

ENES Data Store :

ENES DS data pool & tools

Dissemination level: Public

Disclosing Party: Atef Ben Nasser Institut Pierre Simon Laplace - CNRS

Recipient Party: EGI Webinar



Plan



- ENES Data Space
 - nature of data
 - data in numbers

- Toolset
 - Brief presentation & short term goals
 - Earth Science Documentation tool suite

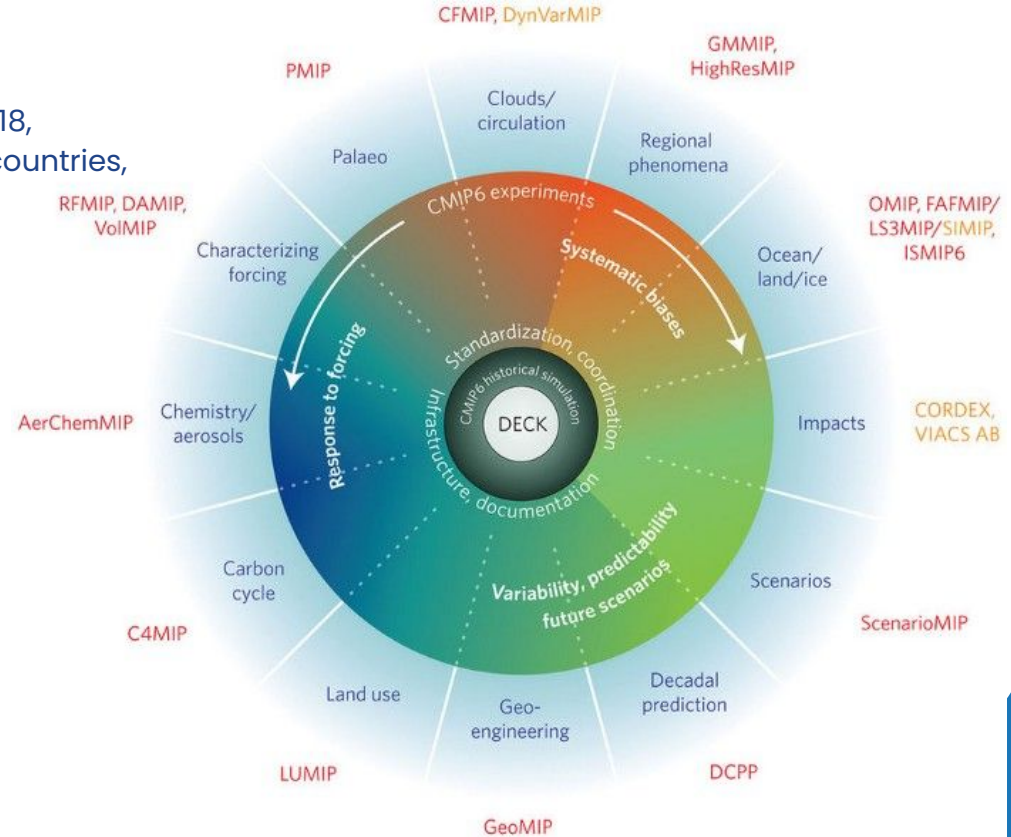
ENES Data Space

CMIP Data



The MIPs in CMIP6. Image source: [Simpkins \(2017\)](#).

- CMIP data (Coupled Model Intercomparison Project)
 - CMIP phase 6 in particular,
 - Planned in 2013, data distribution started in 2018,
 - 23 support MIPs, 33 modeling groups from 16 countries,
 - Available for public use through ESGF network



ENES Data Space

CMIP Data

- Earth System Grid Federation (ESGF for short) network:
 - Peer-to-peer Apache Solr indexes
 - Divided into Data, index and compute nodes in various combinations
 - Operational since 2011



ENES Data Space

CMIP Data



- We made available 1.5Tb of ESGF data to ENES DS users:
 - Data was replicated using Synda software from ESGF nodes across the globe
 - Variables available:
 - tas, pr, uas, vas, huss, hurs
 - In 2 frequencies:
 - Monthly mainly
 - Daily for tas only (data storage constraints)
 - From 28 models:
 - ACCESS-CM2 ACCESS-ESM1-5 AWI-CM-1-1-MR BCC-CSM2-MR CAMS-CSM1-0 CanESM5 CAS-ESM2-0 CESM2-WACCM CMCC-CM2-SR5 CMCC-ESM2 EC-Earth3 EC-Earth3-Veg EC-Earth3-Veg-LR FGOALS-f3-L FGOALS-g3 GFDL-ESM4 IITM-ESM INM-CM4-8 INM-CM5-0 IPSL-CM6A-LR KACE-1-0-G MIROC6 MPI-ESM1-2-HR MPI-ESM1-2-LR MRI-ESM2-0 NorESM2-LM NorESM2-MM TaiESM1
 - From 4 experiments:
 - Historical, ssp245, ssp370, ssp585

ENES Data Space

Toolset



- For this replication effort we used the official ESGF replication and download tool Synda
- Synda: Synchronize your Data
 - Started development at IPSL back in 2011
 - Still in development to keep up with newer MIPs and ESGF evolutions
 - Currently packaged and available for download through conda
 - Official documentation : <https://prodiguer.github.io/synda/>
 - Github repo: <https://github.com/Prodiguer/synda/>

ENES Data Store

Toolset
synda in a nutshell

BASIC
USAGE

- ESGF mirroring
- Command-line interface
- Parameter discovery
- Data discovery
- Optimized, asynchronous, parallel download possibility

ADVANCED
USAGE

- Local Data management
- Use nearest replica
- Downloading supervision

ENES Data Space

Toolset



- To be expected from Synda soon:
 - Docker container image
 - Software citation for your future works
 - Faster replication and download capabilities
 - A smoother UX even for non-power users

ENES Data Space

Toolset



- What if my data is not available in the data pool?
 - New data can still be published on ESGF
 - As well as newer versions
 - We run frequent discoveries on the data samples described earlier
 - Synda has the ability to only add new data to the data pool smoothly, we just need a (reasonable) data request. Storage is not cheap !
 - A daily script also updates the intake-esm catalogue to reflect new data

ENES Data Space

Toolset



- What if I don't understand what these facets mean?
 - Earth System Documentation : <https://es-doc.org/>
 - ES-DOC explorer: <https://explore.es-doc.org/>
 - ES-DOC search: <https://search.es-doc.org/>

ENES Data Space

Toolset



- What if something is wrong with my data?

Generally speaking, publishing a newer version of a dataset is motivated by an errata

- ES-DOC errata DB : <https://errata.es-doc.org/>
- Issue viewer for unauthorized users for the moment
- Will soon enable users to directly report issues that are moderated by data providers
- Can be programmatically queried to ensure file/dataset safety



Contact: abennaser@ipsl.fr

Synda github repo: <https://github.com/Prodiguer/synda/>

Errata github repo: <https://github.com/ES-DOC/>

ES-DOC repo: <https://github.com/ES-DOC/>

ESGF landing page: <https://esgf.llnl.gov/>



Thank you!

Contact: egi-ace-po@mailman.egi.eu
Website: www.egi.eu/projects/egi-ace



[EGI Foundation](#)



[@EGI_eInfra](#)



EGI-ACE receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 101017567.