IS-ENES3 1st General Assembly
25-27th March 2020
Toulouse, France

Status of ESGF for CMIP6

Stephan Kindermann, Katharina Berger, Maria Moreno(DKRZ)
Sandro Fiore, Paola Nassini (CMCC)
Guillaume Levavasseur (IPSL)
Martin Juckes, Ruth Petri (STFC/CEDA)

The IS-ENES3 project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 824084

Overview

1. CMIP6 in ESGF
   a. Overview: Data Nodes → Models
   b. Some statistics
2. IS-ENES3 ESGF related services supporting CMIP6
   a. User support
   b. data citation and PIDs
   c. Errata
   d. Synda and replication
   e. ESMvaltool data products (prototype)
   f. (standards and conventions: CF and DRequ)
   g. (es-doc → separate talk)
3. CMIP6 via ESGF to processing
   a. CMIP6 processing associated to data pools (ESGF and non ESGF)

The IS-ENES3 project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

CMIP6 in ESGF:

Portals:

- 3 tier1 site portals: CEDA, DKRZ, IPSL
- 1 tier2 site portal: LIU/SMHI

Data Nodes: 11 (from 29 worldwide) and **associated institutes** (inst_id)

- Sweden: LIU (inst_id: EC_EARTH)
- Spain: BSC (inst_id: EC_EARTH)
- Norway: (inst_id: NCC)
- France: CNRM (inst_id: CNRM-CERFACS), IPSL (Inst_id: IPSL)
- Ireland: ICHEC (inst_id: EC-Earth-Consortium)
- Italy: CINECA (inst_id: EC_EARTH), CMCC (inst_id: CMCC)
- England: CEDA (Inst_id: CMCC, CNRM-CERFACS, EC-EARTH, ECMWF, MOHC, MPI-M, NERC, NIMS-KMA, NIWA)
- Germany: DWD, DKRZ (Inst_id: AER, AWI, DKRZ, DWD, Hammoz-Consortium, INM, MOHC, MPI-M, RTE-RRTMGP-Consortium, UHH)
- (Denmark: CORDEX only)

Models: 86 worldwide, EU: 40 (3 INM)   (source_id: model identifier)

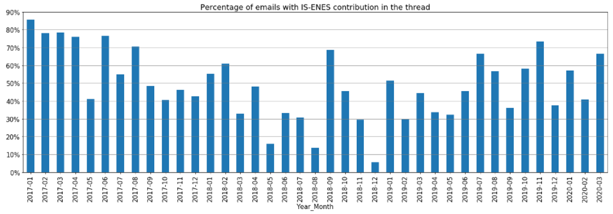
CMIP6 in ESGF: Europe / world

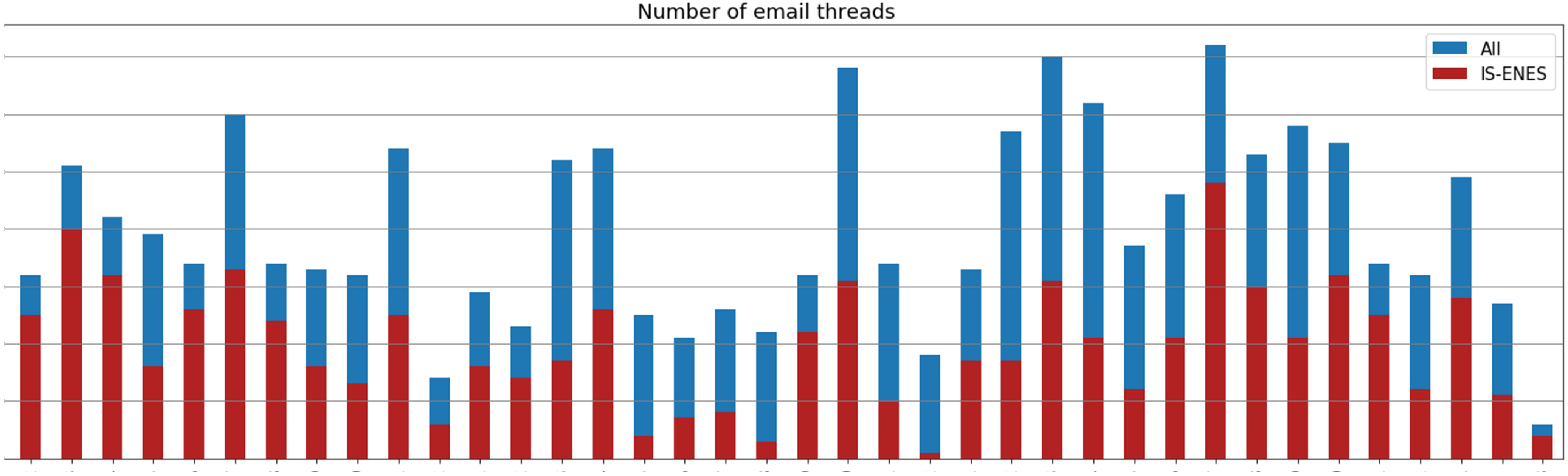
- Number of datasets (worldwide): 2.921.511 (+ 76.431 retracted)
  - Replica: 2.611.182
- Number of datasets (Europe): 1.340.942 (+ 33.215 retracted)
  - Replica: 483.430
- Number of files (worldwide): 10.882.989
  - Replica: 6.531.058
- Number of files (Europe): 6.905.026
  - Replica: 1.326.377
- Size (worldwide): 5.54 PB (+ 3.75 PB published Replica)
- Size (Europe): 3.88 PB (+ 1.3 PB published Replica)

The IS-ENES3 project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

IS-ENES3 services CMIP6/ESGF: User support

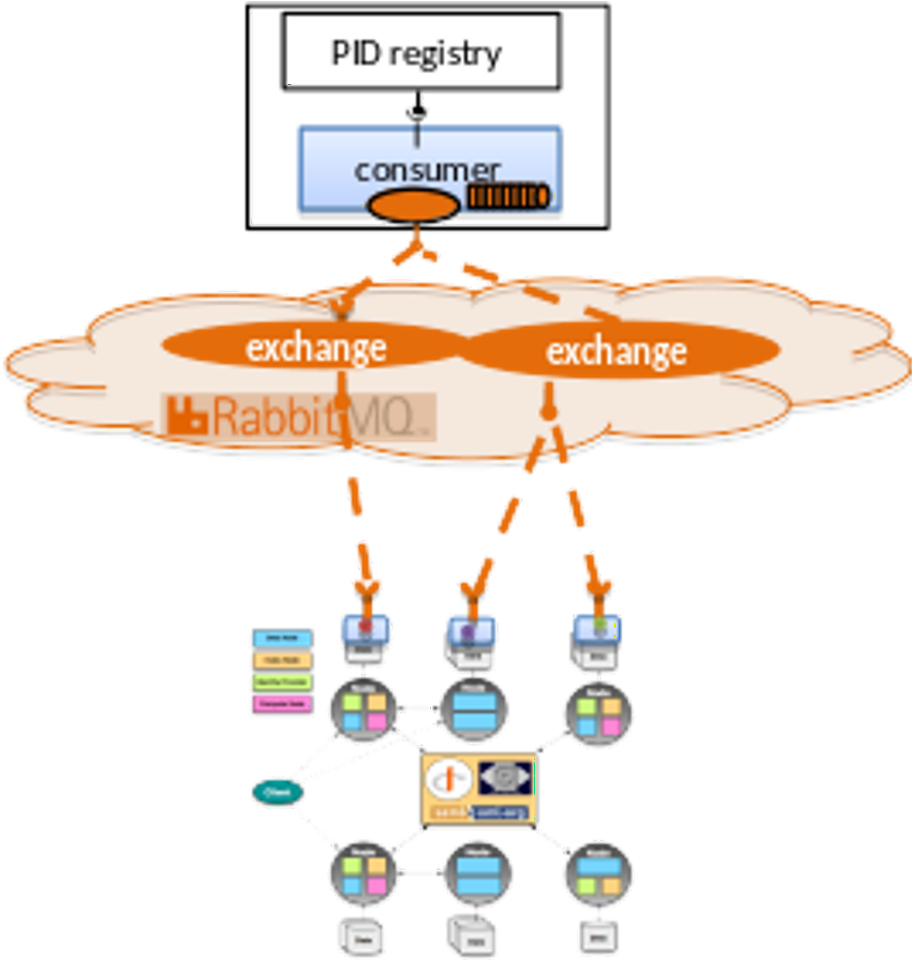
- ESGF user enquiries on ESGF user email list → (distributed) ENES support staff
- No dedicated support system(s) → support forum, ticketing system etc. :-(
- Open access policy reduced end-user download problems significantly ..





The IS-ENES3 project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

IS-ENES3 services CMIP6/ESGF: Persistent identifiers

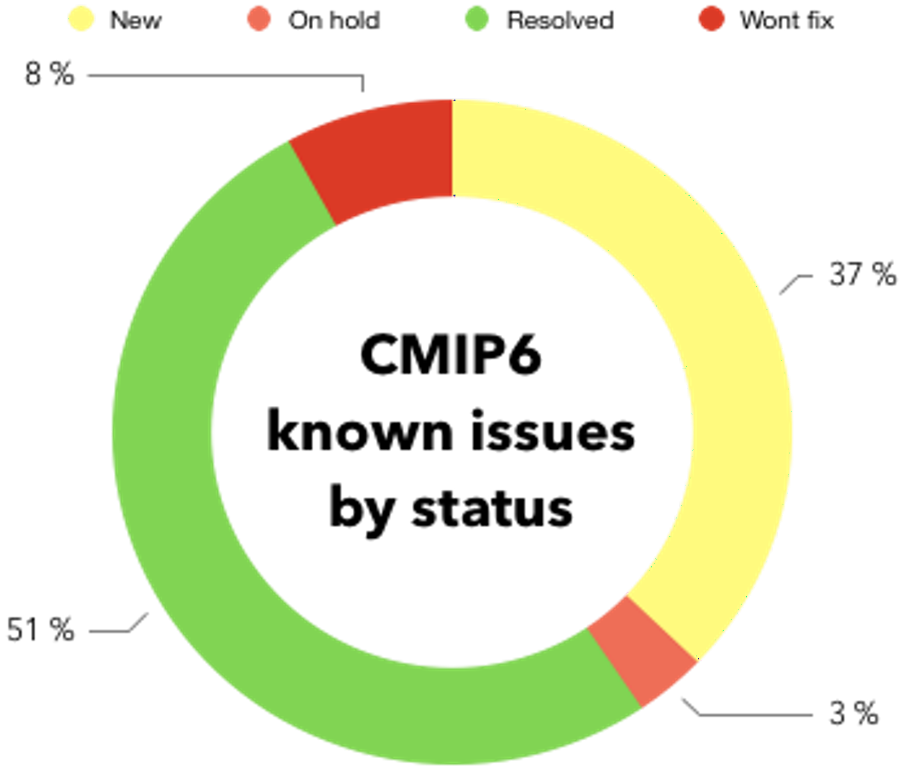


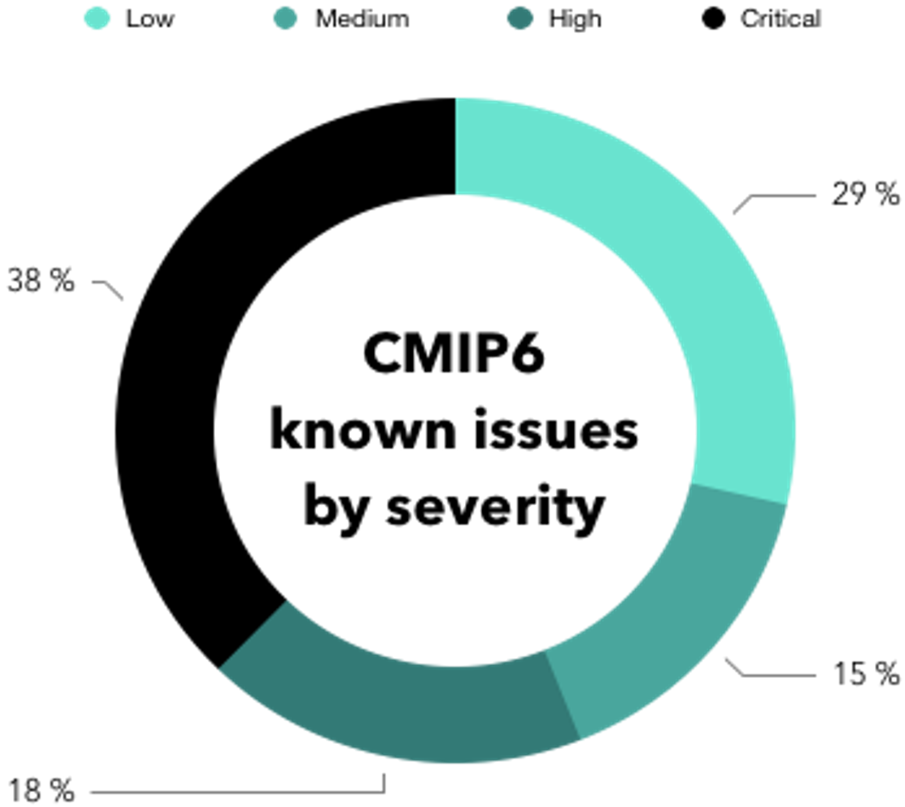
- Successful operation of ESGF/distributed message queue/ Handle infrastructure at global scale
  - Central ~40 GB central registry DB (handle system based) and consumer components
  - Distributed message queue nodes at IPSL, LLNL, DKRZ, GFDL
  - operational agreements as part of CDNOT and WCRP WIP
- Status CMIP6
  - ~ 11 * 10^6 file PIDs + 1.4 * 10^6 dataset PIDs registered
  - overall 16.5 * 10^6 PIDs assigned (including retracted etc.)
  - << 0.1% registration problems (e.g. firewall issues), automatic curation scripts in place
- Cooperation with other e-infra efforts: EOSC-hub, EPIC, EUDAT, ...

The IS-ENES3 project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

IS-ENES3 services CMIP6/ESGF: Errata

- **175 issues** currently registered on the service
- 89,993 datasets affected by at least one issue or **3% of the CMIP6 archive**.



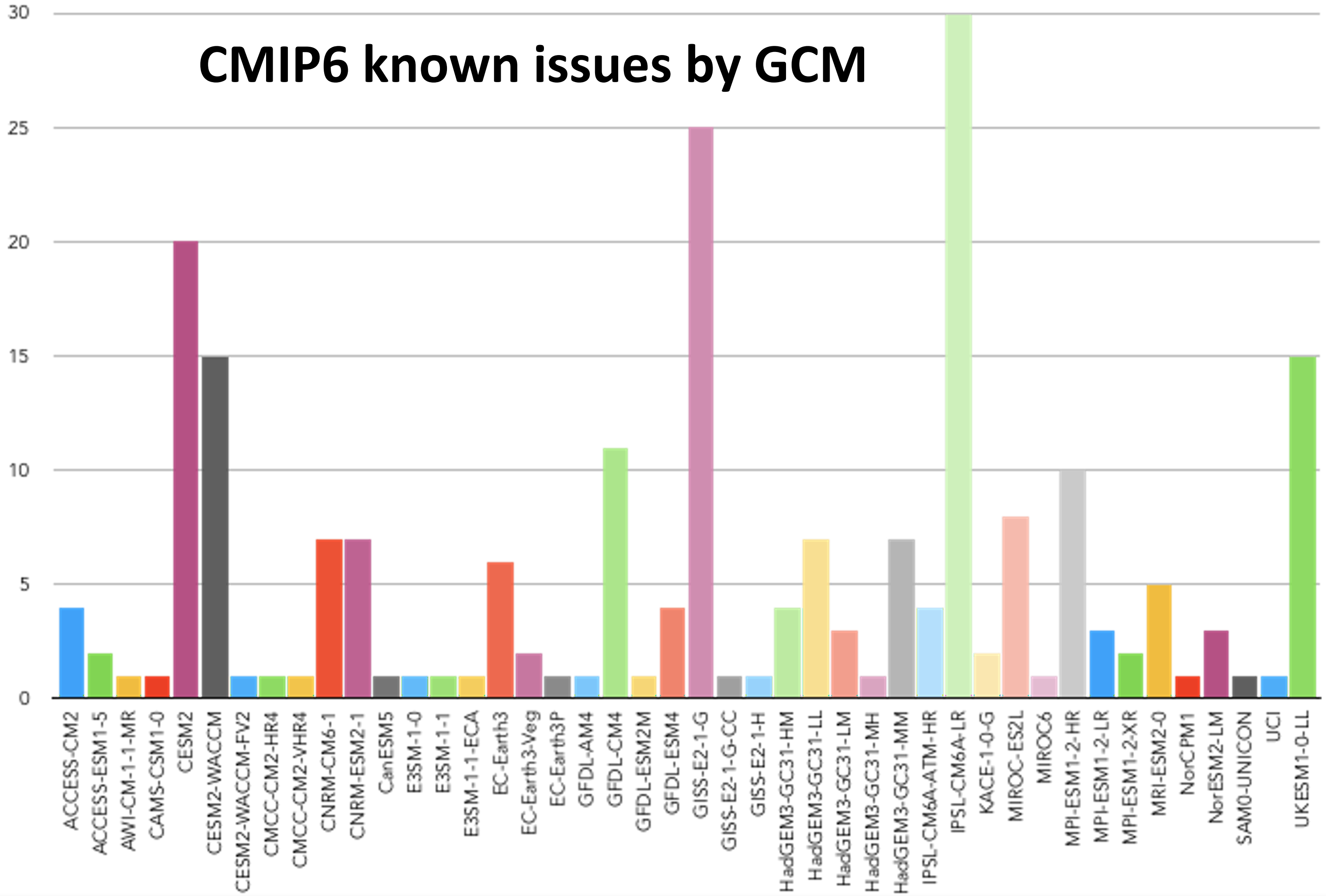


The IS-ENES3 project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

IS-ENES3 services CMIP6/ESGF: Errata

Statistics to be considered **carefully**:

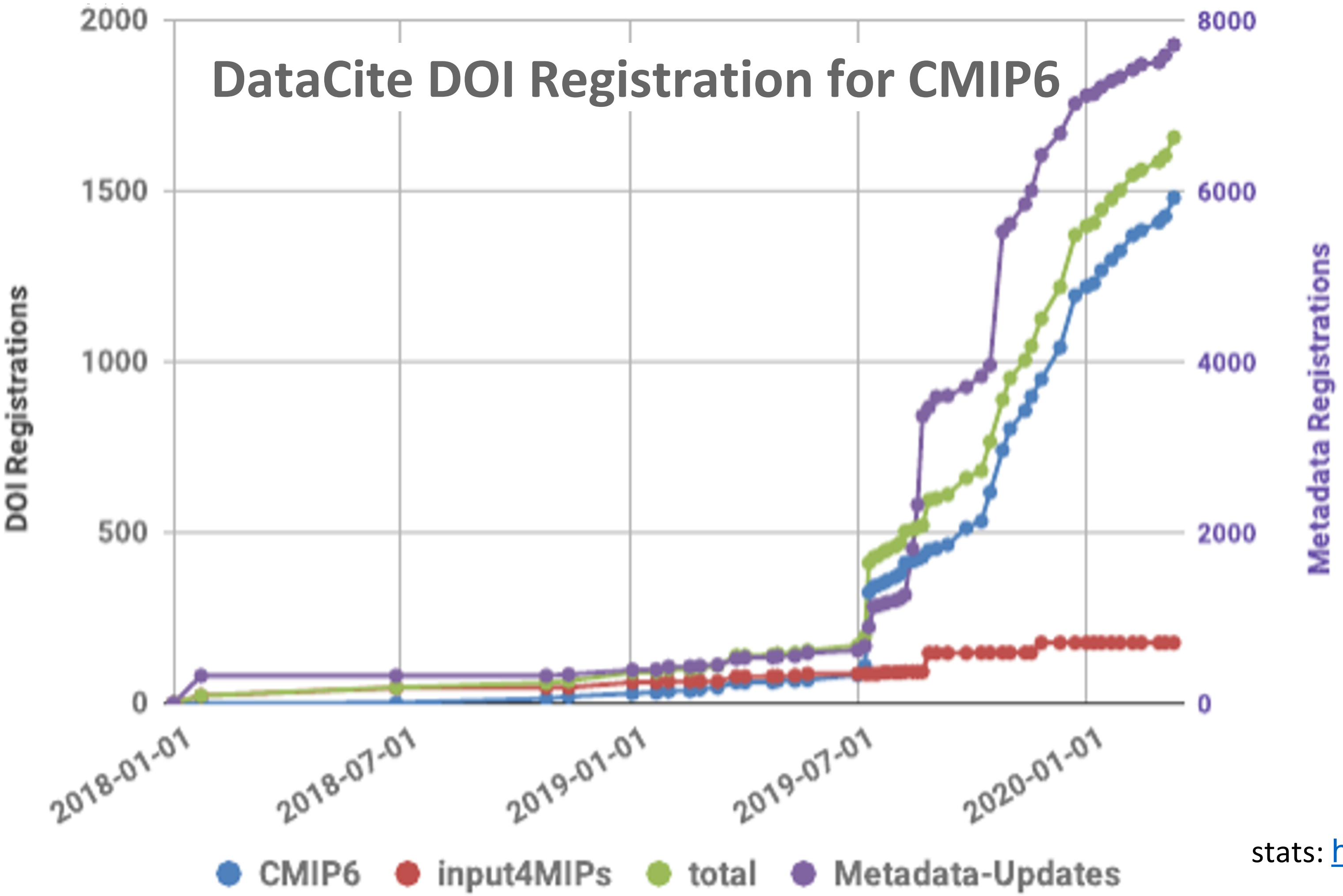
- Some modelling groups do still **NOT** follow CDNOT recommendations and publication workflow
- Most of CMIP6 production has **NOT** yet been analyzed



CMIP6 known issues by GCM

The IS-ENES3 project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

IS-ENES3 services CMIP6/ESGF: Data citation



stats: http://bit.ly/CMIP6_DOI_Statistics

The IS-ENES3 project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

IS-ENES3 services CMIP6/ESGF: synda

- The “Transfer module” (aka “sdt”) is being refactored and migrated to Python 3.7
  - GitHub repo : https://github.com/Prodiguer/synda
    - Issues: 50 open vs. 67 closed
    - Pull requests : 0 open vs. 16 closed
  - 4 branches:
    - `master`: Last functional release with Conda installation and test suite
    - `synda-python3`: incoming Python 3 release with full refactoring
    - `gh-pages`: host documentation
    - `get-file-caching`: additional feature used by PCMDI
  - Releases to date: 33
  - Release scheme: conda package
  - Major points (challenges) ahead:
    - Revise integration with Globus after Lukasz latest coding sprint
    - ORM `sqlalchemy` could facilitate changing db technology (opens door for optimize access to db).
- The “Post-processing module” (aka “sdp”) is dropping to consider integration with other proven pipeline orchestration tool as “Cylc”.

The IS-ENES3 project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 824084

IS-ENES3 services CMIP6/ESGF: Data Pools

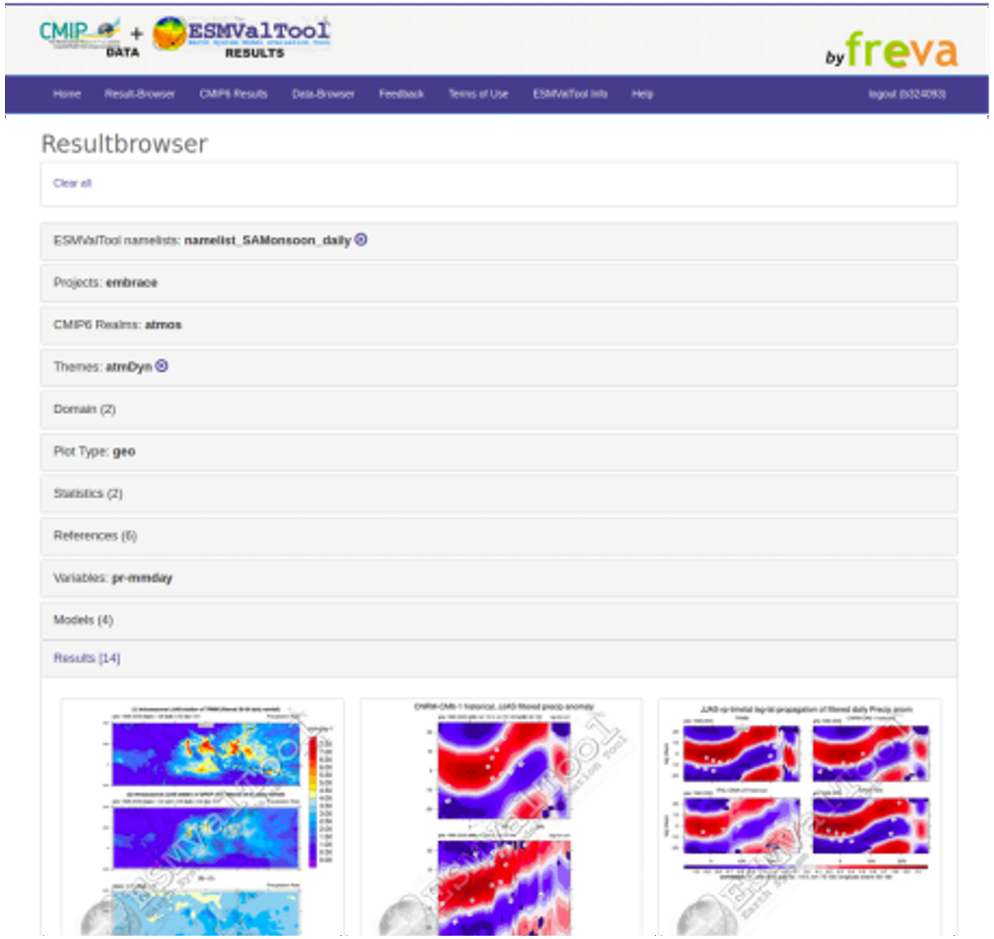
- Replication is based on synda tool and coordinated as part of CDNOT
- DKRZ: ~ 5 PByte disk pool, currently filled with ~2.5 PByte CMIPt data (1 PByte original and 1.5 PByte replica which are only partially republished)
  transfer to archive (CERA, WDC climate) will start in summer
- IPSL: ~ 2PByte CMIP6 disk pool, currently filled with 812.8 TByte, replica partially republished
- CEDA: ~ 2 PByte CMIP6 disk pool, ~ 1 PByte replica
- CMCC: on demand replication from DKRZ

- No access statistics as the data is freely accessible for researchers with accounts at DKRZ, IPSL and CEDA (→ problem for VA/TNA accounting ..)
- Very positive feedback from users (yet service availability not broadly known ..)
- Data pools are basis for TNA and VA compute services
  (different environments, different data cataloguing at centers ..)

The IS-ENES3 project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

CMIP6 via ESGF to processing: ESMValtool data products

pre-operational service running at DKRZ: https://cmip-esmvaltool.dkrz.de


CMIP DATA + ESMValTool RESULTS
by freva
Home
Result-Browser
CMIP6 Results
Data-Browser
Feedback
Terms of Use
ESMValTool Info
Help
logout (b324093)
CNRM-CM6-1 historical, JJAS filtered precip anomaly
JJAS-rp-lineal lag/lat propagation of filtered daily Precip anom

Resultbrowser

Clear all

ESMValTool namelists: **namelist_SAMonsoon_daily** ⊗

Projects: **embrace**

CMIP6 Realms: **atmos**

Themes: **atmDyn** ⊗

Domain (2)

Plot Type: **geo**

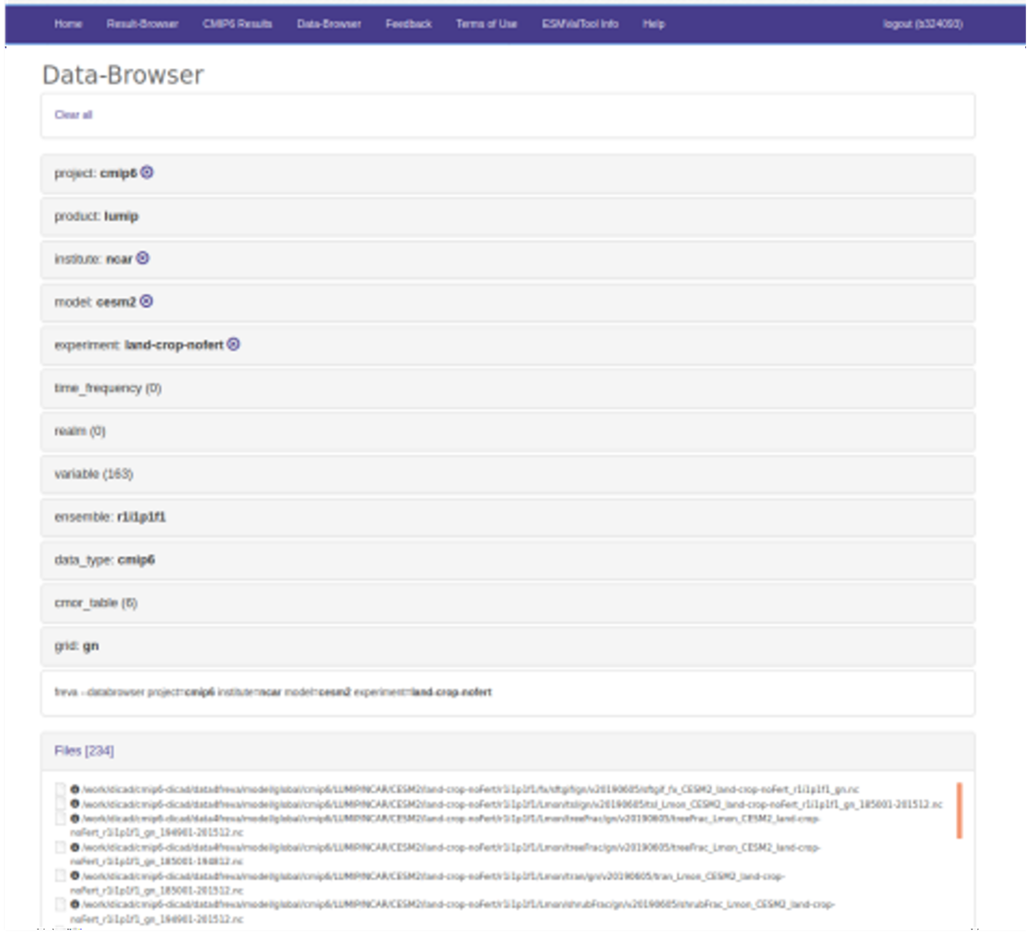
Statistics (2)

References (6)

Variables: **pr-mmday**

Models (4)

Results [14]



Data-Browser

Clear all

project: **cmip6** ⊗

product: **lumip**

institute: **ncar** ⊗

model: **cesm2** ⊗

experiment: **land-crop-nofert** ⊗

time_frequency (0)

realm (0)

variable (163)

ensemble: **r1i1p1f1**

data_type: **cmip6**

cmor_table (6)

grid: **gn**

```
freva --databrowser project=cmip6 institute=ncar model=cesm2 experiment=land-crop-nofert
```

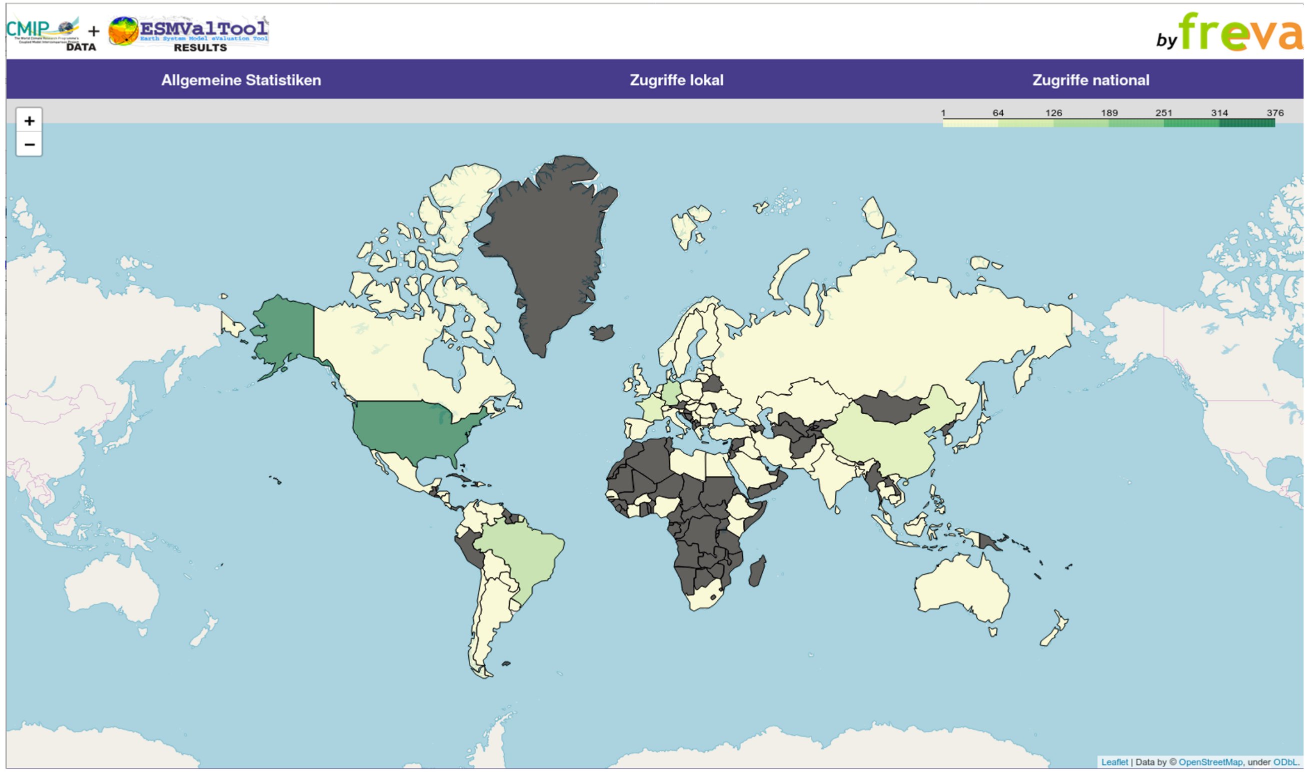
Files [234]

```
/work/dicad/cmip6-dicad/data4freva/model/global/cmip6/LUMIP/NCAR/CESM2/land-crop-noFert/r1i1p1f1/fx/sftgif/gn/v20190605/sftgif_fx_CESM2_land-crop-noFert_r1i1p1f1_gn.nc
/work/dicad/cmip6-dicad/data4freva/model/global/cmip6/LUMIP/NCAR/CESM2/land-crop-noFert/r1i1p1f1/Lmon/tsl/gn/v20190605/tsl_Lmon_CESM2_land-crop-noFert_r1i1p1f1_gn_185001-201512.nc
/work/dicad/cmip6-dicad/data4freva/model/global/cmip6/LUMIP/NCAR/CESM2/land-crop-noFert/r1i1p1f1/Lmon/treeFrac/gn/v20190605/treeFrac_Lmon_CESM2_land-crop-noFert_r1i1p1f1_gn_194901-201512.nc
/work/dicad/cmip6-dicad/data4freva/model/global/cmip6/LUMIP/NCAR/CESM2/land-crop-noFert/r1i1p1f1/Lmon/treeFrac/gn/v20190605/treeFrac_Lmon_CESM2_land-crop-noFert_r1i1p1f1_gn_185001-194812.nc
/work/dicad/cmip6-dicad/data4freva/model/global/cmip6/LUMIP/NCAR/CESM2/land-crop-noFert/r1i1p1f1/Lmon/tran/gn/v20190605/tran_Lmon_CESM2_land-crop-noFert_r1i1p1f1_gn_185001-201512.nc
/work/dicad/cmip6-dicad/data4freva/model/global/cmip6/LUMIP/NCAR/CESM2/land-crop-noFert/r1i1p1f1/Lmon/shrubFrac/gn/v20190605/shrubFrac_Lmon_CESM2_land-crop-noFert_r1i1p1f1_gn_194901-201512.nc
```

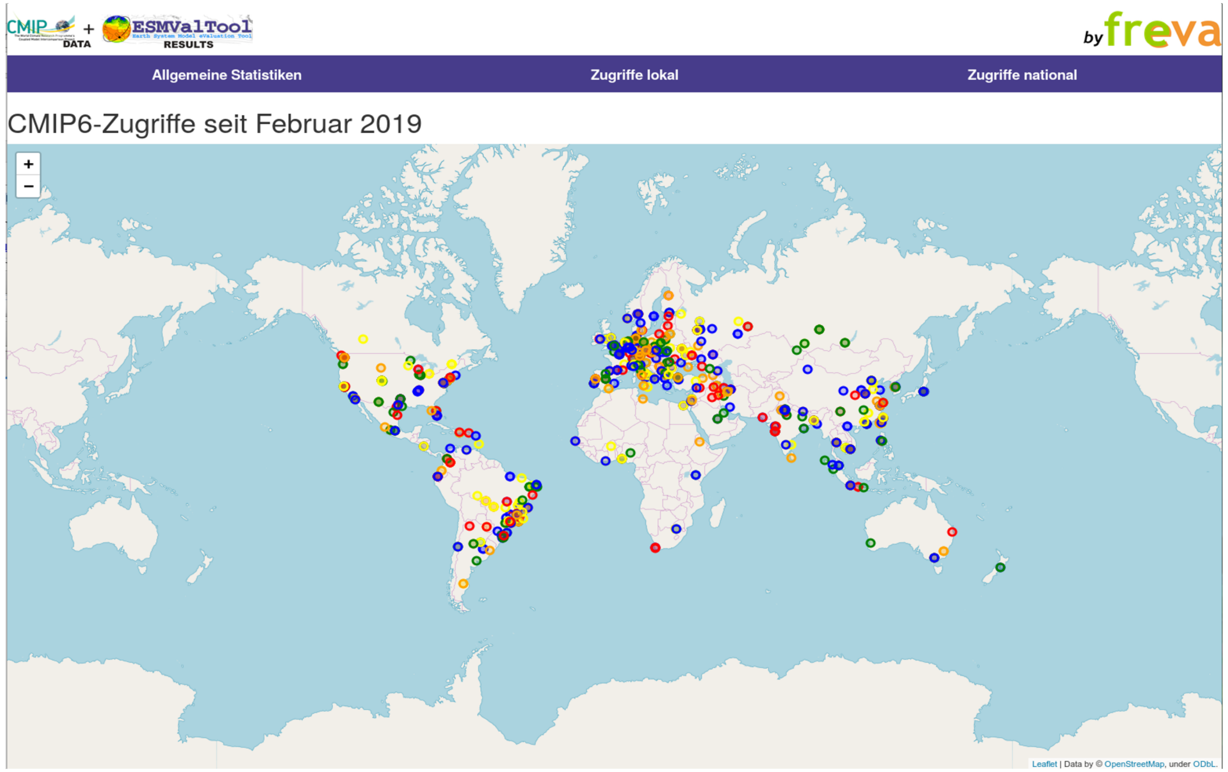
The IS-ENES3 project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 824084

CMIP6 via ESGF to processing: ESMValtool data products

Access from countries (grey: no access)



Access since Feb. 2019



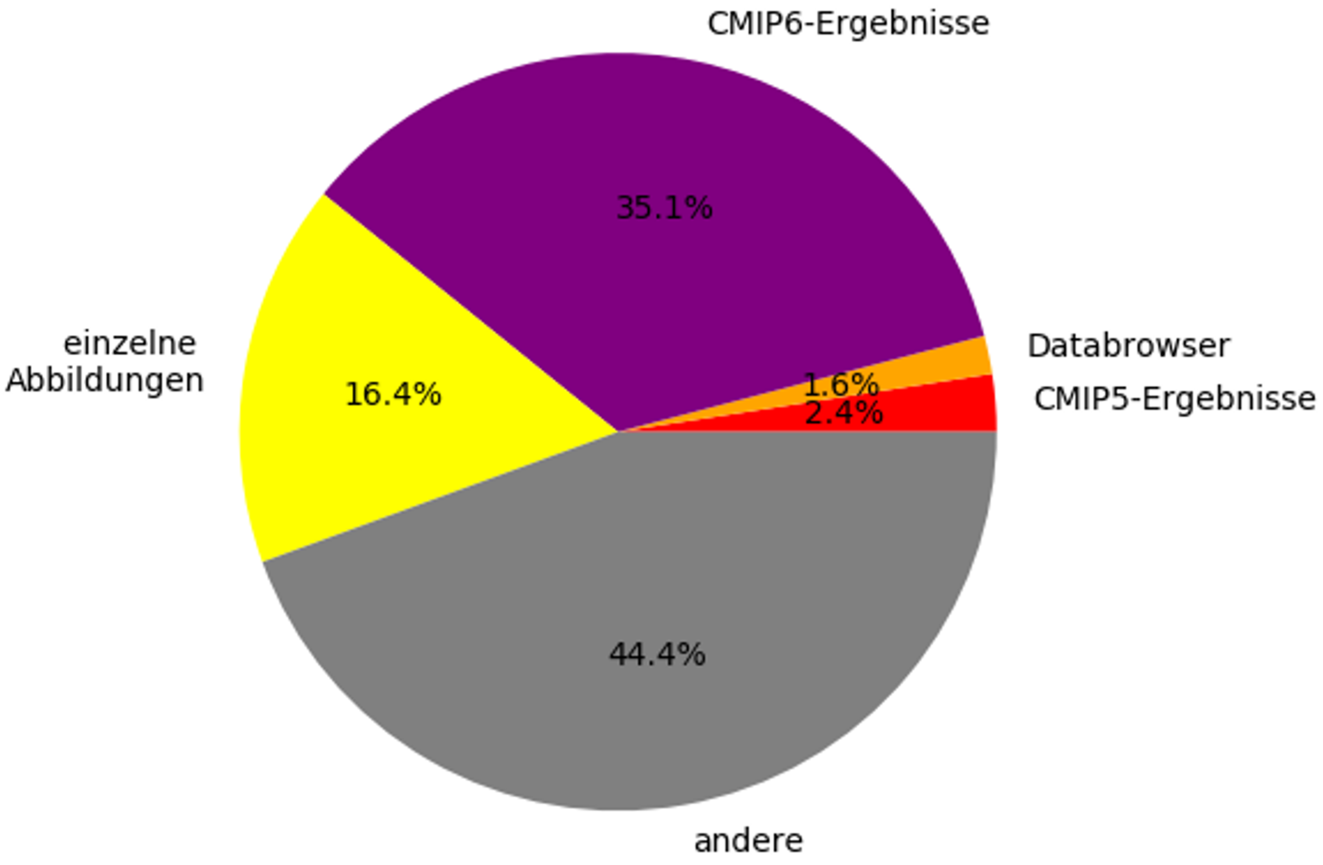
CMIP6-Zugriffe seit Februar 2019

The IS-ENES3 project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

CMIP6 via ESGF to processing: ESMValtool data products

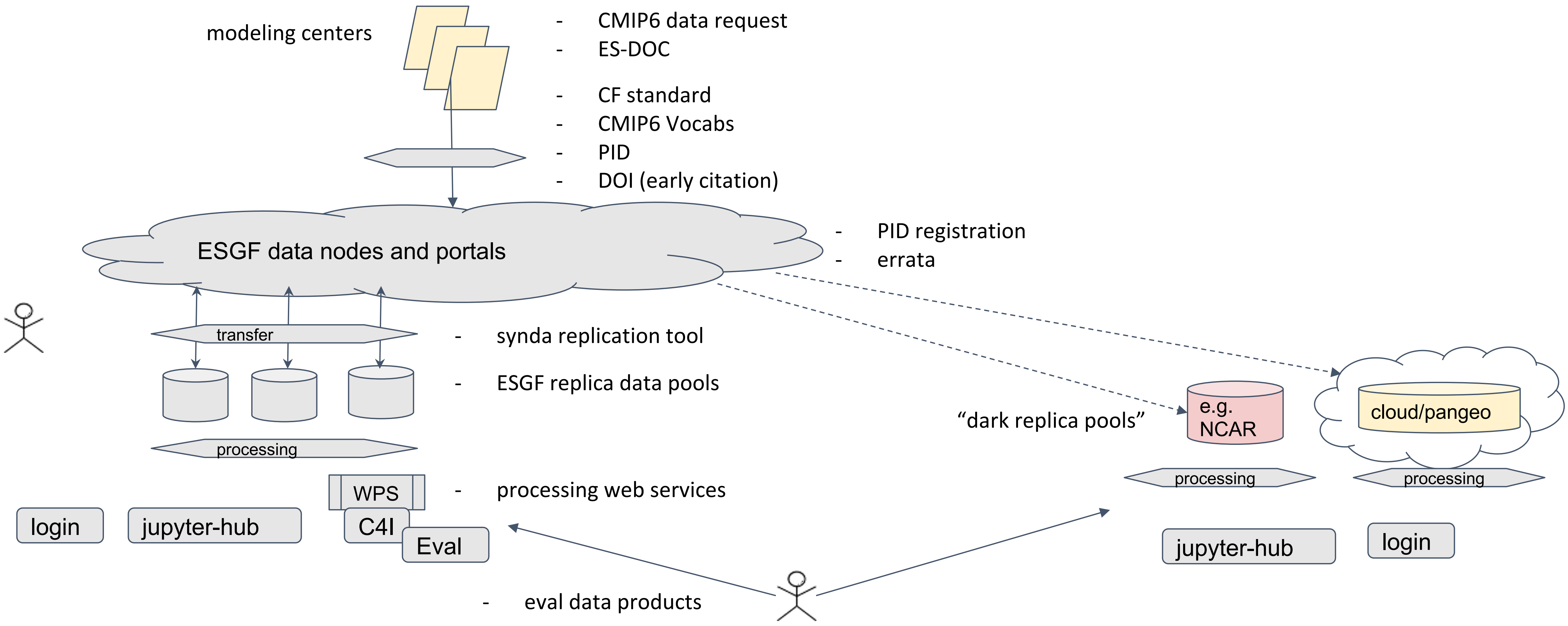
Status:

- operational prototype service
- by now no fully operational monitoring, will start in second half of 2020
- proven interest from research community, yet additional effort needed to operationalize
  - update frequency
  - metadata annotation / result search
  - integration roadmap with C4I open



The IS-ENES3 project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

CMIP6 via ESGF to processing



modeling centers

- CMIP6 data request
- ES-DOC

- CF standard
- CMIP6 Vocabs
- PID
- DOI (early citation)

ESGF data nodes and portals

- PID registration
- errata

transfer

- synda replication tool

- ESGF replica data pools

processing

WPS

- processing web services

login

jupyter-hub

C4I

Eval

- eval data products

“dark replica pools”

e.g. NCAR

cloud/pangeo

processing

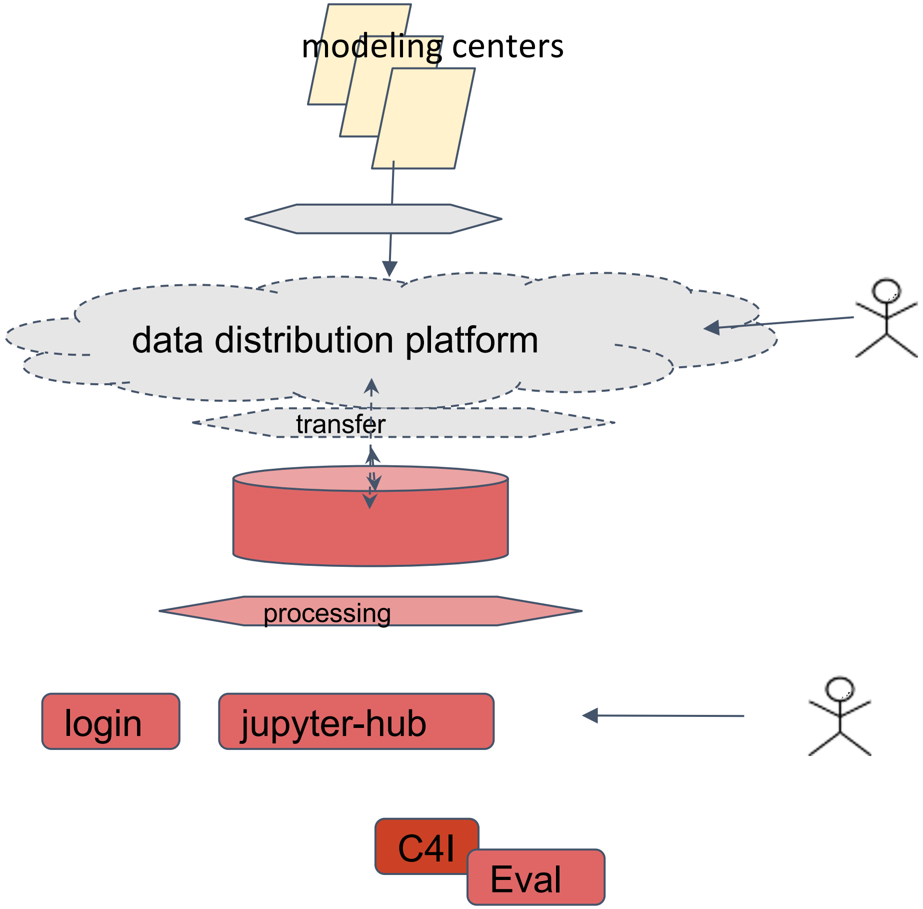
processing

jupyter-hub

login

The IS-ENES3 project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 824084

CMIP6 via ESGF to processing



Centrally managed data pools decouple usage scenarios from "core ESGF data distribution infrastructure":

- direct and jupyterhub based file access
- access to processing ressources

→ Changing requirements for "CORE ESGF" infrastructure and services

Key ENES CDI services quite independent of "CORE ESGF":

- Standards, Vocabs
- PID service, citation service → versioning info, citation info
- errata service
- es-doc service

The IS-ENES3 project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

Thank you

The IS-ENES3 project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

CMIP6 via ESGF to processing: Viewpoints

- end user view
  - (ESGF) search, (ESGF) download, process
- data pool view on CMIP6
  - search, locate, process
- downstream view: e.g. Evaluation portal, C4I

The IS-ENES3 project has received funding from the European Union’s Horizon 2020
research and innovation programme under grant agreement No 824084

is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK
FOR EARTH SYSTEM MODELLING

The Consortium

Coordinated by **CNRS-IPSL**, the IS-ENES3 project gathers **22 partners** in **11 countries**

with

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N°824084

Our website
https://is.enes.org/

Follow us on Twitter !
@ISENES_RI

Contact us at
is-enes@ipsl.fr

Join the community on ZENODO !