Objective 3: exploitation of Earth system models

General talk: Involvement of ENES in EOSC and other data projects

DKRZ: Stephan Kindermann
CMCC: Sandro Fiore
CERFACS: Christian Pagé
CNRS-IPSL: Sebastein Denvil

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084".

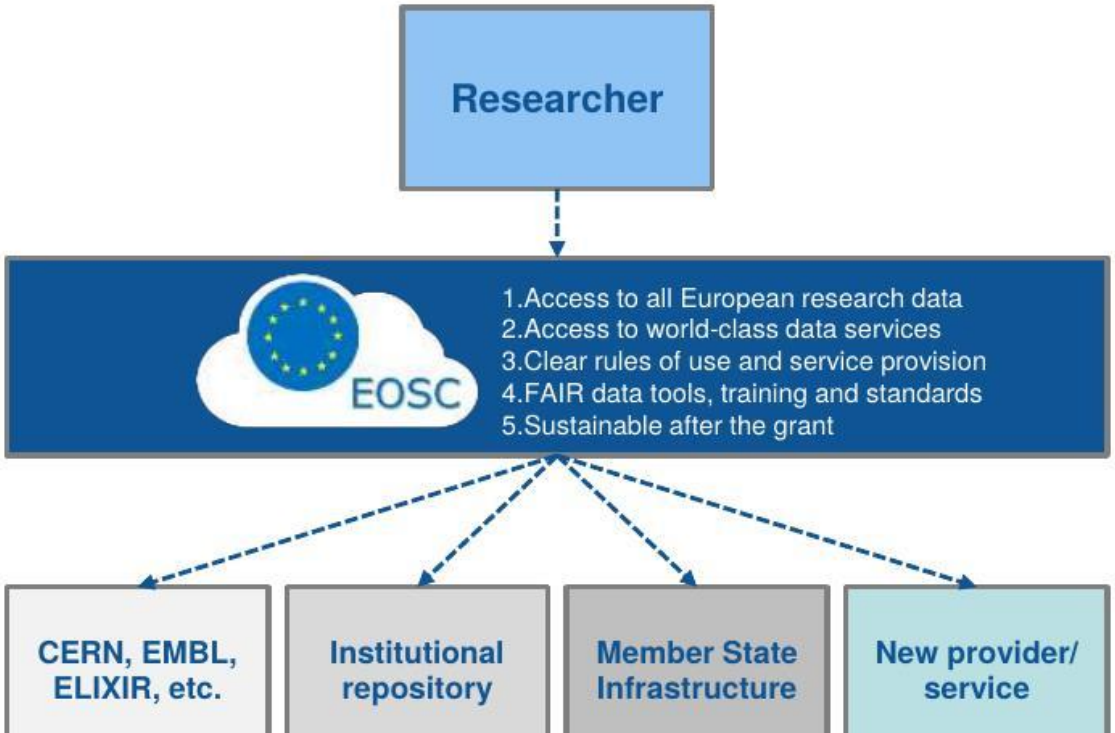
Overview

- EOSC: overview and main objectives
- EOSC involvement:
  - EOSC-hub
  - EOSC thematic service: ECAS
  - EOSCpilot
- Other data projects:
  - SeaDataNet
  - DARE
- New EOSC related project proposals

European Open Science Cloud (EOSC)

Notably:

- Not only European
- Not fully open
- Not just for Science
- Not a cloud



- Easy access through a universal access point for ALL European researchers
- Cross-disciplinary access to data unleashes potential of interdisciplinary research
- Services and data are interoperable (FAIR data)
- Data funded with public money is in principle open (as open as possible, as closed as necessary)
- EOSC will help increase recognition of data intensive research and data science

EOSC: many infrastructures to federate and varying levels of activities and infrastructures among Member States

Source: [eosc strategic implementation roadmap]

EOSC-hub

- EOSC-hub mobilises providers from 20 major digital infrastructures, EGI, EUDAT-CDI and INDIGo-DataCloud jointly offering services, software and data for advanced data-driven research and innovation.
  - 100 Partners, 76 beneficiaries (75 funded)
  - ENES partners: DKRZ, CMCC (CNRS, STFC,..)
  - ~30.000.000 EUR funding
  - 36 months, starting in 2018
- What does EOSC-hub provide ?
  - Provides an EOSC service integrator and federator
  - Catalog of resources and services
  - Develops the know-how and prototype procurement and purchasing framework to acquire digital services from publicly funded infrastructures and commercial providers
  - Corpus of policies, processes and federation services
  - Quality assurance reviews
  - Competences centres and joint digital innovation hub

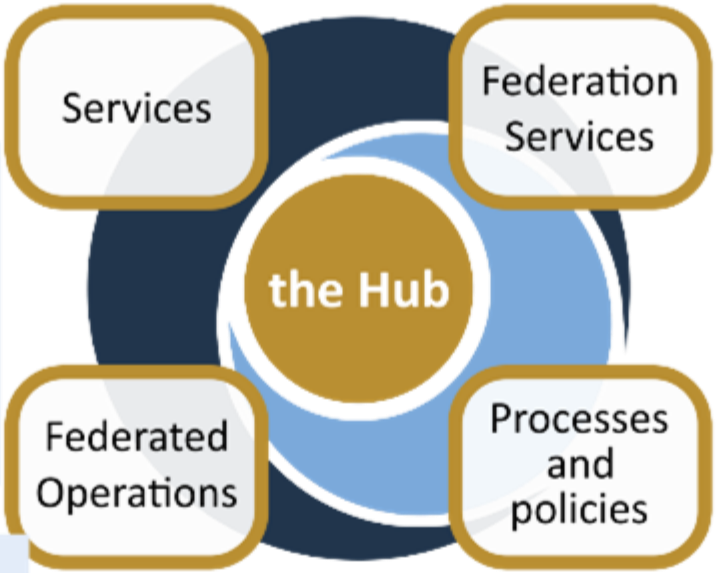
EOSC-hub Mission

A federated integration and management system for EOSC

EOSC-hub

From the consortium AND from external contributors

- Data
- Applications & tools
- Baseline services (storage, compute, connectivity)
- Training, consultants



Usage according to Principles of engagement

- Marketplace
- AAI
- Accounting
- Monitoring

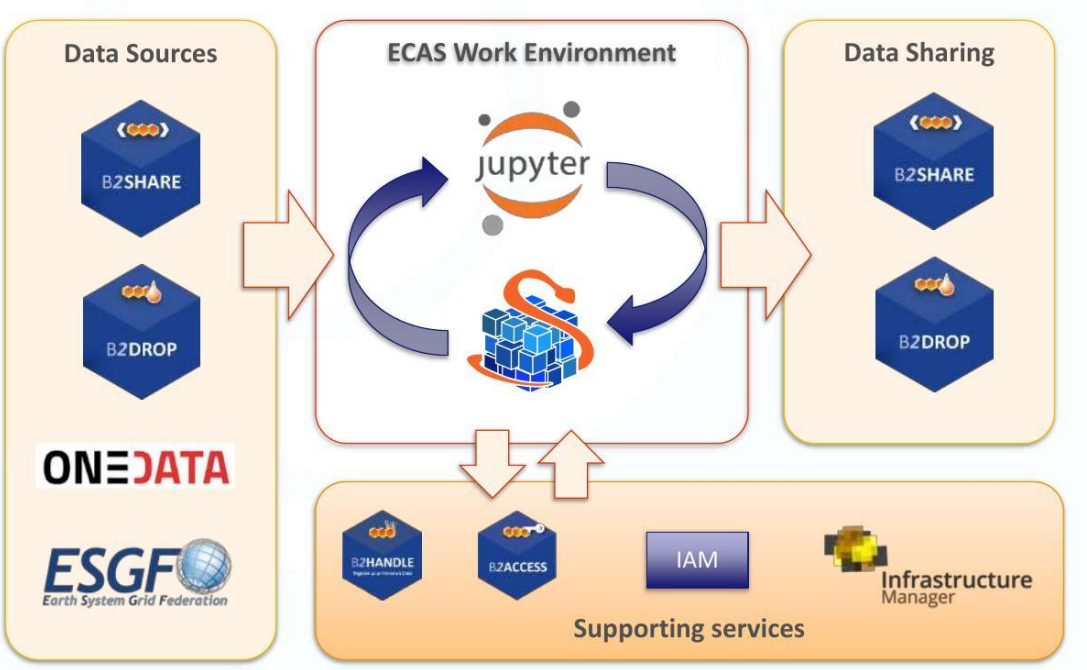
- Lightweight certification of providers
- SLA negotiation
- Customer Relationship Management

Based on FitSM

- Security regulations,
- Compliance to standards,
- Terms of use,
- FAIR implementation guidelines

EOSC thematic service: ENES Climate Analytics Service (ECAS)

- ENES partners: DKRZ, CMCC
- ECAS provides a server-side data analysis environment
  - Ophidia backend, Jupyter notebook, PyWPS and data sharing services
  - The main activities are related to integration effort between ECAS and EGI/EUDAT/INDIGO services, exploitation by other research communities, training events and dissemination



Work is done as part of EOSC-hub work package 7: thematic services

The core services of ECAS are not funded in EOSC-hub (only integration effort with respect to other EOSC-related services)

ECAS proposal was ranked 1st out of 63 Thematic Service proposals

ENES contribution to EOSC-hub services

B2Find EOSC service (DKRZ, EUDAT): Interdisciplinary data catalog

B2Handle EOSC service (DKRZ, EUDAT): PID APIs, operations

The ECAS thematic EOSC service

DKRZ and CMCC are current ECAS service providers

- CMCC: https://ecaslab.cmcc.it/,DKRZ: https://ecaslab.dkrz.de/
- Sharing of analytics workflows and python applications on github
- Access policies and funding models
  - Excellence driven / market driven (+for training and education)
  - Details need to be defined in collaboration with the relevant EOSC-hub actors

ENES in EOSC and other data projects

EOSCpilot:

- Funded to support the first phase in the EOSC development (1FTE for 1 year)
- ENES partners (IPSL, DKRZ (unfunded))

Science Demonstrators

- 15 science demonstrators selected
  - 5/5/5 demonstrators selected, 12 months each
  - ENES/ICOS demonstrator in first 5 selected (2017), ended Nov. 2018
  - Large delays (technical + non-technical issues, “real work” started summer 2018)
- Science demonstrator title: ERFI - ENVRI Radiative Forcing Integration
- Partners: ICOS ERIC, ENES (IPSL, DKRZ), EGI (“shepherd”)
- Idea - Approach:
  - Use IS-ENES data to drive land surface/ecosystem models
  - IS-ENES climate data → EGI OneData ← → LPJ-GUESS model on EGI VM
- Problems
  - Onedata availability, stability - EGI-VM / Onedata interoperability
  - “Real” ENES/ICOS data interoperability issues not addressed

ENES in EOSC and other data projects

Other data projects:

- SeaDataCloud (DKRZ)
- DARE (KNMI, CERFACS)

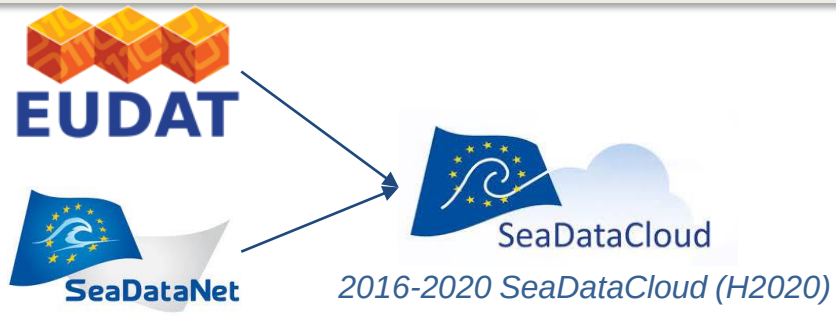
New EOSC related projects

- INFRAEOSC-02-2019:
  - Prov4R (DKRZ)
  - KOINE (DKRZ)
  - Indigo-Next (CMCC, CERFACS)
- INFRAEOSC-2018-2020
  - EOSC-Pillar (DKRZ, CMCC)
  - EOSC-Nordic (DKRZ)
- BG-07-2019-2020: BlueCloud (DKRZ)

  - Future of the Seas and Oceans Flagship Initiative

SeaDataCloud

- 4 year, started 2016
- 10 M Euros, 56 partners, 32 countries
- ENES partner: DKRZ
- 5 EUDAT members are SeaDataCloud partners: CINECA, CSC, DKRZ, GRNET and STFC



SeaDataNet: A pan-European infrastructure set up and operated for managing marine and ocean data in cooperation with the NODCs and data focal points of 35 countries bordering European seas

SeaDataCloud:

- Improving and innovating services and products (the data portal, data ingestion, …)
- Adopting and elaborating new technologies (cloud environments, etc.)
- Implementing a strategic and operational cooperation between the SeaDataNet consortium (marine and ocean data centres) and the EUDAT consortium (e-infrastructure service providers).

DKRZ SeaDataCloud focus (1): Developing and deploying a **Virtual Research Environment (VRE),**

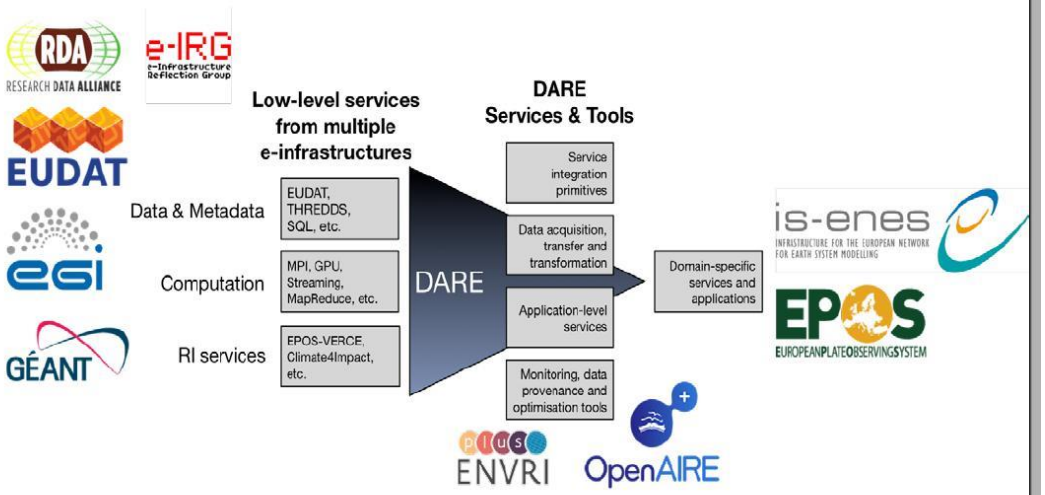
- a collaborative web- and cloud-based workspace
- services for researchers: Finding, handling, curating, quality controlling, transforming and processing marine and ocean data into data products which can be integrated, visualized and published using OGC and high level services

DKRZ SeaDataCloud focus (2): Ingesting and storing marine data in the cloud (instead of decentralized and mainly off-line storage at (sub-)national data centres)

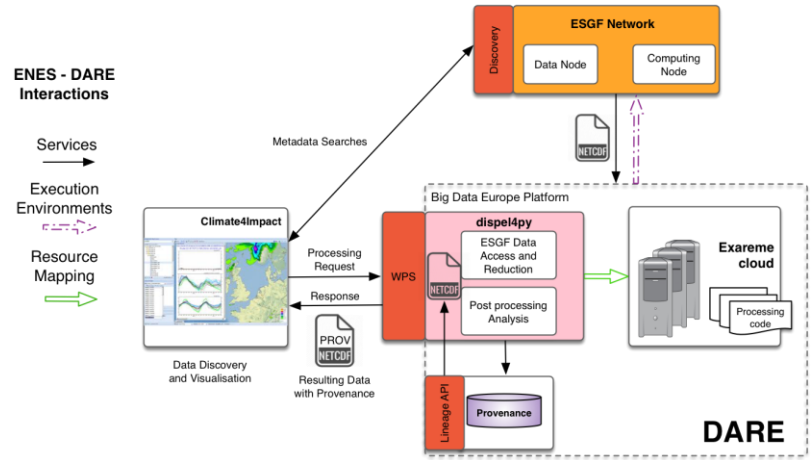
DARE (H2020)

- 9 partners, ENES: KNMI, CERFACS, other: NCSR-Demokritos (Coord), INGV, UEDIN, GRNET, Fraunhofer, ATHENA, KIT (two RI's: EPOS, IS-ENES)
- Main objectives: Abstraction and bridges of e-infra and Infra architectures' tools and resources (parallel computing). Full Lineage implementation, researcher user interfaces (Jupyter Notebooks, ...), platform developer APIs (C4I -> DARE).

Delivering Agile Research Excellence on European infrastructures



DARE ENES Use Case Integration Prototype



New EOSC related project proposals with ENES participation

PROV4R: INFRAEOSC-02-2019 action on provenance services

to be submitted, ~500 PMs → ½ IS-ENES3

Focus:

- Generic provenance support services for communities (PID, PROV based)
- Concrete community use cases
- Integration into EOSC

Partners:

- DKRZ: Project lead, Middleware services hub - WP1, WP6
- KCL: Provenance modeling and research - WP7
- EAA: Community workflows - WP5
- CINECA: Community engagement - WP4
- GWDG: Sustainability - WP3
- Maris: Communication - WP2
- KNMI, GFZ, KNAW, BSC, UKAEA, STFC, EUDAT Ltd., OpenAIRE, LU, MI

Communities:

- ENES, LTER, SeaData, Clarin, Dariah, Elixir, Fusion, OpenAIRE, ICOS, EPOS

KOINE: INFRAEOSC-02-2019 action on terminology services

to be submitted, ~500 PMs

Focus:

- Generic terminology support services for communities
- Enabling semantic data interoperability, discovery and exploitation
- Enhancing B2FIND and community portals
- Integration into EOSC

Partners:

- TIB: Project lead
- DKRZ (B2FIND, IS-ENES)
- IFREMER
- ...

Communities:

- ENES, LTER, Clarin, ACTRIS, SIOS, LifeWatch, ...

INDIGO-Next: INFRAEOSC-02-2019

to be submitted

Partners:

- INFN (Proposal coordinator)
- CMCC and CERFACS are involved jointly with several other partners

- Follow up of INDIGO-DataCloud
- Still under preparation

- A strict NDA has been signed by all the proposal participants, so only limited info can be disclosed

ENES in EOSC and other data projects

EOSC-Pillar: INFRAEOSC-2018-2020

ENES partners: CMCC, DKRZ, (CNRS, CINES) - submitted

Partner roles:

- DKRZ: defining procedures/service to foster data provenance for thematic communities
- DKRZ, CMCC Agile FAIR data curation (environmental and ENES community)

EOSC-Nordic: INFRAEOSC-2018-2020

ENES partners: DKRZ - submitted

Partner roles: B2Find adoption/integration

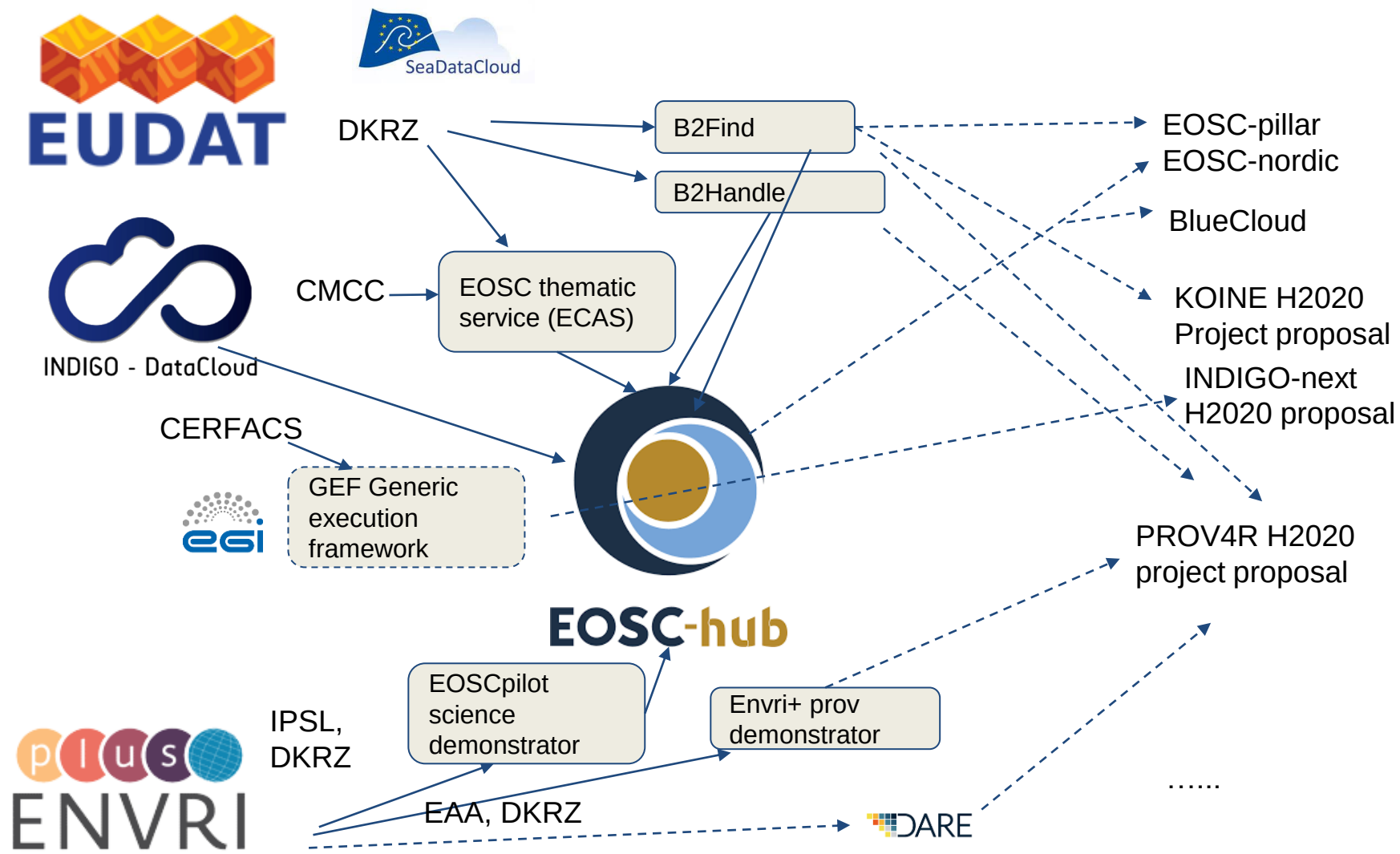
BlueCloud: BG-07-2019-2020 - Future of the Seas and Oceans Flagship Initiative

ENES partners: DKRZ - submitted

Partner roles: VRE provisioning, service/data hosting

ENES in EOSC related projects

(Some) Relations between projects ..



previous activities

EOSC related activities

Planned activities

Additional material

- **Additional slides**
- **Past projects with influence on current project involvements**

EXDCI2 (H2020 FETHPC CSA)

- 2 partners: PRACE and ETP4HPC
  - 20+ Third parties
  - Involvement of ENES HPC Task Force representatives
    - J. C. André - T3.1 Roadmap of HPC applications and usages
    - G. Aloisio - T3.2 Engagement with HPC users communities and CoEs

- Main objectives: **This project will coordinate the update and implementation of a common strategy for the European HPC Ecosystem**

- The two main strategic goals of the proposed CSA are:
  - **Development and advocacy of a competitive European HPC Exascale Strategy**
    - EXDCI-2 will support the implementation of a common European HPC strategy, open to synergistic areas including **HPDA** and **Artificial Intelligence (AI)**
  - **Coordination of the stakeholder community for European HPC at the Exascale**
    - EXDCI-2 will promote joint community structuring and synchronisation

B. Germany has made significant progress over the last years to become EOSC ready

Illustrative

Policy and strategy

- Driving Go Fair together with The Netherlands.
- The BMBF released its Open Access Strategy entitled "Open Access in Germany" in September 2016 which contains a clear commitment to the principles of open access and open science.
- DFG, the main research funder, has tied open access into its funding policy.

National facilities & EU participation in RI

- Diverse and well established national Research Infrastructures, host and member of numerous pan-European Research Infrastructures on the ESFRI roadmap and very strong participation in main European e-Infrastructures.

Funders and funding streams

- Germany has a federal system with shared funding responsibilities between federal and state governments.
- Together they fund about one third of all R&D expenditure in Germany.
- DFG provides the main source of project funding for university research. It acts similarly to a research council in that most of its funds are competitively awarded following a process of peer review of applications.

eInfras information

- Member/delegate of GEANT, EGI Council, PRACE, EUDAT, e-IRG
- National Research and Education Network, Registry of Research Data Repositories (re3data.org), a national research data infrastructure (NFDI).
- Has an OpenAIRE national OA desk.

13

Source: eosc strategic implementation roadmap

EOSC services

Open Collaboration services

- Applications Database
- Repositories

WP5

Federation services

- Accounting
- ARGO
- Check-in
- GGUS
- GOCDB
- Marketplace
- Operations Portal
- RC Auth
- SPMT
- DPMT
- B2ACCESS
- TTS
- SYMON

Basic infrastructure and added-value services

- EGI High-Throughput Compute
- EGI Cloud Compute
- EGI Cloud Container
- DIRAC4EGI
- EGI Online storage
- EGI DataHub
- B2HANDLE
- B2FIND
- B2DROP
- B2SAFE
- B2STAGE
- B2NOTE
- ETDR
- Sensitive Data Service
- Advanced IaaS
- TOSCA for Heat
- OPIE

Thematic services

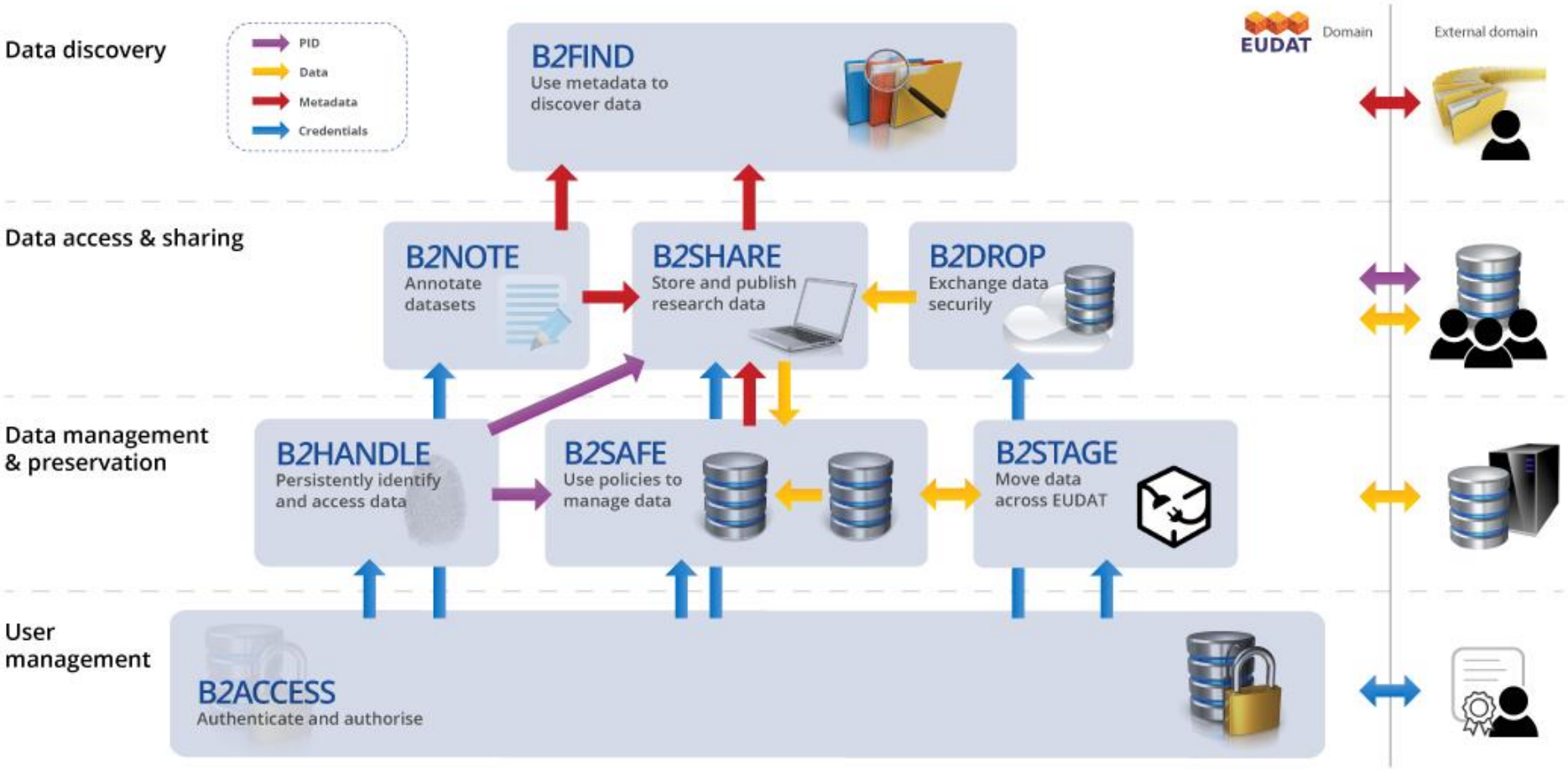
- ECAS
- DARIAH Gateway
- OPENCoastS
- GEOSS
- EO Pillar
- WeNMR
- DODAS
- LifeWatch
- CMI

From month 19:

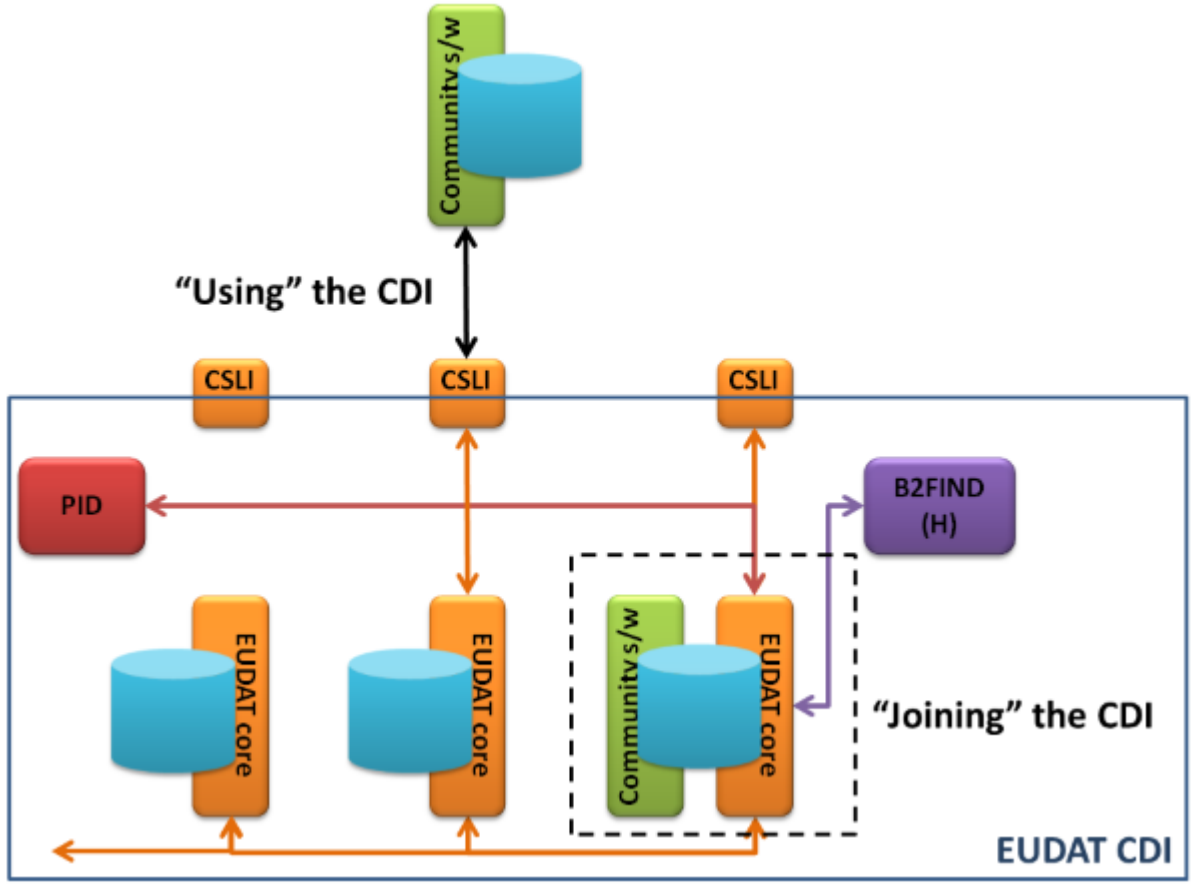
- IFREMER
- EISCAT_3D Portal

+ new service from outside the consortium

Services


Data discovery
PID
Data
Metadata
Credentials
B2FIND
Use metadata to discover data
EUDAT
Domain
External domain
Data access & sharing
B2NOTE
Annotate datasets
B2SHARE
Store and publish research data
B2DROP
Exchange data securily
Data management & preservation
B2HANDLE
Persistently identify and access data
B2SAFE
Use policies to manage data
B2STAGE
Move data across EUDAT
User management
B2ACCESS
Authenticate and authorise

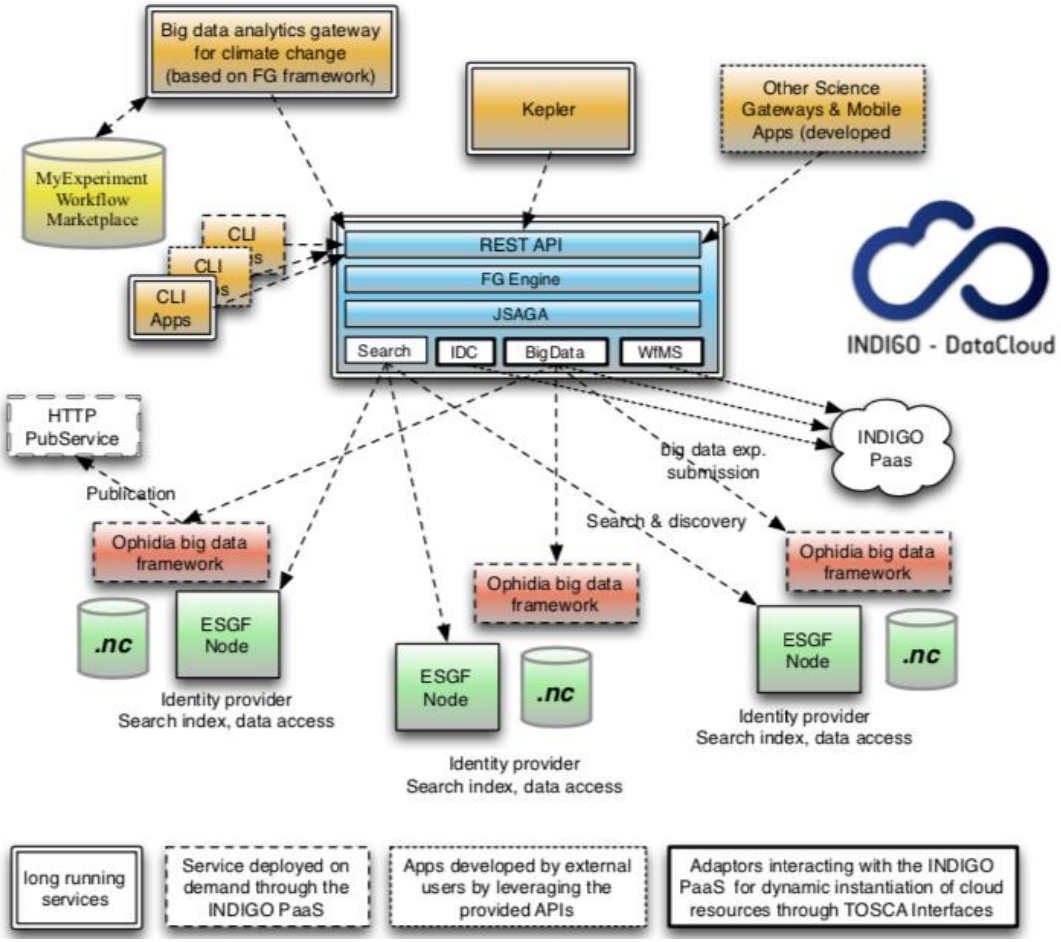
Using the CDI versus joining the CDI



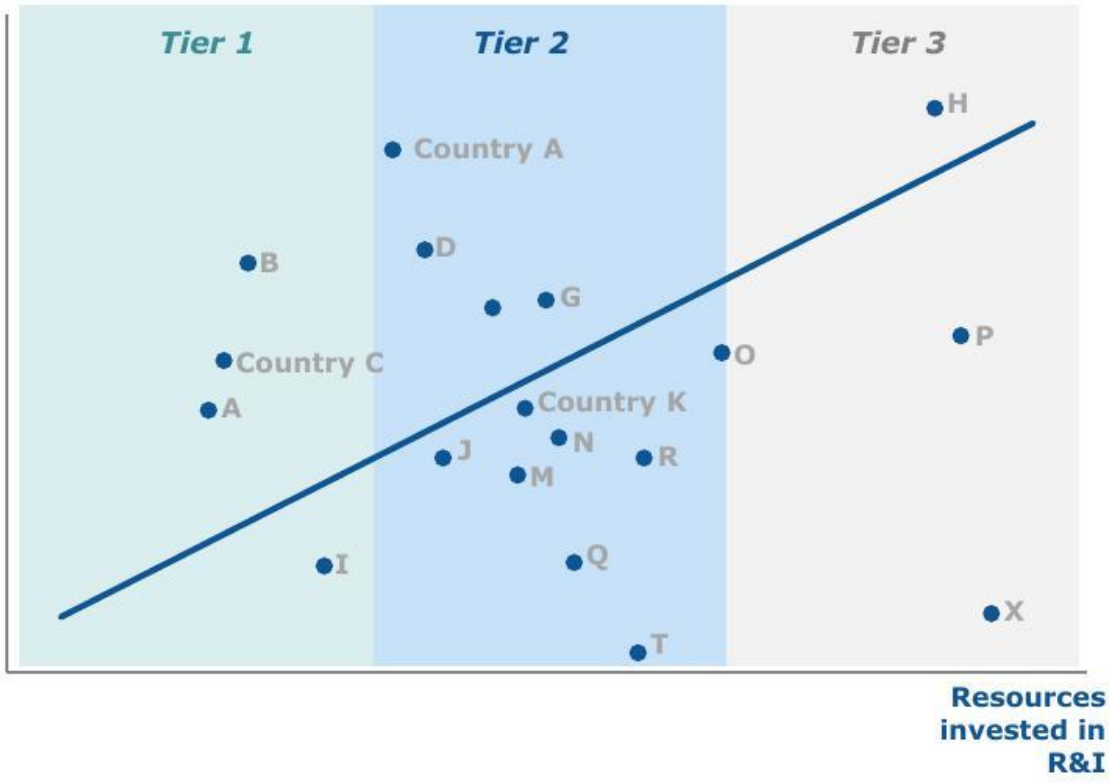
INDIGO-DataCloud

ENES partner: CMCC - Ophidia based processing

- Distributed experiments for climate data analysis
- Server-side, parallel processing
- Two-level workflow strategy to orchestrate large scale experiments
- Interoperability with ESGF
- Access through different clients
  - Kepler
  - Science Gateway
- Interactive and batch scenarios



Each member state starting point is different → customized implementation strategy shall be adopted



Conceptual

The EC should – e.g. via the H2020 Policy Support Facility - support each MS to self-assess its starting position and the implications for the EOSC implementation roadmap.

11

Source: [eosc strategic implementation roadmap]