Workflow manager)
- Ecopotential VRE (related to GEO/GEOSS)

Conclusions analysis

- 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 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.

SDC VRE: Combining lessons and EUDATs infrastructure (draft)



A skeleton: some technical options

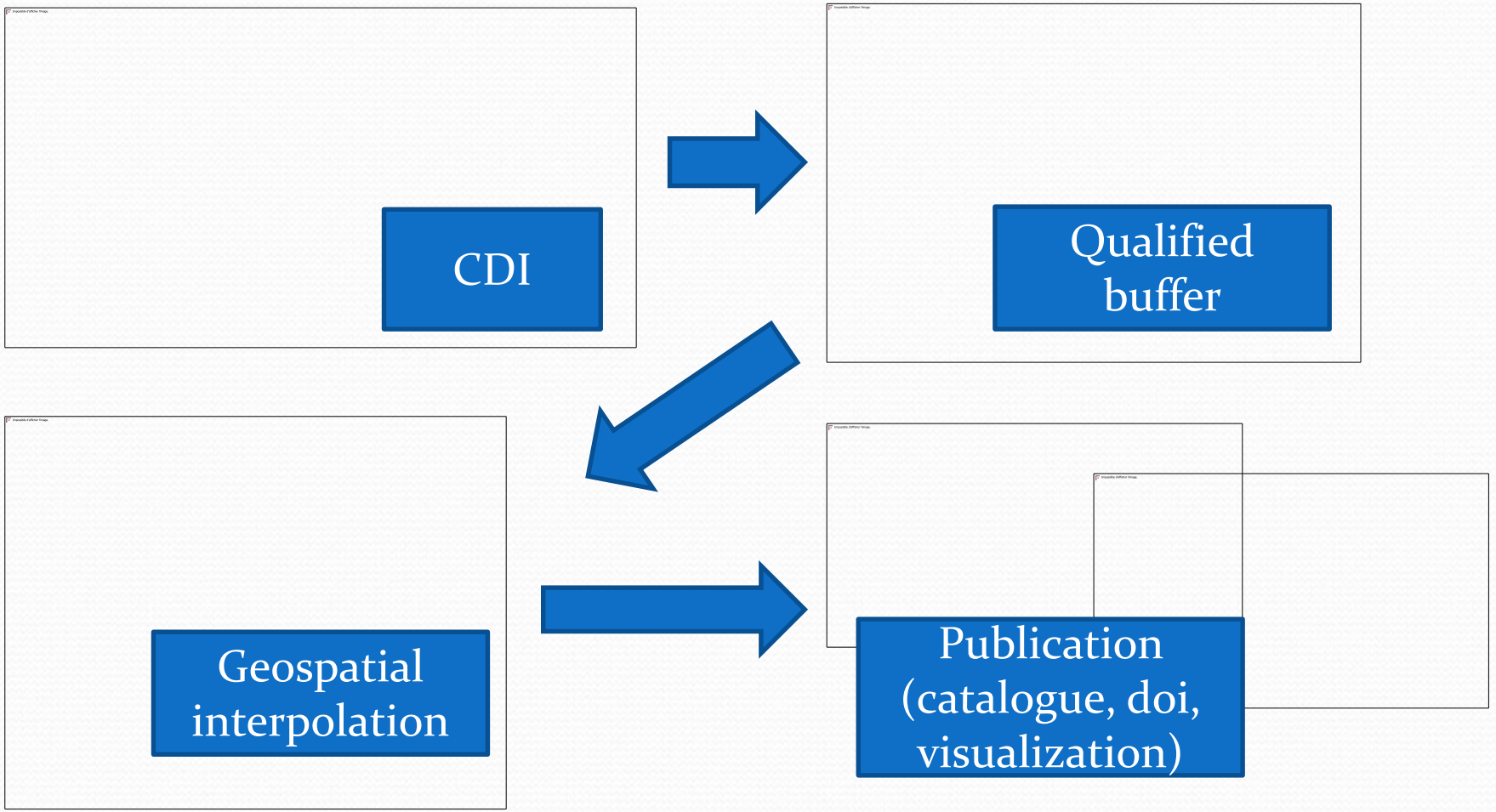
- Authentication: Oauth2 protocol (B2ACCESS + Marine-ID)
- Integrated menu: Application in php + Javascript library
- Private file system, sharing: B2DROP (Owncloud/NextCloud)
- Write, execute code: Jupyter notebook
- Predefined processing: OGC/WPS
- Workbenches, ie applications fit for a specific purpose: web applications deployed with VM or docker, e.g. webODV
- Reference datasets
- Communication (chat, forum…)

A flexible framework for versatile use cases

- **SeaDataNet**, T/S qualification and optimal interpolation, biology statistical control
- **EMODNET-Chemistry**, same for bio-geo-chem
- **EMODNET-Bathymetry**, DTM processing
- **EMODNET-Ingestion**, convert files
- **Marinet2**, Marine Renewable Energies prototype test analysis
- And much more, Research is innovation…

First use case: Workbench T+S

First use case: SeaDataCloud T/S products

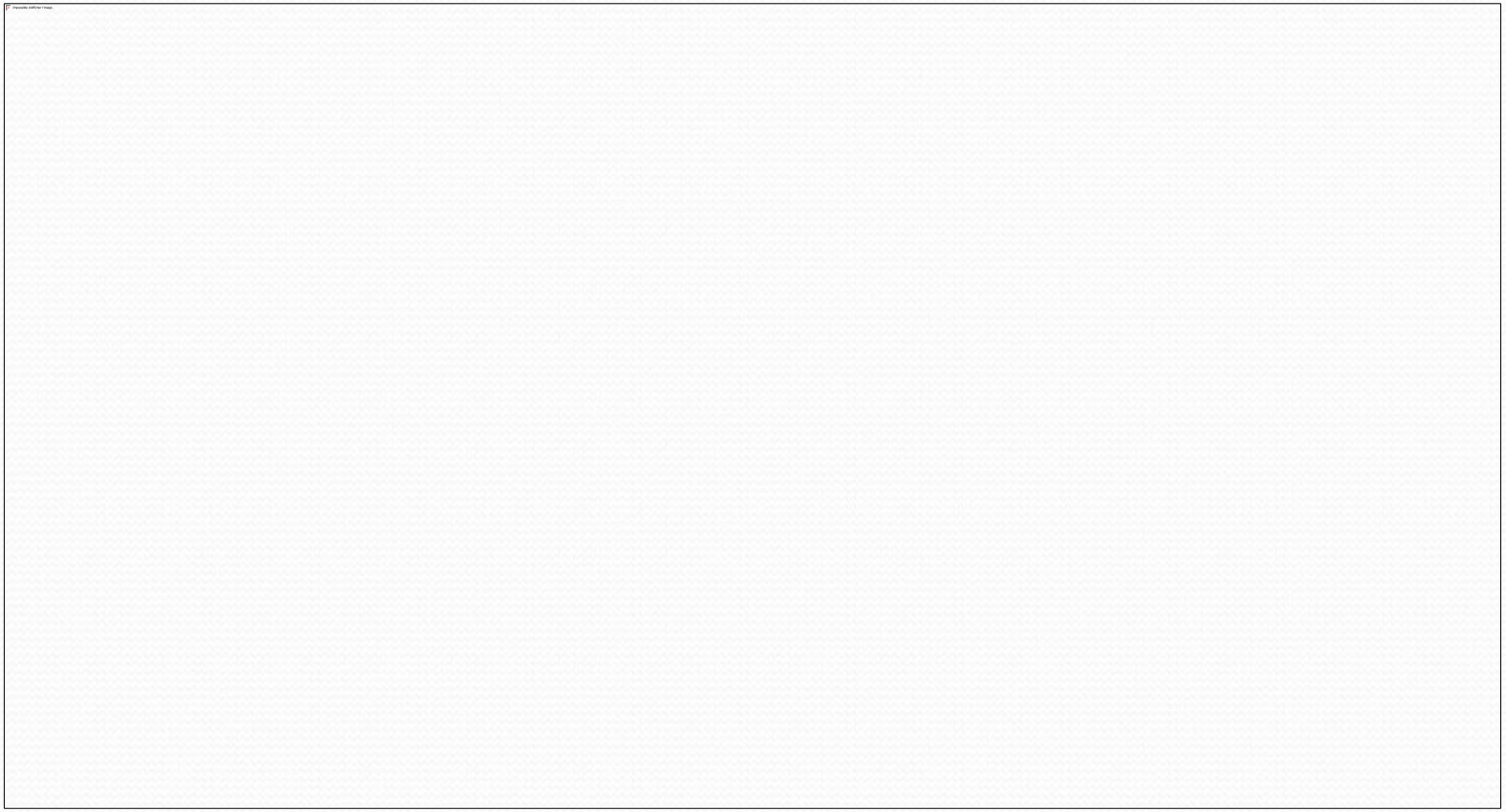


First functions targeted

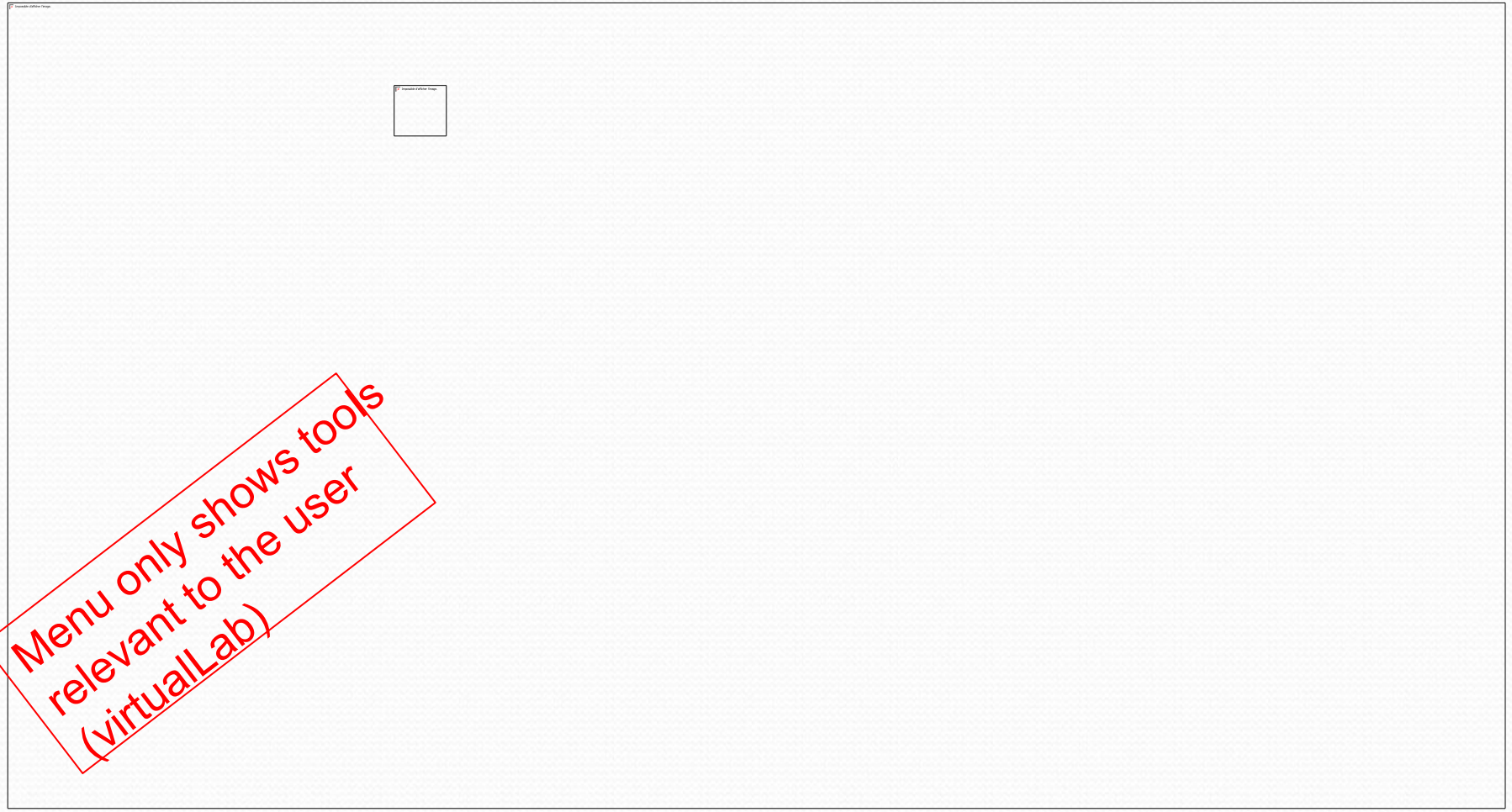
log in with single sign on	B2ACCESS + Marine-ID
integration GUI development	Javascript library
apply water column obs quality control with friendly data editor and save result,	webODV
advise data centre of the regional quality control	
be advised of quality control result (email of log of changes/anomalies 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+oceanotron +sextant-dataCite

First sketches of how this could look

Authentication

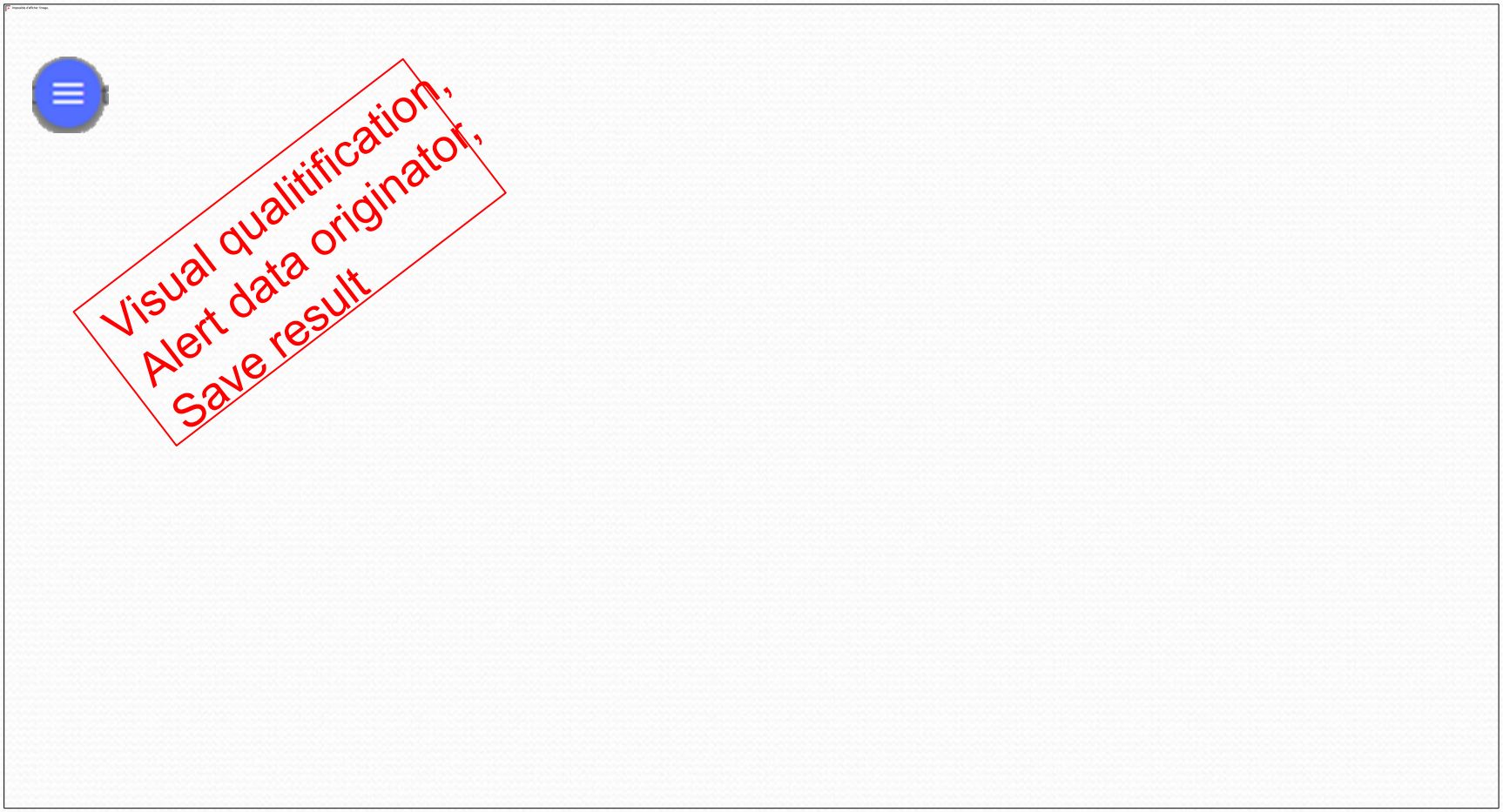


Integrated menu, sliding navigation bar,

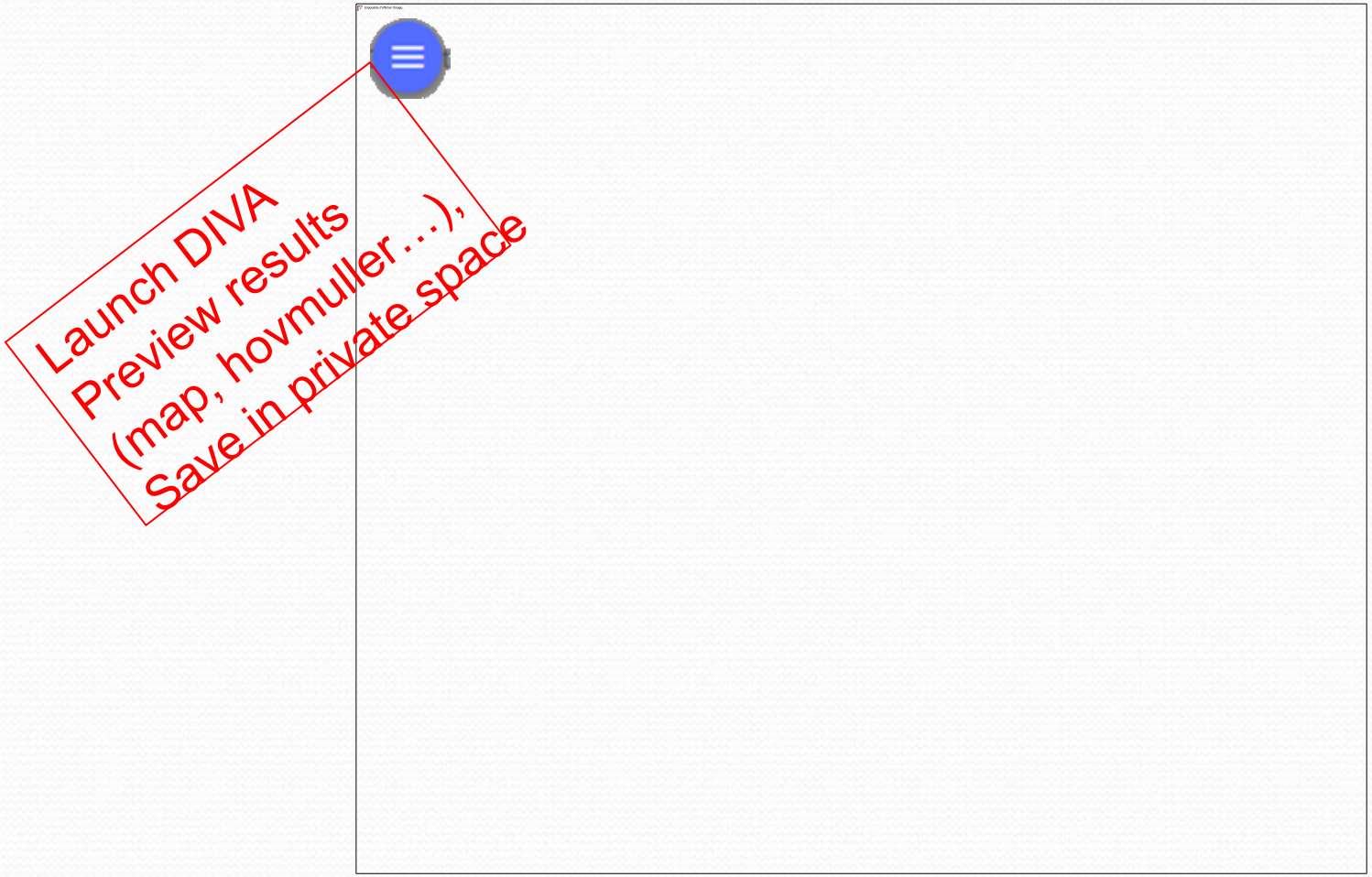


Private, shareable space

Data Visual Qualification

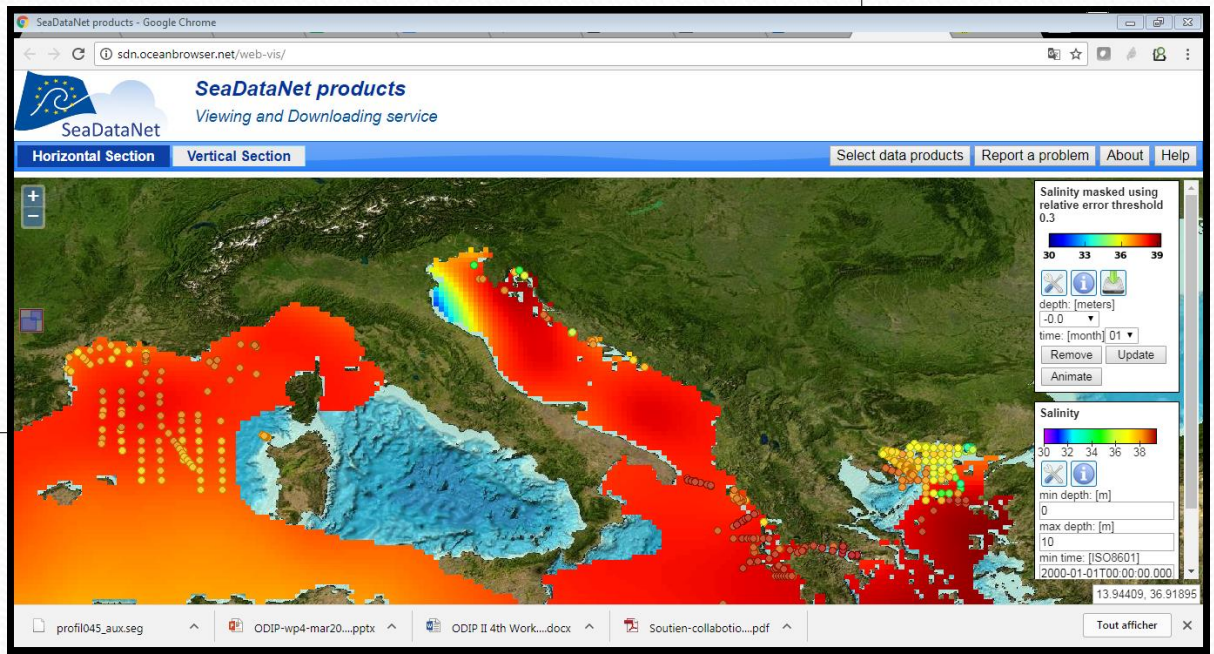


Compute, visualize results, save in private space



Publication

Publish by transfering Metadata and file to data manager



Discussion topics

- Architecture: Which is (most) appropriate and what are experiences from others?
- Workflow engine: Which options exist and experiences?
- How to handle applications/services developed by different partners? Docker? Other solutions?

http://www.odip.org