



SeaDataCloud

Developing the
SeaDataCloud **Virtual Research Environment (VRE)**

Merret Buurman (German Climate Computing Centre, DKRZ),
EUDAT Conference, Porto, Portugal

23 January 2018



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

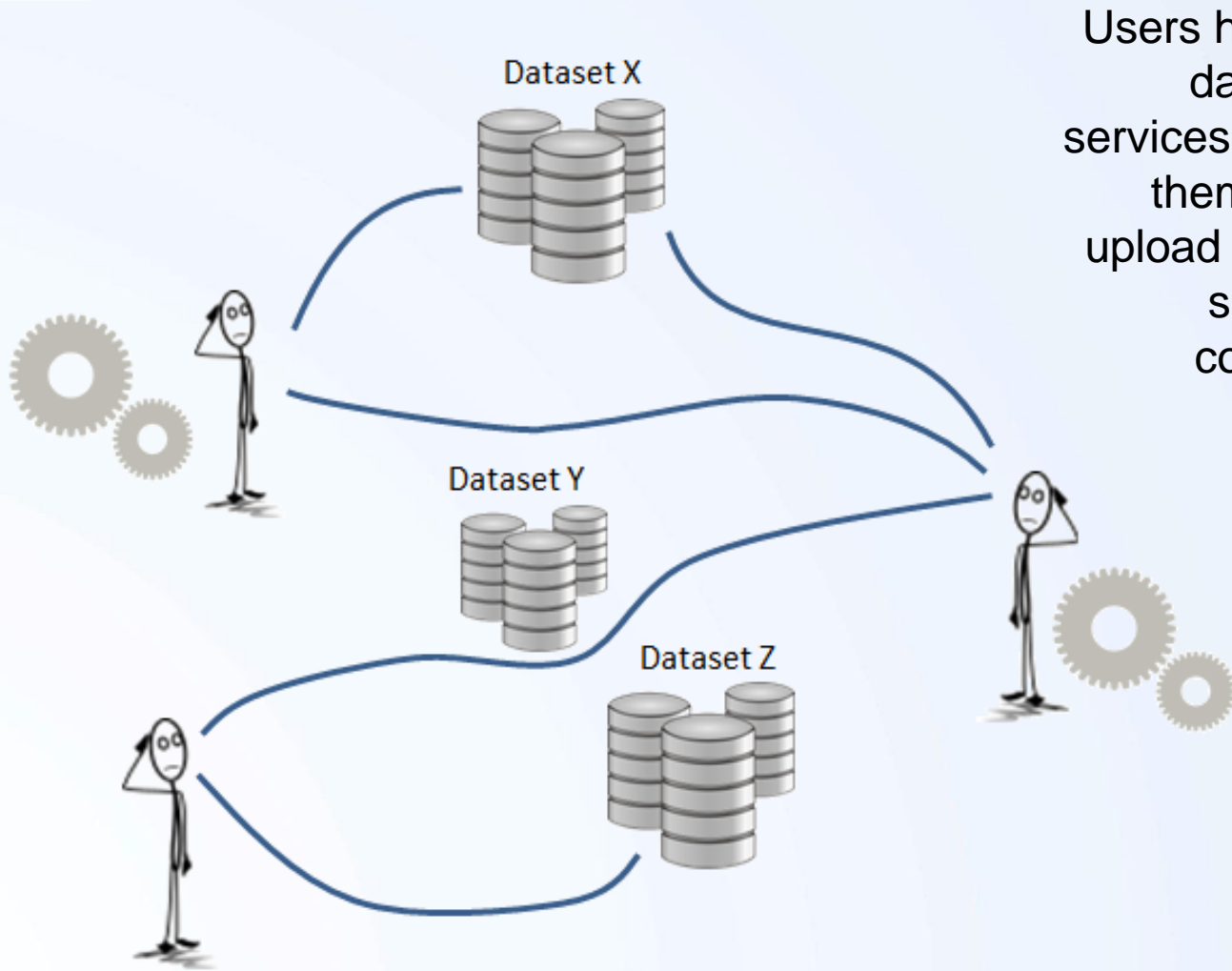
Marine research now...

- **Data** is getting bigger
- **Processing** is getting more resource-intensive
- Cumbersome: Download – Process – Upload
- (Potential) use of **outdated** data & software
- LBNT: Research is getting more **cooperative**

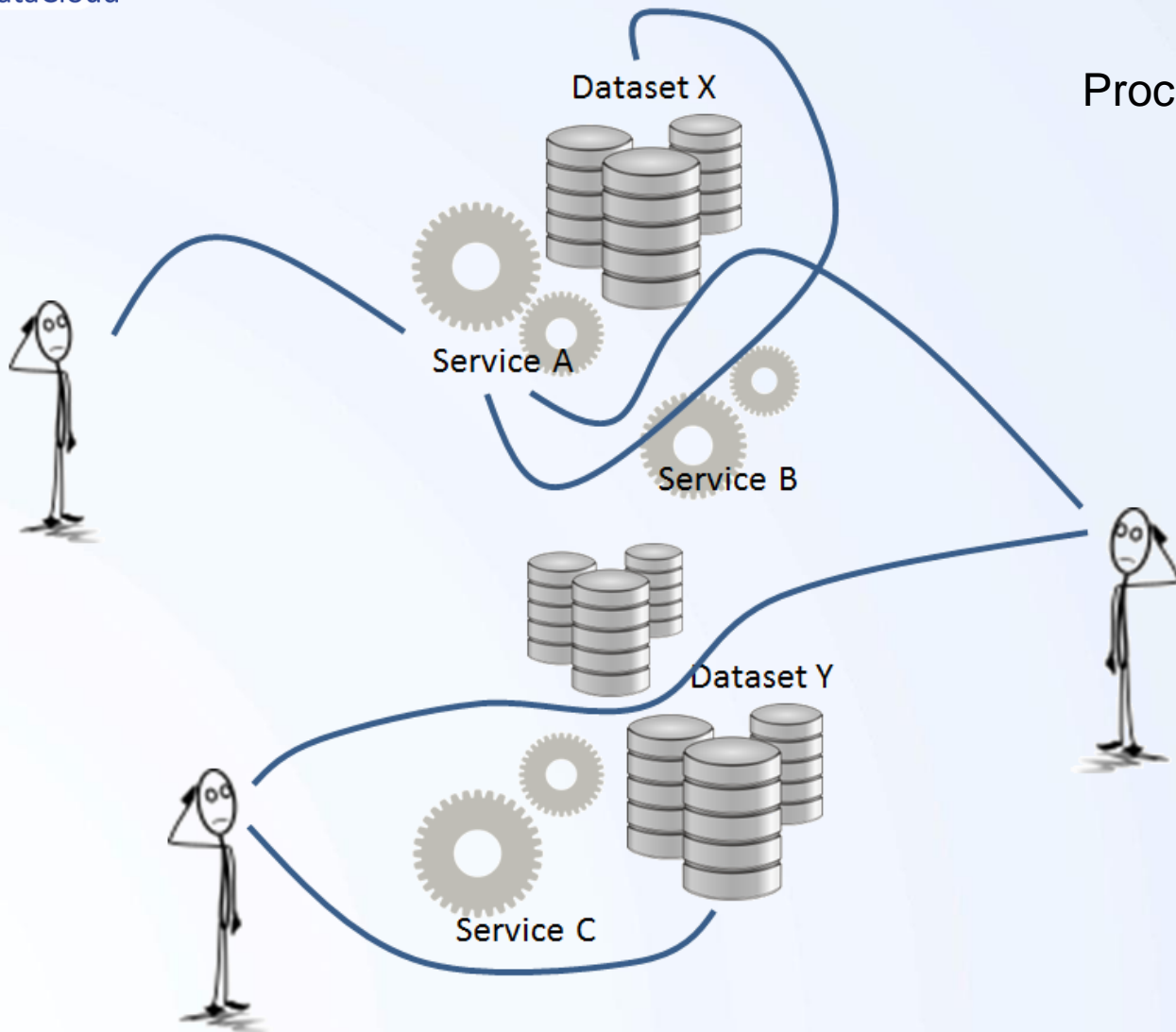
Virtual Research Environment

... a web-based workspace providing seamless access to all services a researcher needs to do her work and collaborate with her community.

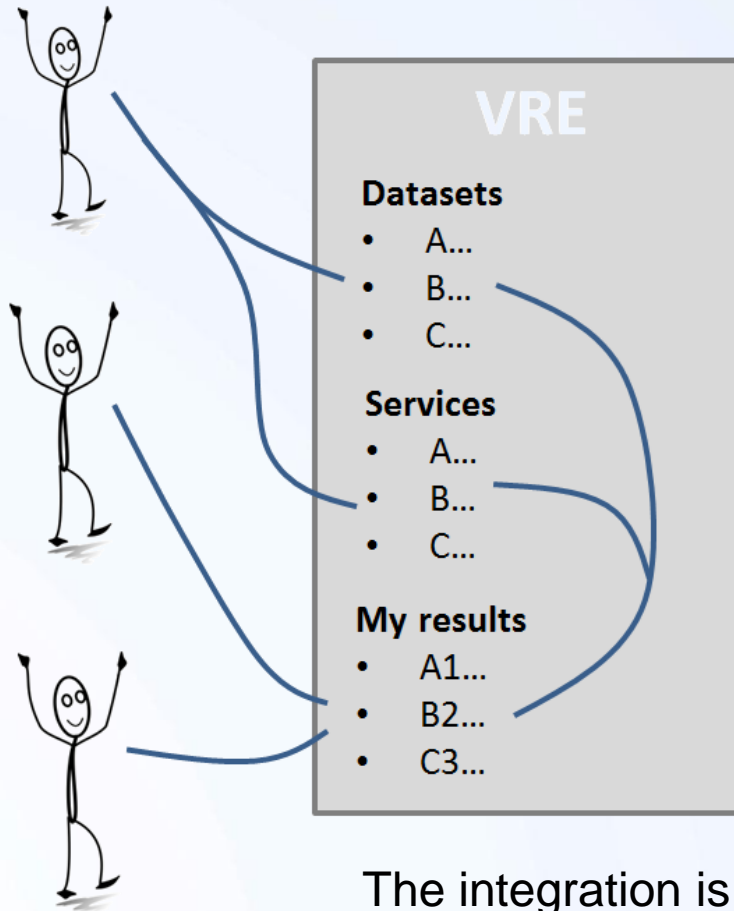
- Finding data
- (Centralized) access to data
- Processing of the data
- Visualisation of data/results
- Sharing of results with colleagues and/or with a wider public



Users have to find datasets and services, download them, process, upload the data to share it with colleagues...

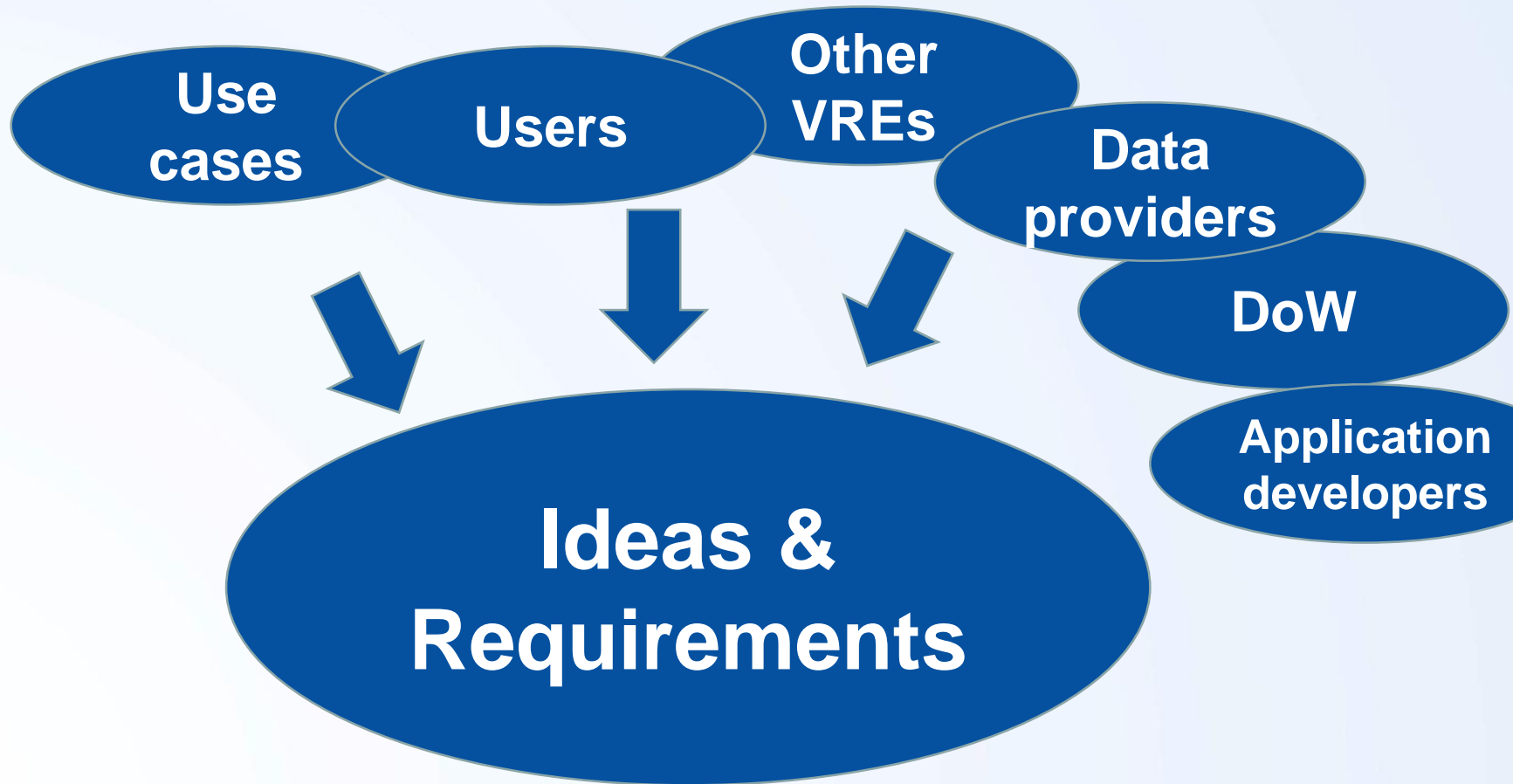


Next step:
Processing in the
cloud..



The integration is done by the VRE...





Use cases

- **SeaDataNet**, T/S quality control and optimal interpolation (climatologies), biology statistical control
- **EMODNET-Chemistry**, same for bio-geo-chemistry
- **EMODNET-Bathymetry**, DTM processing
- **EMODNET-Ingestion**, converting files
- **Marinet2**, Marine Renewable Energies prototype test analysis
- And much more...

Requirements (excerpt)

Processing:

- Functionality that is required to cover the **use cases**
- Access to **JupyterHub** incl. data and services
- **Visualisation** services for results

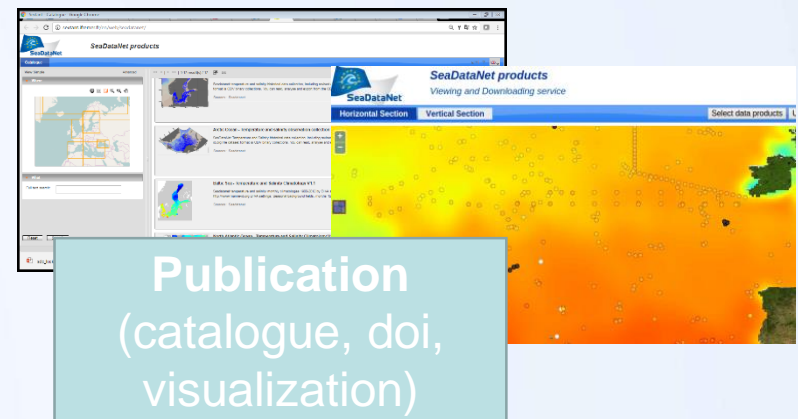
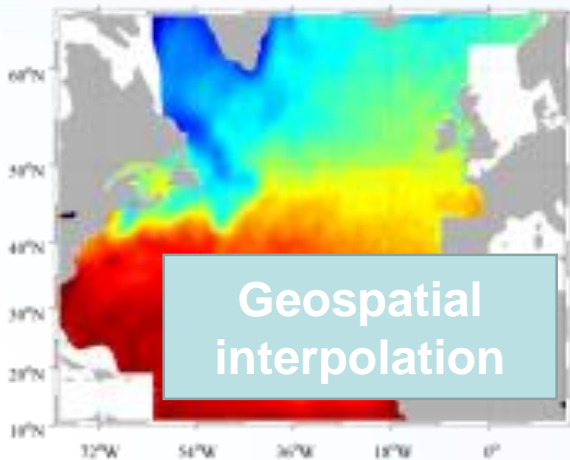
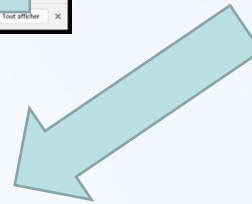
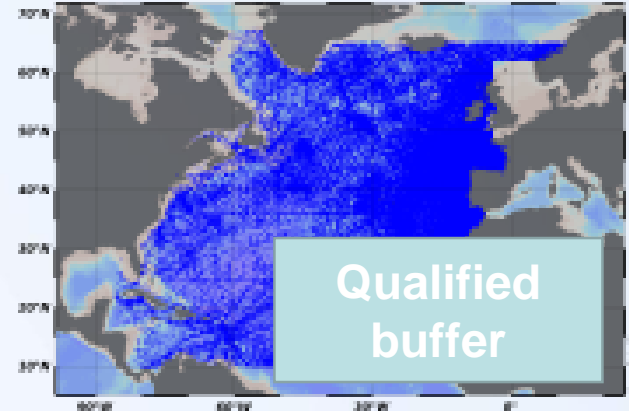
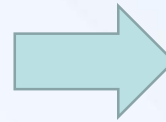
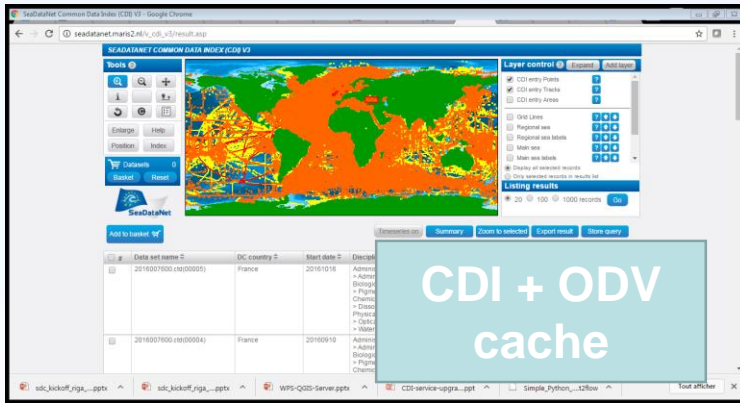
Data:

- Upload **custom input** data
- Access to **CDI data + reference datasets**
- **View, store, download, re-use results**
- **Sharing** of outcomes to public/custom/private
- Respect **privacy** and different data policies! (e.g. restricted data)

Other:

- Different **Virtual Labs** (data pools, tool compositions, custom entry pages...)
- **Single Sign On** with Marine-ID
- **User permission** management
- **Communication** to facilitate collaborative research (forum, tools, ...)
- **Performance** + capacity (etc.)

First use case: SeaDataCloud T/S products





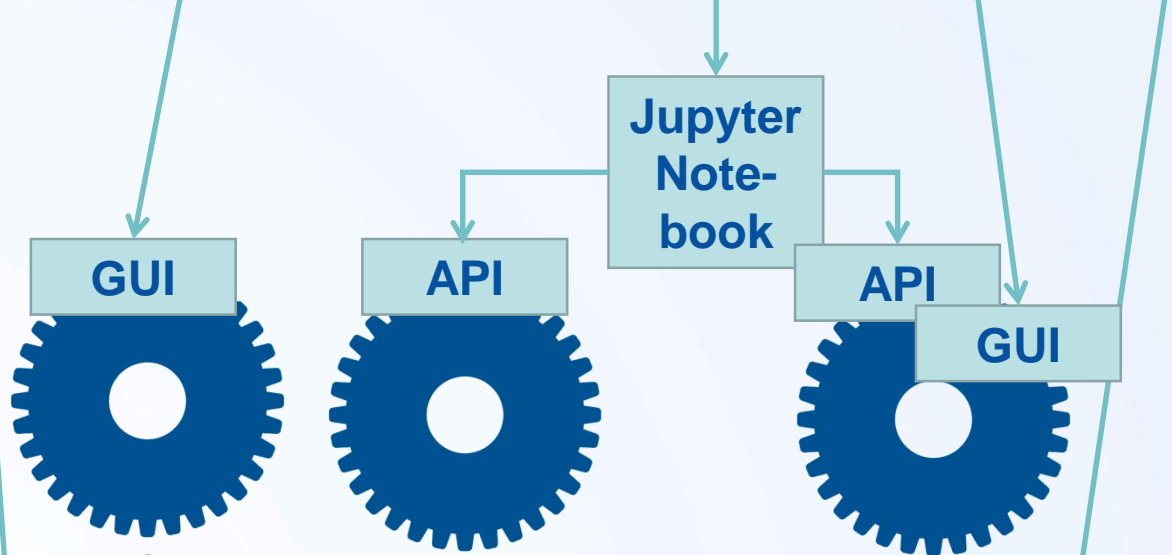
VL1



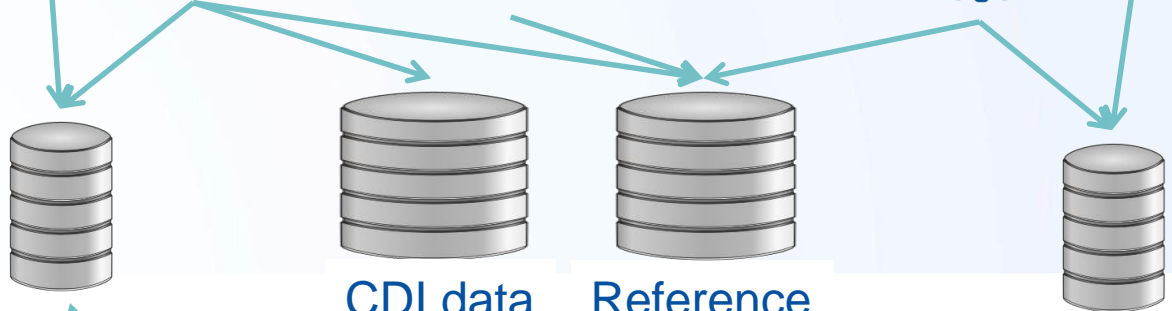
VL2



Frontend



Processing layer



Data access

Data layer

Custom data Alice

CDI data

Reference datasets

Custom data Bob



VL1



VL2

Frontend (+Single Sign On)

Frontend

Bla?

Bla!

Bing!

GUI

API

Jupyter notebook

API

GUI

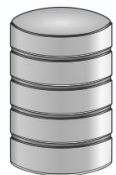
Workflows

layer

Data access

Data layer

Logs, Acc

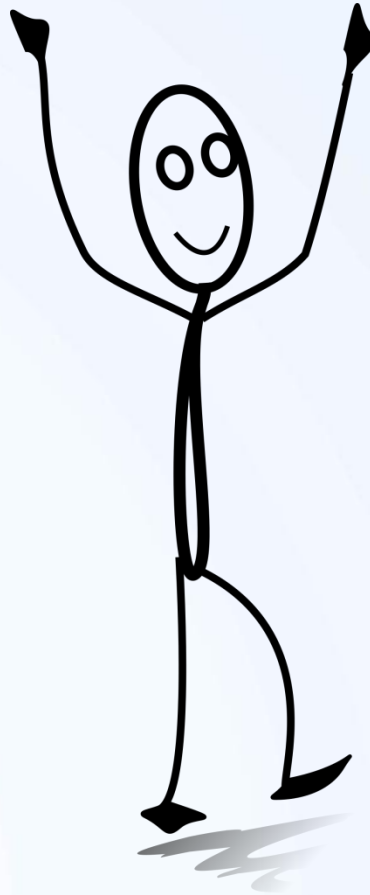


Custom data Alice

CDI data

Reference datasets

Custom data Bob



Quite a challenge!

- Complex requirements
- High ambition
- Team members/developers from very varied backgrounds
- Very diverse experiences and skills to build on
- Gaining more and more insights into each other's domains and concerns
- Use of existing EUDAT services
+ plus customization and addition of features

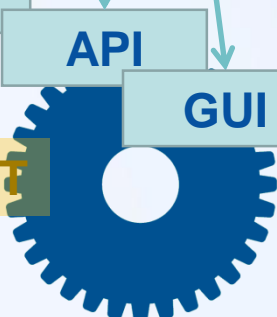
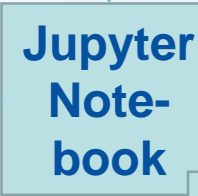
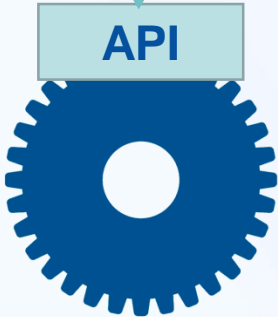
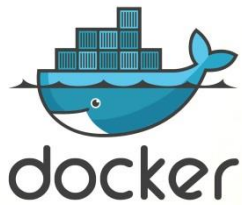


VL1

VL2



Frontend



B2HOST

Processing layer



Data access

Data layer

Custom data Alice

CDI data

Reference datasets

Custom data Bob

B2DROP

B2SAFE

B2DROP

Frontend

(+Single Sign B2ACCESS)

GUI

API

API

GUI

B2DROP

B2SAFE

B2DROP

Development approach...

- Highly distributed teams (spatially and thematically)
- Small steps:
 - Four month cycles with concrete goals
 - Followed by coding sprints

Development Status & Outlook

Very early state!

- First start with few applications (first use case)
- Integrate more and more
- Later: Add features (such as communication means, version update notifications, complex workflows, ...)

→ Technical Kick-Off next week in Delft!

Thank you! *Obrigada!*

(Time for questions...)

Requirements (excerpt)

Processing:

- Functionality that is required to cover the **use cases**
- Access to **JupyterHub** incl. data and services
- **Visualisation** services for results

Data:

- Upload **custom input** data
- Access to **CDI data + reference datasets**
- **View, store, download, re-use results**
- **Sharing** of outcomes to public/custom/private
- Respect **privacy** and different data policies! (e.g. restricted data)

Other:

- Different **Virtual Labs** (data pools, tool compositions, custom entry pages...)
- **Single Sign On** with Marine-ID
- **User permission** management
- **Communication** to facilitate collaborative research (forum, tools, ...)
- **Performance** + capacity (etc.)