

# Climate Data Analysis: Enhancing Usage of Research Data with Climate4impact, EUDAT and DARE

Christian Page<sup>1</sup>, Wim Som de Cerff<sup>2</sup>, Maarten Plieger<sup>2</sup>, Alessandro Spinuso<sup>2</sup>, Ernst de Vreede<sup>2</sup>, and Niels Drost<sup>3</sup>

<sup>1</sup>CERFACS European Centre for Research and Advanced Training in Scientific Computation

<sup>2</sup>Royal Netherlands Meteorological Institute

<sup>3</sup>Netherlands eScience Center

November 21, 2022

## Abstract

Researchers using climate data are facing challenge to analyze the data they need. Data volumes are increasing very rapidly, and the ability to download all needed data is often not a possibility anymore. A platform called climate4impact (C4I) has been designed and developed to enhance the use of research data, to support researchers with analytics and to support other climate portals. It is currently under development within the European Project IS-ENES3 and builds on previous developments from previous IS-ENES projects, CLIPC and C3S-Magic. C4I offers a front-end and standard services (with APIs) on top of the climate data infrastructure, and it can be visited at <https://climate4impact.eu>. The current version provides processing services include climate indicator calculations, country based statistics and polygon extraction. C4I makes use of the DKRZ Birdhouse framework, which is an extendable and modular processing framework based on PyWPS. Data is obtained from various ESGF nodes using secure OpenDAP. C4I provides a personal basket where users can upload their own data and do research with the provided tools. The software is open, reusable, modular and packaged. Components are available via docker containers to enable easy re-use. The on-demand calculations are taking place on the front-end server, and this is not scalable and can lead to performance problems. Within the DARE project, delegation of the calculations on the DARE Platform using the DARE API has been implemented and tested in a prototype, using EUDAT B2DROP as an intermediate storage service. It is to be noted that the DARE Platform as well as the EUDAT B2 Services should be interoperable with the European Open Science Cloud (EOSC). This prototype service delegation will be made operational during the upcoming year. In the IS-ENES3 project, the web portal will be redesigned with a completely new architecture using a micro-services and containerized approach, building on experience gained during the previous projects. The next version of the portal will be built using the React framework, which allows for creating large web applications which can change data, without reloading the page. We are actively seeking input from current as well as potential users at this time, to make the next version of C4I useful to as many people as possible. The material presented here is made possible because the IS-ENES3 project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°824084.

## I New Challenges for Science

	CMIP5	CMIP6	CMIP7
Year	2012	2017	2022
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)

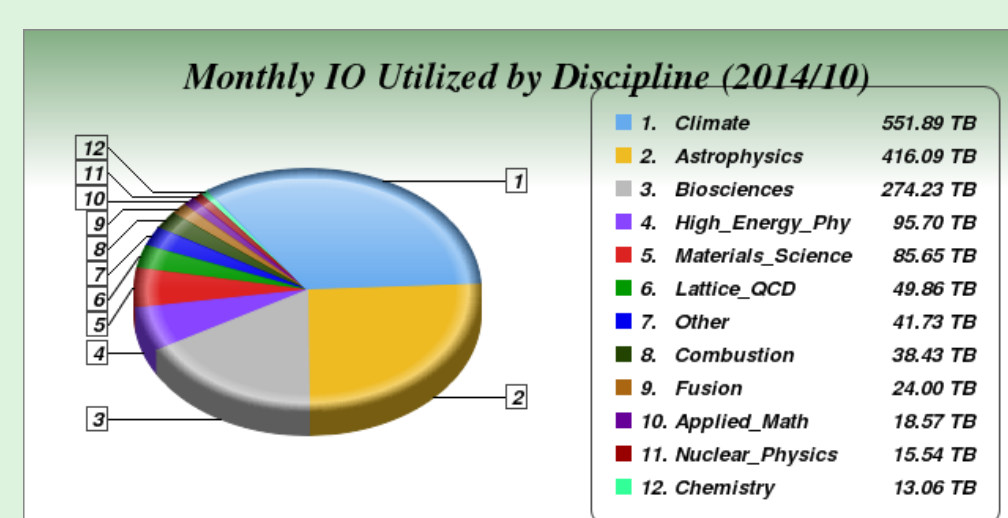
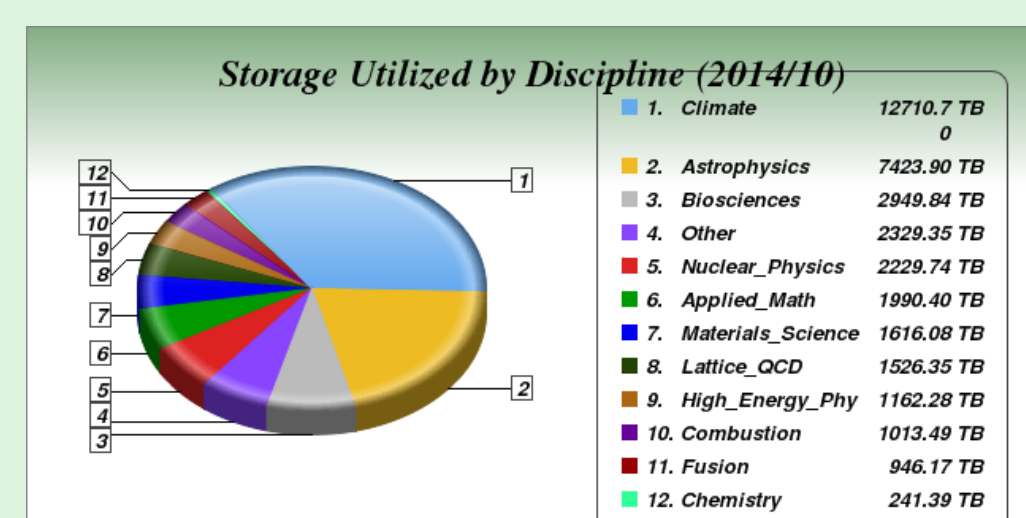


FIG 2: National Energy Research Scientific Computing Center (NERSC) Storage and I/O by Discipline

- 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

1. Current Data Analysis Workflow is no longer possible: processing delegation is needed
2. Heterogeneous Processing Backends are available
3. It is necessary to hide underlying complexity
4. Provenance & Lineage is essential
5. 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)

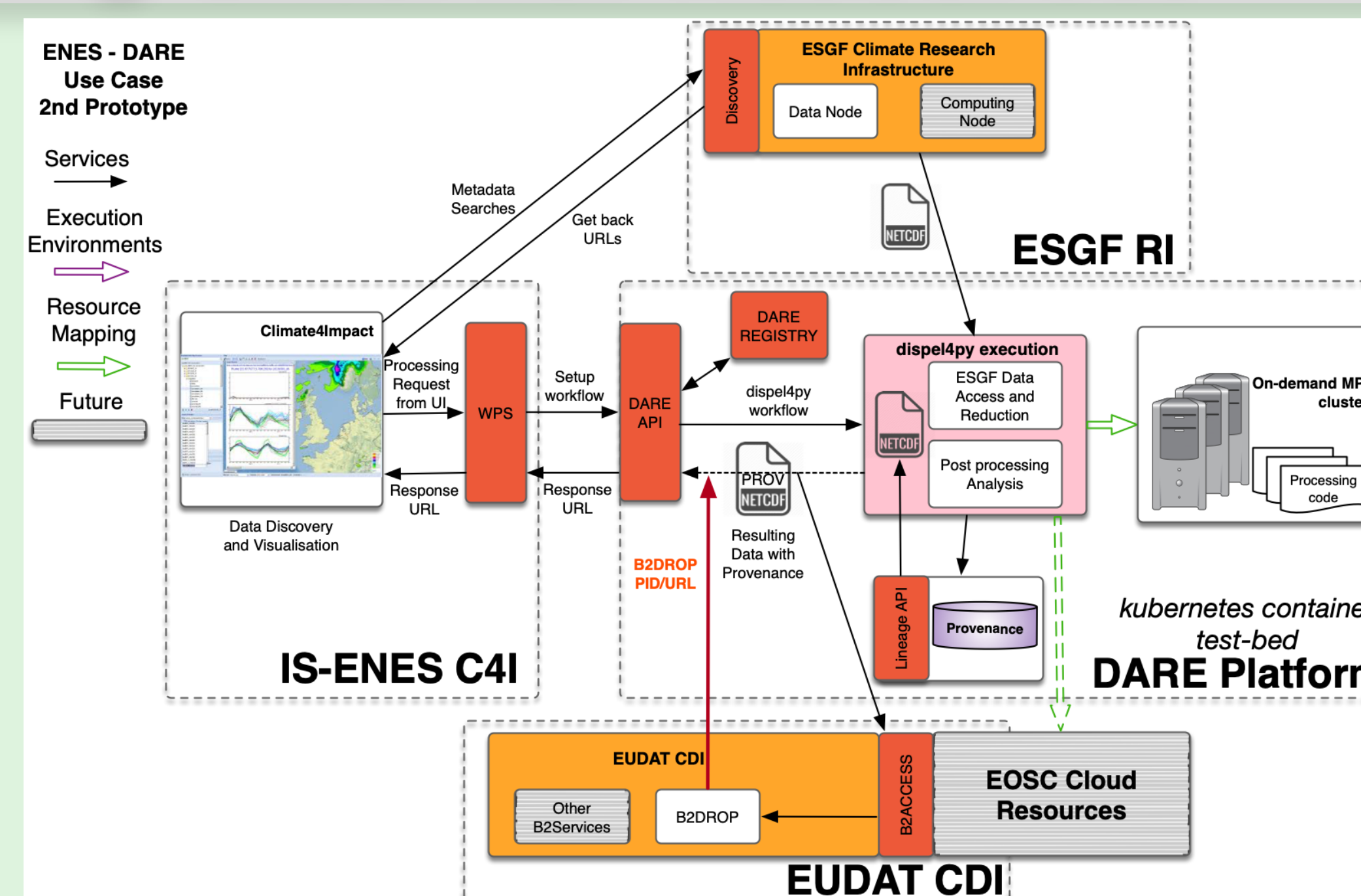


FIG 4: Prototype Integration

## 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

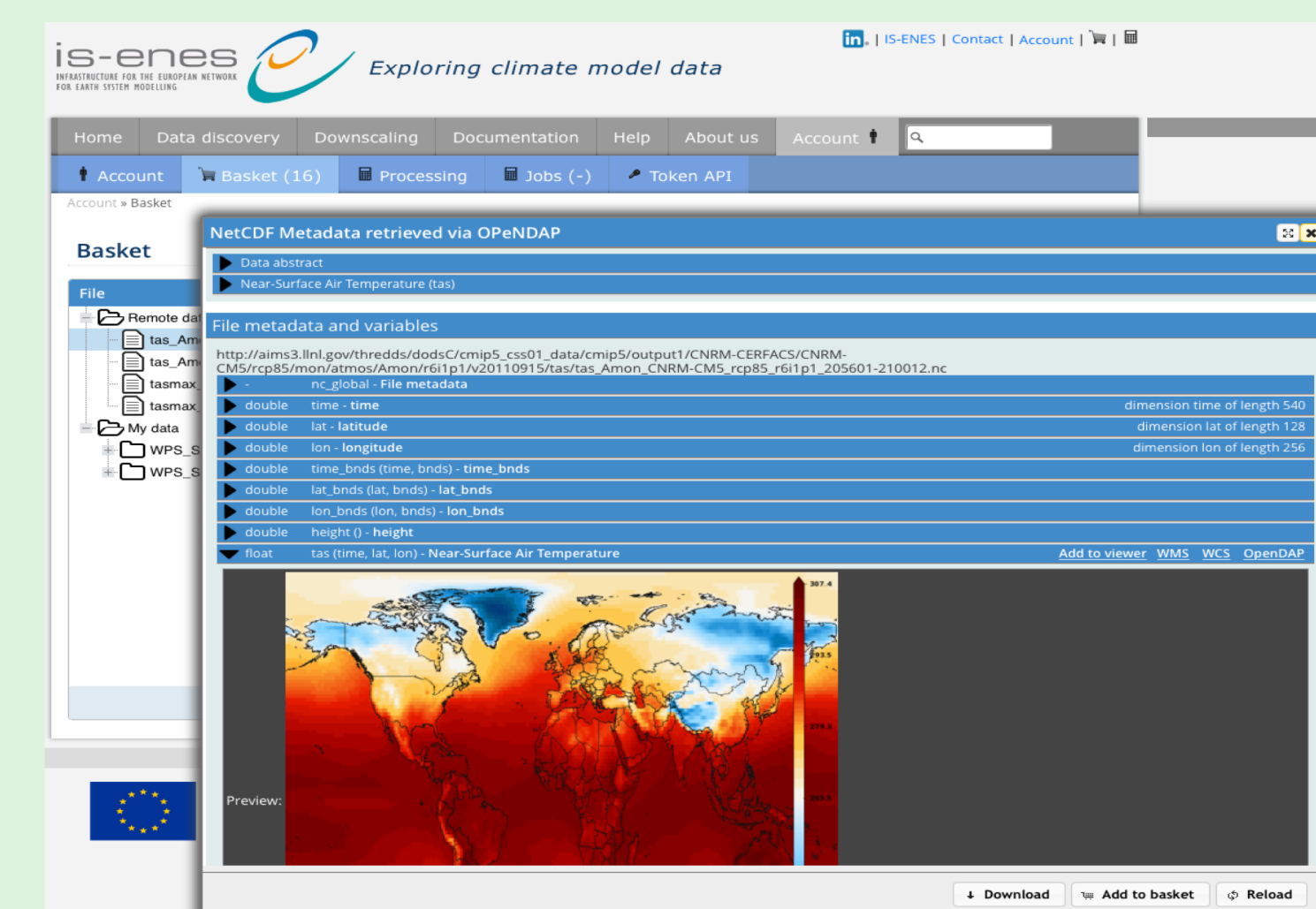
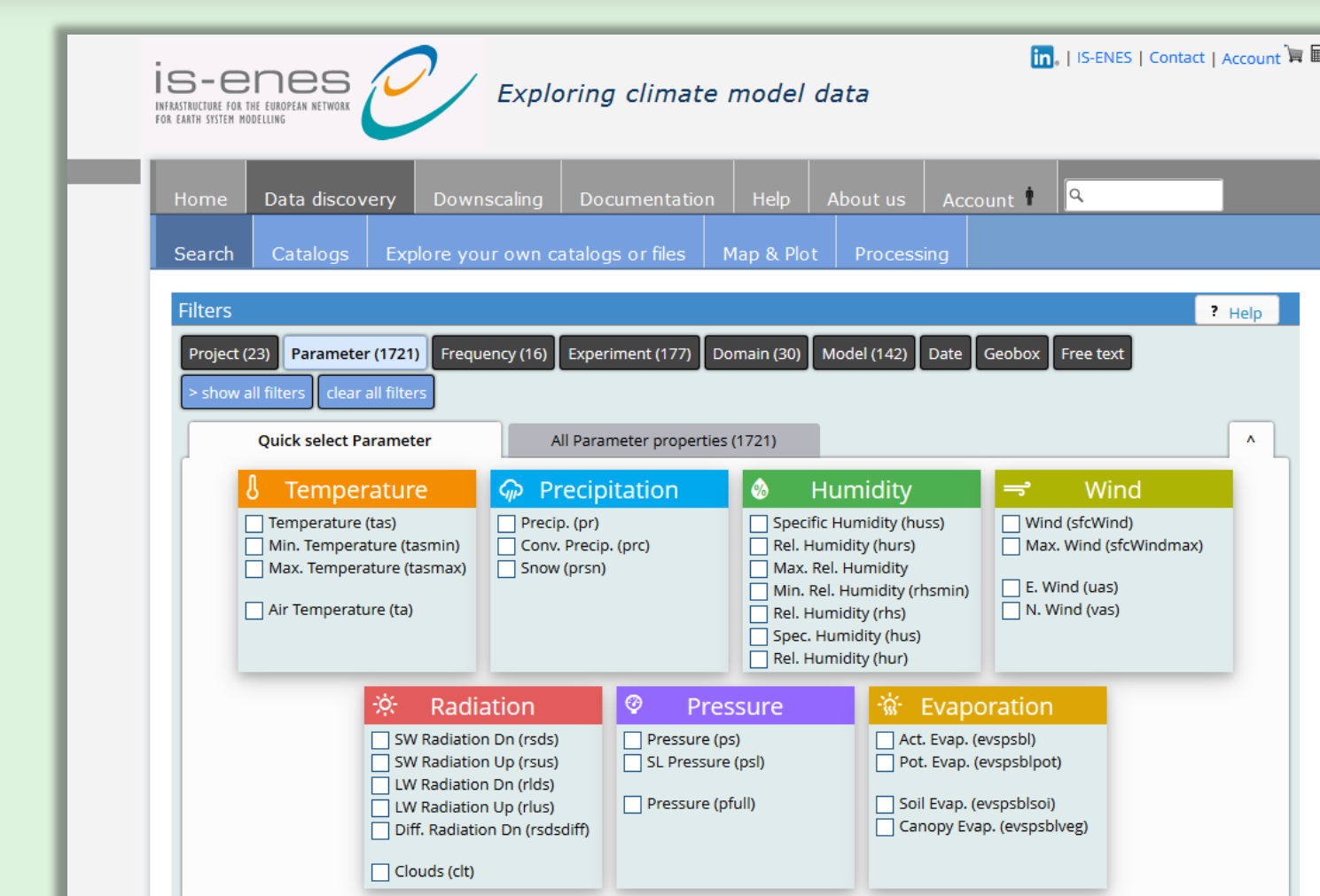


FIG 3: C4I Faceted Search and Interface

## 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**