Data Analysis made easy with the ENES Climate Analytics Service (ECAS)

Sofiane Bendoukha, Donatello Elia, Fabrizio Antonio, Sandro Fiore, Tobias Weigel, Alessandro D'Anca

April 10, 2019 @ EGU 19
SC1.22

eosc-hub.eu

@EOSC_eu

Dissemination level: Public

EOSC-hub receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 777536.

Useful links

- Training materials
  - https://github.com/ECAS-Lab/ecas-training
- ECASLab / JupyterHub
  - **ECASLab @ DKRZ** https://ecaslab.dkrz.de
  - **ECASLab @ CMCC** https://ecaslab.cmcc.it
- Ophidia framework documentation
  - http://ophidia.cmcc.it/documentation/users/index.html

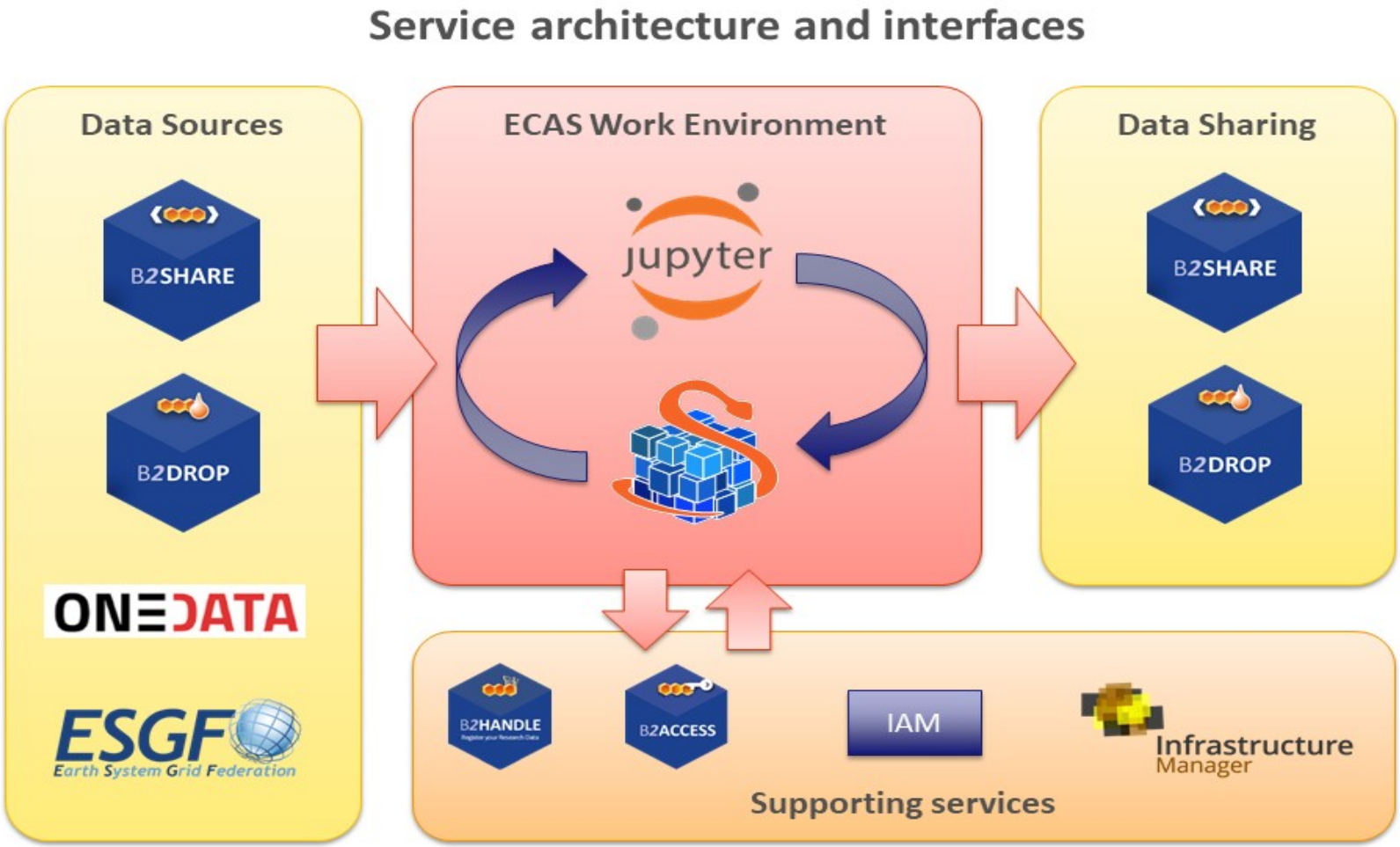
ENES Climate Analytics Service

- ECAS is part of the **EOSC-HUB** service catalogue
  - ECAS enables scientific end-users to perform data analysis experiments
- Server-based
  - Computation @ **CMCC** or **DKRZ**
  - Avoid data transfer (download)
  - Improved reusability of data and workflows (FAIR approach)
- ECAS supports different Auth* providers
  - Local and external AAI providers supported (LDAP, B2ACCESS, EGI Check In)
  - Additional AAI providers can be integrated (e.g. INDIGO IAM)

Data overview

- ECAS provides data access via ESGF
- Coordinated Regional Climate Downscaling Experiment
  - ~ 100 Tbyte Cordex
- Coupled Model Intercomparison Project 5
  - ~ 1.2 Pbyte CMIP5 Data
- Coupled Model Intercomparison Project 6
  - ~ 250 Tbyte CMIP6 Data from the 1PByte published
- Other Data pools can be mounted on demand
  - MPI Grand ensemble (**MPI - GE**)
  - data collection exposed in the Federated Data Archive (e.g. through **OneData**)

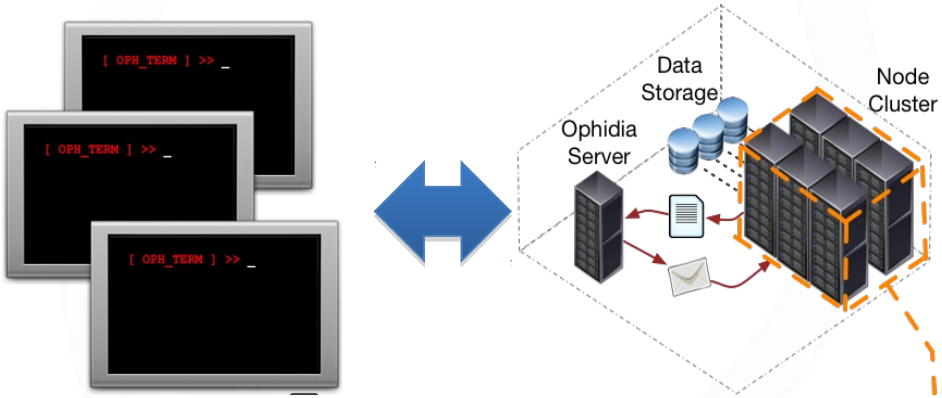
ECAS in a Nutshell

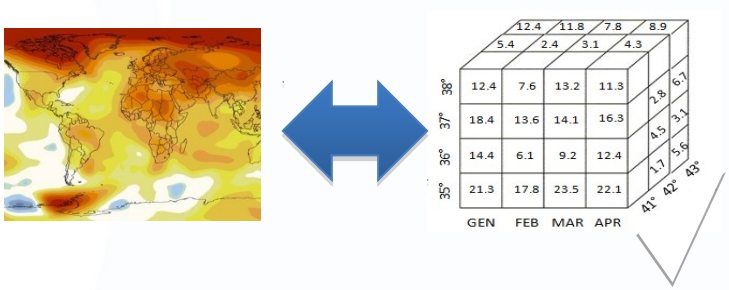


Ophidia framework overview

- The Ophidia framework addresses big data challenges for eScience
  - support for declarative, parallel, server-side data analysis exploiting parallel computing techniques
  - end-to-end mechanisms to support complex experiments and large processing workflows on scientific multi-dimensional datacubes
- Ophidia supports both **batch** & **interactive** data analytics
  - More than 50 datacube-oriented **operators** are available, including: data reduction and subsetting, data intercomparison, metadata and provenance management, time series analysis with array-based primitives
  - A wide set of (low-level) array-based primitives (over 100) to perform, e.g. data summarization, algebraic expressions, predicates evaluation, statistical analysis
  - Support for complex workflows and Python applications execution

Server-side paradigm and datacube abstraction in Ophidia



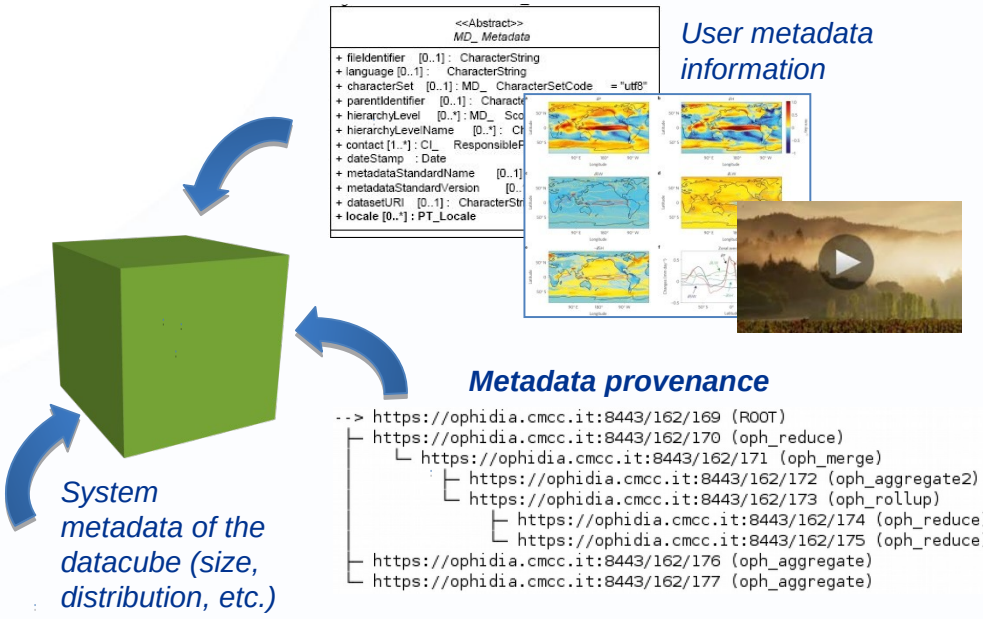


Oph_Term: a terminal-like commands interpreter serving as a client for the Ophidia framework

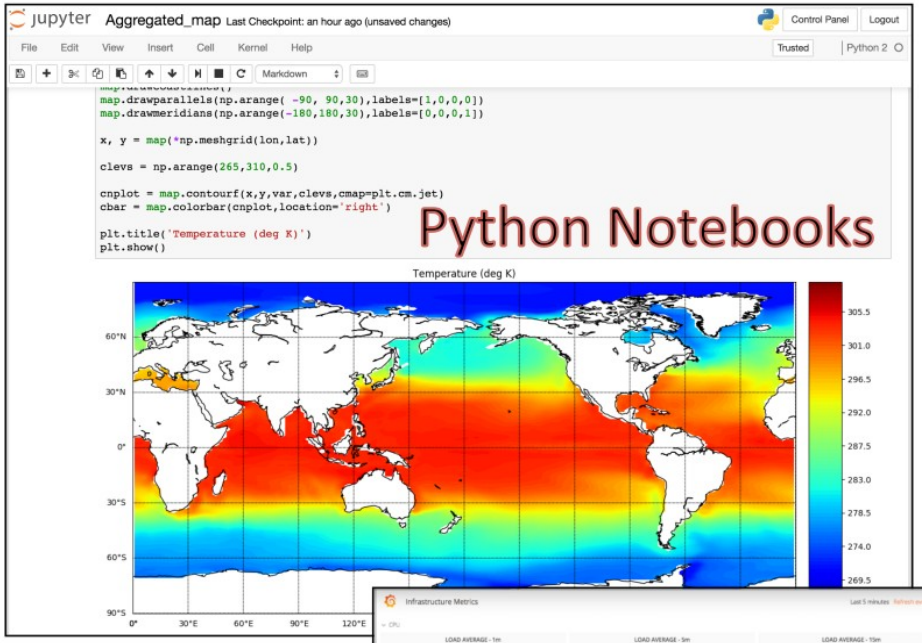
PyOphidia: a Python interface for datacube management & analytics with Ophidia

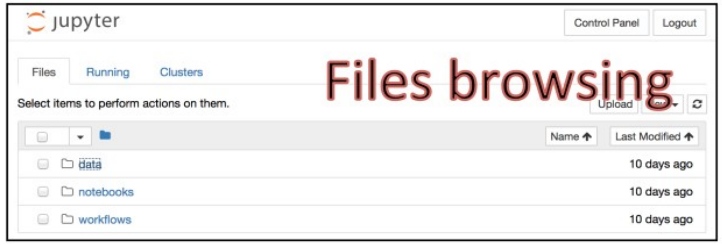
Ophidia framework: declarative, parallel server-side processing

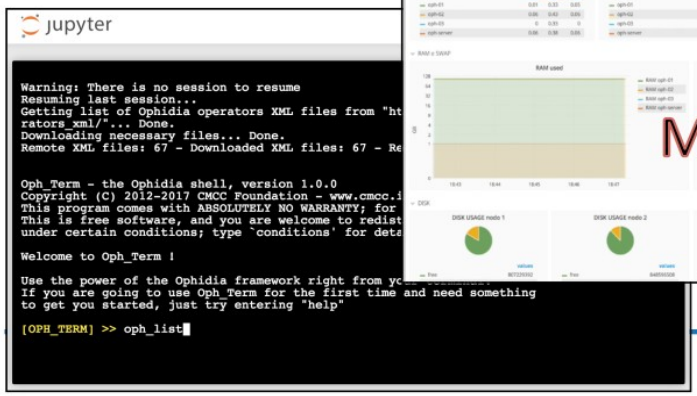
Through **oph_term/PyOphidia** the user run ("send") commands ("operators") to the Ophidia framework to manipulate datasets ("datacubes")

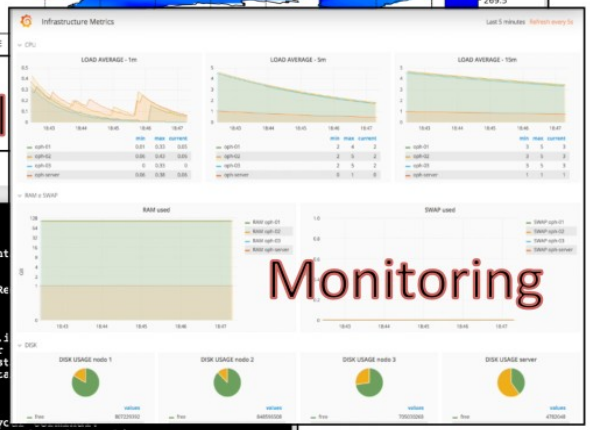


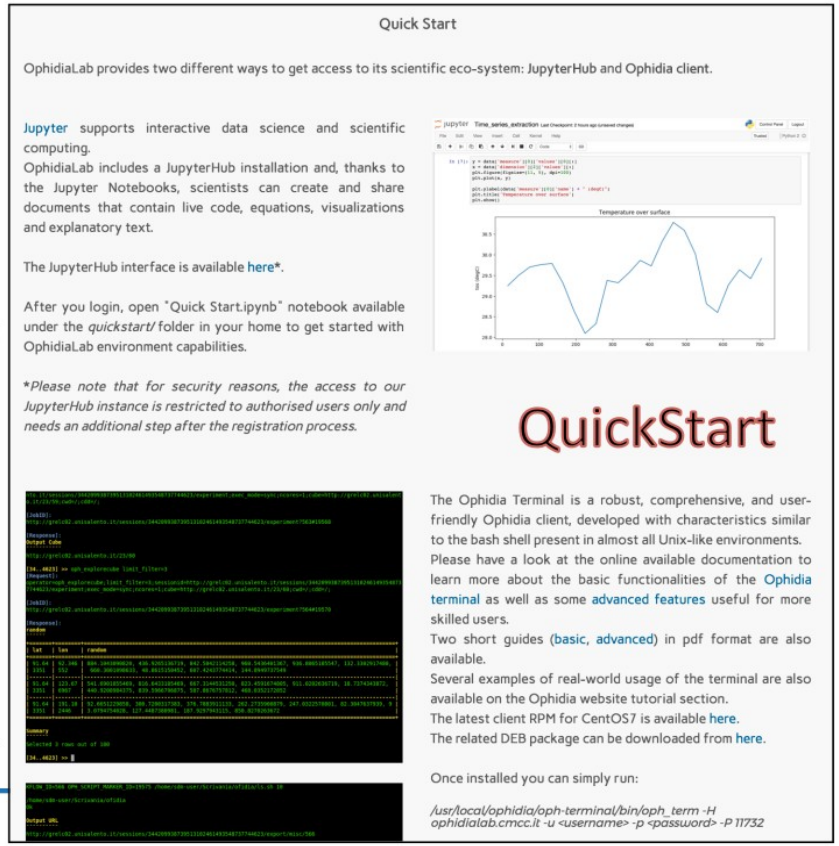
ECASLab











Source: Sandro Fiore

Thank you for your attention!

Questions?

Sofiane Bendoukha bendoukha@dkrz.de
Donatello Elia donatello.elia@cmcc.it
Fabrizio Antonio fabrizio.antonio@cmcc.it

EOSC-hub

eosc-hub.eu @EOSC_eu

Useful links

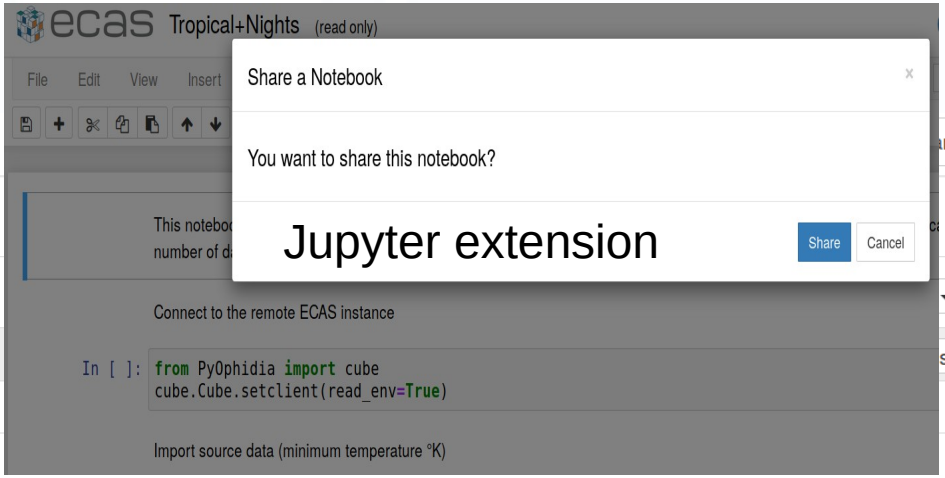
- Training materials
  - https://github.com/ECAS-Lab/ecas-training
- ECASLab / JupyterHUb
  - **ECASLab @ DKRZ** https://ecaslab.dkrz.de
  - **ECASLab @ CMCC** https://ecaslab.cmcc.it
- Ophidia framework documentation
  - http://ophidia.cmcc.it/documentation/users/index.html

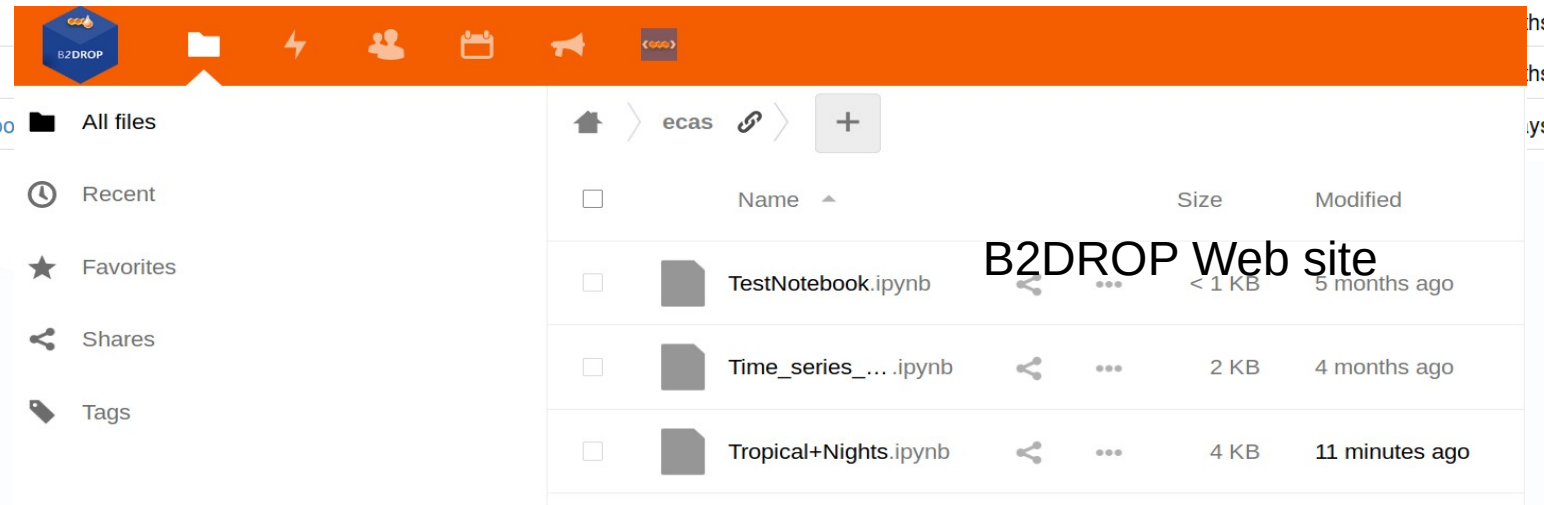
Data sharing with ECAS

- Share your results with your team or with researchers from a broader community
  - Post-processing Datasets
  - Notebooks
- Cloud-based storage services integrated within ECAS
  - **B2DROP** online storage for external colloborators as well as for collaborators who uses B2DROP too. Keep data synchronized and up-to-date.
  - **B2SHARE** store and publish research data from diverse contexts. Data assigned a **Persistent Identifier** for better findability.
  - **DataHub (OneData)** a global data access solution for eScience.

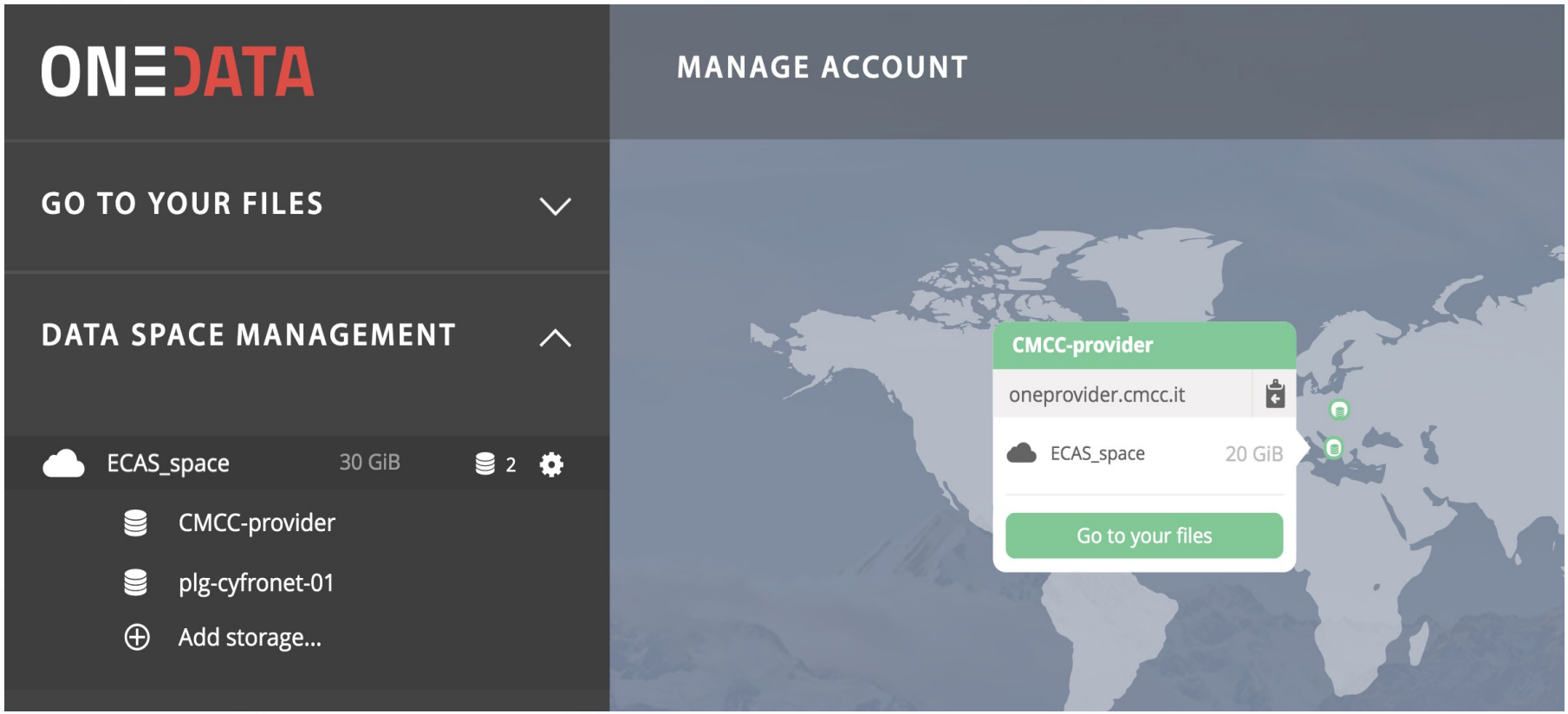
B2DROP shared

https://b2drop.eudat.eu/s/gDyJjMeJ2Xiapwi





OneData &ECAS



Check the notebook here for more details