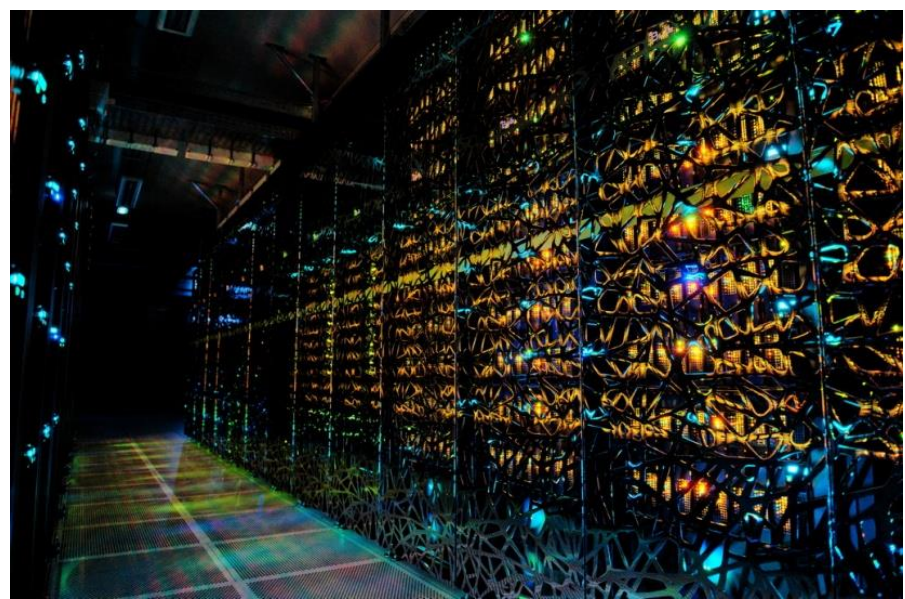
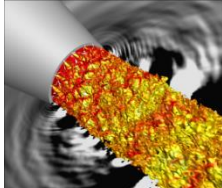
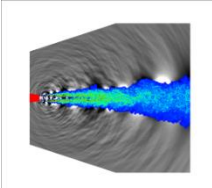
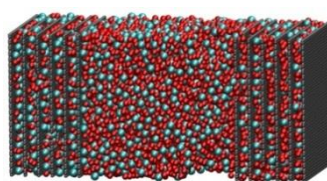
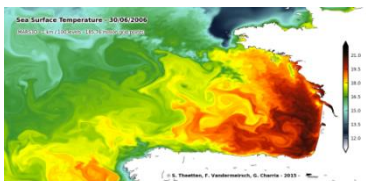
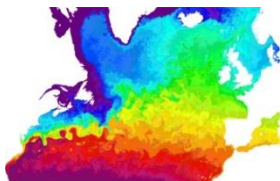


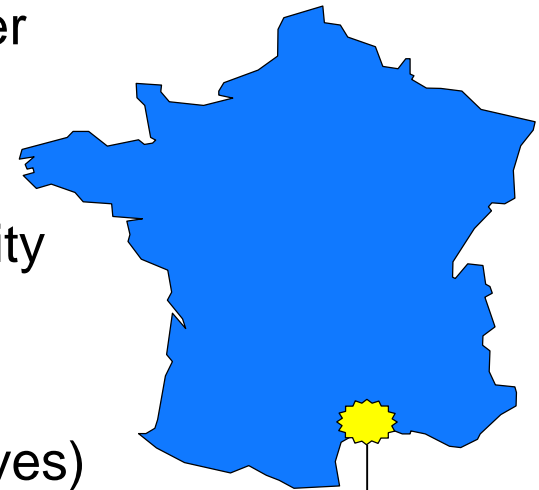
Centre Informatique National de l'Enseignement Supérieur



CINES is located in Montpellier:

- CINES is a public organisation (EPA) under the supervision of the French Ministry in charge of Higher Education, Research and Innovation (MESRI).
- CINES provides the french public research community with computing resources and IT services.
- ~60 people (technicians, engineers and administratives) are working in CINES; budget ~ 9 M€/yr

More informations: <https://www.cines.fr/>





IT hosting = sharing the infrastructures with partners
 Hosting of national public partners strategic platforms (high school + research communities):


- Racks of IT servers; invoices based on the number of racks + electric power use.
- Classical IT (DSI): ABES, INSERM, IRD, Rectorat Montpellier, CIRAD, ISSN,....
- HPC: Meso@LR, IRSTEA, IRD...
- Mid 2018: total=65 racks




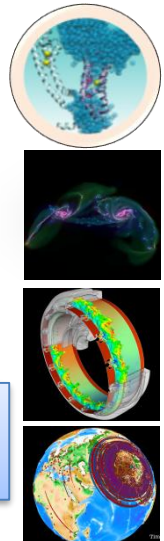
Long term preservation of data and digital documents
 Approval from the « Service Interministériel des Archives de France », SIAF

3 main national strategic missions

Shared Infrastructure

Data
 EUDAT 

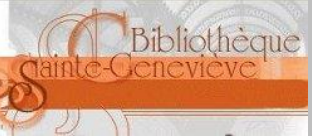
Computing




High Performance Computing
 OCCIGEN 3.5 Pflops, 70th position@TOP500 (2018-06)

Contribution to european projects
HPC-Europa2
 Pan-European Research Infrastructure on High Performance Computing

Our current partners



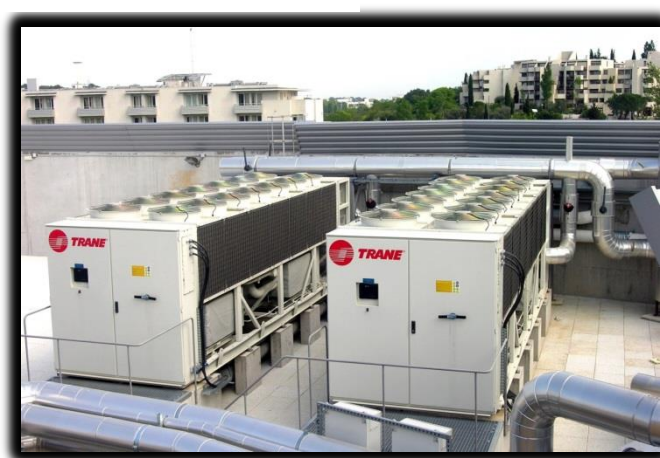
BORIS DINTRANS - CINES



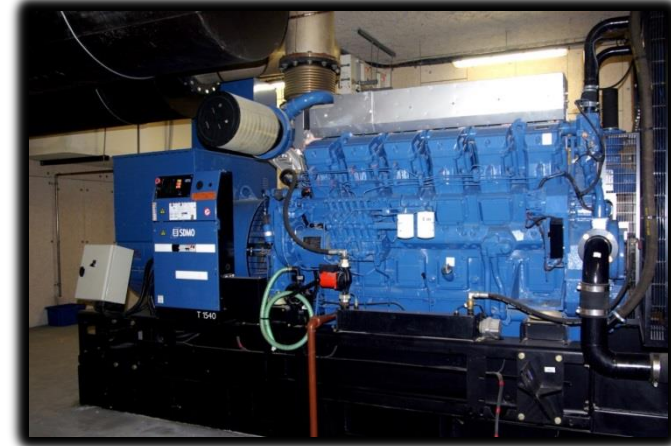
COUR DES COMPTES



The need of Tier3 N+1 infrastructures to provide high-quality IT services (99.98%, down~2h/yr)



*Cooling units
(6xcold and 4xhot water)*



Power generators (x3)

Highly-secured datacenter, available 24h/24, 7d/7

- 1500 m² on 5 IT rooms
- 2000 m² of technical areas
- 2 electric sources: 2,5 MW and 10 MW
- Redundant inverters and generator sets
- Broadband network accesses (20 Gbits/s)
- ZRR: high-level security (physical+virtual)



*Network access (RENATER,
HDMON, R3LR, TH'DOC...)*



Dedicated machine room for Occigen: 600 m²

2017-01: 3.5 Pflop/s



OCCIGEN2
+1,260 nodes Broadwell (2x14 cores)
2,106 nodes Haswell (2x12 cores)
85,824 cores, 64 or 128 GB/node,
IB 4xFDR, 5PB scratch Lustre, 105 Go/s
Elec. 1000 kW

2015-01: 2.1 Pflop/s



OCCIGEN: ATOS/BULL DLC B720
2,106 bi-proc nodes (2x12 cores)
Intel Haswell, IB 4xFDR,
50,544 cores, 64 or 128 GB/node,
5PB scratch Lustre, 105 Go/s
**Elec. 700 kW, Warm water directly on cpus
(35°) + air ; PUE < 1,1**

2010-07: 267 Tflop/s



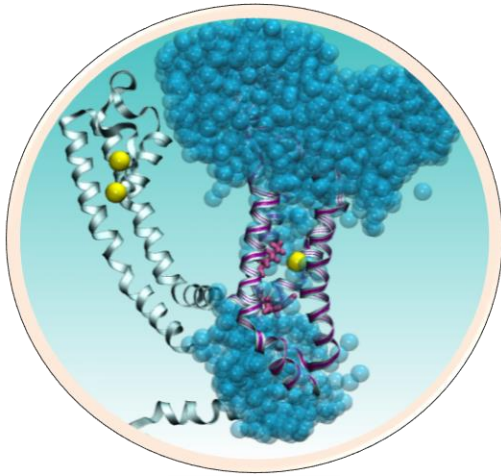
JADE 2
2,880 bi-proc nodes
+1,344 nodes Nehalem-EP, IB QDR
1,536 nodex Harpertown, IB DDR,
23,040 cores, 4 GB/core, 700 TB Lustre
Elec. 1000 kW

2008-09: 147 Tflop/s

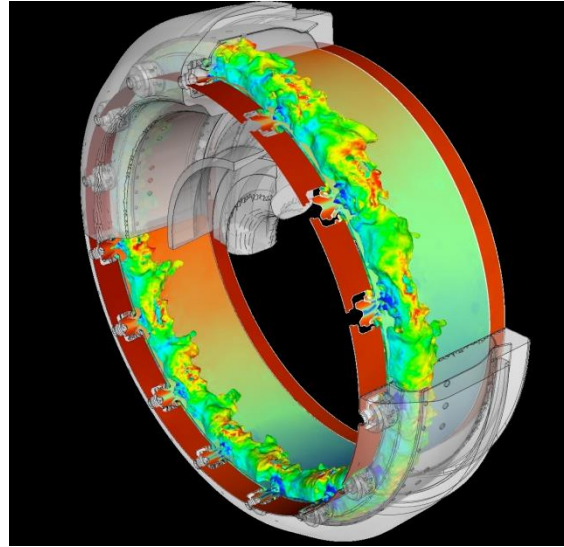


JADE: SGI Altix ICE 8200 EX
1,536 nodes bi-proc (2x4 cores)
INTEL Xeon – Harpertown
Elec. 600 kW, Cold water cooling inside rack doors (12°)

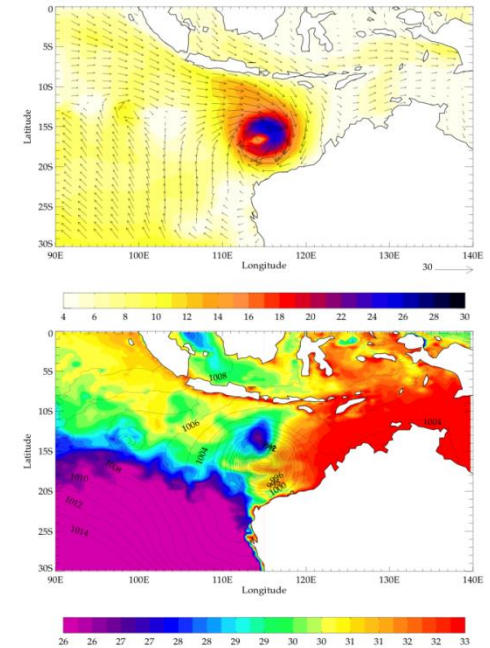
Massively-parallel simulations done in many scientific domains



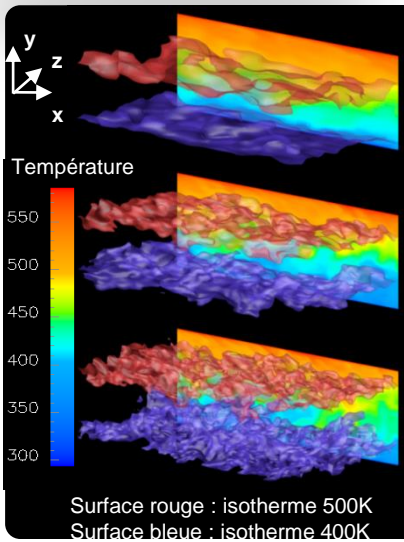
Chemistry, biology



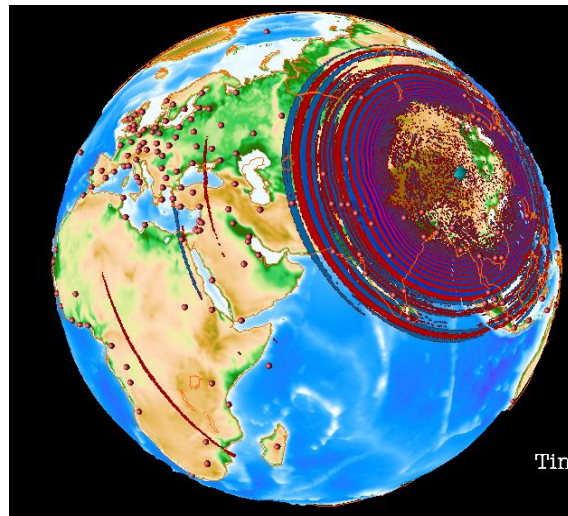
CFD w/o combustion



Climatology: ocean/atmosphere

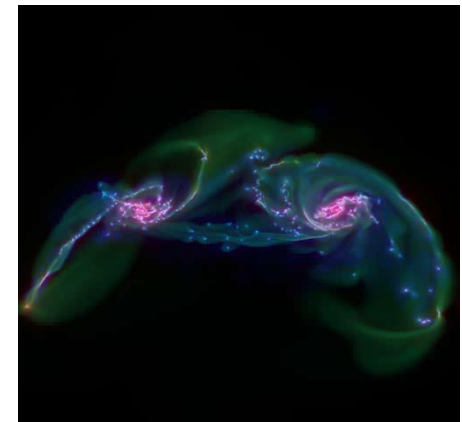


Physics



Geophysics

Boris DINTRANS - CINES



Astrophysics

HPC environment

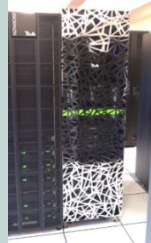
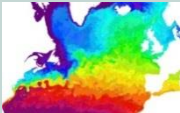
Pre/post processing resources

CRISTAL: Bull s6030 (Visualization)

2 nodes s6030 3 Tflops

- ❑ 32 cores : Intel x7560 @ 2.27 GHz
- ❑ 128 GB node1
- ❑ 256 GB node2
- ❑ 2 GPUs Nvidia

Quadro Plex D2 (FX5800)



NFS



20 Gb/s Backbone

HPC resources

OCCIGEN: Bull DLC

3,336 nodes bullx B720: 3.5 Pflops

- ❑ 85,824 cores Intel Haswell/Broadwell @ 2.6 GHz - 12 or 14 cores - 64 and 128 GB/node
- ❑ Infiniband FDR network between nodes
- ❑ Lustre 5 PB @ 100 GB/s (/scratch)
- ❑ Panasas 260 TB @ 10 GB/s (/home)
- ❑ Racks: 38 compute + 7 disks/network
- ❑ DLC cooling technology (~32°C)



FDR Infiniband gateways

Storage / data resources

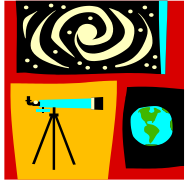


2 x IBM TS3500 lib. ~4 PB,
2 x 1000 cartridges, 9 x Jaguar4 + 8 x LTO6 + 2 x LTO4



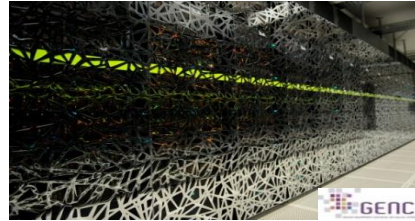
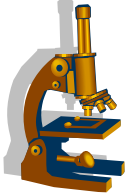
Lustre HSM 2 PB @ 50 GB/s, DMF

The Big Data era



Observations

- Particle accelerator
- Genome sequencer
- Meteorology/climatology sensors
- Shot point (oil seismic) or spatial data
- Etc.



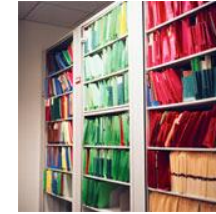
HPC results

- Multi-physical simulations, multi-scale simulations
- Model resolution increase
- New thematic



Born digital or digitalized data

- Old manuscripts
- Digitalized scientific reviews
- Digital thesis
- Etc.



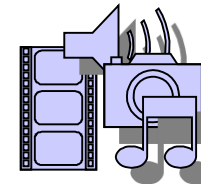
Management records on database



emails

Issues: the 3 V's of Big Data

- *Volume*: data size (file sizes *and* number of files)
- *Velocity*: Facebook=900,000,000 images uploaded/day!
- *Variety*: file format diversity (text, image, sound, 2D/3D...)



Objectives

Disseminate

Distribute to user communities

Process

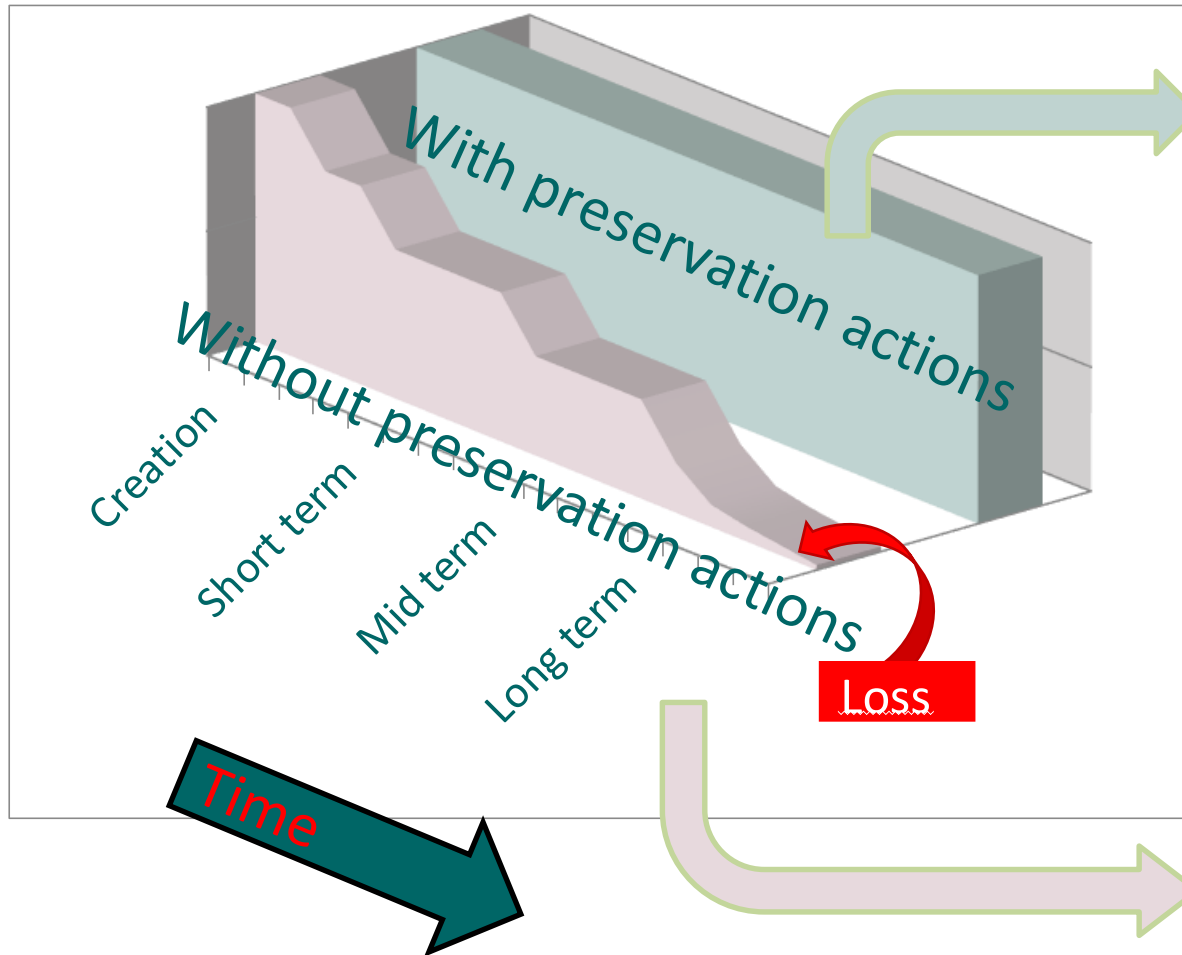
Create values from data

Preserve

Long-term
archiving

Mid-term
archiving

Long-term preservation challenges



Main concerns:

- Metadata
- File format
- Information representation
- Data storage (media evolution)
- Access / read on support
- Technology watch

Main risks:

- comprehension loss
- integrity loss
- usability loss
- valorisation loss



DATA → DMP needed!

Réaliser un plan de gestion de données “ FAIR ” : modèle

Nathalie Reymonet, Magalie Moysan, Aurore Cartier, Renaud Délémontez

► To cite this version:

Nathalie Reymonet, Magalie Moysan, Aurore Cartier, Renaud Délémontez. Réaliser un plan de gestion de données “ FAIR ” : modèle . Ce document a pour vocation d'accompagner les chercheurs et chargés de projets dans la rédaction .. 2018. <sic_01690547v2>

HAL Id: sic_01690547

https://archivesic.ccsd.cnrs.fr/sic_01690547v2

Submitted on 15 Feb 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial | 4.0 International License

PAC: the main archiving platform



1. Receipt



2. Check

the quality of the received data



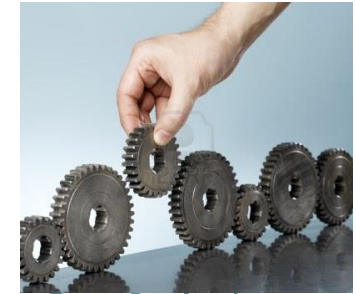
3. Add

informations (PID, hash, archive date, ...)



4. Additional

processing (get related informations, etc.)



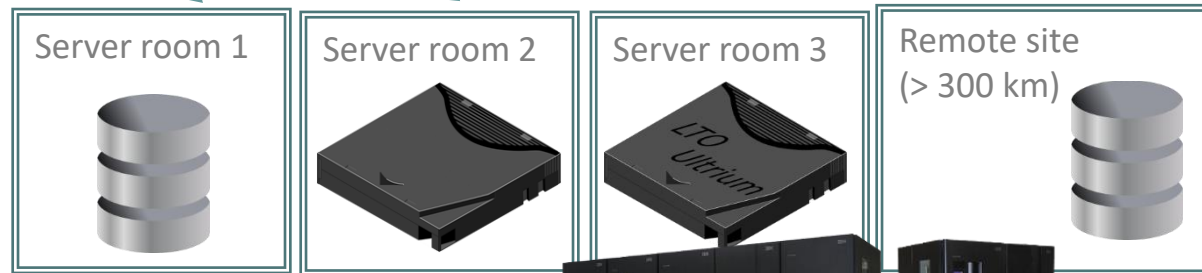
Transferring Agency



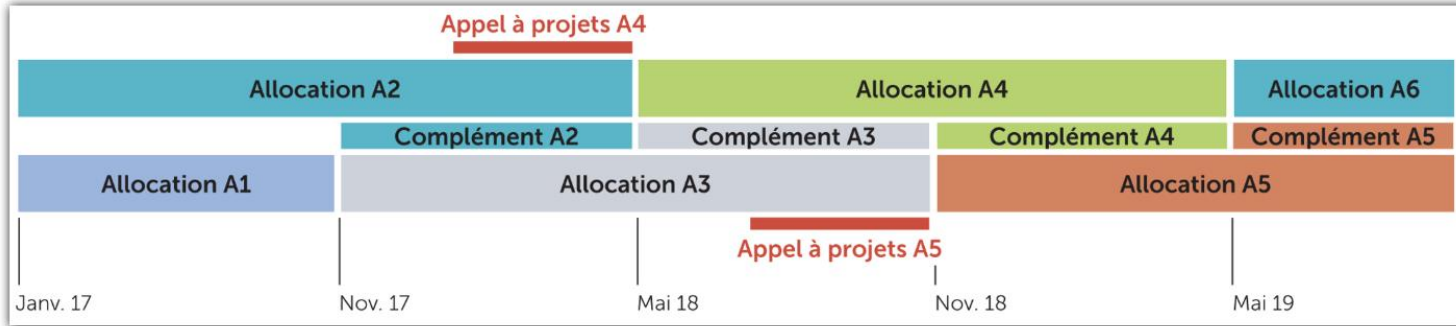
6. Inspect periodically all archived copies



5. Store the archive in multiple copies

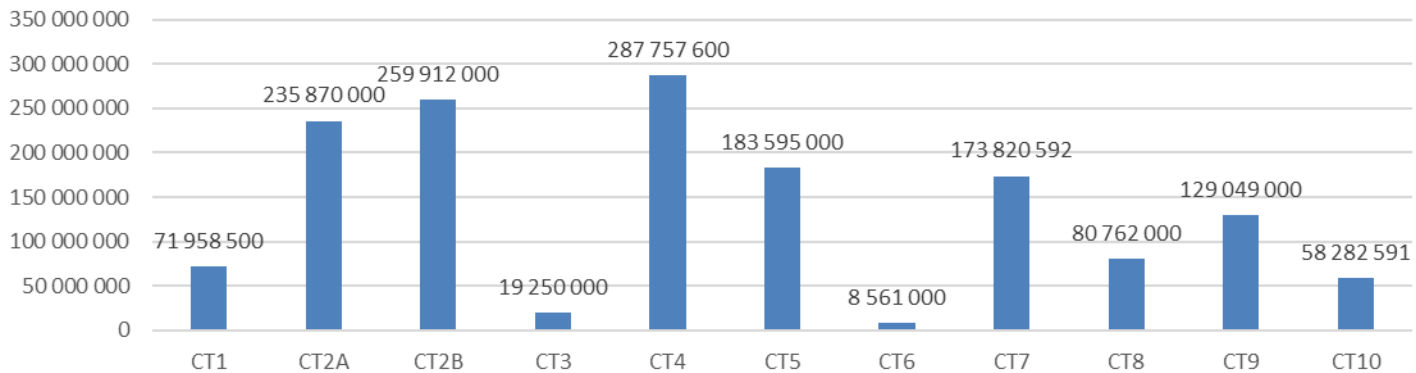


Occigen: DARI statistics since A1 (2017-01)



Somme de CINES-Occigen-heures accordées

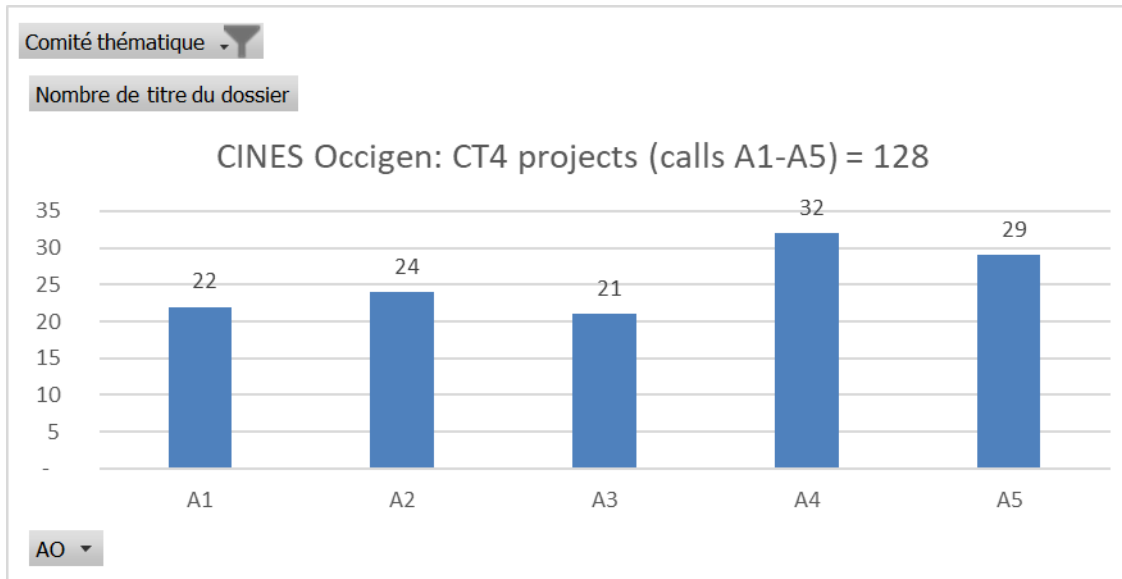
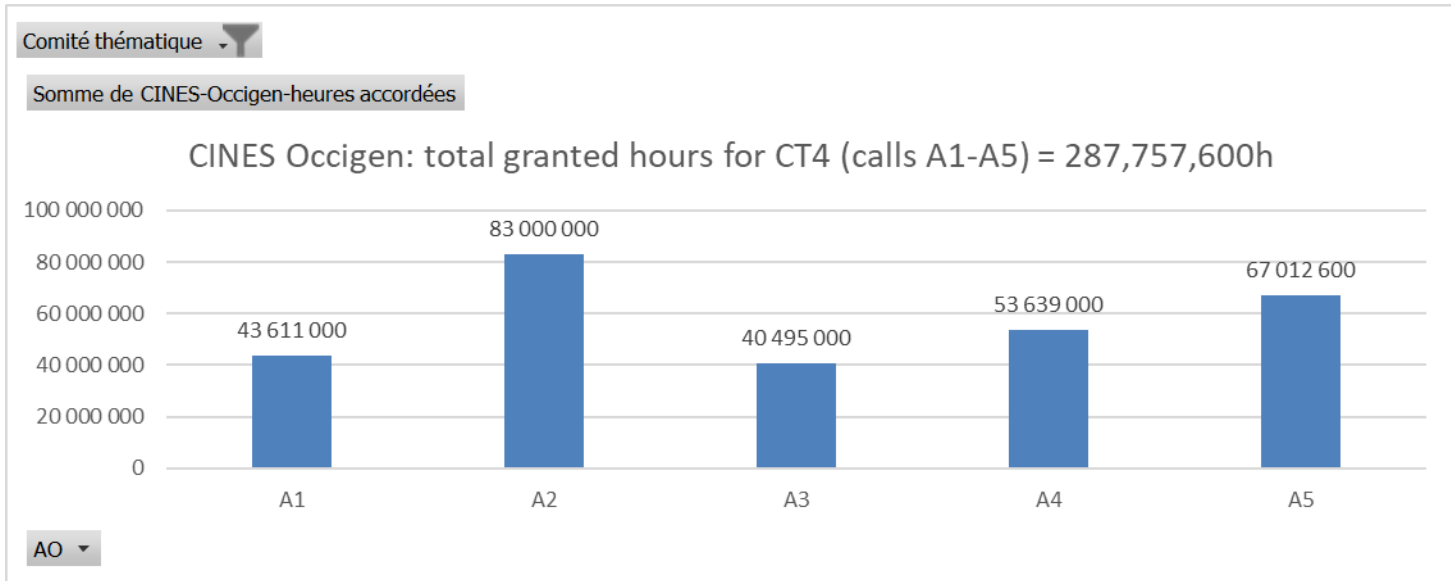
CINES Occigen: total granted hours by CT (calls A1-A5) = 1,508,818,283h



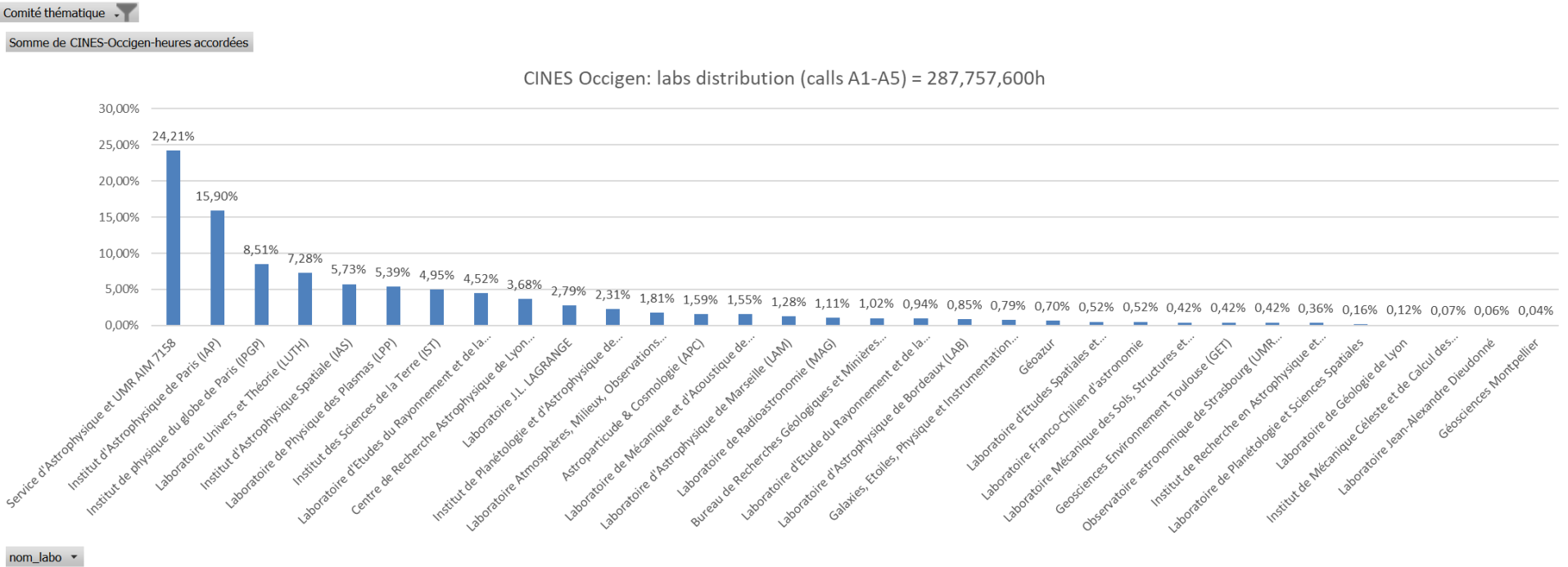
Comité thématique ▾

| Comité thématique | Libellé |
|-------------------|-------------------------------------------------------|
| CT1 | ENVIRONNEMENT |
| CT2A | ECOULEMENTS NON REACTIFS |
| CT2B | ECOULEMENTS REACTIFS OU/ET MULTIPHASIQUES |
| CT3 | SIMULATION BIOMEDICALE ET APPLICATION A LA SANTE |
| CT4 | ASTROPHYSIQUE ET GEOPHYSIQUE |
| CT5 | PHYSIQUE THEORIQUE ET PHYSIQUE DES PLASMAS |
| CT6 | INFORMATIQUE, ALGORITHMIQUE ET MATHÉMATIQUES |
| CT7 | SYSTEMES MOLECULAIRES ORGANISES ET BIOLOGIE |
| CT8 | CHIMIE QUANTIQUE ET MODELISATION MOLECULAIRE |
| CT9 | PHYSIQUE, CHIMIE ET PROPRIETES DES MATERIAUX |
| CT10 | NOUVELLES APPLICATIONS ET APPLICATIONS TRANSVERSALLES |

Occigen: DARI statistics for CT4 by call



Occigen: DARI statistics for CT4 by labs



HPC@CINES: the future



STRATÉGIE NATIONALE
DES INFRASTRUCTURES DE RECHERCHE
ÉDITION 2018 - N°2



HPC@CINES: the future

- One still needs strong HPC resources: Occigen renewal
 - 2015: Occigen1
 - 2017: Occigen2
 - 2020: next Occigen
- What kind of machine?
 - New communities are running on the three national tier1: IA (GPU), bioinformatics (HTC), humanities (HPDA)...
 - Moore's law is ``dead'' and new HPC architectures are much more complicated to use (GPU, manycore cpus, ARM, FPGA...)
 - CINES philosophy: be as close as possible to the uses
- HPC-HPDA-HTC must all be adressed
 - Convergence of HPC & Big Data on a single IT infrastructure
 - Diffusion of results is also very important → Science DMZ, Open Science, Open Data...

One solution: HPC + HPDA + Data storage + Diffusion through a Science DMZ

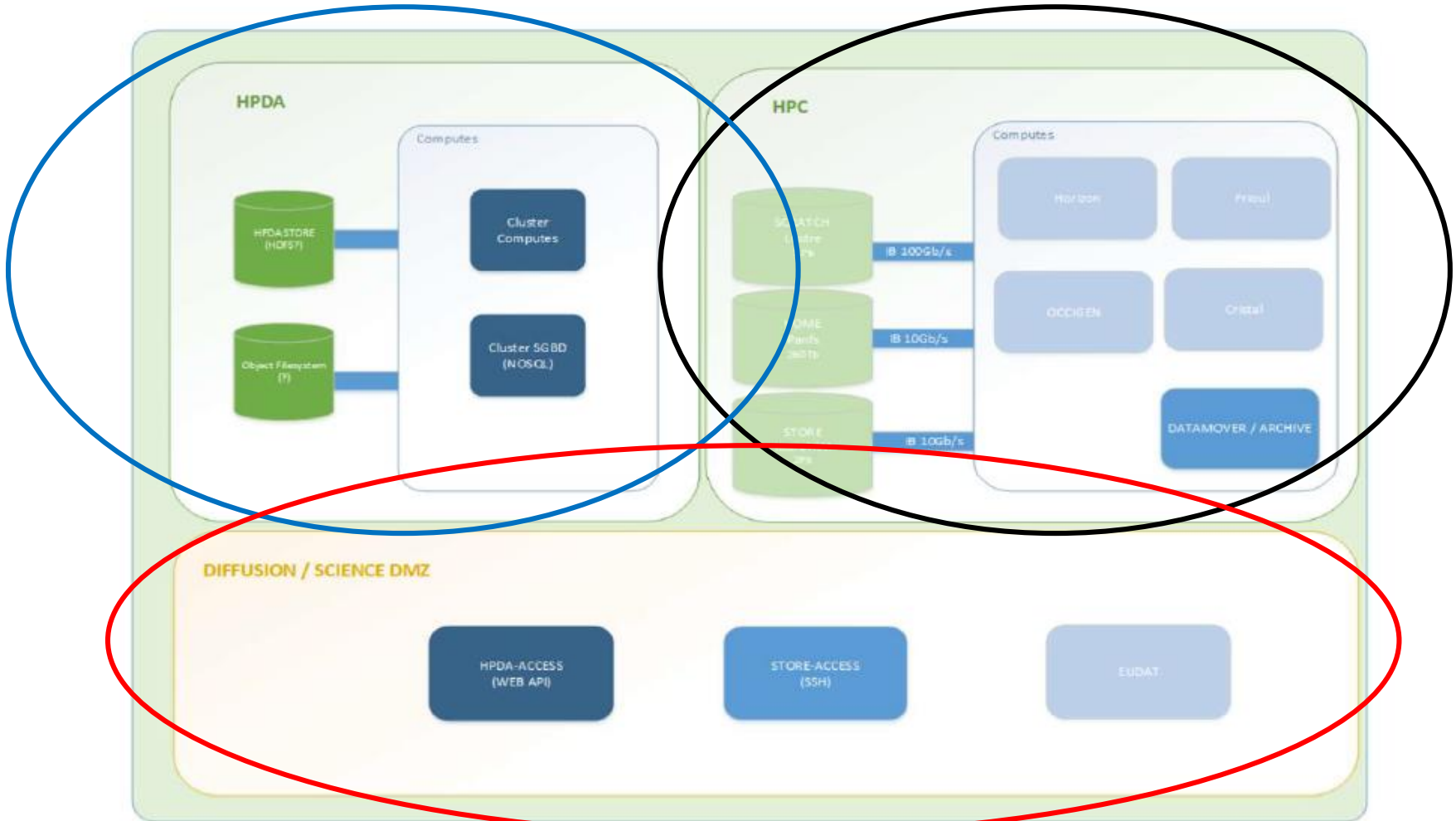


Figure 16 : les futurs services autour du stockage *datacentrique* (HPDA et Science DMZ).

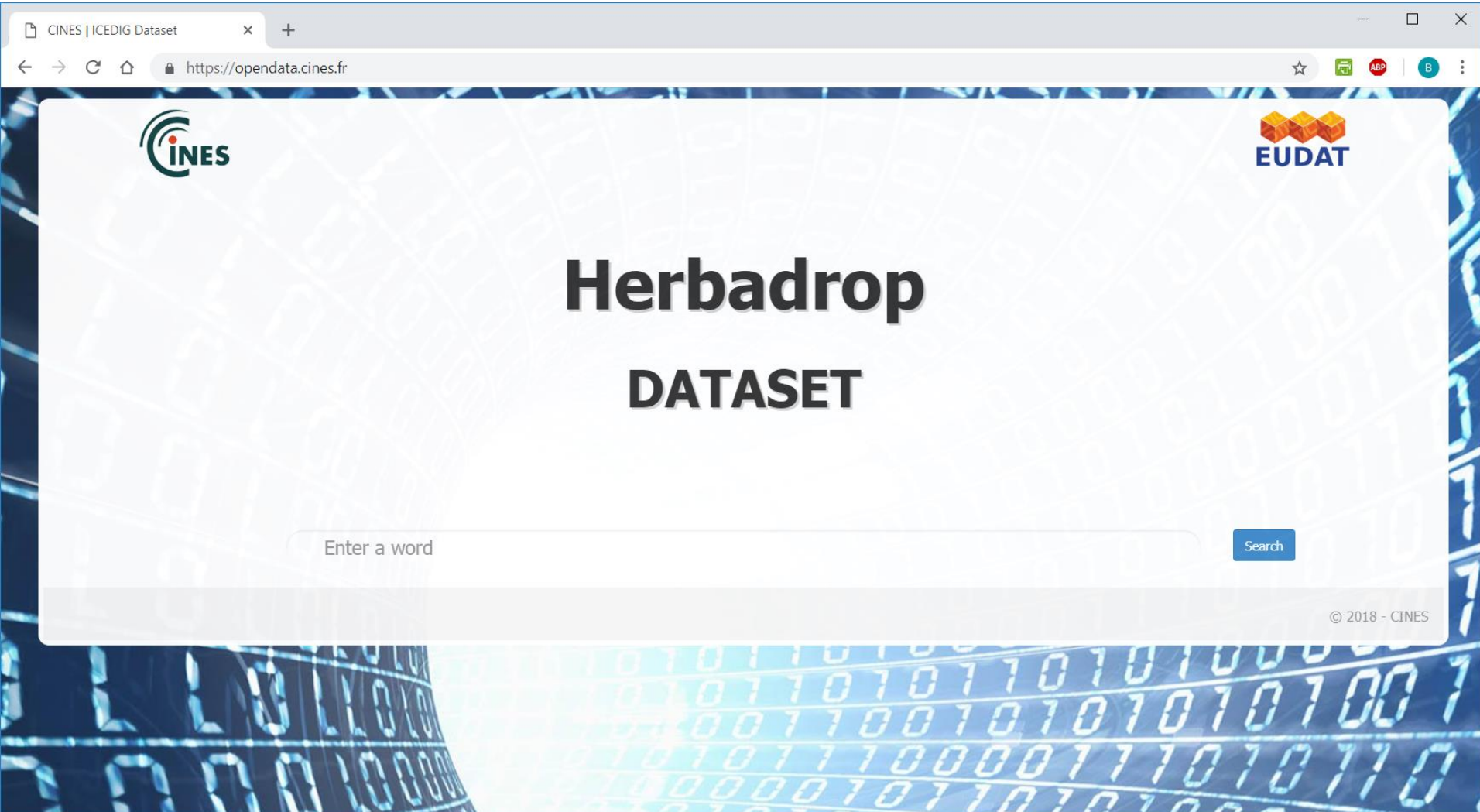
The Herbadrop POC: using HPC *before* archiving



Herbadrop in Europe



Data (scans) → HPC for OCR → archiving raw data + metadata (OCR) → web portal created for diffusion across Europe



The screenshot shows a web browser window with the URL <https://opendata.cines.fr>. The page features the INES logo in the top left and the EUDAT logo in the top right. The main heading is "Herbadrop DATASET". Below the heading is a search bar with the placeholder text "Enter a word" and a "Search" button. The background of the page is a blue and white pattern of binary code (0s and 1s).