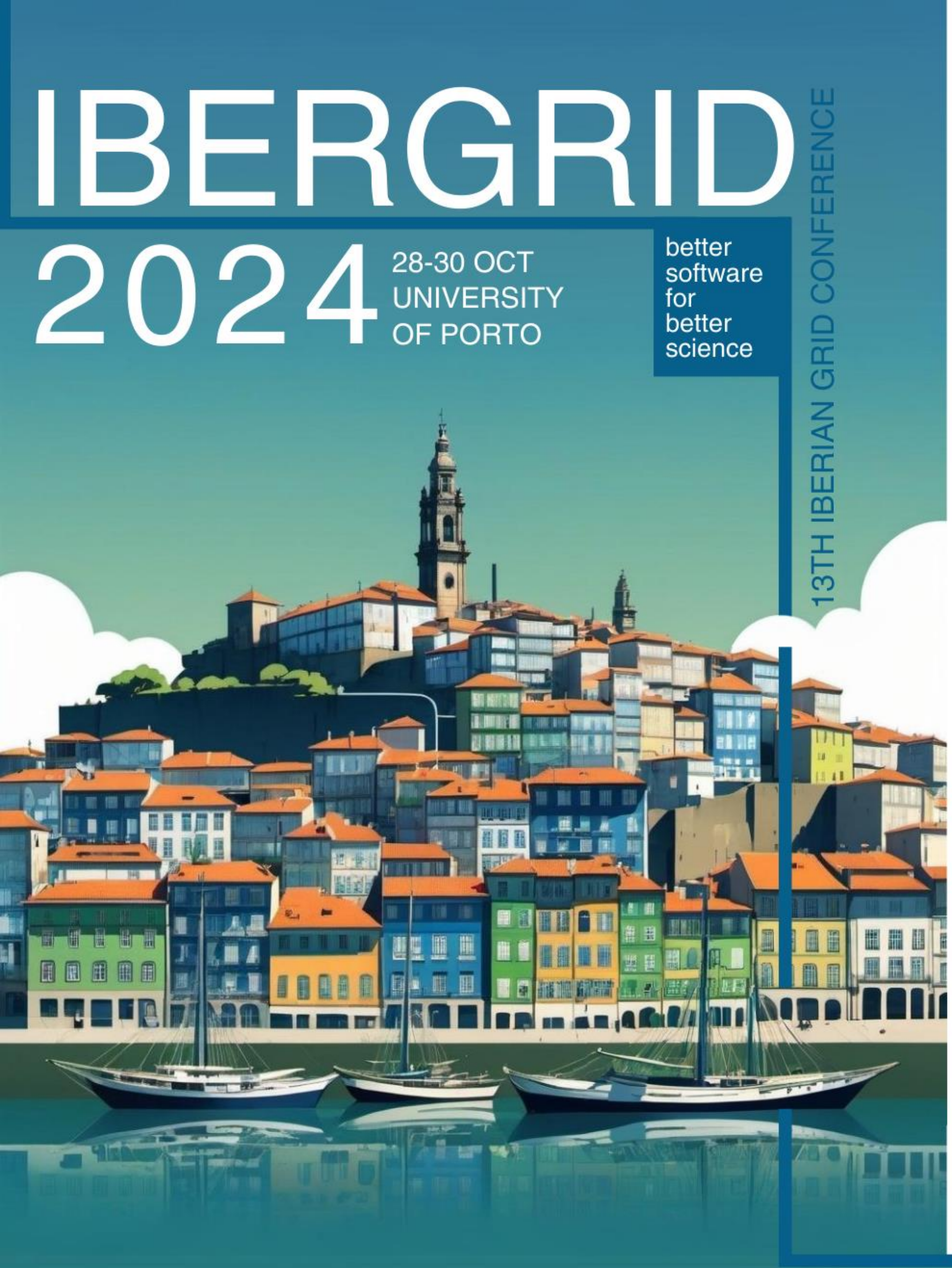


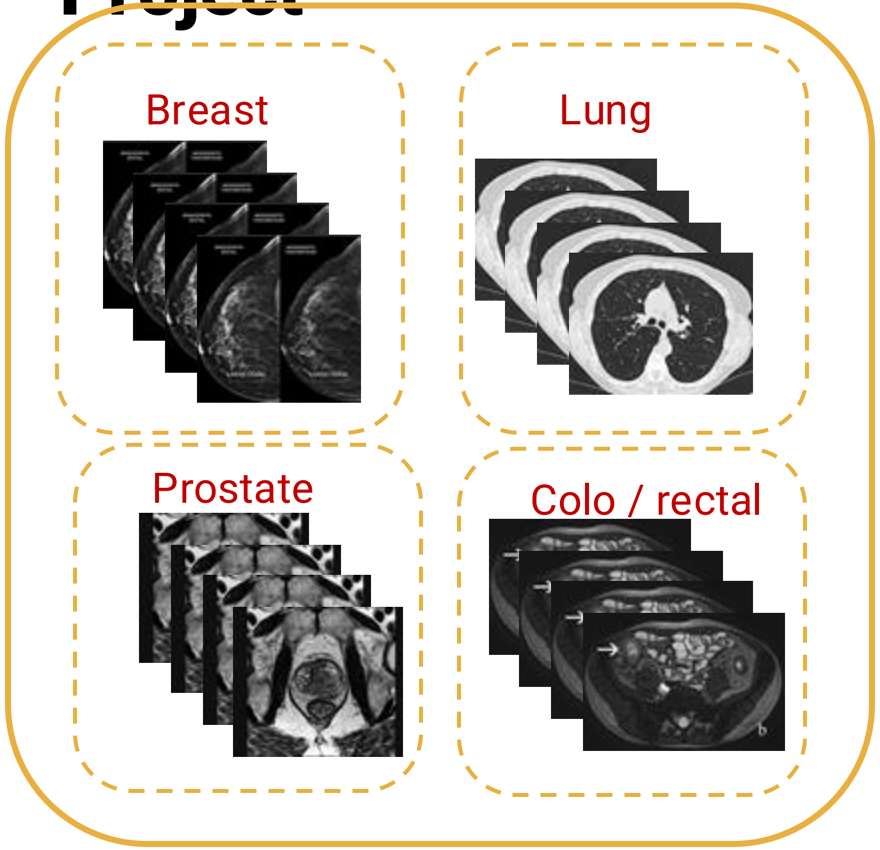
In situ processing of medical imaging data

(An Open Challenge Experience)

Andrei S. Alic
Ignacio Blanquer
Pau Lozano
Damià Segrelles-Quilis

Institute of Instrumentation for Molecular Imaging
Universitat Politècnica de València

Accelerating the lab to market transition of AI tools for cancer management: CHAIMELEON Project



Images + Related clinical data (e-form)

A Cloud-based cancer imaging repository as an online resource for the AI community working on the development of cancer management solutions

Not just a data warehouse…

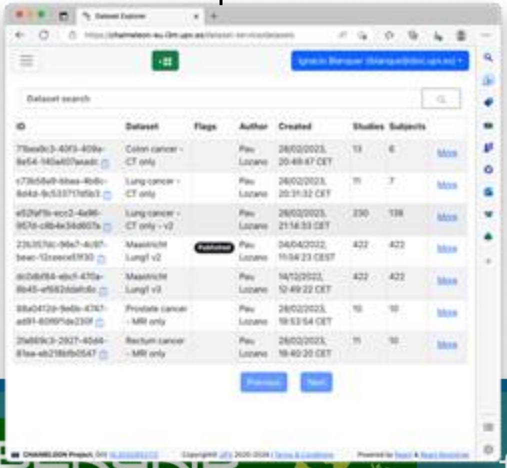
- Incorporating all necessary functionalities to allow AI experimentation on the cloud (without downloading the data).
- Powered with automation tools.
- Interoperable with other existing initiatives.

Use Cases and Design Principles

- Use Cases
  - Create, publish and explore a Dataset.
  - Access to the Images and clinical data from a Dataset through an interactive application.
  - Submit a processing job on the dataset to the infrastructure.
  - Publish a processing application on the Marketplace.

- Design principles
  - The data is anonymised and uploaded to a central repository.
  - Data cannot be downloaded.
  - Data is organized into datasets (a coherent set of annotated image studies and the associated clinical data that have a persistent identifier - a FAIR citable research object).
  - Published Datasets have their metadata publicly accessible.
  - Data can be processed "in situ" using the tools available in the platform on the cloud resources of the platform.

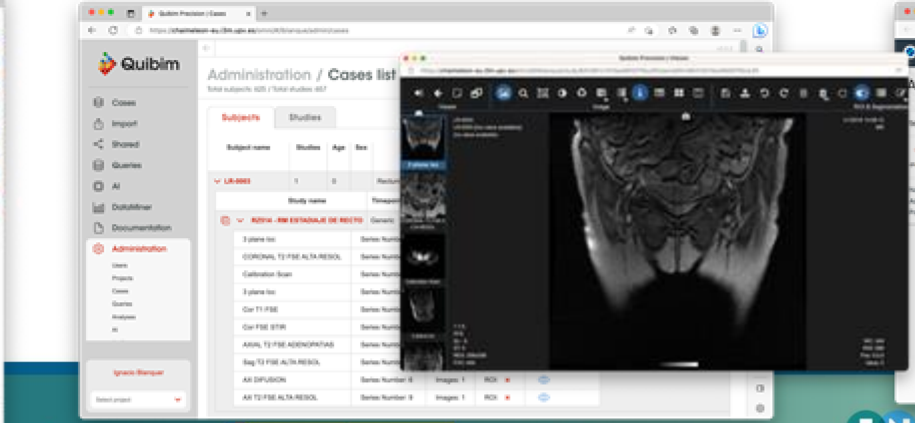
Dataset Explorer



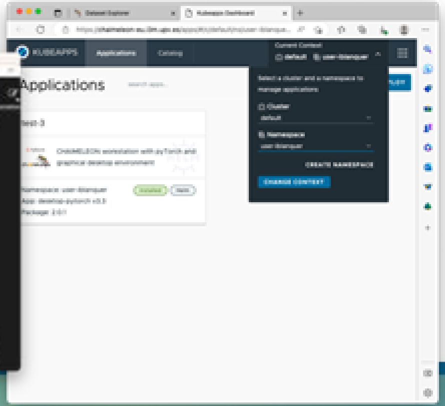
Dataset search

ID	Dataset	Flags	Author	Created	Studies	Subjects	
[illegible]	Colon cancer - CT only		Pau Lozano	28/02/2023, 20:49:47 CET	13	6	More
[illegible]	Lung cancer - CT only		Pau Lozano	28/02/2023, 20:31:32 CET	11	7	More
[illegible]	Lung cancer - CT only - v2		Pau Lozano	28/02/2023, 21:14:33 CET	230	138	More
[illegible]	Maastricht Lung1 v2	Published	Pau Lozano	04/04/2022, 11:04:23 CEST	422	422	More
[illegible]	Maastricht Lung1 v3		Pau Lozano	14/12/2022, 12:49:22 CET	422	422	More
[illegible]	Prostate cancer - MRI only		Pau Lozano	28/02/2023, 19:53:54 CET	10	10	More
[illegible]	Rectum cancer - MRI only		Pau Lozano	28/02/2023, 18:40:20 CET	11	10	More

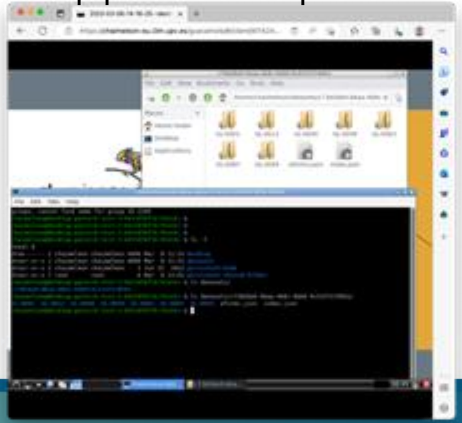
Case Explorer



Application Dashboard

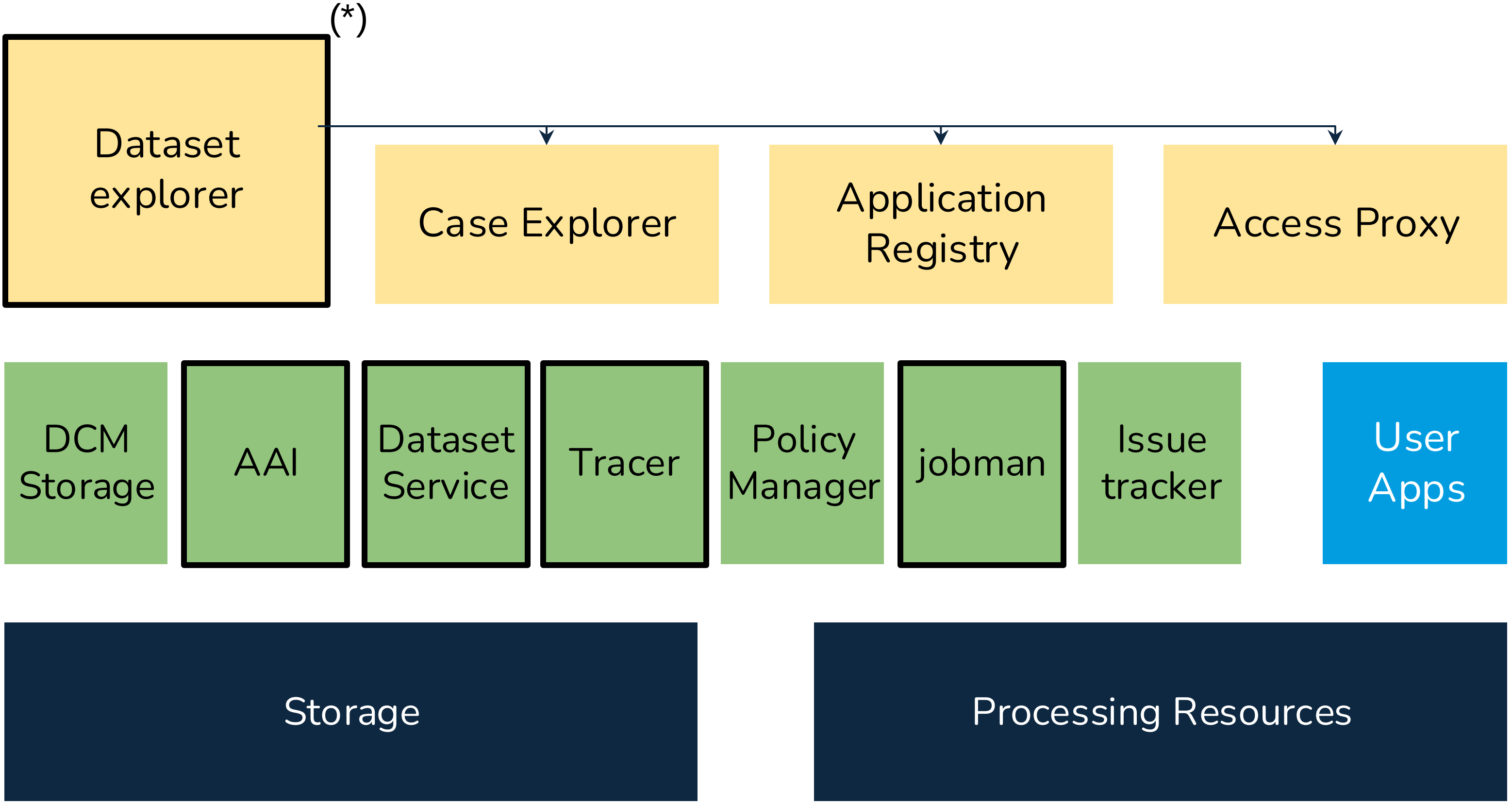


App Desktop Access



Central Platform

- All services managed as Kubernetes deployments and IaC recipes.
- **Secure**. All services are secured and access to data is restricted to the boundaries of the platform.
- **Auditable**. Access to data is registered in a Blockchain.
- **FAIR compliant**. It provides Findability, Accessibility, Interoperability and Reusability for the datasets.
- **High-Performance**. Integrates GPUs.
- **Reliable**. It uses Kubernetes deployments.
- **Reproducible**. It uses IaC, Open Source technologies, containers and cloud backends.

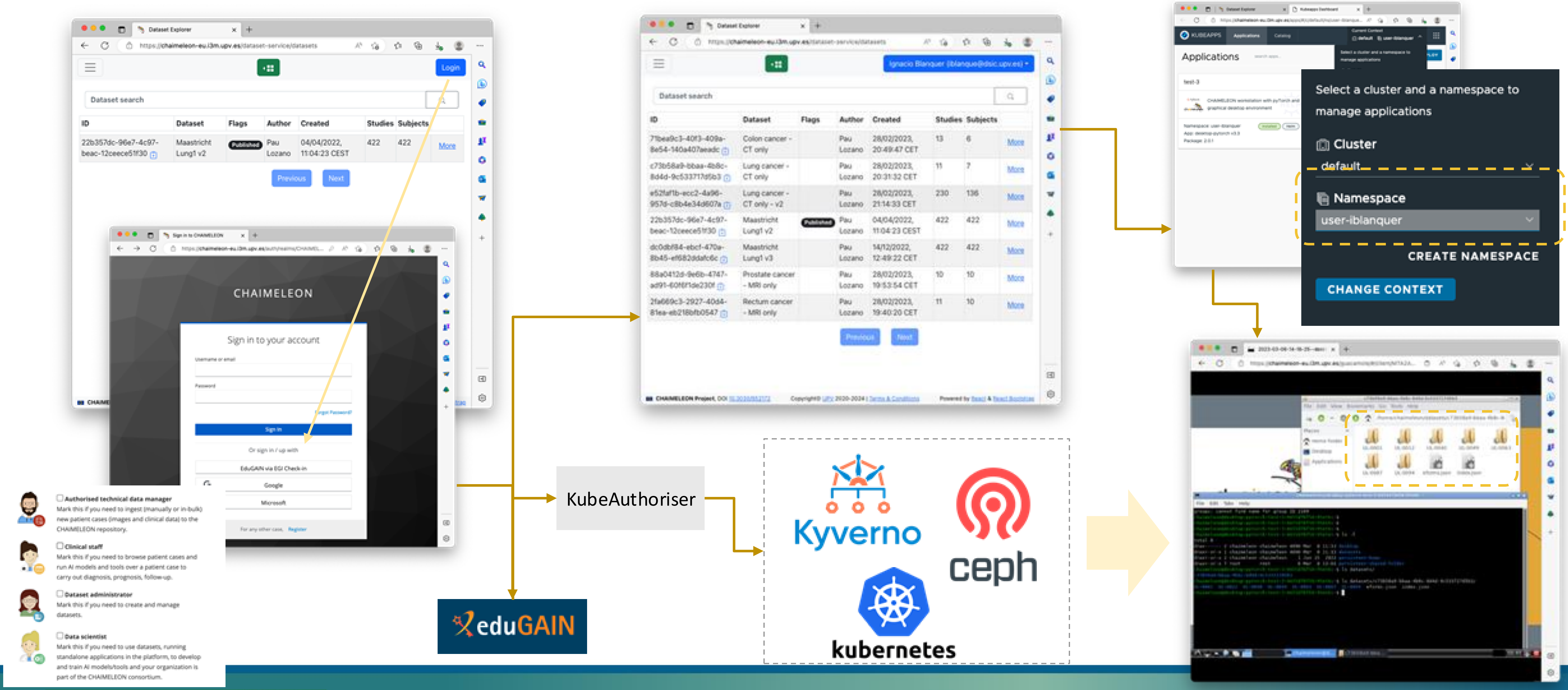


A Kubernetes elastic platform on top of a Cloud-based platform with 128 cores, 1,5 TB RAM, 3 V100 and 4 A30 GPUs (192 GB of GPU RAM)

(*) Fully developed in CHAIMELEON

Authentication and Authorisation

https://chaimeleon-eu.i3m.upv.es/auth/realms/CHAIMELEON/account/#/



Dataset Explorer — https://chaimeleon-eu.i3m.upv.es/dataset-service/datasets

Login

Dataset search

ID	Dataset	Flags	Author	Created	Studies	Subjects	
22b357dc-96e7-4c97-beac-12ceece51f30	Maastricht Lung1 v2	Published	Pau Lozano	04/04/2022, 11:04:23 CEST	422	422	More

Previous Next

Sign in to CHAIMELEON

CHAIMELEON

Sign in to your account

Username or email

Password

Forgot Password?

Sign In

Or sign in / up with

EduGAIN via EGI Check-in

Google

Microsoft

For any other case, Register

☐ **Authorised technical data manager**
Mark this if you need to ingest (manually or in-bulk) new patient cases (images and clinical data) to the CHAIMELEON repository.

☐ **Clinical staff**
Mark this if you need to browse patient cases and run AI models and tools over a patient case to carry out diagnosis, prognosis, follow-up.

☐ **Dataset administrator**
Mark this if you need to create and manage datasets.

☐ **Data scientist**
Mark this if you need to use datasets, running standalone applications in the platform, to develop and train AI models/tools and your organization is part of the CHAIMELEON consortium.

eduGAIN

Dataset Explorer — https://chaimeleon-eu.i3m.upv.es/dataset-service/datasets

Ignacio Blanquer (iblanque@dsic.upv.es)

Dataset search

ID	Dataset	Flags	Author	Created	Studies	Subjects	
71bea9c3-40f3-409a-8e54-140a407aeadc	Colon cancer - CT only		Pau Lozano	28/02/2023, 20:49:47 CET	13	6	More
c73b58a9-bbaa-4b8c-8d4d-9c533717d5b3	Lung cancer - CT only		Pau Lozano	28/02/2023, 20:31:32 CET	11	7	More
e52faf1b-ecc2-4a96-957d-c8b4e34d607a	Lung cancer - CT only - v2		Pau Lozano	28/02/2023, 21:14:33 CET	230	136	More
22b357dc-96e7-4c97-beac-12ceece51f30	Maastricht Lung1 v2	Published	Pau Lozano	04/04/2022, 11:04:23 CEST	422	422	More
dc0dbf84-ebcf-470a-8b45-ef682ddafc6c	Maastricht Lung1 v3		Pau Lozano	14/12/2022, 12:49:22 CET	422	422	More
88a0412d-9e6b-4747-ad91-60f6f1de230f	Prostate cancer - MRI only		Pau Lozano	28/02/2023, 19:53:54 CET	10	10	More
2fa669c3-2927-40d4-81ea-eb218bfb0547	Rectum cancer - MRI only		Pau Lozano	28/02/2023, 19:40:20 CET	11	10	More

Previous Next

CHAIMELEON Project, DOI 10.3030/952172 Copyright© UPV 2020-2024 | Terms & Conditions Powered by React & React Bootstrap

KubeAuthoriser

KUBEAPPS Applications Catalog

Applications

Select a cluster and a namespace to manage applications

Cluster
default

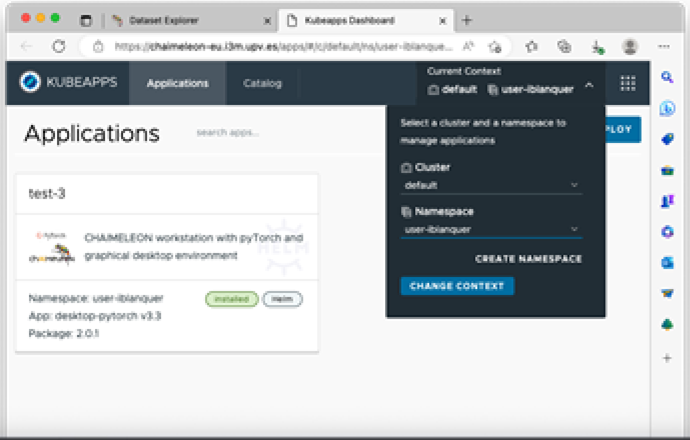
Namespace
user-iblanquer

CREATE NAMESPACE

CHANGE CONTEXT

Processing Model

- Processing is performed in situ by means of interactive and batch applications.
- Interactive applications
  - Accessed through a VNC proxy, to avoid the risk of downloading data.
  - Coded as Helm charts registered in KubeApps.
- Batch applications
  - Run unattended from the interactive virtual environments through a command line.
  - Run seamlessly in the Kubernetes infrastructure, accessing the data in the same way as the interactive applications.
  - Use a collection of verified Docker containers from an internal repository.



https://chaimeleon-eu.i3m.upv.es/apps/#/c/default/ns/user-iblanque...

Applications

test-3

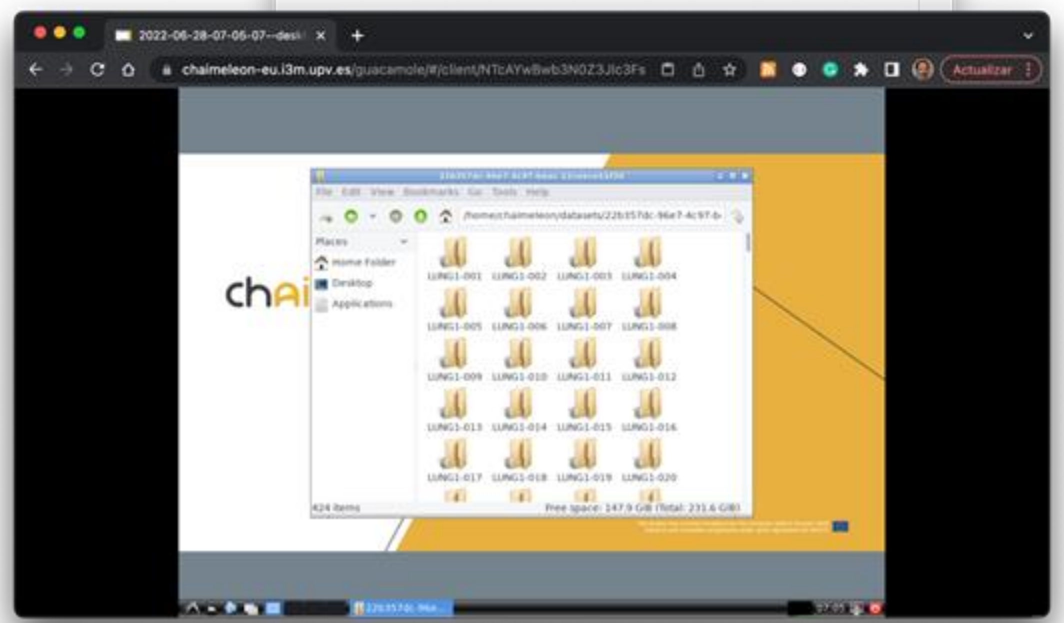
CHAIMELEON workstation with pyTorch and graphical desktop environment

Namespace: user-iblanquer
App: desktop-pytorch v3.3
Package: 2.0.1

Select a cluster and a namespace to manage applications

Cluster: default

Namespace: user-iblanquer



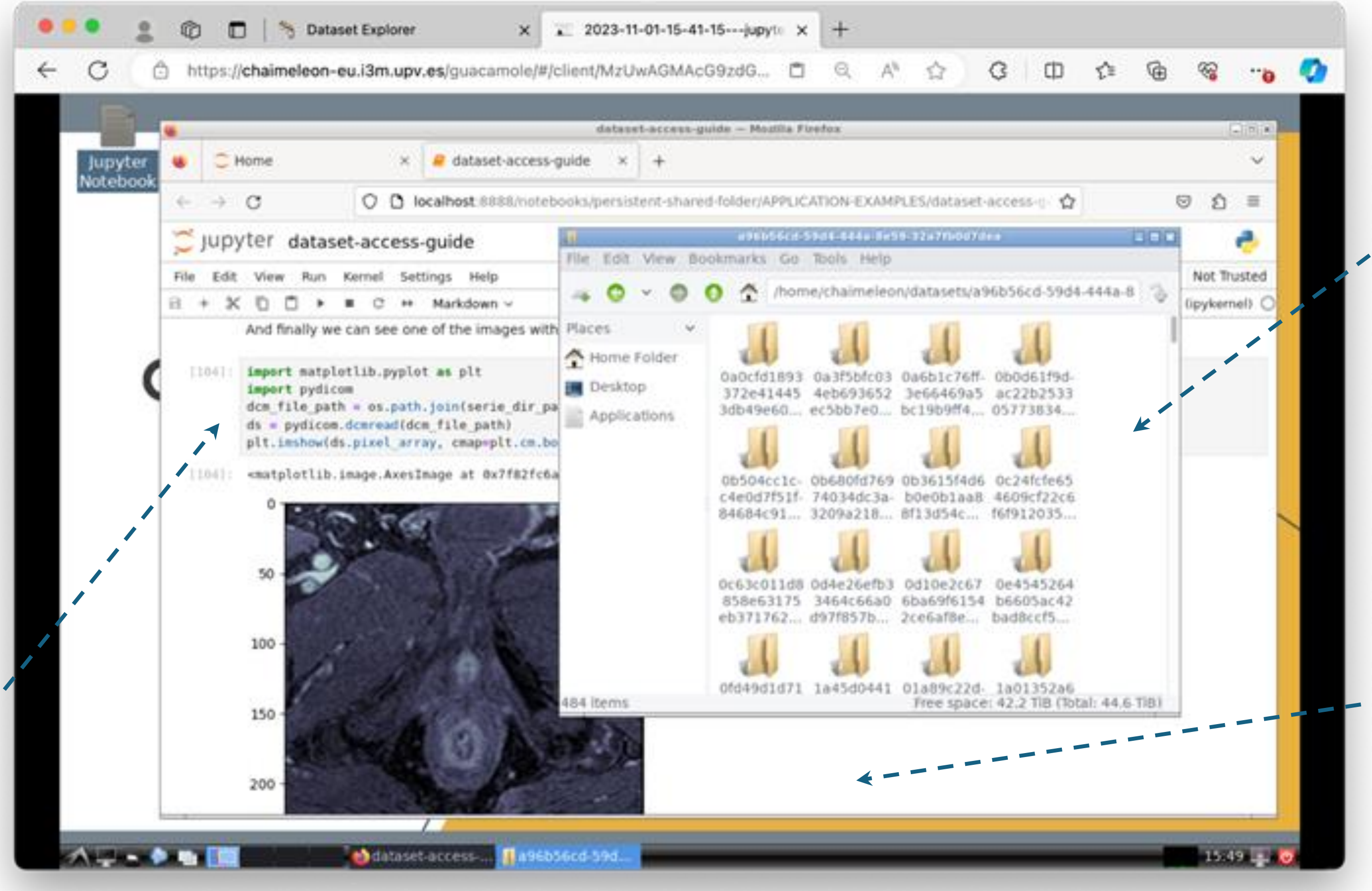
chaimeleon-eu.i3m.upv.es/guacamole/#/client/NTcAYwBwb3N0Z3Jlc3Fs

/home/chaimeleon/datasets/22b157dc-96e7-4c97-b...

Virtual Research Environments

A GUI with an ubuntu container in a network-restricted environment

Data Analytic SW libraries


Dataset Explorer
2023-11-01-15-41-15---jupyte
https://chaimeleon-eu.i3m.upv.es/guacamole/#/client/MzUwAGMAcG9zdG...
Jupyter Notebook
dataset-access-guide — Mozilla Firefox
Home
dataset-access-guide
localhost:8888/notebooks/persistent-shared-folder/APPLICATION-EXAMPLES/dataset-access-g
jupyter dataset-access-guide
File Edit View Run Kernel Settings Help
Markdown
Not Trusted
(ipykernel)
a96b56cd-59d4-444a-8e59-32a7fb0d7dea
File Edit View Bookmarks Go Tools Help
/home/chaimeleon/datasets/a96b56cd-59d4-444a-8
Places
Home Folder
Desktop
Applications
0a0cfd1893 372e41445 3db49e60...
0a3f5bfc03 4eb693652 ec5bb7e0...
0a6b1c76ff- 3e66469a5 bc19b9ff4...
0b0d61f9d- ac22b2533 05773834...
0b504cc1c- c4e0d7f51f- 84684c91...
0b680fd769 74034dc3a- 3209a218...
0b3615f4d6 b0e0b1aa8 8f13d54c...
0c24fcfe65 4609cf22c6 f6f912035...
0c63c011d8 858e63175 eb371762...
0d4e26efb3 3464c66a0 d97f857b...
0d10e2c67 6ba69f6154 2ce6af8e...
0e4545264 b6605ac42 bad8ccf5...
0fd49d1d71
1a45d0441
01a89c22d-
1a01352a6
484 items
Free space: 42.2 TiB (Total: 44.6 TiB)
0
50
100
150
200
dataset-access-...
a96b56cd-59d...
15:49

And finally we can see one of the images with

```
[104]: import matplotlib.pyplot as plt
       import pydicom
       dcm_file_path = os.path.join(serie_dir_pa
       ds = pydicom.dcmread(dcm_file_path)
       plt.imshow(ds.pixel_array, cmap=plt.cm.bo

[104]: <matplotlib.image.AxesImage at 0x7f82fc6a
```

Mounts the studies of the datasets as a POSIX volume.

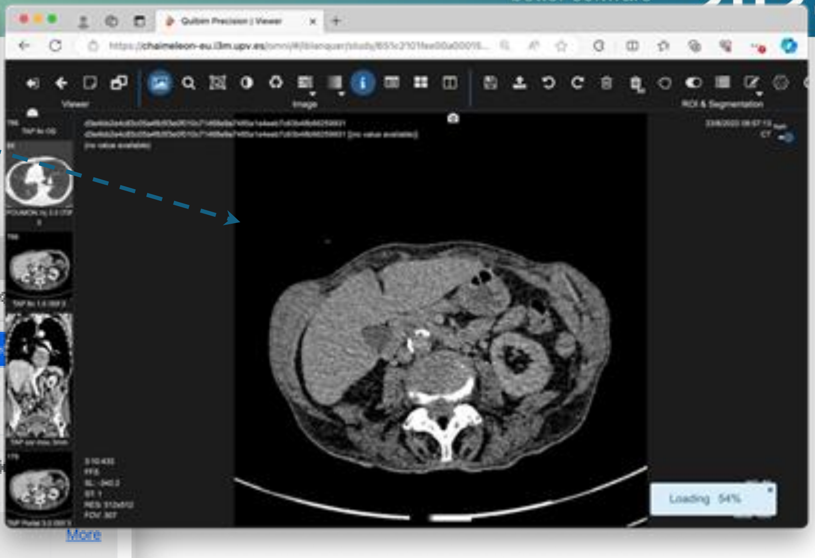
Link to a batch queue system with GPUs and powerful resources.

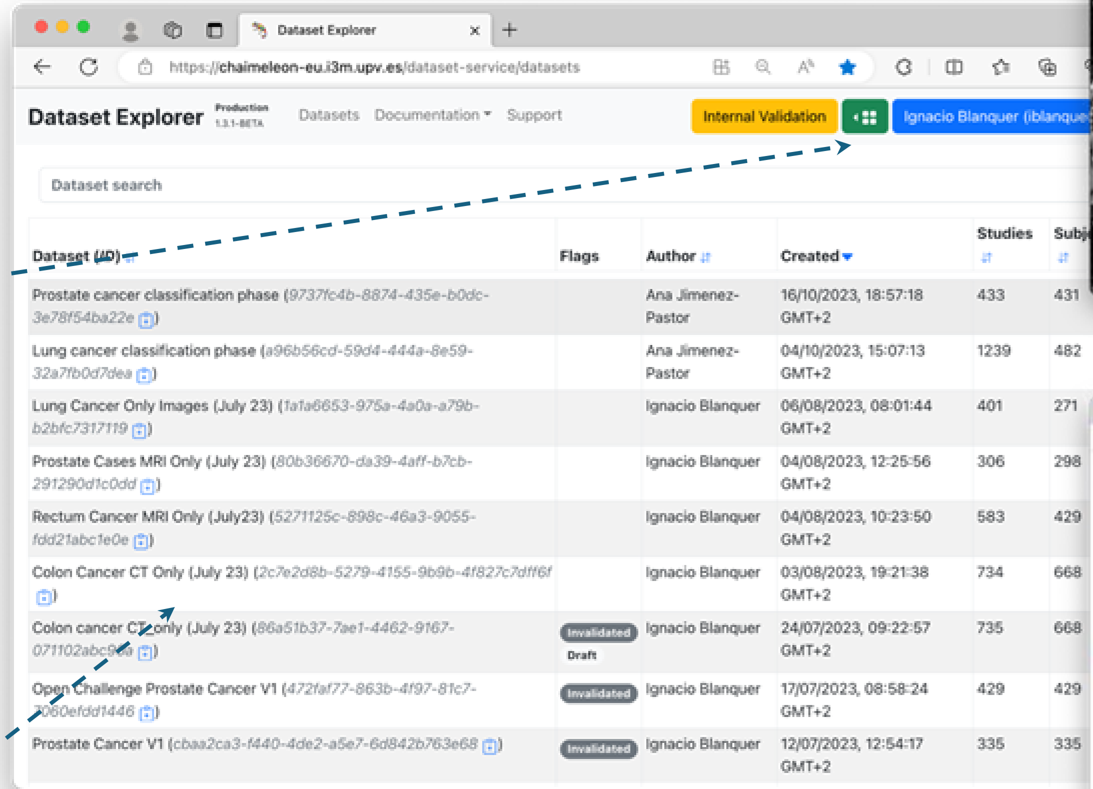
Dataset explorer

Links to the Application Catalogue, Case Explorer and Application access proxy

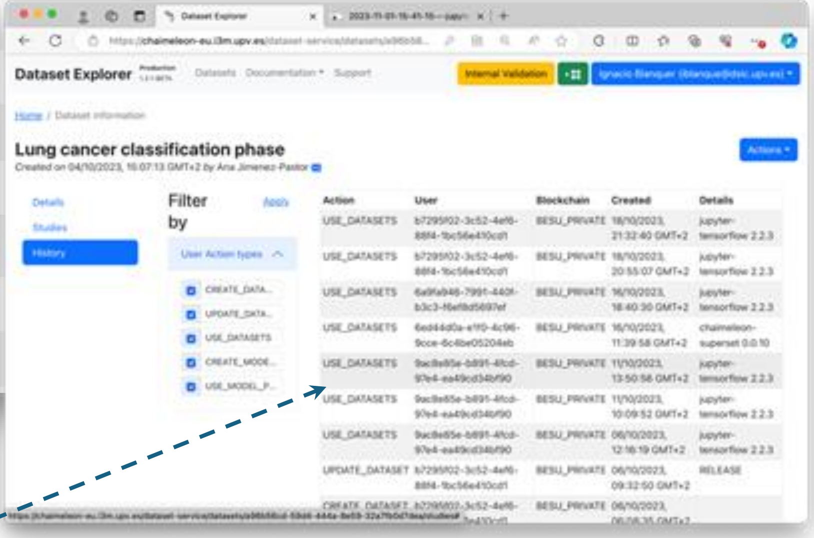
List of the available collections, with PIDs and basic metadata

Explore images and DICOM metadata.


Quibim Precision | Viewer
https://chaimeleon-eu.i3m.upv.es/omni/#/blanquer/study/
Viewer
Image
ROI & Segmentation
Loading 54%



Dataset (ID)	Flags	Author	Created	Studies	Subj	
Prostate cancer classification phase (9737fc4b-8874-435e-b0dc-3e78f54ba22e)		Ana Jimenez-Pastor	16/10/2023, 18:57:18 GMT+2	433	431	More
Lung cancer classification phase (a96b56cd-59d4-444a-8e59-32a7fb0d7dea)		Ana Jimenez-Pastor	04/10/2023, 15:07:13 GMT+2	1239	482	More
Lung Cancer Only Images (July 23) (1a1a6653-975a-4a0a-a79b-b2bfc7317119)		Ignacio Blanquer	06/08/2023, 08:01:44 GMT+2	401	271	
Prostate Cases MRI Only (July 23) (80b36670-da39-4aff-b7cb-291290d1c0dd)		Ignacio Blanquer	04/08/2023, 12:25:56 GMT+2	306	298	
Rectum Cancer MRI Only (July23) (5271125c-898c-46a3-9055-fdd21abc1e0e)		Ignacio Blanquer	04/08/2023, 10:23:50 GMT+2	583	429	
Colon Cancer CT Only (July 23) (2c7e2d8b-5279-4155-9b9b-4f827c7dff6f)		Ignacio Blanquer	03/08/2023, 19:21:38 GMT+2	734	668	
Colon cancer CT_only (July 23) (86a51b37-7ae1-4462-9167-071102abc96a)	Invalidated Draft	Ignacio Blanquer	24/07/2023, 09:22:57 GMT+2	735	668	
Open Challenge Prostate Cancer V1 (472faf77-863b-4f97-81c7-7060efdd1446)	Invalidated	Ignacio Blanquer	17/07/2023, 08:58:24 GMT+2	429	429	
Prostate Cancer V1 (cbaa2ca3-f440-4de2-a5e7-6d842b763e68)	Invalidated	Ignacio Blanquer	12/07/2023, 12:54:17 GMT+2	335	335	



Lung cancer classification phase

Created on 04/10/2023, 15:07:13 GMT+2 by Ana Jimenez-Pastor

☑ CREATE_DATA...
☑ UPDATE_DATA...
☑ USE_DATASETS
☑ CREATE_MODE...
☑ USE_MODEL_P...

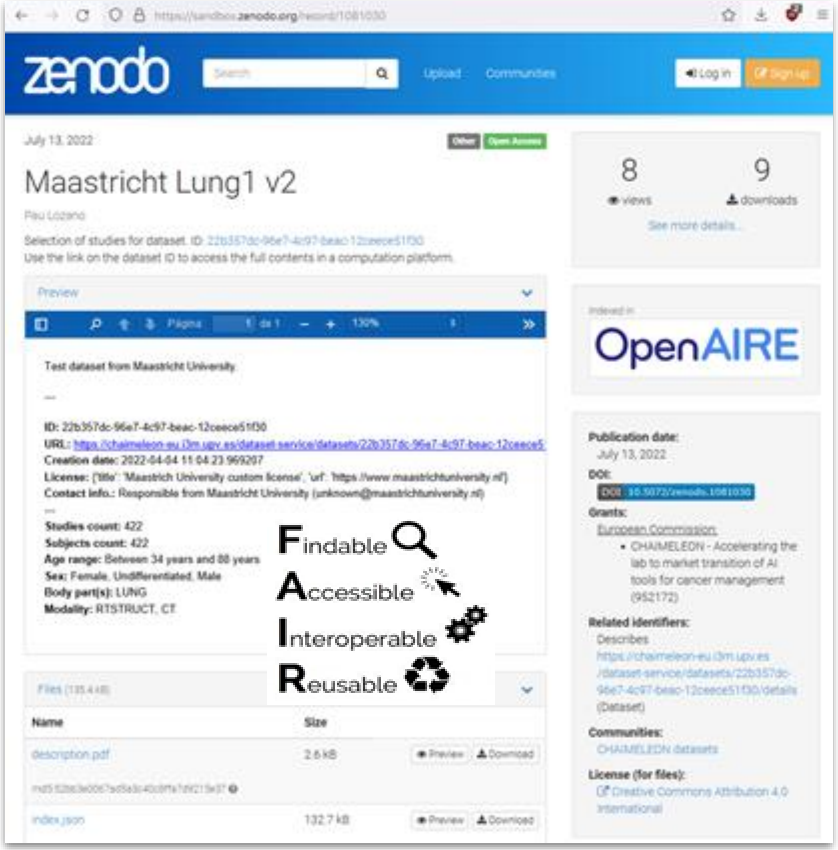
Action	User	Blockchain	Created	Details
USE_DATASETS	b7295f02-3c52-4ef6-88f4-1bc56e410cd1	BESU_PRIVATE	18/10/2023, 21:32:40 GMT+2	jupyter-tensorflow 2.2.3
USE_DATASETS	b7295f02-3c52-4ef6-88f4-1bc56e410cd1	BESU_PRIVATE	18/10/2023, 20:55:07 GMT+2	jupyter-tensorflow 2.2.3
USE_DATASETS	6a9fa946-7991-440f-b3c3-f6ef8d5697ef	BESU_PRIVATE	16/10/2023, 18:40:30 GMT+2	jupyter-tensorflow 2.2.3
USE_DATASETS	6ed44d0a-e1f0-4c96-9cce-6c4be05204eb	BESU_PRIVATE	16/10/2023, 11:39:58 GMT+2	chaimeleon-superset 0.0.10
USE_DATASETS	9ac8e85e-b891-4fcd-97e4-ea49cd34bf90	BESU_PRIVATE	11/10/2023, 13:50:56 GMT+2	jupyter-tensorflow 2.2.3
USE_DATASETS	9ac8e85e-b891-4fcd-97e4-ea49cd34bf90	BESU_PRIVATE	11/10/2023, 10:09:52 GMT+2	jupyter-tensorflow 2.2.3
USE_DATASETS	9ac8e85e-b891-4fcd-97e4-ea49cd34bf90	BESU_PRIVATE	06/10/2023, 12:16:19 GMT+2	jupyter-tensorflow 2.2.3
UPDATE_DATASET	b7295f02-3c52-4ef6-88f4-1bc56e410cd1	BESU_PRIVATE	06/10/2023, 09:32:50 GMT+2	RELEASE
CREATE_DATASET	b7295f02-3c52-4ef6-...e410cd1	BESU_PRIVATE	06/10/2023, 06:08:35 GMT+2	

https://chaimeleon-eu.i3m.upv.es/dataset-service/datasets/a96b56cd-59d4-444a-8e59-32a7fb0d7dea/studies#

Explore additional metadata and access logs.

FAIR Compliance and Traceability

- A Dataset is a FAIR citable research object
  - Datasets have a metadata that contains aggregated information, which could be made public (just metadata).
  - Published Datasets have their metadata publicly accessible.
- Traceability is provided through a Web service
  - Logs users' actions (create / update / use datasets) in the CHAIMELEON repository (the traces)
  - Stores traces in private blockchain(s) (support for BigchainDB & Hyperledger Besu)
  - Anti-tampering, redundancy, and distribution of the complete set of traces
  - It does not store repository's users/patients private/sensitive information
  - API @ https://app.swaggerhub.com/apis/UPV-CHAIMELEON/Trace[illegible]



July 13, 2022

Maastricht Lung1 v2

Pau Lozano

Selection of studies for dataset. ID: 22b357dc-96e7-4c97-beac-12ceece51f30
Use the link on the dataset ID to access the full contents in a computation platform.

Test dataset from Maastricht University

—

ID: 22b357dc-96e7-4c97-beac-12ceece51f30
URL: https://chaimeleon-eu.i3m.upv.es/dataset-service/datasets/22b357dc-96e7-4c97-beac-12ceece5
Creation date: 2022-04-04 11:04:23.969207
License: {'title': 'Maastrich University custom license', 'url': 'https://www.maastrichtuniversity.nl'}
Contact info.: Responsible from Maastricht University (unknown@maastrichtuniversity.nl)
—
Studies count: 422
Subjects count: 422
Age range: Between 34 years and 88 years
Sex: Female, Undifferentiated, Male
Body part(s): LUNG
Modality: RTSTRUCT, CT

Files (135.4 kB)

Name	Size	
description.pdf md5:[illegible]	2.6 kB	Preview Download
index.json	132.7 kB	Preview Download

8 views
9 downloads
See more details...

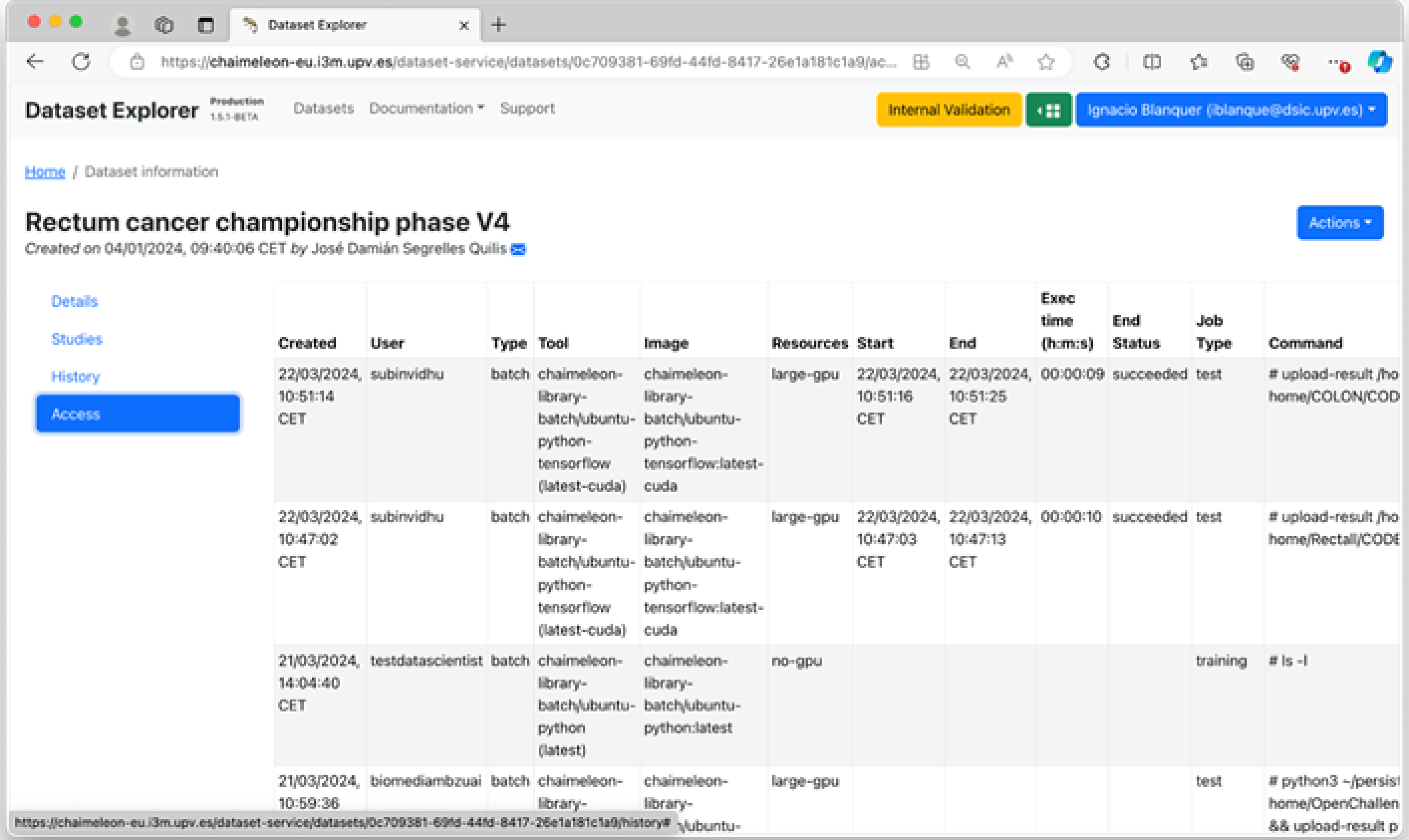
Indexed in
OpenAIRE

Publication date:
July 13, 2022
DOI:
DOI 10.5072/zenodo.1081030
Grants:
European Commission:
- CHAIMELEON - Accelerating the lab to market transition of AI tools for cancer management (952172)

Related identifiers:
Describes
https://chaimeleon-eu.i3m.upv.es/dataset-service/datasets/22b357dc-96e7-4c97-beac-12ceece51f30/details (Dataset)
Communities:
CHAIMELEON datasets
License (for files):
Creative Commons Attribution 4.0 International

Logging workloads

- Along with the usage of the datasets (through the Interactive Applications), access through the batch processing tools is also registered.
- Detailed information is persisted on a Blockchain
  - Dataset creation (User and Type)
  - Dataset Access (user, environment, start and end)
  - Job (Execution time, End Status, Job Type, Command)



Rectum cancer championship phase V4

Created on 04/01/2024, 09:40:06 CET *by* José Damián Segrelles Quilis

Created	User	Type	Tool	Image	Resources	Start	End	Exec time (h:m:s)	End Status	Job Type	Command
22/03/2024, 10:51:14 CET	subinvidhu	batch	chaimeleon-library-batch/ubuntu-python-tensorflow (latest-cuda)	chaimeleon-library-batch/ubuntu-python-tensorflow:latest-cuda	large-gpu	22/03/2024, 10:51:16 CET	22/03/2024, 10:51:25 CET	00:00:09	succeeded	test	# upload-result /ho home/COLON/COD
22/03/2024, 10:47:02 CET	subinvidhu	batch	chaimeleon-library-batch/ubuntu-python-tensorflow (latest-cuda)	chaimeleon-library-batch/ubuntu-python-tensorflow:latest-cuda	large-gpu	22/03/2024, 10:47:03 CET	22/03/2024, 10:47:13 CET	00:00:10	succeeded	test	# upload-result /ho home/Rectal/CODE
21/03/2024, 14:04:40 CET	testdatascientist	batch	chaimeleon-library-batch/ubuntu-python (latest)	chaimeleon-library-batch/ubuntu-python:latest	no-gpu					training	# ls -l
21/03/2024, 10:59:36	biomediambzuai	batch	chaimeleon-library-	chaimeleon-library- /ubuntu-	large-gpu					test	# python3 ~/persis home/OpenChallen && upload-result p

https://chaimeleon-eu.i3m.upv.es/dataset-service/datasets/0c709381-69fd-44fd-8417-26e1a181c1a9/history#

Testing the platform – Open Challenges

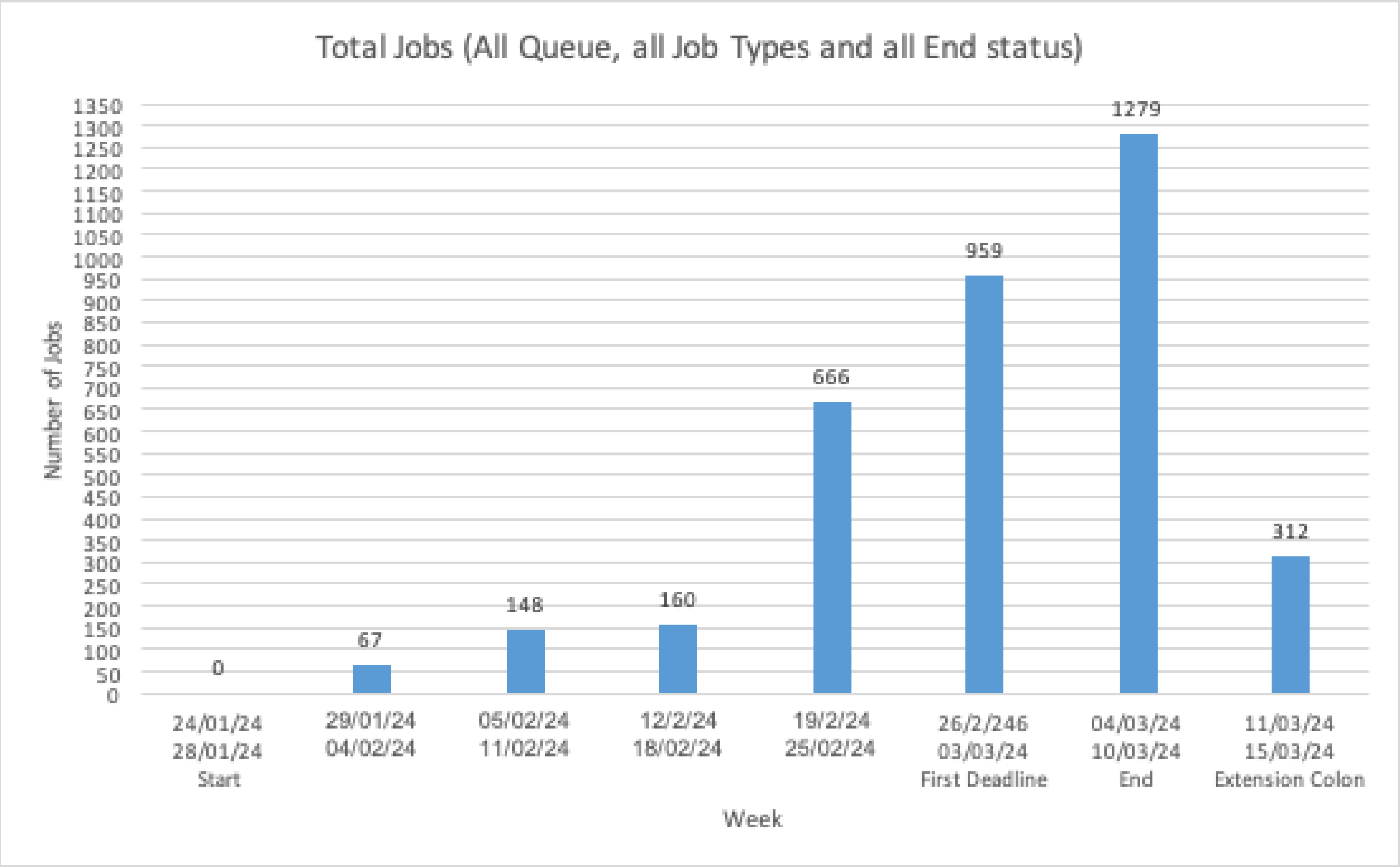
https://chaimeleon.eu/open-challenges/

- Five clinical questions
  - Related to Lung, Prostate, Colon, Breast and Rectal Cancer
- Two-stage championship
  - Classification phase
    - 161 participants sent the organizers the T&C signed (240 got interest).
    - Provided with synthetic data for Prostate and Lung cancer.
  - Championship pase
    - 30 Qualified, 20 active participated.
    - In-situ Access.
    - Access provided for the full 5 datasets.

VMs	Description	Jobs total	GPU	VRAM	Cores Total	RAM total	Disk for docker images	Disk for jobs containers	Total ephemeral disk
7	(2022.XLarge) for the master node, core services	-	-	-	56	224 GB	280 GB	280 GB	560 GB
3	(2022.XLarge) for the CEPH master and storage nodes	-	-	-	24	96 GB	120 GB	120 GB	240 GB
5	(2023.Chaimeleon.XLarge) for desktops	40	-	-	40	320 Gb	400 GB	300 GB (60 each job)	700 GB (140 each VM)
5	(2023.Chaimeleon.XLarge-V100) for large-gpu jobs	5	5 x V100 32GB	5x32 GB	40	320 GB	160 GB	300 GB (60 each job)	700 GB (140 each VM)
2	(2023.Chaimeleon.XXL128-A30) for small-gpu jobs	8	2 x A30 24GB	8x6 GB	32	256 GB	160 GB	360 GB (45 each)	520 GB
2	(2023.Chaimeleon.XXL128-A30) for medium-gpus jobs	4	2 x A30 24GB	4x12 GB	32	256 GB	160 GB	360 GB (60* each job)	520 GB
1	(2023.CHAIMELEON.Ceph) for Ceph persistent home	-	-	-	8	128 GB	80 GB	-	80 GB
25		**57**			**232**	**1.600 GB**	**1.360 GB**	**1.720 GB**	**3.320 GB**

Total VCPU Capacity: 284.801,93 VM/hours for the Open Challenge period

Open Challenge workload

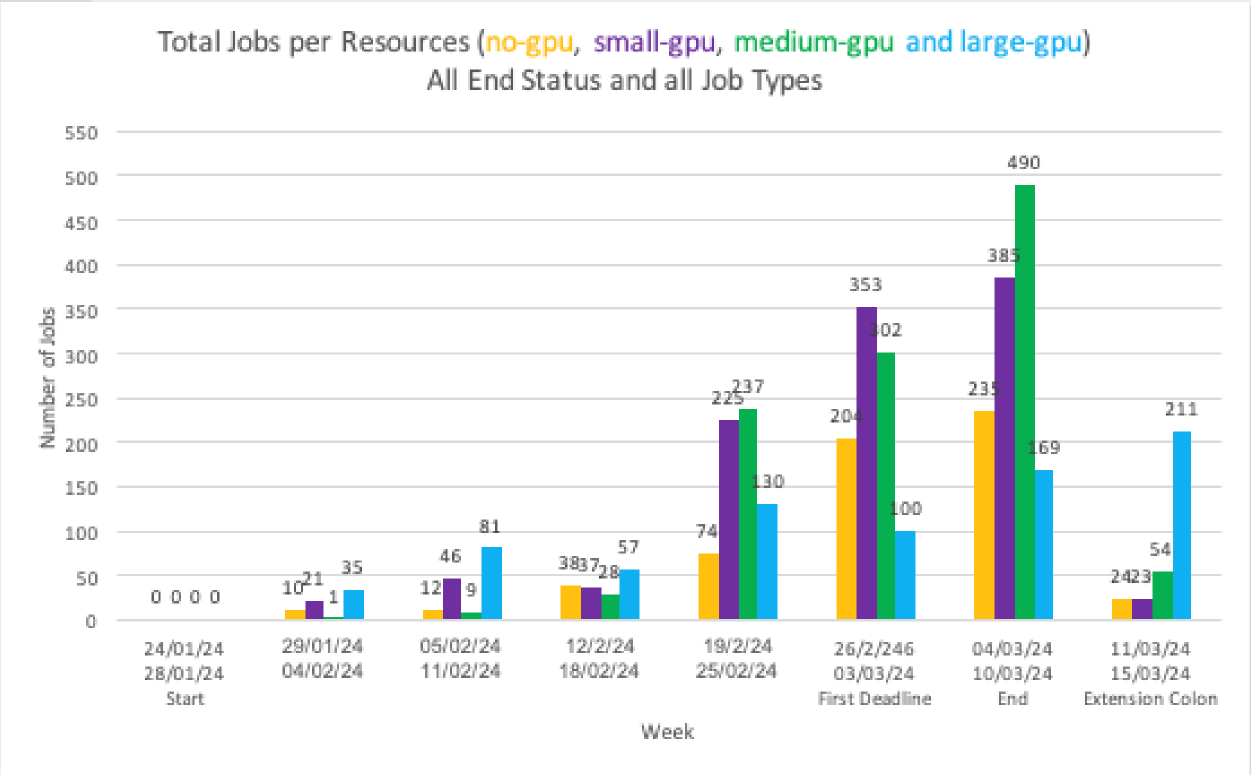


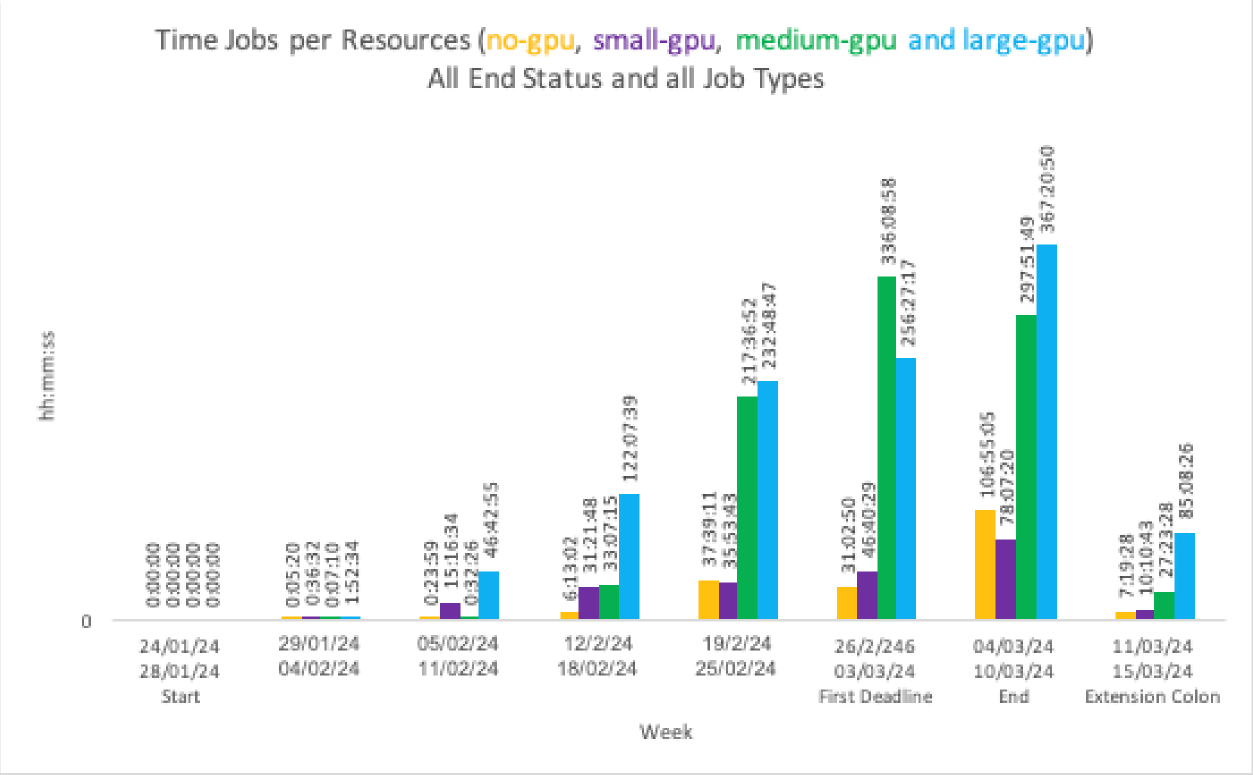
- **A total of 3.591 Batch Jobs** have been submitted during the championship phase. These corresponds to more than **2.433 hours.**

- The increase in weeks of batch jobs may be explained by the fact that users have gained experience with the platform and the closeness to the championship deadline.

- Significant increase after weeks 4 and 5.

Open Challenge workload





- GPUs resources are the ones with the highest number of submissions and running time.
- “large-gpu” has a lowest number of jobs but the highest computational time.
    - This can be explained by the fact that final training has been performed in the on “large-gpu” meanwhile preliminary adjustments have been done on “small-gpu” or “medium-gpu”.

Conclusions

Strengths

- The amount of resources has been over-dimensioned
  - We expected 40 users but 29 were selected and 20 were active.
  - The queueing time has been negligible.
  - We estimate that with the current computing resources, at least 60 users could have been supported.
- The platform has very few technical incidences
  - Just 1 resource had network issues which were solved without losing the running workload.
  - The platform support worked well.
- The number of jobs submitted was reasonably high (over 3.500 jobs, mostly on three weeks)

Weaknesses

- The learning curve is steep
  - Most of the complaints were addressed in the documentation (including those related to the data format)
  - We need to consider a proactive learning process and a better documentation, as well as a mandatory training period.
- The number of failed jobs is considerably high
  - Many of them, according to the information in the forum, were due to a wrong understanding on how the storage system works.