

Christian Pagé (Cerfacs, France)

christian.page@cerfacs.fr <http://linkedin.com/in/pagechristian> https://www.researchgate.net/profile/Christian_Page <http://cerfacs.fr/~page>
CECI, Université de Toulouse, CNRS, Cerfacs, Toulouse, France

Wim Som de Cerff & Maarten Plieger & Alessandro Spinuso & Ernst de Vreede (KNMI, Netherlands) Niels Drost (Netherlands eScience Center)

Download Poster!

I New Challenges for Science

Year	CMIP5	CMIP6	CMIP7
Power factor	1	30	1000
Npp	200	357	647
Resolution [km]	100	56	31
Number of mesh points [millions]	3.2	18.1	108.4
Ensemble size	120	214	388
Number of variables	800	1068	1439
Interval of 3-dimensional output (hours)	6	4	3
Years simulated	90000	120170	161898
Storage density	0.00002	0.00002	0.00002
Distributed Archive Size (Pb)	3.19	86.05	2260.20

FIG 1: Climate Model Intercomparison Projects (CMIP) Archive Size (PB)

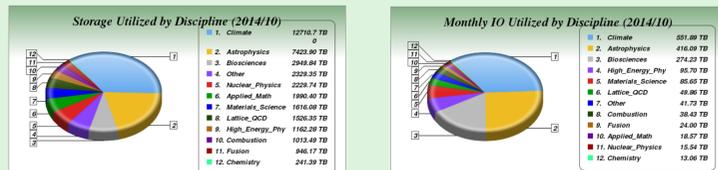


FIG 2: National Energy Research Scientific Computing Center (NERSC) Storage and I/O by Discipline

- Large needs for Storage and I/O
- Heterogeneous communities of users

IV Needs for Evolution

- Modernize the front-end: Drupal → ReactJS
- Separate the front-end and the processing back-end
- Improve the ergonomics and the responsiveness of the Search Interface
- Improve the users' experience
 - Hide complexity
 - Move from file to data-oriented approach
- Incorporate more complete Provenance & Lineage
- Have users control and build their own workflows, and share them
- Bring advanced functionalities such as extra-tropical/tropical cyclone tracking

II Common Users' Needs

- Guidance/tools for data and scenarios subsetting: selecting a subset of representative climate scenarios
- Lower significantly the total data download size
- Calculate as much as possible remotely
- Reformat/Repackage the data into easier formats
- Access full Provenance and Lineage
- Proper Metadata description (derived data)
- Variety of Access Interfaces: GUI, OGC, REST APIs, Jupyter Notebooks, ...



{RESTful API}

Take Home Messages

- Current Data Analysis Workflow is no longer possible: processing delegation is needed
- Heterogeneous Processing Backends are available
- It is necessary to hide underlying complexity
- Provenance & Lineage is essential
- Precise (Metadata-)Standards are mandatory

V Connecting External Resources

- Connect to external computing/storage resources:
 - Clouds (AWS, etc.)
 - e-infrastructures:
 - EUDAT CDI
 - European Science Cloud (EOSC)
 - DARE Platform
 - ESGF Computing Nodes (CWT)

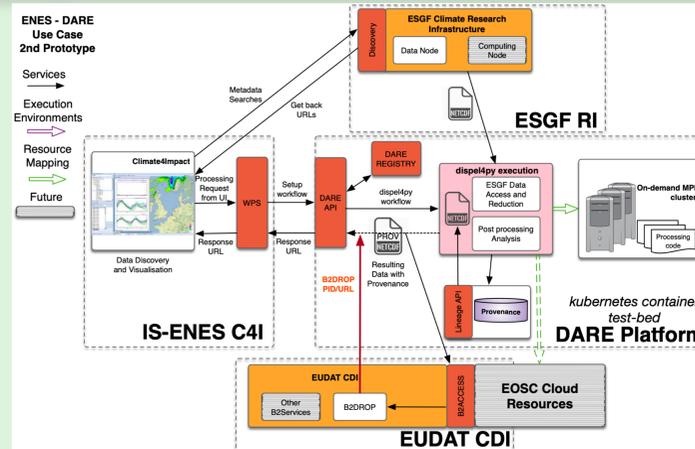


FIG 4: Prototype Integration

III climate4impact 1.0

<https://climate4impact.eu>



- Developed and managed by IS-ENES since 2010
- Not only UI, but also Services (WPS, WCS,...)
- Tailored for end-users
- Supports on-demand data processing
- Now containerized version
 - docker
 - docker-compose

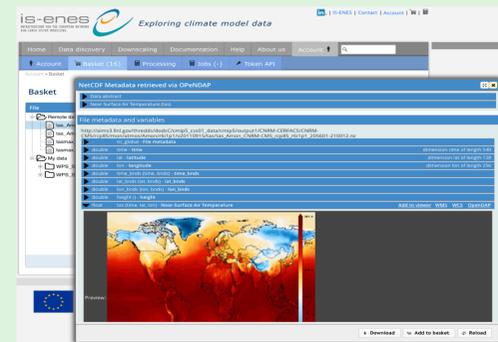


FIG 3: C4I Faceted Search and Interface

VI Current Work: C4I 2.0

- Possible micro-services approach
 - Python/Flask-based
 - Reuse old java code from version 1.0 if needed
- Refactor whole documentation and guidance using S3 Bucket for content storage
- Implement a Vocabulary Service
- Restructure and optimize icclim backend processing
- Evaluate possible C4I/WPS Proxy
- Services with external APIs
 - MyCollection (Basket)
 - OGC-WPS using Birdhouse Framework
- Support for Climate Infrastructure (ESGF) Computing Nodes (CWT) for Pre-Processing Data
- All code will be released open-source**