

Objectives/Benefits

The **ENES Climate Analytics Service (ECAS)** will enable scientific end-users from climate data community and other disciplines to perform data analysis experiments on large volumes of climate data. It follows a **PID-enabled, server-side** and **parallel** approach.

Benefits:

- ▶ Reduce the need for local data downloads
- ▶ Reduce the effort of maintaining client-side tools
- ▶ Reduce the complexity of workflows for simple users
- ▶ Encourage flexible and open data sharing
- ▶ Enable PID-based provenance support
- ▶ More efficient use of computing resources

Integration with EUDAT Service Suite

The ECAS relies on services that are part of the EUDAT service catalogue (B2HANDLE, B2ACCESS, B2DROP and B2SHARE)

- ▶ **B2DROP:** Through B2DROP, ECAS provides a secure and trusted data exchange service to store and exchange processed data outputs
- ▶ **B2SHARE:** ECAS offer users to receive, store and publish processed data outputs
- ▶ **B2HANDLE:** Through the B2HANDLE service, ECAS assigns Persistent Identifiers (PID) to Ophidia output and records kernel information according to a community profile
- ▶ **B2ACCESS:** As a common AAI solution^a

^aThe EGI AAI Check-in Service is also a possible candidate for this purpose.

Data Provenance Management

The integrated data analytics service will enable basic data provenance tracking by exploiting PID capabilities through the whole processing chain. It relies on B2HANDLE operations to cover the following aspects:

- ▶ manage the identifier namespaces (Handle prefixes)
- ▶ operate Handle servers and technical services
- ▶ provide a user-friendly Python library for general interaction with Handle servers

Interaction with the Research Data Alliance (RDA)



ECAS will showcase the application and interaction between multiple relevant RDA efforts:

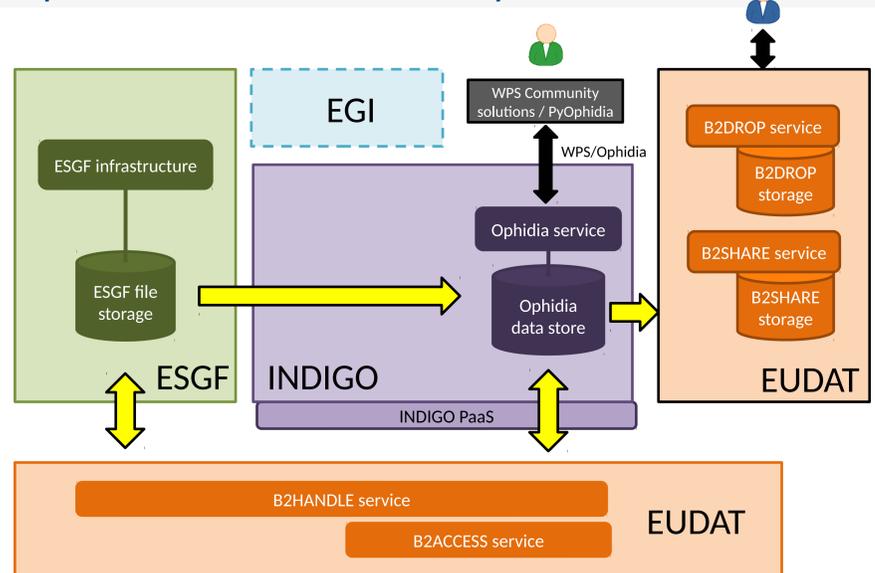
- ▶ ECAS employs multiple components whose roles are expressed under a Data Fabric architecture to deliver a modular, integrated data service
- ▶ ECAS provides a first example of comprehensive usage of PID Kernel Information for provenance tracking
- ▶ The server-side processing capabilities of ECAS can become a cornerstone to achieve further automation, driven by persistently identified digital objects with registered data types and associated services

References

- ▶ **ESGF**, <https://esgf.llnl.gov/>.
- ▶ **EUDAT Service Suite**, <https://sp.eudat.eu/ui/catalogue/services/>.
- ▶ **Ophidia**, ophidia.cmcc.it.
- ▶ **INDIGO-DataCloud**, <https://www.indigo-datacloud.eu/>.
- ▶ **European Network for Earth System Modelling (ENES)**, <https://verc.enes.org/>.
- ▶ S. Fiore, A. D'Anca, D. Elia, C. Palazzo, D. N. Williams, I. T. Foster, and G. Aloisio. Ophidia: A full software stack for scientific data analytics. In *International Conference on High Performance Computing & Simulation, HPCS 2014, Bologna, Italy, 21-25 July, 2014*, pages 343–350. IEEE, 2014.

Service Architecture

Ophidia / ENES Climate Analytics Service

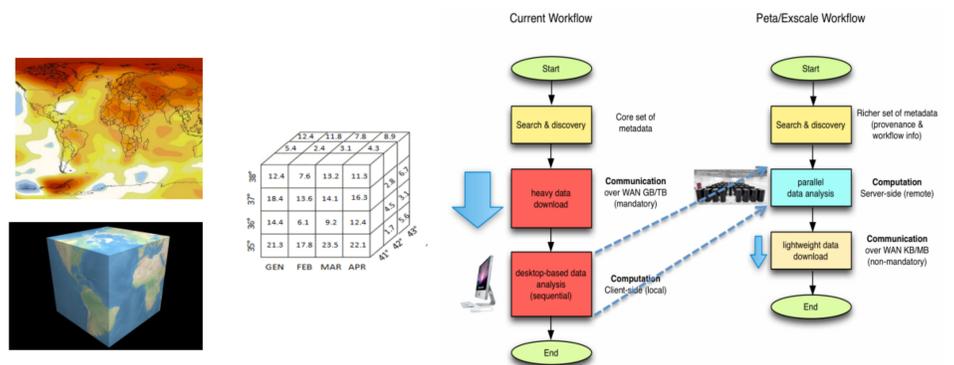


Key integration milestones:

- ▶ Enhance ENES/ESGF PID management services to be fully integrated into the EUDAT B2HANDLE service concept
- ▶ Provide/enable connectors to allow data transfer between ESGF nodes and Ophidia instances
- ▶ Integration of Ophidia with EUDAT B2HANDLE
- ▶ Make Ophidia final output available through EUDAT B2DROP and B2SHARE
- ▶ Setup and configuration of workflow repository for publishing and sharing workflows

Data Analytics Framework

Ophidia is a research effort carried out at the Euro Mediterranean Centre on Climate Change (CMCC) to address big data challenges, issues and requirements for climate change data analytics.



- ▶ Server-side, parallel and distributed
- ▶ Extensible
- ▶ Client application available as user interface
- ▶ Workflow support

Service Components

Service components enabling ECAS are:

- ▶ **Ophidia** framework: provides declarative, server-side, and parallel data analysis, jointly with an internal storage model able to efficiently deal with multidimensional data and a hierarchical data organization to manage large data volumes.
- ▶ **ENES data infrastructure (Earth System Grid Federation (ESGF) software stack)**: supports the European contribution to coordinated numerical experiments like CMIP
- ▶ **EUDAT service suite**: B2HANDLE, B2ACCESS, B2DROP and B2SHARE
- ▶ **EGI and INDIGO-DataCloud**: plug-ins (extensions) like FutureGateways for scientific workflow systems

