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

CERFACS

ESGF F2F Workshop
Washington, DC, December 2016

Server-side computing services provided by IS-ENES through the climate4impact.eu platform

CERFACS, KNMI, University of Cantabria, SMHI, Wageningen University & Research, CNRS-IPSL, CMCC, STFC, DMI, INHGA

Christian Pagé

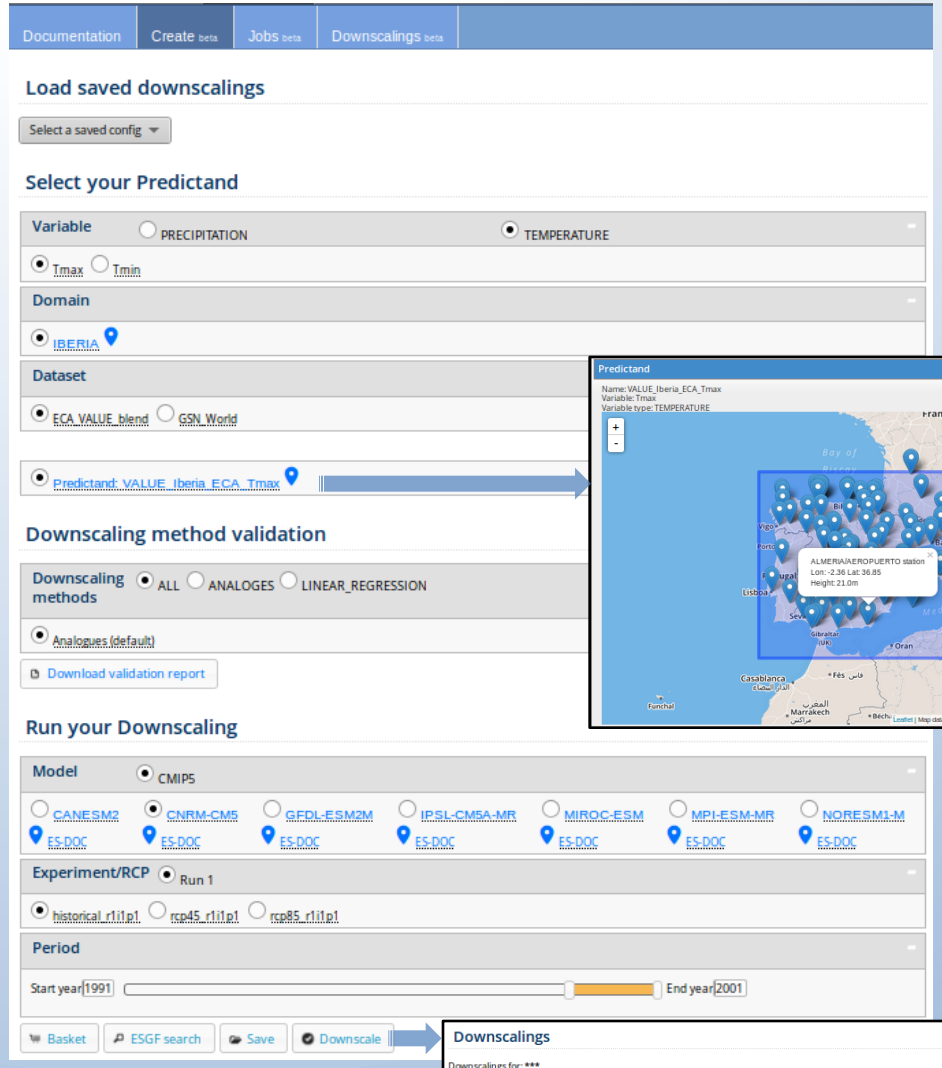
- Platform for impact researchers to explore climate data and perform analysis
- In-depth documentation and guidance
- Use cases from impact researchers
- **Perform calculations / Data processing – WPS**

http://climate4impact.eu/

http://climate4impact.eu/

Online on-demand calculations

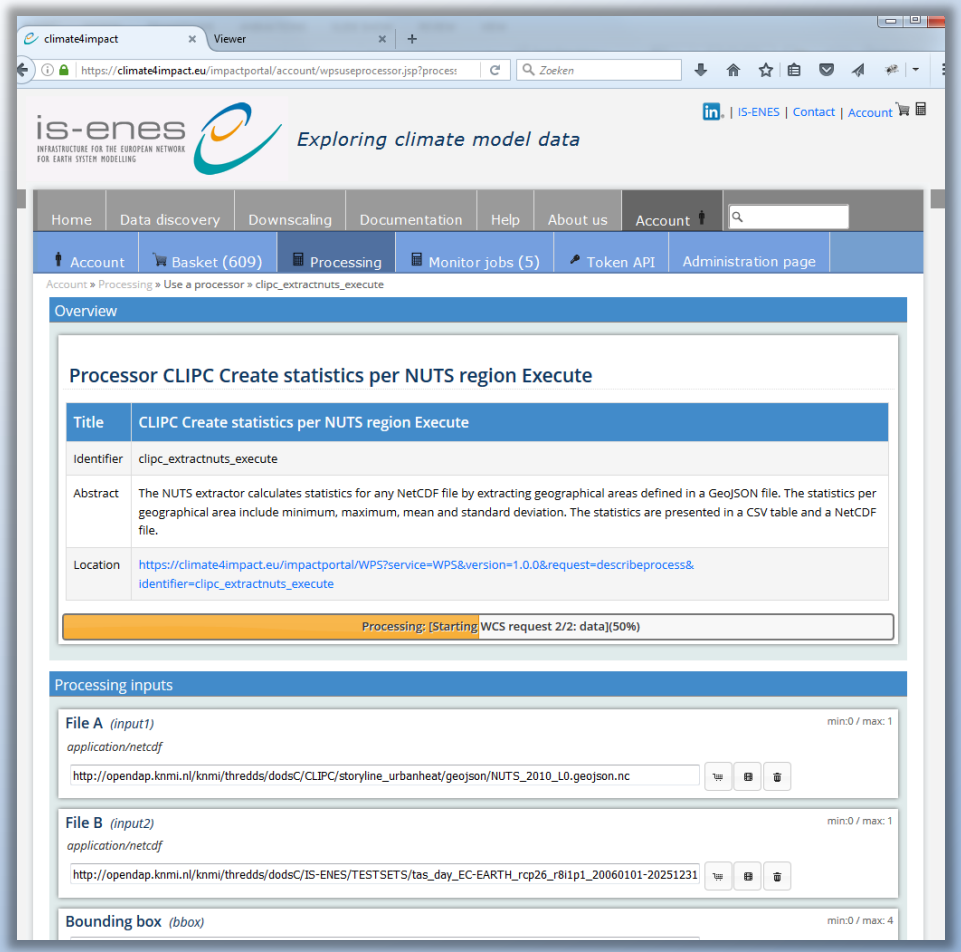
- C4I Statistical Downscaling REST API
  - Services provided by the University of Cantabria servers
  - Connected to ESGF
  - Friendly user interface on C4I
- C4I Climate Indices
  - All ETCCDI indices and simple statistics available
  - Native Python open-source ICCLIM software (fully validated against R.climdex)
  - Expandable to climate indicators as well



ETCCDI: The joint CCI/CLIVAR/JCOMM Expert Team (ET) on Climate Change Detection and Indices

Web Processing Service

- Generic WPS UI
  - Based on Describe Process XML file.
  - Links to basket
  - Selections populated dynamically
- Main WPS Processes
  - Climate indices calculation
  - Subsetting over a large time period
  - Regrid and reformat, extract regions
  - Combine climate indicators (CLIPC)
- Wizards
  - Indices calculation
  - Subsetting / file conversions (planned)



http://climate4impact.eu/

Web Processing Service

Job progress can be viewed from anywhere (also in other portals using the services, e.g. CLIPC)

Account | Basket (621) | Processing | Monitor jobs (10) | Token API | Administration page

Account » Monitor jobs

Submitted processing jobs

Created at	Name	Status location	Progress	View	X
2016-12-06 20:36:36Z	clipc_extractnuts_execute	pywps-aec7c400-bbf3-11e6-ac70-78e3b502c7d8.xml	ready	view	X
2016-12-06 20:35:32Z	clipc_simpleindicator_execute	pywps-8835e09c-bbf3-11e6-8874-78e3b502c7d8.xml	failed	view	X
2016-12-06 14:02:50Z	clipc_extractnuts_execute	pywps-ac876b46-bbbc-11e6-86ec-78e3b502c7d8.xml	ready	view	X
2016-12-05 10:32:43Z	clipc_extractnuts_execute	pywps-27bdecbe-bad6-11e6-96df-78e3b502c7d8.xml	ready	view	X
2016-12-05 10:32:00Z	clipc_extractnuts_execute	pywps-0ddf0c60-bad6-11e6-92df-78e3b502c7d8.xml	ready	view	X
2016-12-03 19:25:45Z	knmi_advanced_combine	pywps-4984287a-b98e-11e6-8d4b-78e3b502c7d8.xml	ready	view	X
2016-12-03 19:24:13Z	clipc_extractnuts_execute	pywps-1266efc6-b98e-11e6-a9ae-78e3b502c7d8.xml	ready	view	X
2016-12-02 13:16:55Z	clipc_combine_identify	pywps-98511b00-b891-11e6-9307-78e3b502c7d8.xml	ready	view	X
2016-12-02 12:55:03Z	knmi_advanced_combine	pywps-8a6e920e-b88e-11e6-b96f-78e3b502c7d8.xml	ready	view	X
2016-12-02 10:28:06Z	clipc_extractnuts_execute	pywps-03058ed0-b87a-11e6-a30f-78e3b502c7d8.xml	ready	view	X

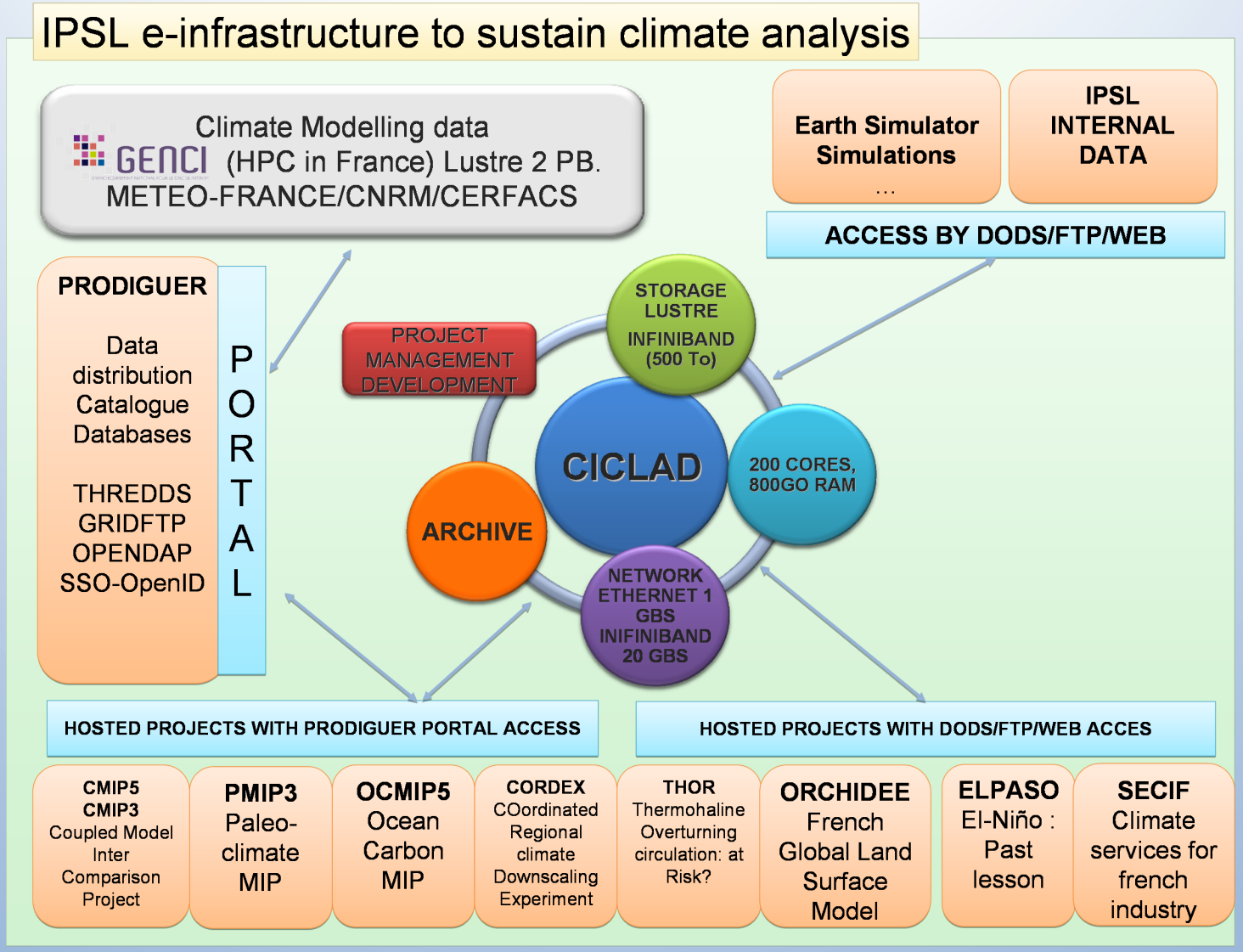
Calculations near storage (or away from user's servers)

- **Dedicated high performance systems**
  - Replicate all needed data onto one large data service: all local data only
  - Expose services through standard protocols (http REST, WPS, OpenDAP, ...)
  - Give (selected) users command-line access

- **Remote Computing Services**
  - Expose computing services
  - Grab remote data and perform calculations on remote high capacity servers/clusters
  - Only send results to users

- **Central Computing Platform**
  - Portal/platform with standard services (WPS, ...)
  - Orchestrate calculations with, when possible, delegations to external services (computations, storage)
  - Grab results, and make available to users with more services (visualization, quicklook, download, ...)

http://climate4impact.eu/



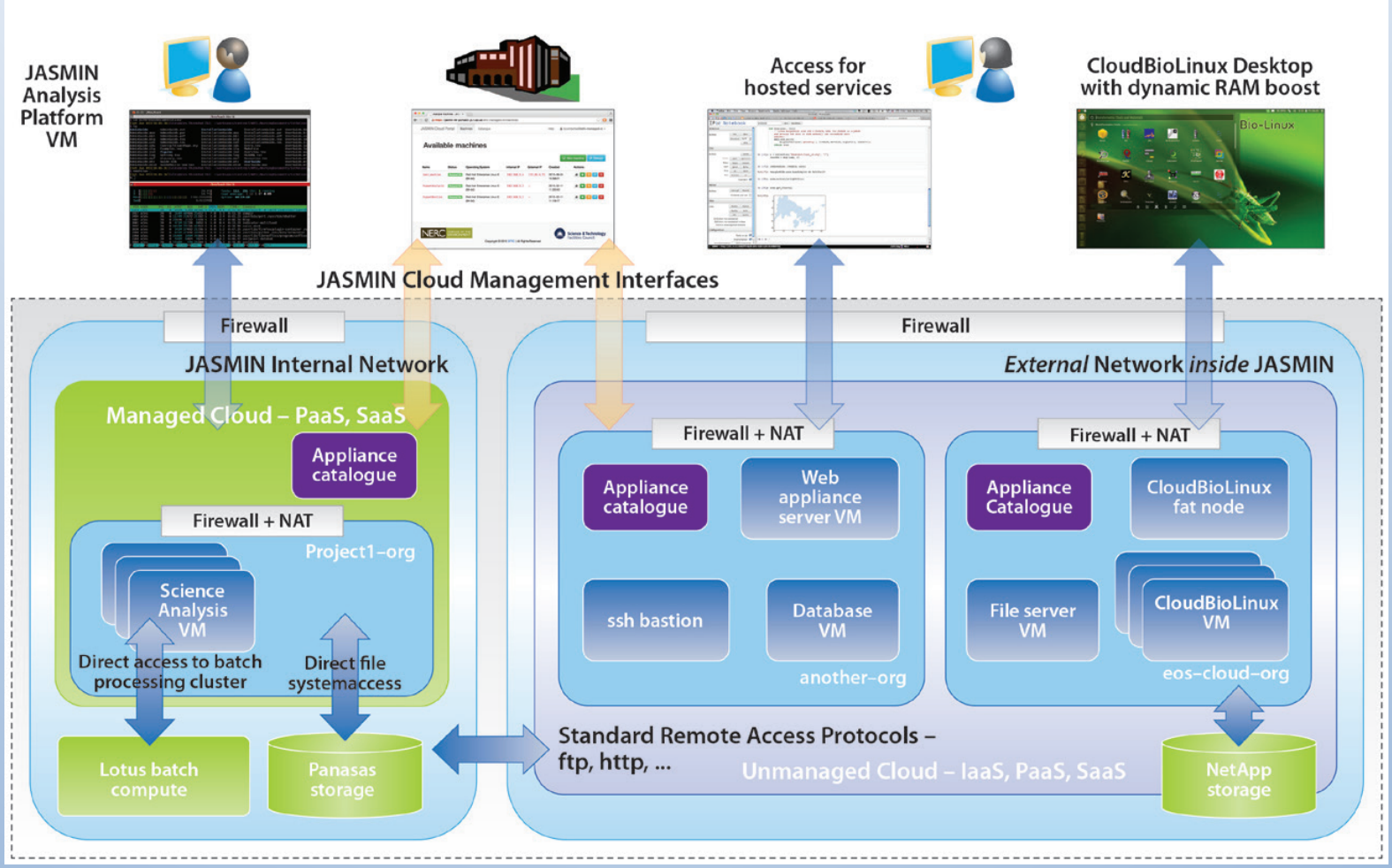
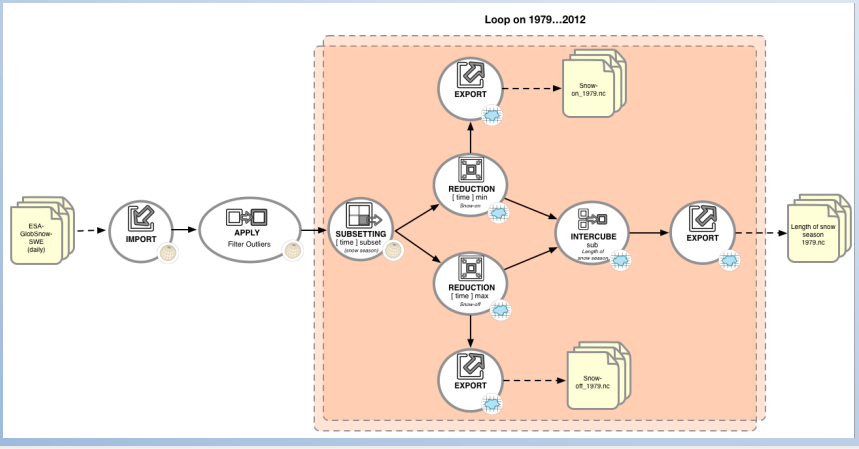


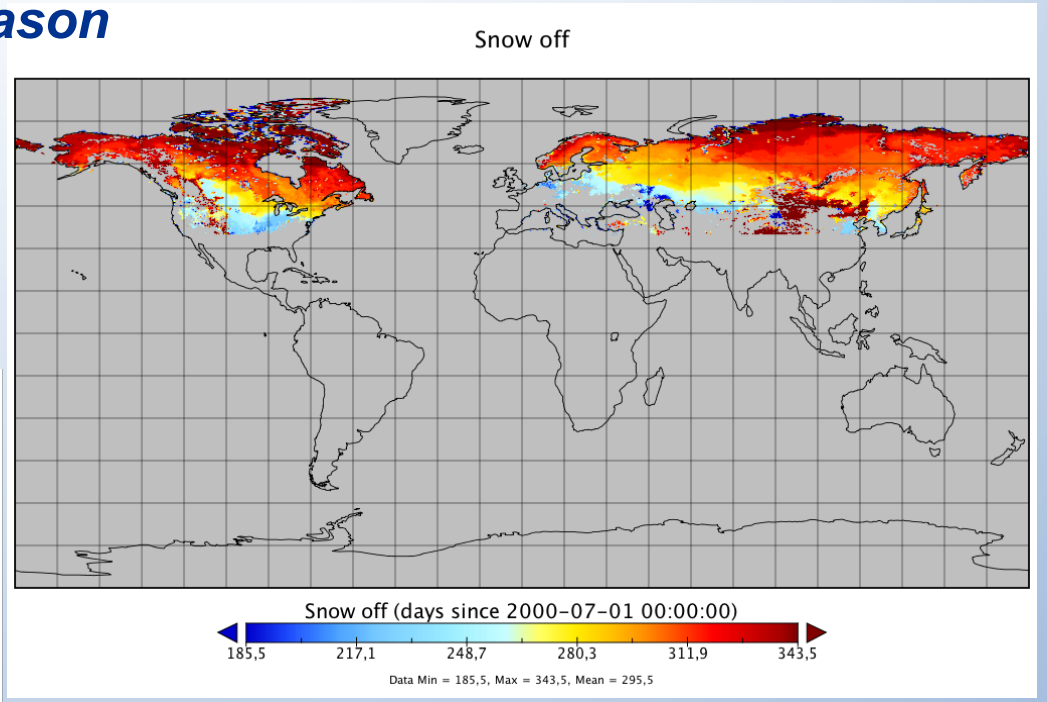
Fig. 13. CEDA's JASMIN analysis platform. *JASMIN integrates cloud architecture, container technologies, and virtual machines to improve flexibility and performance and track maintenance.*

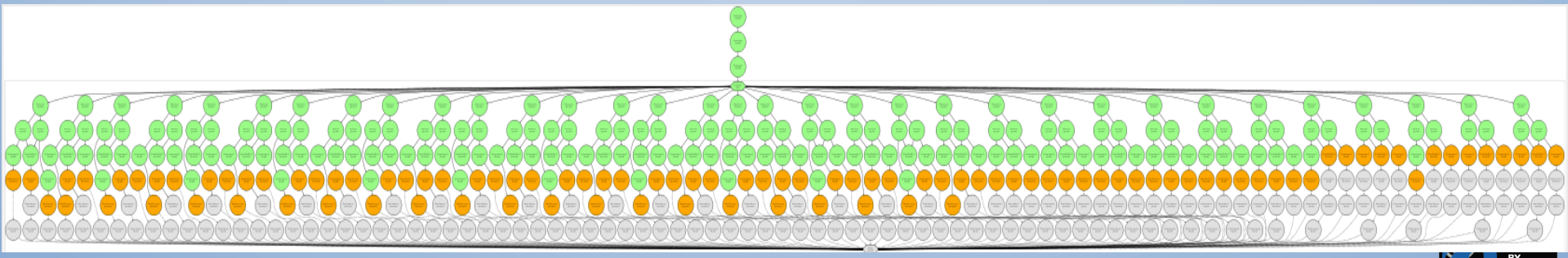
Tier1 indicators using Ophidia (CLIPC)

Snow on/off – Length of snow season

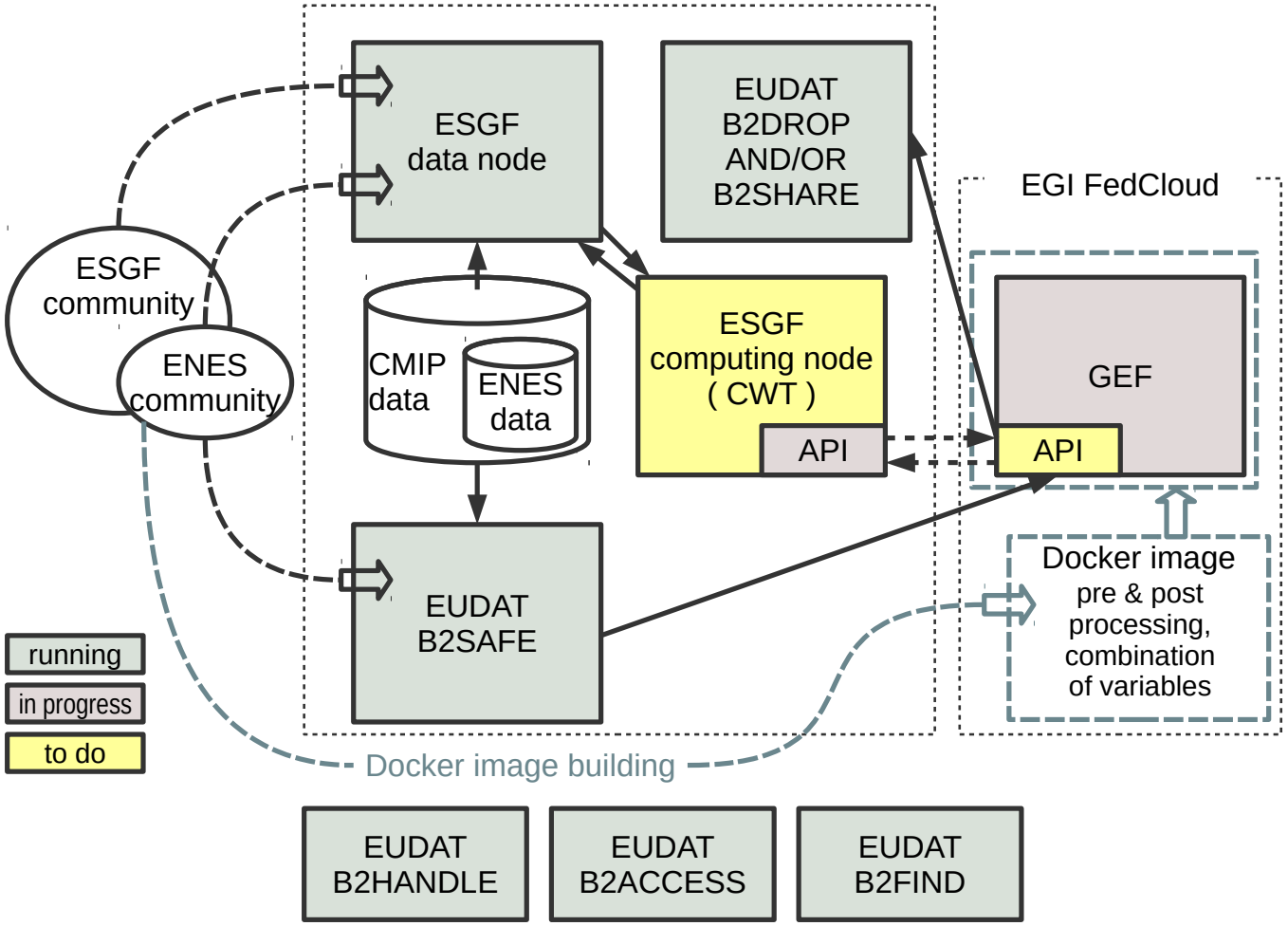
- ✔ *Dataset time range: 1979-2012*
- ✔ ***50 GB** of input data*
- ✔ ***434 tasks** performed*
- ✔ *99 NetCDF output files*







Delegations of some calculations; integration of services



Background: Integration of European and International projects

IS-ENES2 (FP7)
23 partners

Infrastructure for the European Network of Earth System Modelling

Access to large climate datasets for climate impact researchers **C4I**
http://climate4impact.eu/

Objective: **C4I** Delegates selected processing/storage to EUDAT B2 Services

EUDAT2020 (H2020)
35 partners

Collaborative Pan-European infrastructure

Developing Generic Execution Framework: Providing data processing on EGI Federated Cloud using EUDAT B2 Services

Objective: **C4I** backend for on-demand calculations: **icclim**

CLIPC (FP7)
22 partners

Copernicus Climate Information Platform

Developing icclim for climate indices and indicators calculation
Defining metadata standards for automated processing

Objective: **C4I** Delegates selected processing to future ESGF Computing Nodes

ESGF-CWT
Compute Working Team

Worldwide (mainly US-Europe) Collaboration on Data Processing of climate data near data storage

Design of API and Implementation of Processing Capabilities for ESGF

Conclusion and next steps

- ESGF + climate4impact enables impact research
  - Requires ESGF Search API, OpenDAP and THREDDS catalogs
  - Climate4impact processing, search and visualization is a layer over ESGF
  - Support of Downscaling ESGF datasets on the fly

- Climate4impact is flexible due to applied technologies and standards
  - PyWPS with ICCLIM as generic processing framework for climate indices
  - ADAGUC WMS can be used to visualize local and remote files
  - OPeNDAP can be used to access small bits of large files over the internet
  - Many APIs developed to integrate C4I services into CLIPC portal https://dev.knmi.nl/projects/impactportal/wiki/API

- Next steps
  - Implement more use cases from climate impact researchers
  - Climate indices calculation wizard
  - File abstraction: focus on physical parameters

Sustainability Issues

- CLIPC and IS-ENES2 ending
  - Currently 2-3 Coding Sprints per year for development and implementation: needed to improve but not only.
  - Very good core of dev people within IS-ENES2/CLIPC
  - Long to-do list based on user requirements
- What will happen after?
  - New project proposals to further develop and use the C4I platform
  - Limited staff, but we want to continue working together (own institutes funding, related projects)
  - Will need to adapt the platform and portal to future ESGF changes
  - Involvement in ESGF CWT will continue
  - Will need to tackle new datasets, new projects, CMIP6 and +
  - More dissemination needed
  - Support to users
  - Improvements
  - Delegate calculations to ESGF Computing Nodes, External Computing Resources (EGI, Ophidia, etc.)

http://climate4impact.eu/

http://icclim.readthedocs.org/

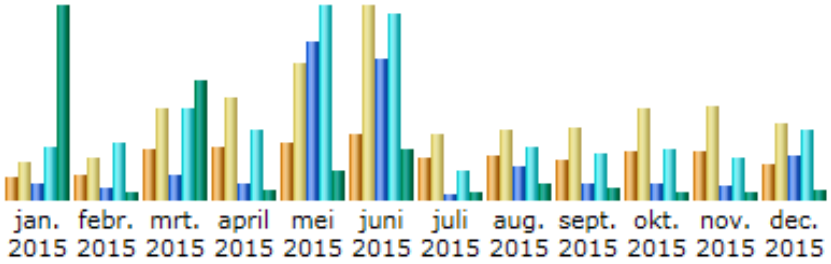
Samenvatting

Rapportageperiode	Maand jan. 2015
Eerste bezoek	01 jan. 2015 - 01.16
Laatste bezoek	31 jan. 2015 - 23.24

	Unieke bezoekers	Aantal bezoeken	Pagina's	Hits	Bytes
Bekeken verkeer *	**1002**	**1644** (1.64 bezoeken/bezoeker)	**43149** (26.24 Pagina's/bezoek)	**142263** (86.53 Hits/bezoek)	**48.08 GB** (30668.11 KB/bezoek)
Niet-bekeken verkeer *			**80506**	**82125**	**3.66 GB**

* „Niet bekeken" is verkeer dat gegenereerd werd door robots of wormen, of respons met een speciale HTTP-statuscode.

Maandelijkse historie



Maand	Unieke bezoekers	Aantal bezoeken	Pagina's	Hits	Bytes
jan. 2015	1002	1644	43149	142263	48.08 GB
febr. 2015	1059	1855	32931	153179	1.90 GB
mrt. 2015	2186	3923	64718	247721	29.51 GB
april 2015	2255	4430	41533	187479	2.46 GB
mei 2015	2490	5920	429910	525534	7.25 GB
juni 2015	2843	8358	383313	505838	12.60 GB
juli 2015	1806	2819	15752	81211	2.01 GB
aug. 2015	1907	3027	91892	145005	4.20 GB
sept. 2015	1688	3099	42509	123899	2.82 GB
okt. 2015	2074	3924	43705	139625	2.08 GB
nov. 2015	2049	4019	39164	115474	1.85 GB
dec. 2015	1497	3317	122402	191622	2.50 GB
Totaal	22856	46335	1350978	2558850	117.26 GB

Samenvatting

Rapportageperiode	Maand jan. 2016
Eerste bezoek	01 jan. 2016 - 00.00
Laatste bezoek	31 jan. 2016 - 23.54

	Unieke bezoekers	Aantal bezoeken	Pagina's	Hits	Bytes
Bekeken verkeer *	**1556**	**3063** (1.96 bezoeken/bezoeker)	**193052** (63.02 Pagina's/bezoek)	**319625** (104.35 Hits/bezoek)	**4.21 GB** (1441.05 KB/bezoek)
Niet-bekeken verkeer *			**159418**	**161593**	**3.08 GB**

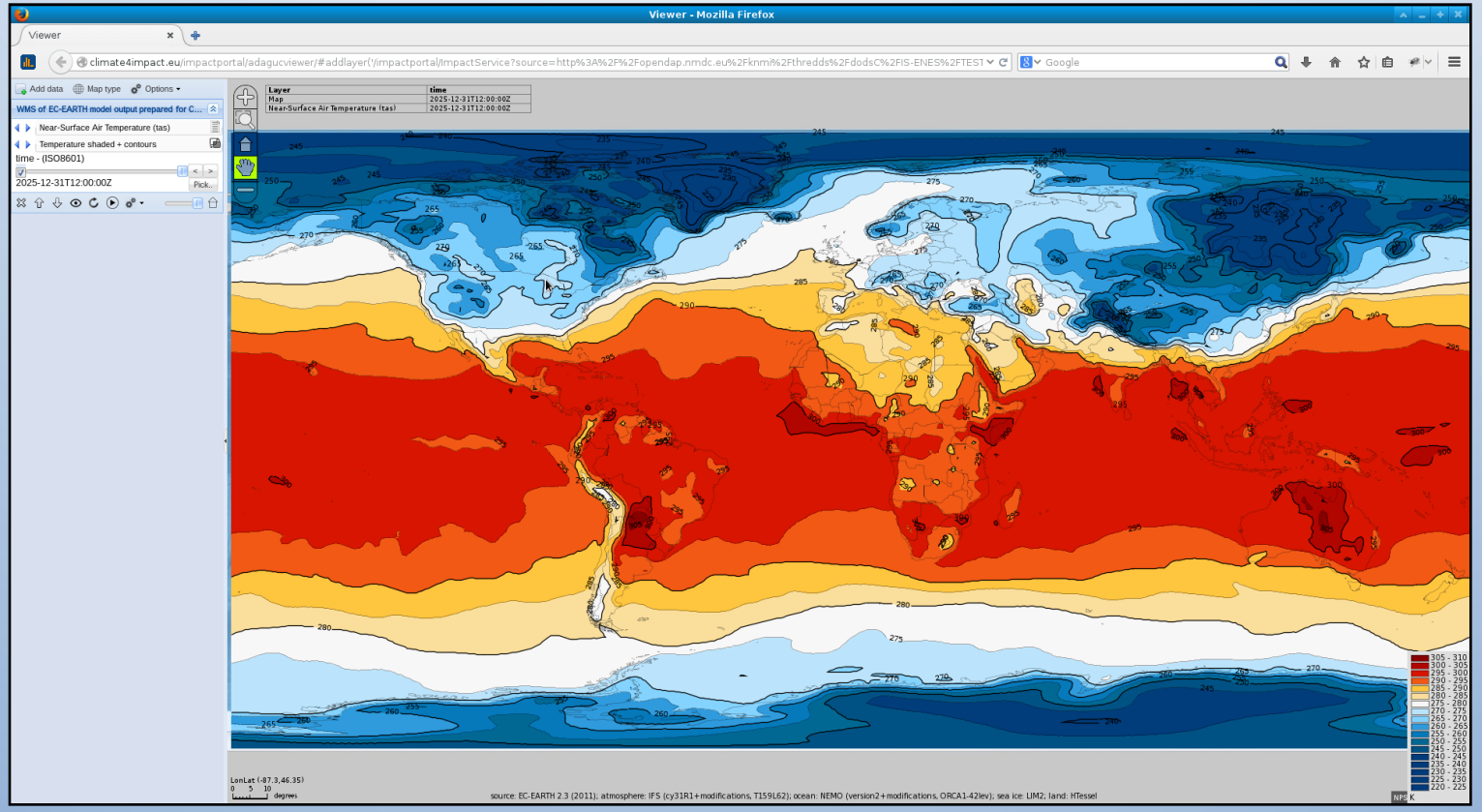
* „Niet bekeken" is verkeer dat gegenereerd werd door robots of wormen, of respons met een speciale HTTP-statuscode.

Maandelijkse historie

Maand	Unieke bezoekers	Aantal bezoeken	Pagina's	Hits	Bytes
jan. 2016	1556	3063	193052	319625	4.21 GB
febr. 2016	955	1684	73453	154556	2.22 GB
mrt. 2016	0	0	0	0	0
april 2016	0	0	0	0	0
mei 2016	0	0	0	0	0
juni 2016	0	0	0	0	0
juli 2016	0	0	0	0	0
aug. 2016	0	0	0	0	0
sept. 2016	0	0	0	0	0
okt. 2016	0	0	0	0	0
nov. 2016	0	0	0	0	0
dec. 2016	0	0	0	0	0
Totaal	2511	4747	266505	474181	6.43 GB

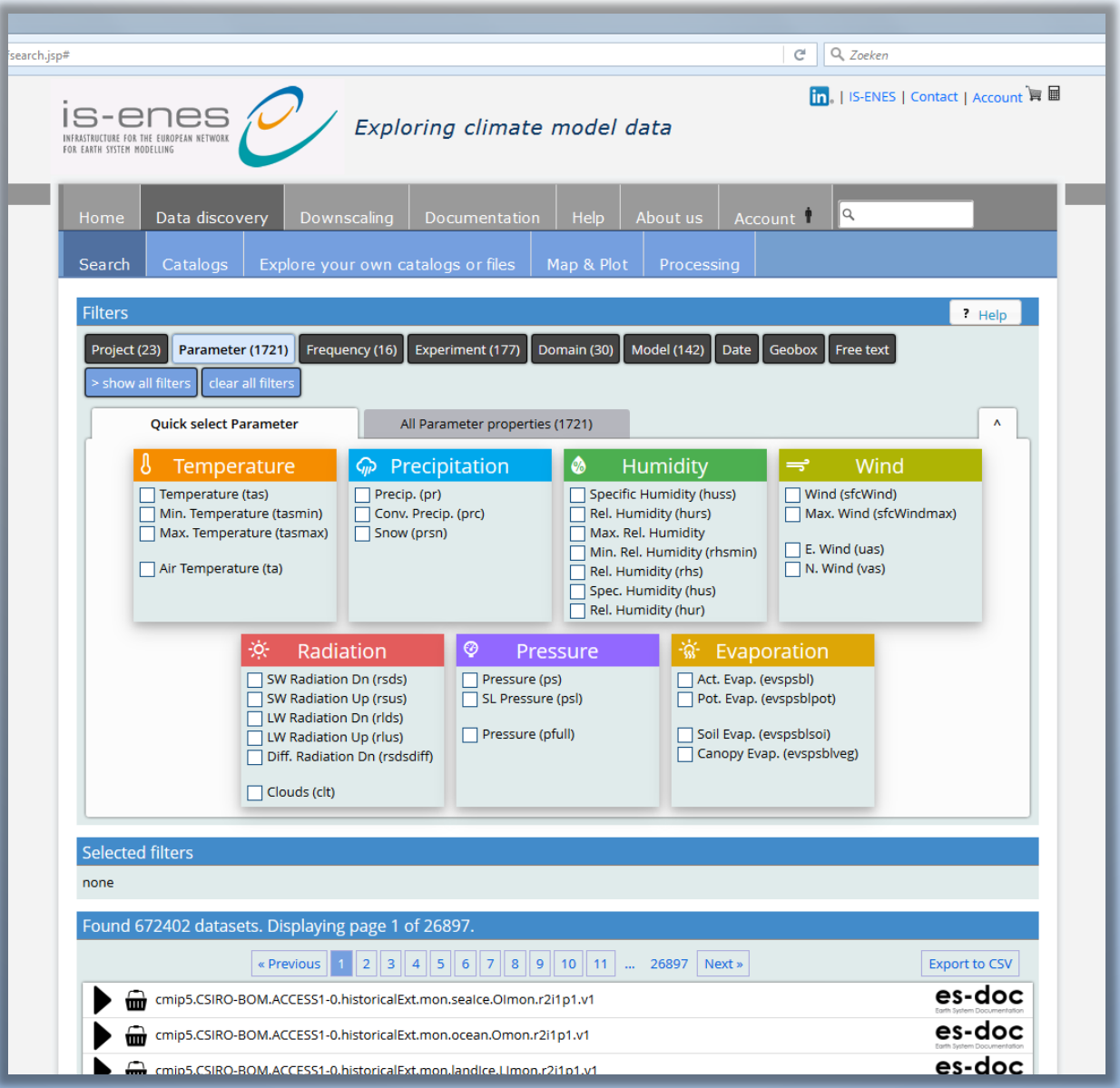
Use of open standards and open source software

- Data access over OPeNDAP → THREDDS
- Online analysis using WPS → PyWPS and CERFACS ICCLIM
- Online visualization using WMS → KNMI ADAGUC
- Single Sign On using OpenID, and OAuth2, delegation using MyProxy X509



http://climate4impact.eu/

Search using ESGF Search API – **New** faceted search



Search using ESGF Search API – **New** faceted search

- Search interface has been improved based on feedback from impact users
- Search interface speed has greatly been improved by using short lived (1 minute) caches
- Improved error handling in C4I for ESGF data nodes, catalogs are checked in advance for availability. Catalog status is clearly indicated to the user.
- Started with handling search queries as an aggregated dataset, e.g. use a search query as input for your processing.

Download

- By default the basket contains:
  - “Remote data” for links
  - “My data” for your own data
- Script based download: select and batch download multiple files
- The basket allows for uploading your own files
  - Can be used in processing or visualization
  - A per-user OpenDAP server on files in user's basket
- In development: **abstraction of the file concept**. **Packaging results**.