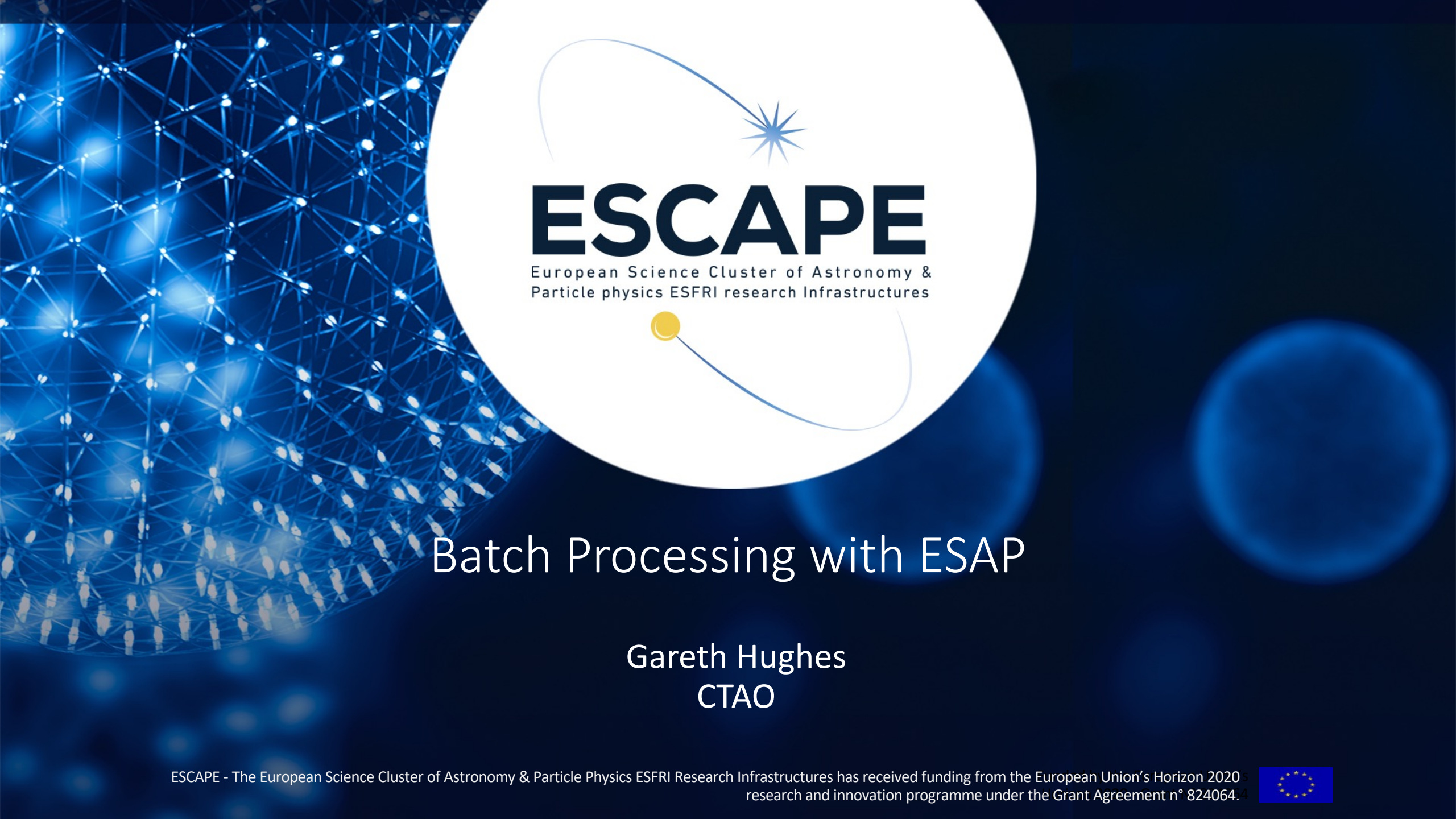


Batch Processing with ESAP

Gareth Hughes
CTAO

ESCAPE - The European Science Cluster of Astronomy & Particle Physics ESFRI Research Infrastructures has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement n° 824064.

Outline

- What and why we want batch computing
  - general use cases
- The plan we developed
- What we built
- What still can be done (currently)
- The future possibilities

- A gamma-ray perspective

- Thanks to:
  - Klass Kliffen
  - John Swinbank
  - Hugh Dickinson
  - Dave Morris

https://www.cta-observatory.org/

Funded by the European Union's Horizon 2020 - Grant N° 824064

Use Cases

- Batch processing is when a computer processes **a number of tasks that it has collected in a group**. It is designed to be a completely **automated** process, without human intervention. It can also be called workload automation (WLA) and **job scheduling**.

- Simulations (Monte Carlo)
  - Used in counting experiments to characterize the instrument
- Data Analysis (Quality checking / Science)
  - Set of observations taken over a long period of time
  - Spread over many files
  - But the same workflow to be applied (with some caveats)

Funded by the European Union's Horizon 2020 - Grant N° 824064

When is Workload Management useful?

Data Lake | **Science Archive**

Staff Scientists — Automated

Batch

Science User — Interactive

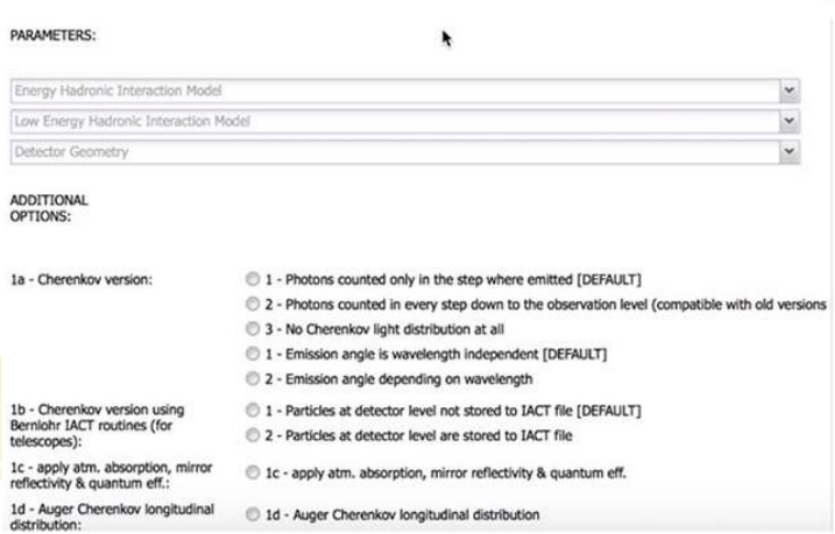
Data Level 0 Raw → **DL1/2** → **DL3** Science → **DL4/6** → **DL6** High

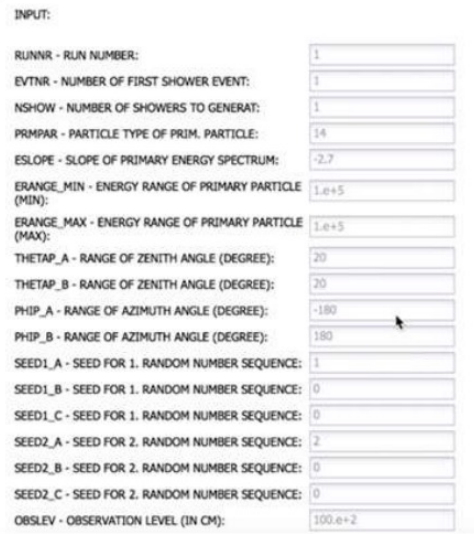
File size

Funded by the European Union's Horizon 2020 - Grant N° 824064

CONCORDIA (WP3/OSSR Task 3.3)

- ConCORDIA (Container Corskia DIRAC)
- Provide a common simulation tool between experiments (CTA & KM3NeT)
  - Singularity containers to be deployed on the GRID
- Integrated as a DIRAC WebApp
  - Access to GRID job submissions and existing resources utilities
- EGI resources for DIRAC access
  - First developments in the EGI DIRAC-client docker
  - First tests on EGI-connected resources
- Containers can be tweaked on-demand
  - GUI for container creation
  - Tune the simulation parameters according to needs
  - CORSIKA setup and runcards
- Running the containers:
  - Scripting access and management
  - GUI access and management

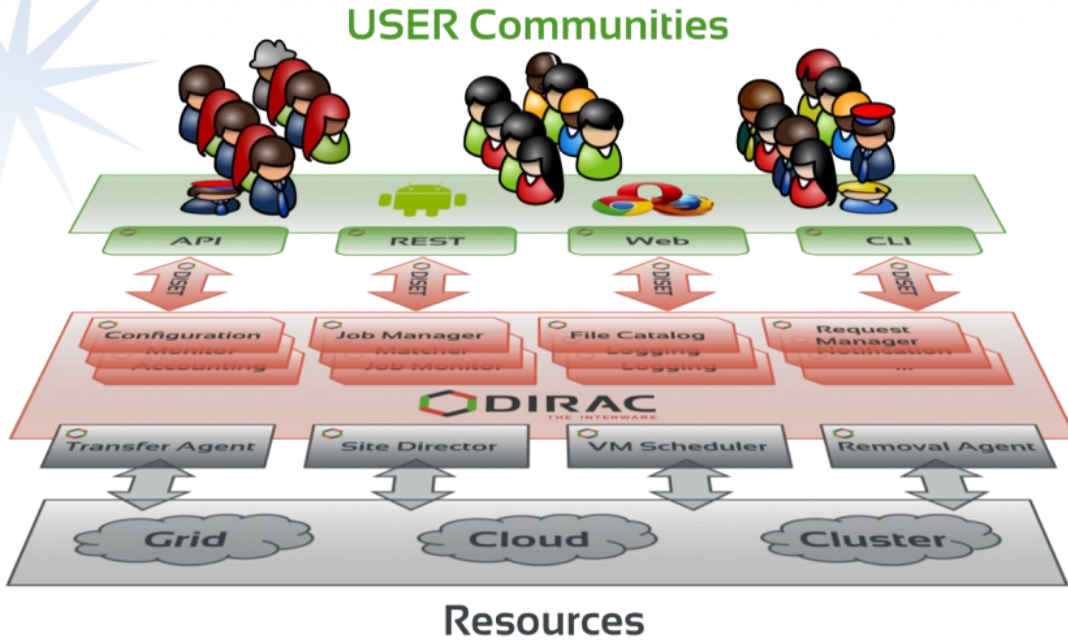




Funded by the European Union's Horizon 2020 - Grant N° 824064

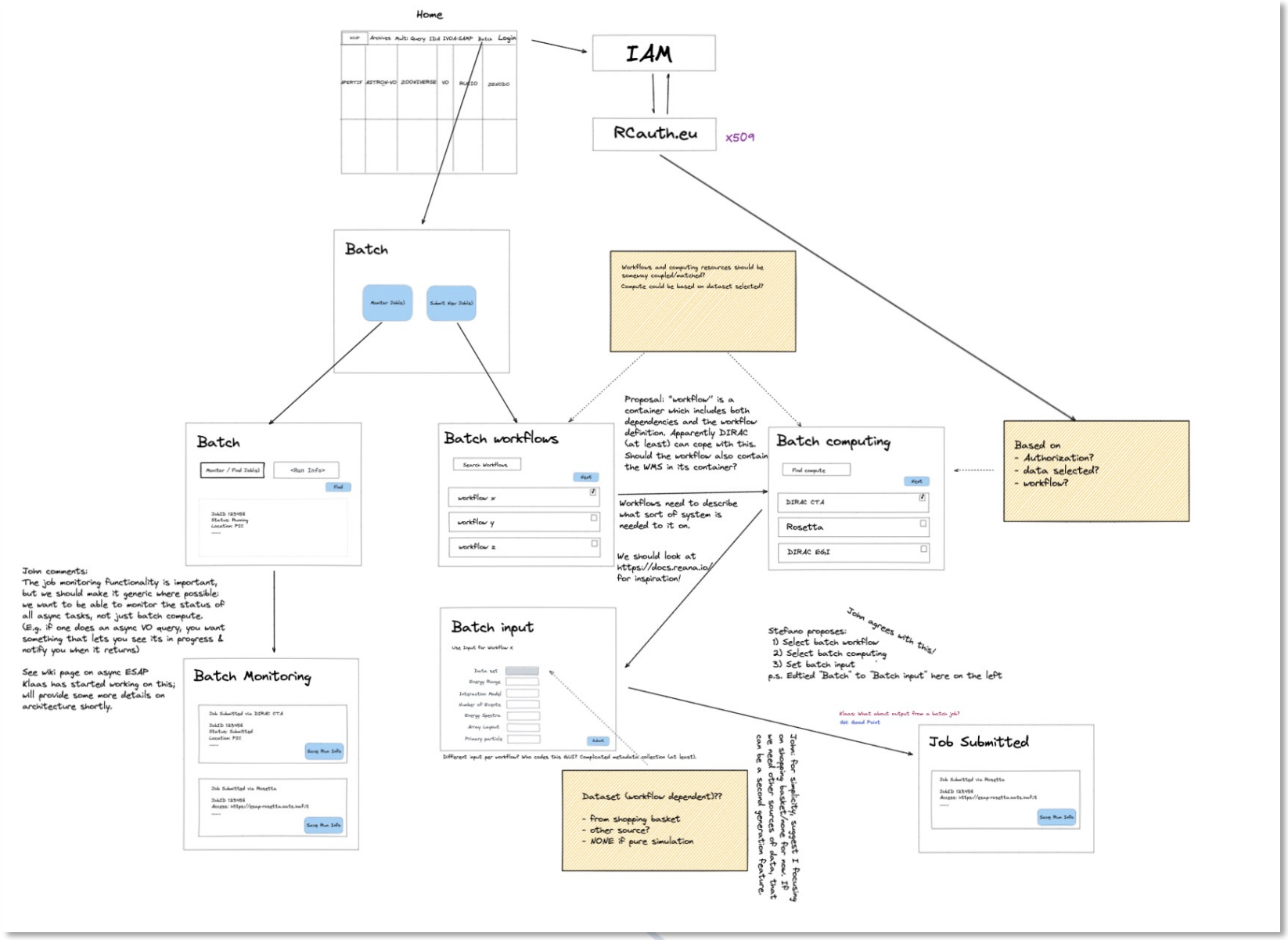
What is DIRAC

- DIRAC: **D**istributed **I**nfrastructure with **R**emote **A**gent **C**ontrol
- An open source **software framework** for distributed computing
  - http://diracgrid.org/
- Started at CERN, LHCb
  - Used by a large number of high energy and astronomy experiments
- Systems include:
  - workload management
  - data management
  - Job management API
  - accounting (provenance)
  - +++ …. Much more
- DIRAC Workload Management System (WMS)
  - uses novel approach of pilot jobs
  - allows for detailed job monitoring
  - can submit to a wide range of computing centers (HPC/HTC/Cloud)
  - able to create workflows: full automatization of multi-step workflow execution
- CTA has its own CTADIRAC instance



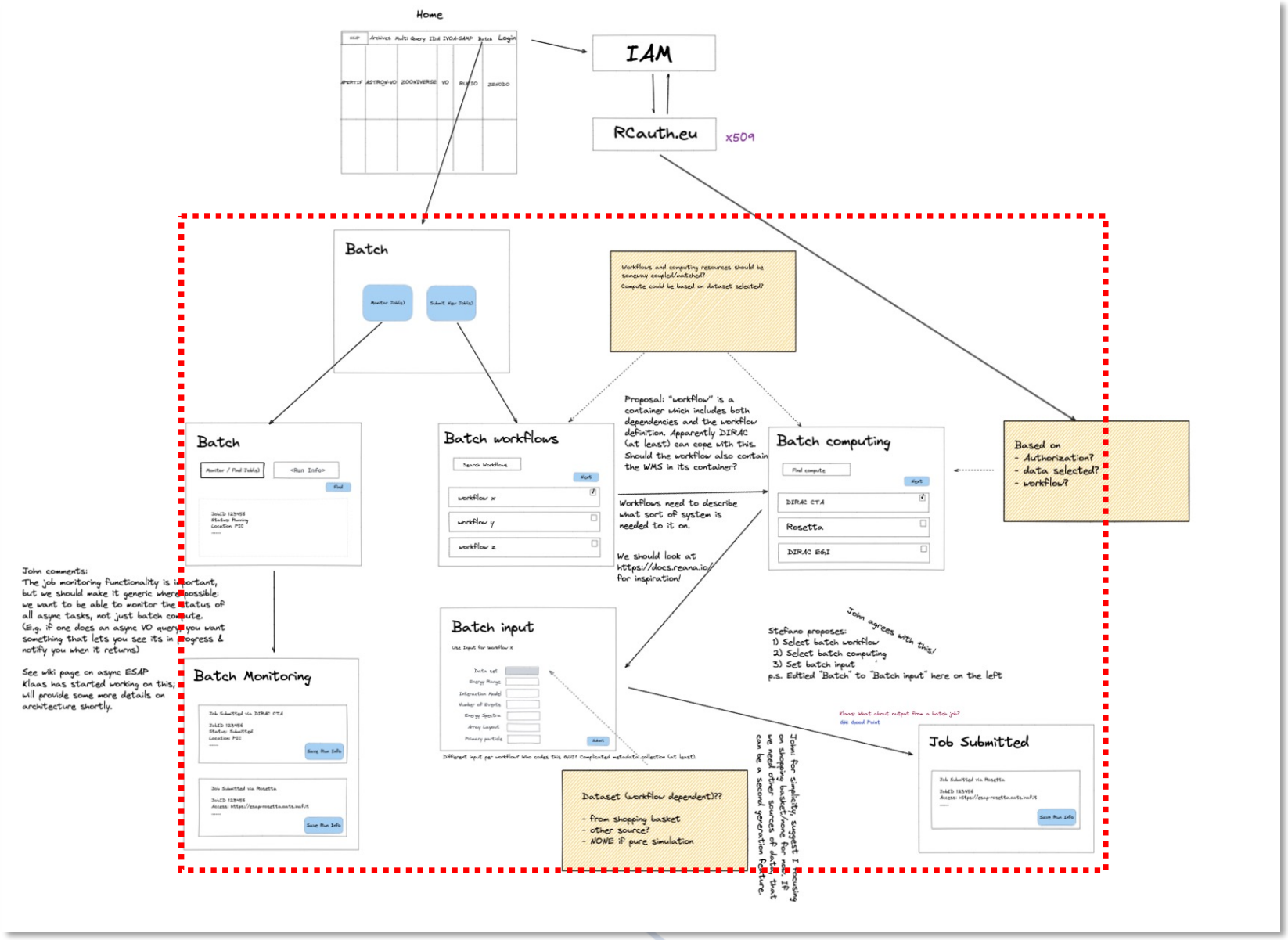
Funded by the European Union's Horizon 2020 - Grant N° 824064

Sketch



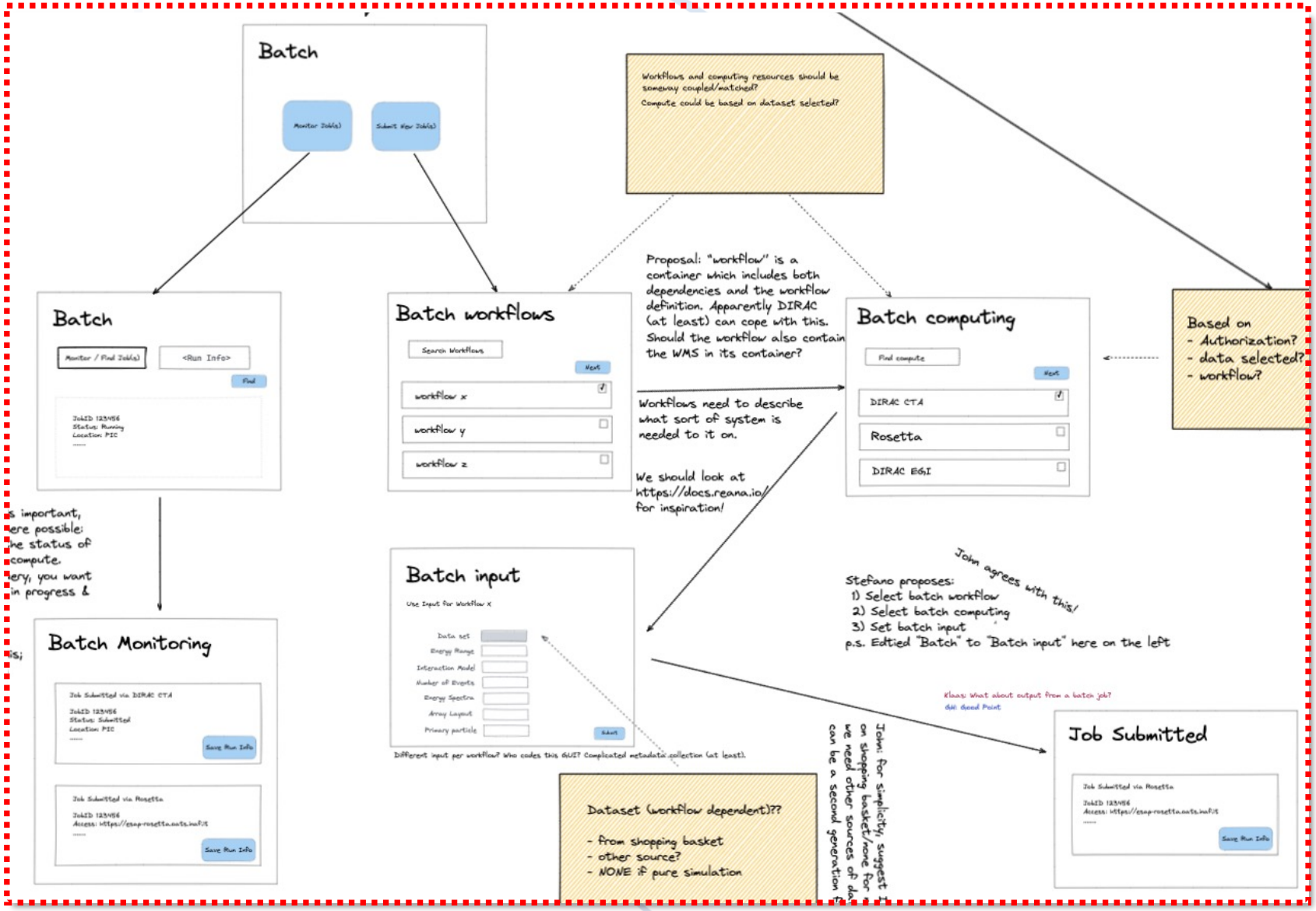
Funded by the European Union's Horizon 2020 - Grant N° 824064

Sketch



Funded by the European Union's Horizon 2020 - Grant N° 824064

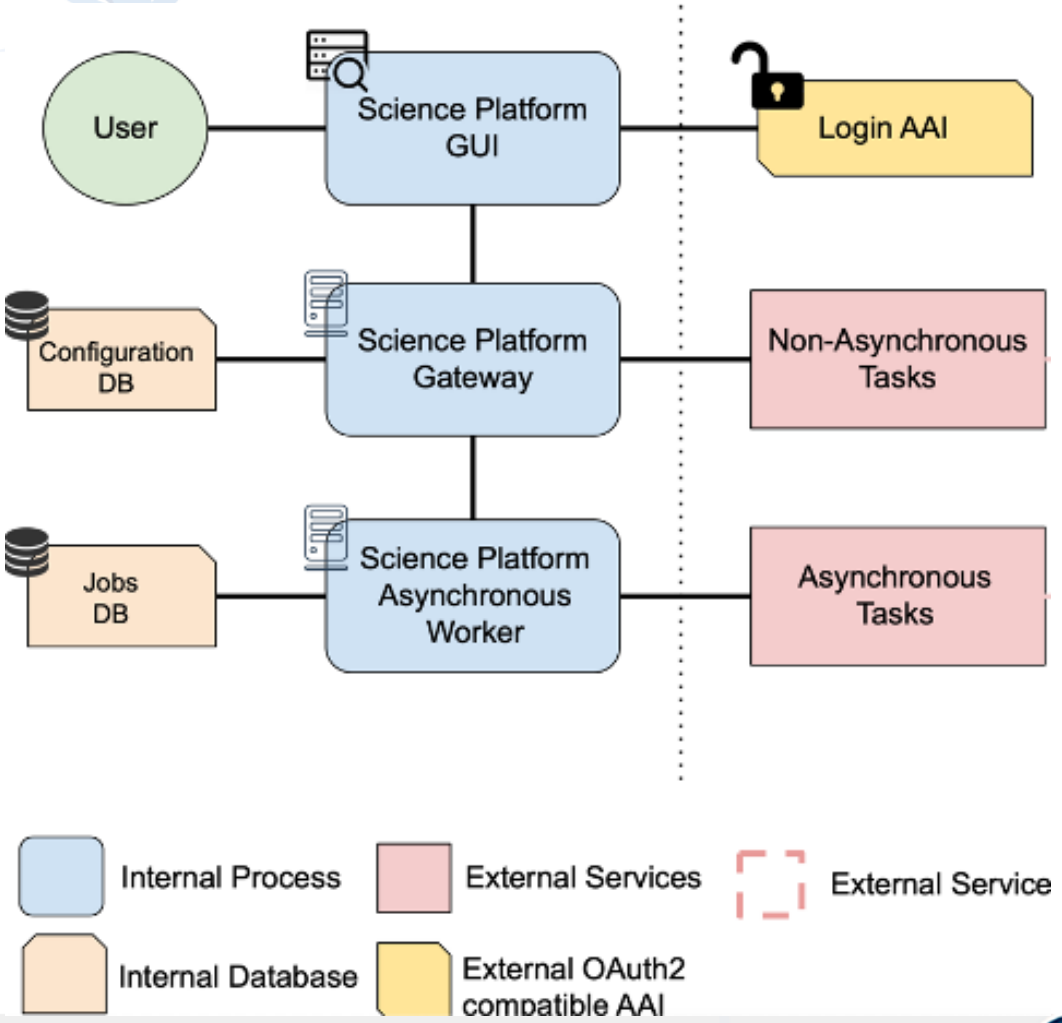
Sketch



Funded by the European Union's Horizon 2020 - Grant N° 824064

Components

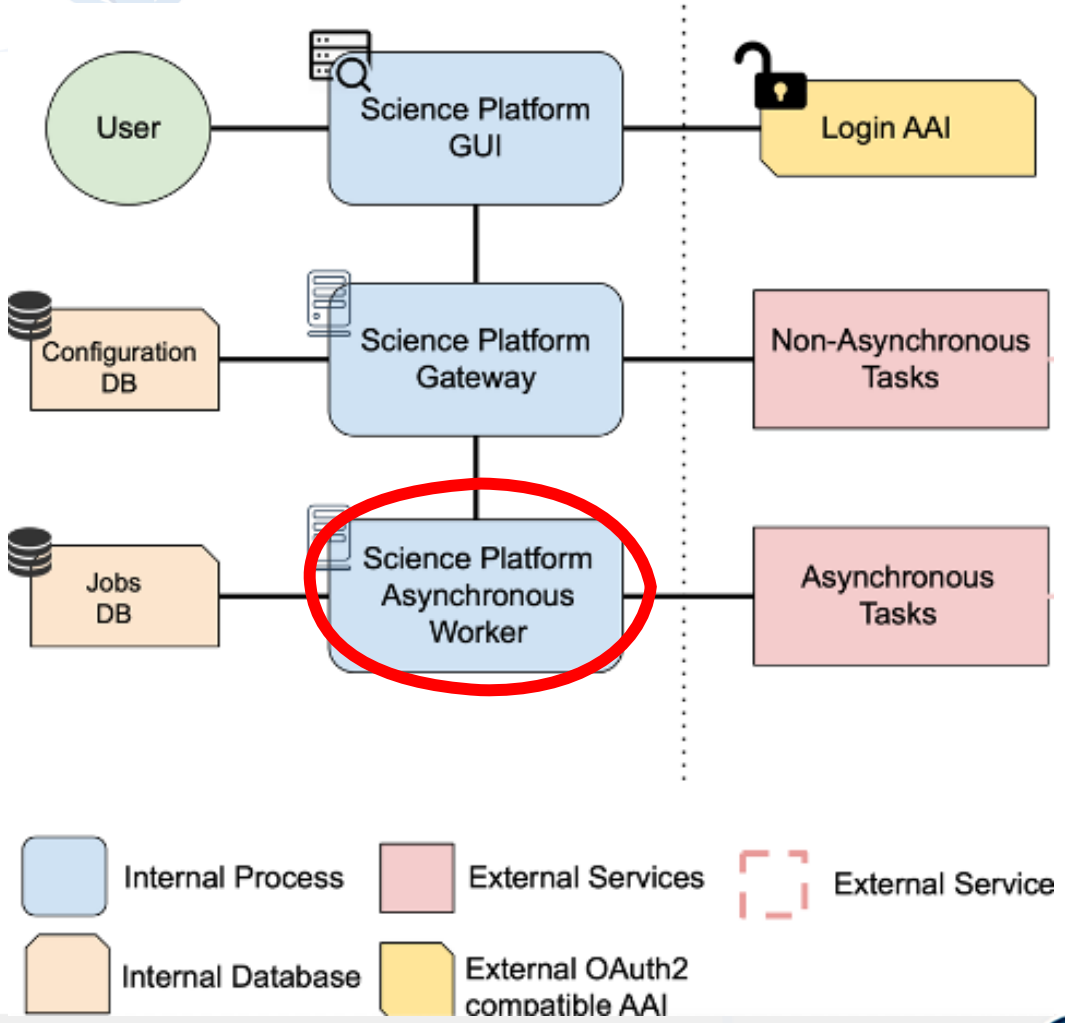
- Involves all three major components of ESAP
- Authorization will be mentioned at the end
- Main part lives in the Asynchronous Worker



Funded by the European Union's Horizon 2020 - Grant N° 824064

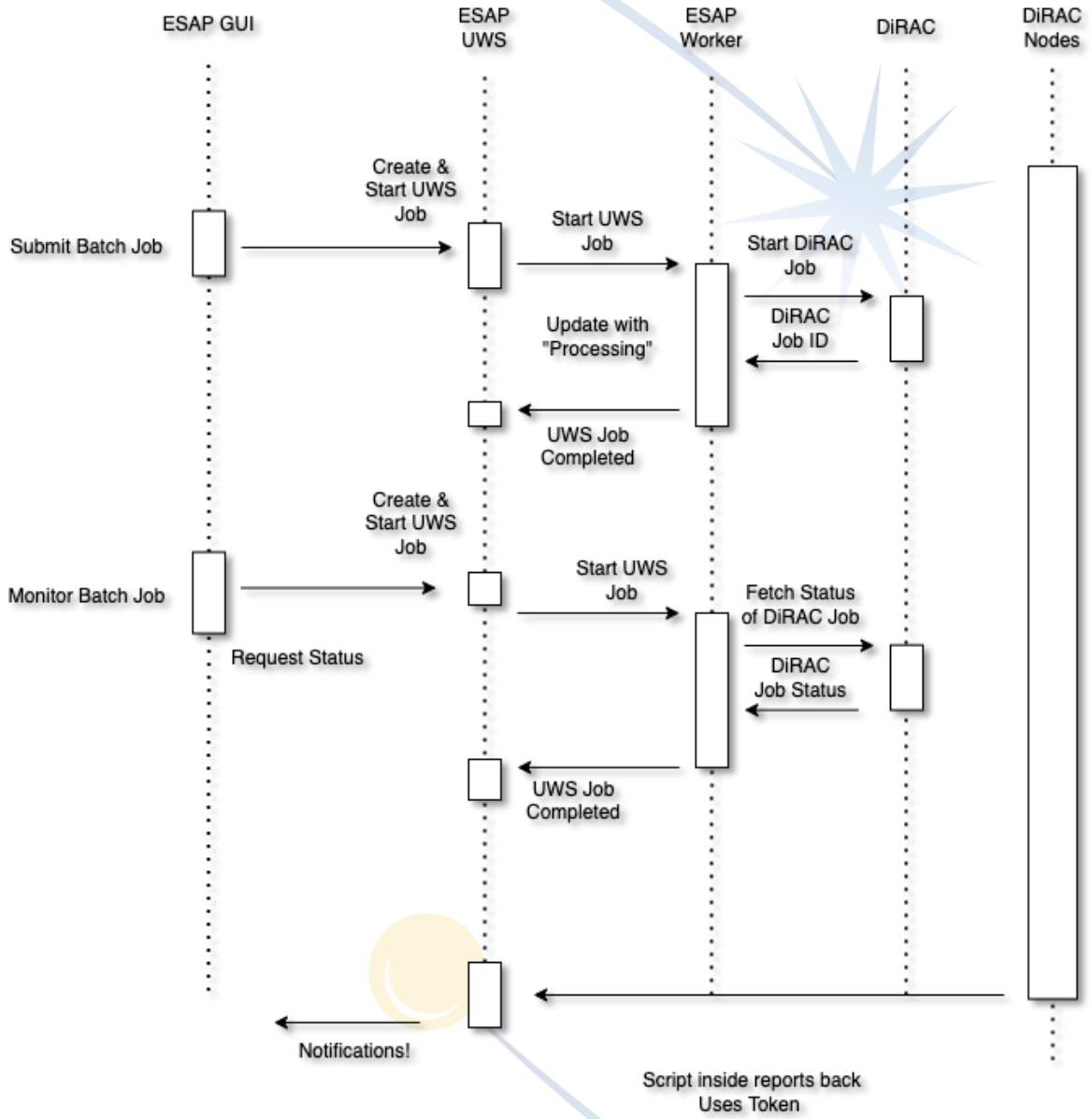
Components

- Involves all three major components of ESAP
- Authorization will be mentioned at the end
- Main part lives in the Asynchronous Worker



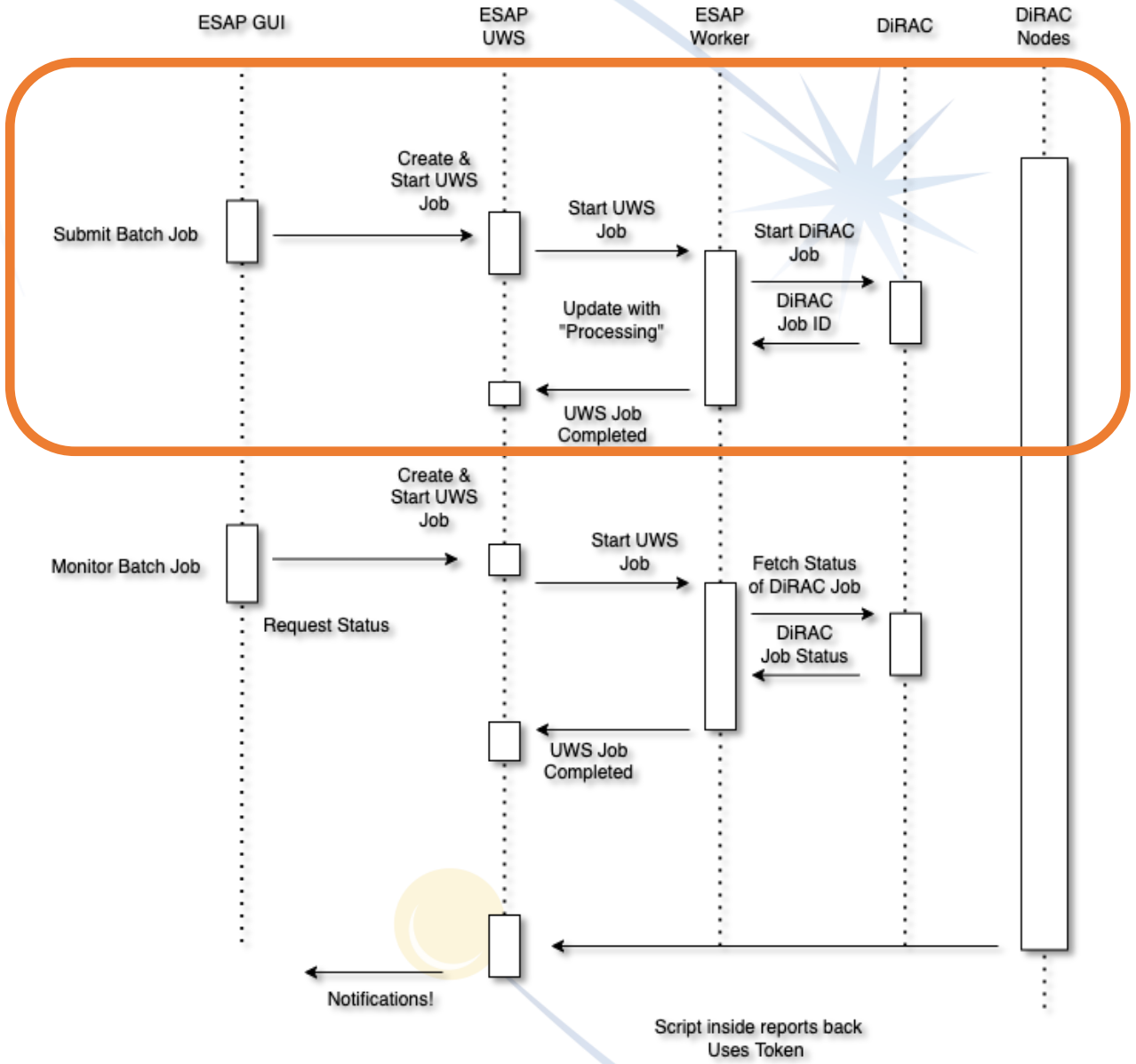
Funded by the European Union's Horizon 2020 - Grant N° 824064

The Plan



Funded by the European Union's Horizon 2020 - Grant N° 824064

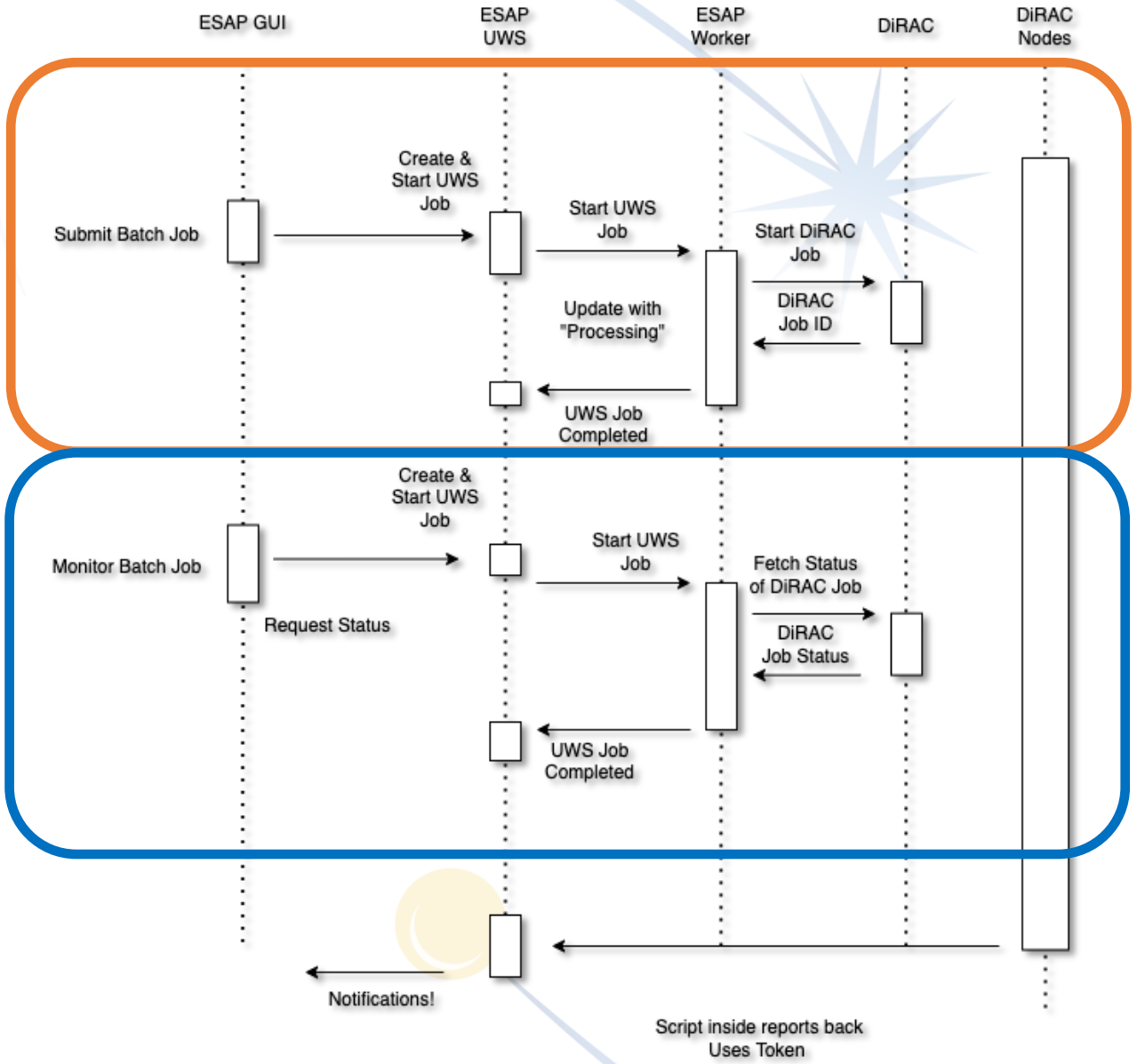
The Plan



Submission

Funded by the European Union's Horizon 2020 - Grant N° 824064

The Plan

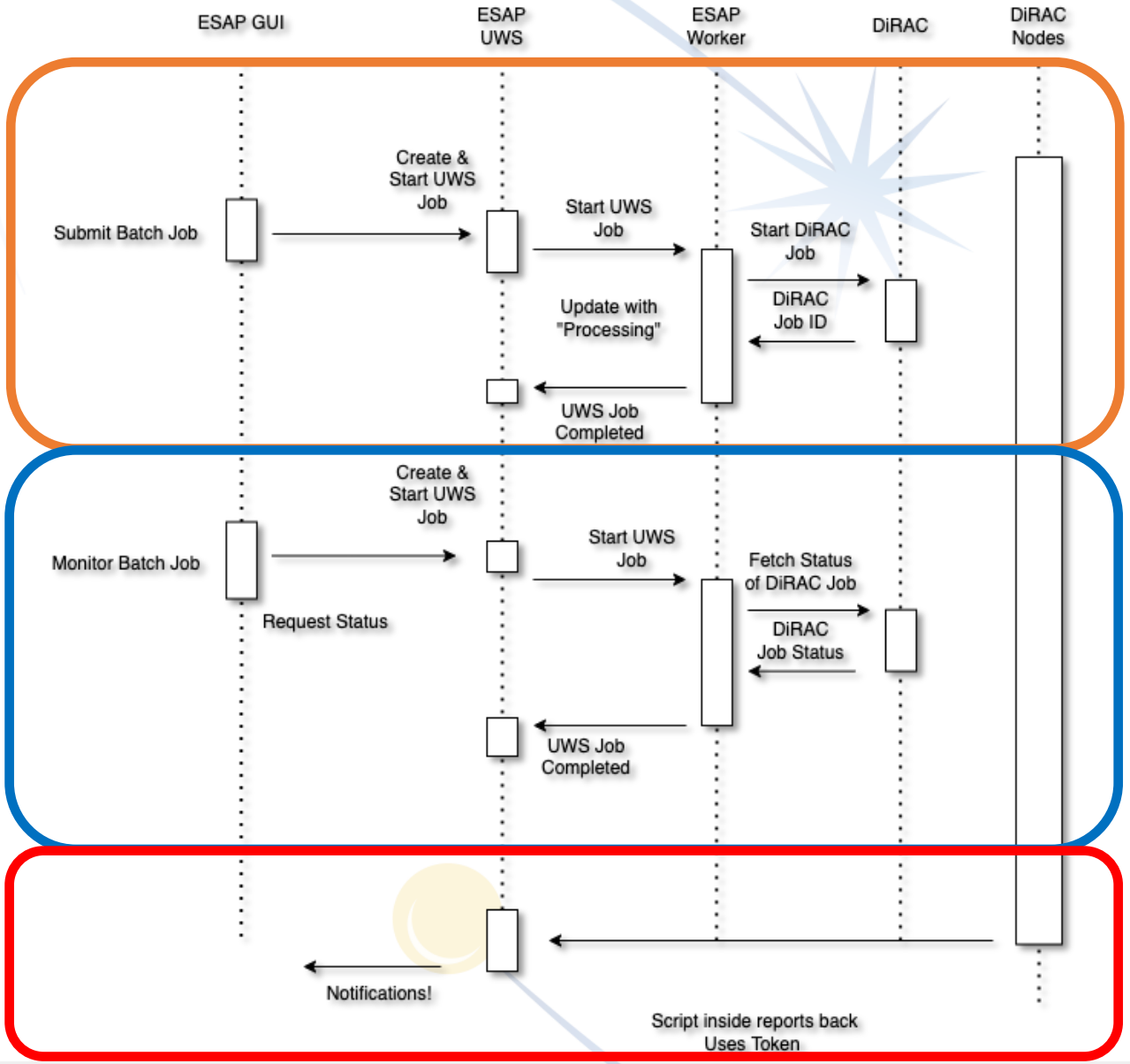


Submission

Monitoring

Funded by the European Union's Horizon 2020 - Grant N° 824064

The Plan



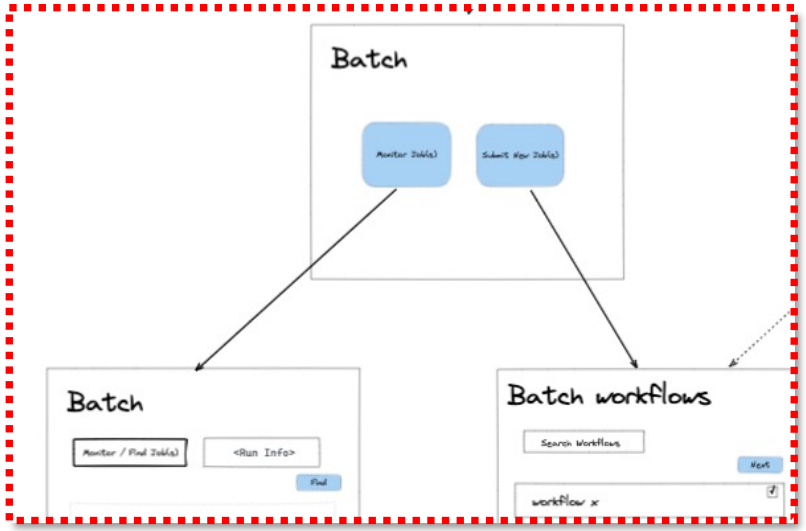
Submission

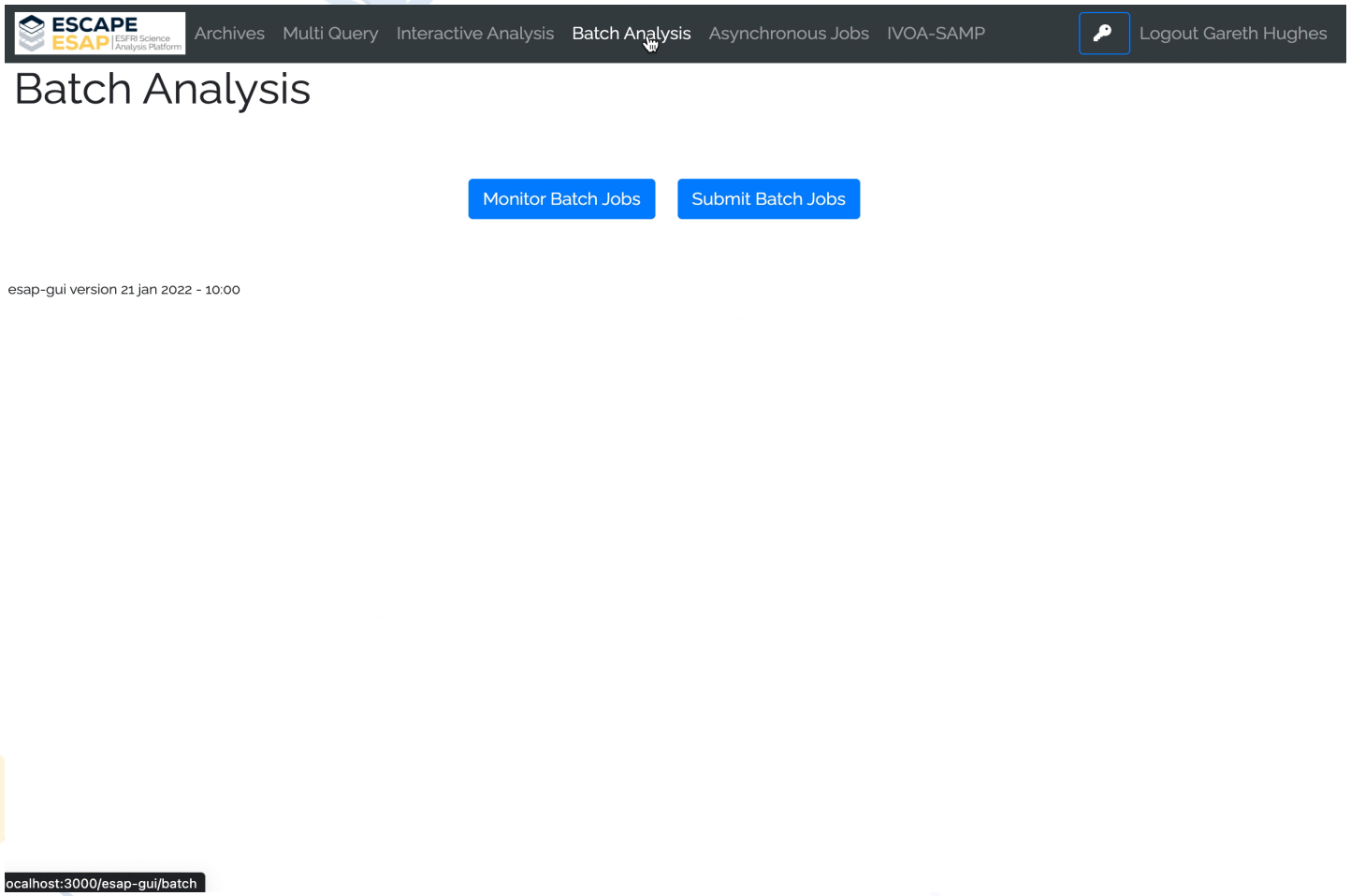
Monitoring

Reporting

Funded by the European Union's Horizon 2020 - Grant N° 824064

Batch/DIRAC on ESAP



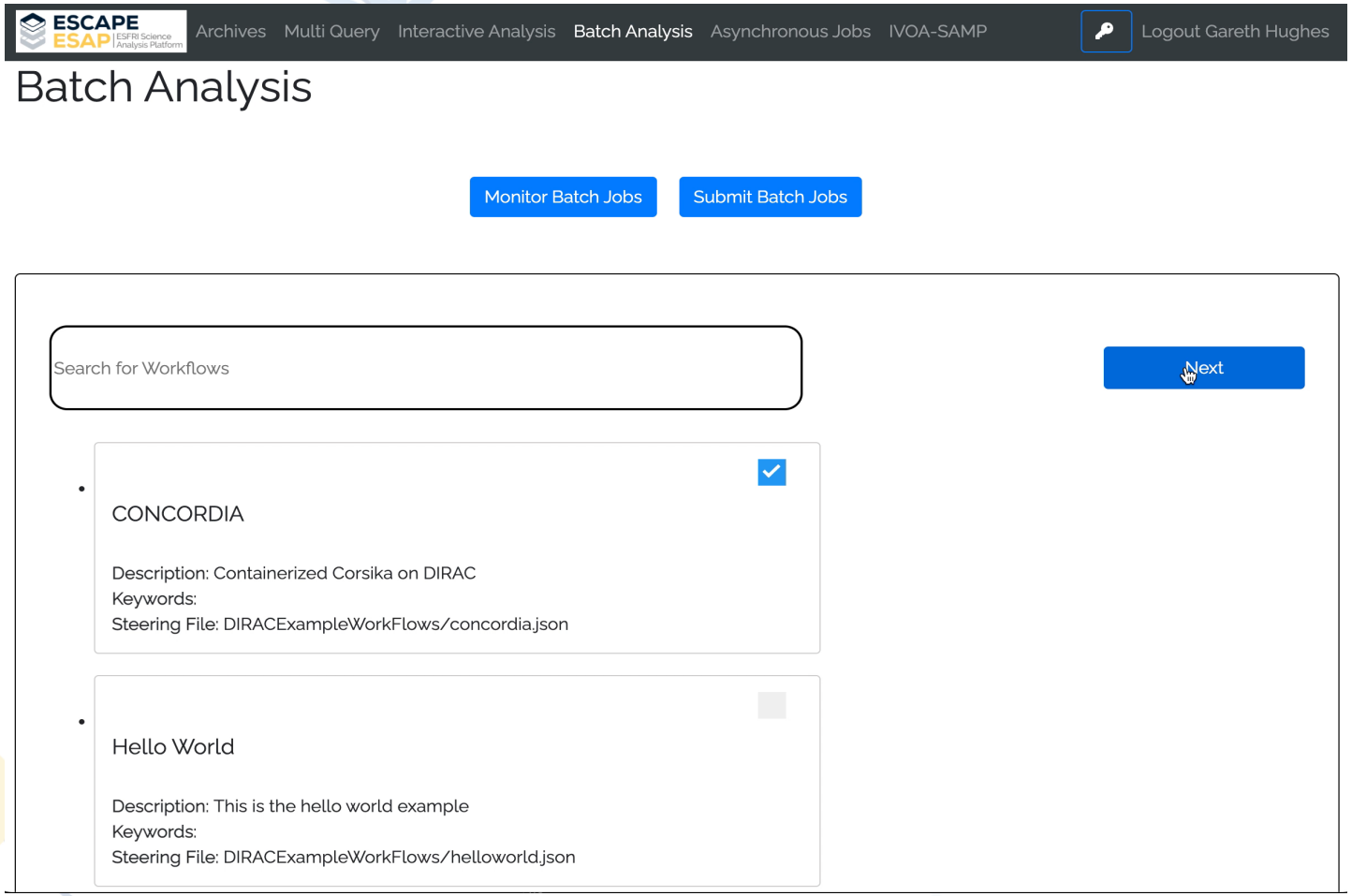


https://youtu.be/4kAskIWFuFk

Funded by the European Union's Horizon 2020 - Grant N° 824064

Batch/DIRAC on ESAP

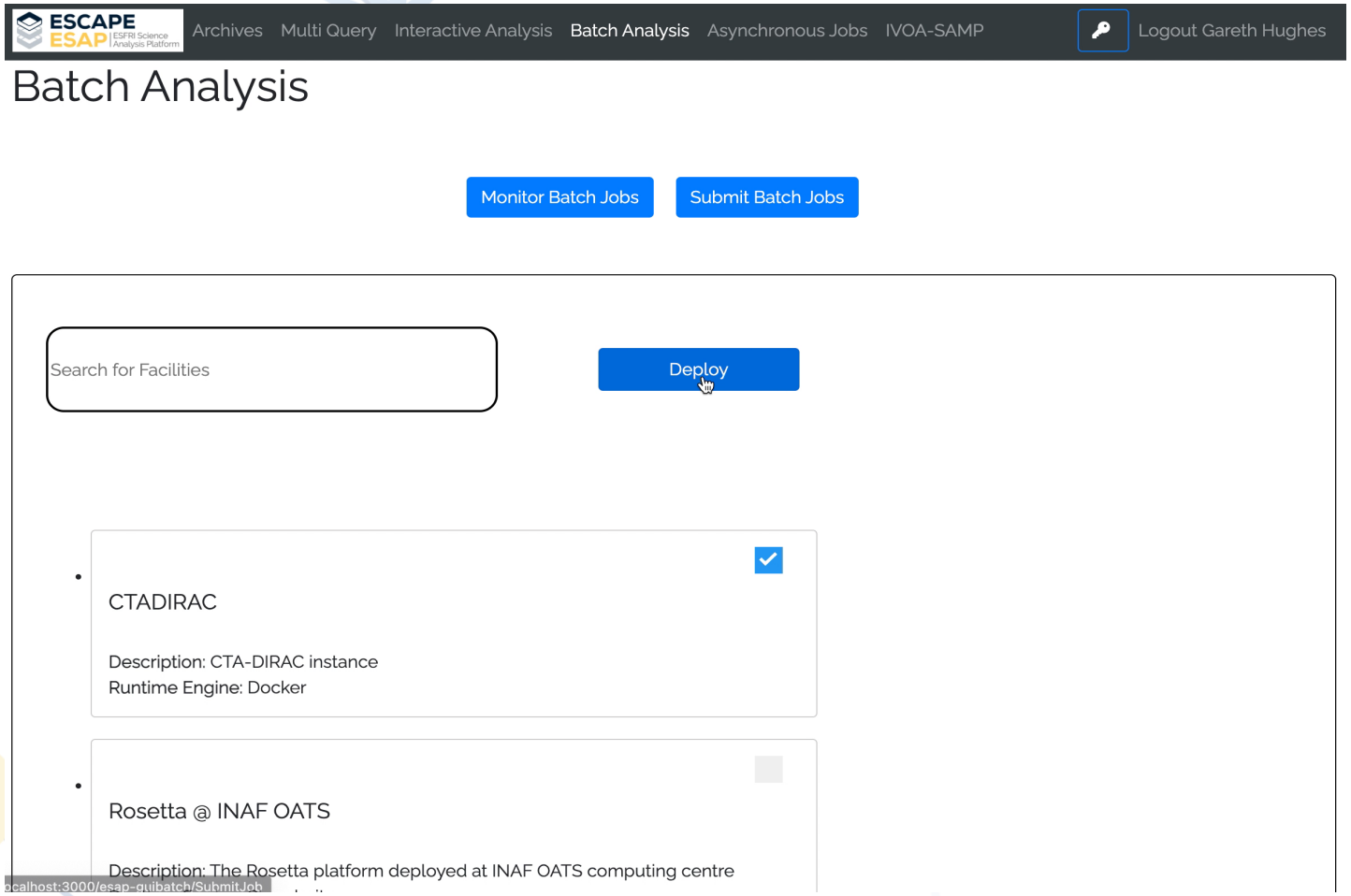
```
[16:56:ghughes@DIRACExampleWorkFlows]$ more concordia.json
{
   "uuid": "fa2b5ebc-ee25-10ec-7ea1-0247ac120006",
   "type": "URL",
   "title": "ESAP CONCORDIA",
   "description": "This workflow is the example",
   "workflow": "CONCORDIA",
   "Parameters": {
     "singularity_image": "http://doublehi5.com/helloworld.sif",
     "output_file_name ": "output.tar",
     "cputime": "500",
     "job_name": "ESAPTEST",
     "group_name": "ESAPTEST_GROUP",
     "dirac_path": "/vo.cta.in2p3.fr/user/g/ghughes/",
     "dirac_SE": "CC-IN2P3-USER",
     "arguement1": "1",
     "arguement2": "2",
     "arguement3": "3"
   }
}
```



Funded by the European Union's Horizon 2020 - Grant N° 824064

Batch/DIRAC on ESAP

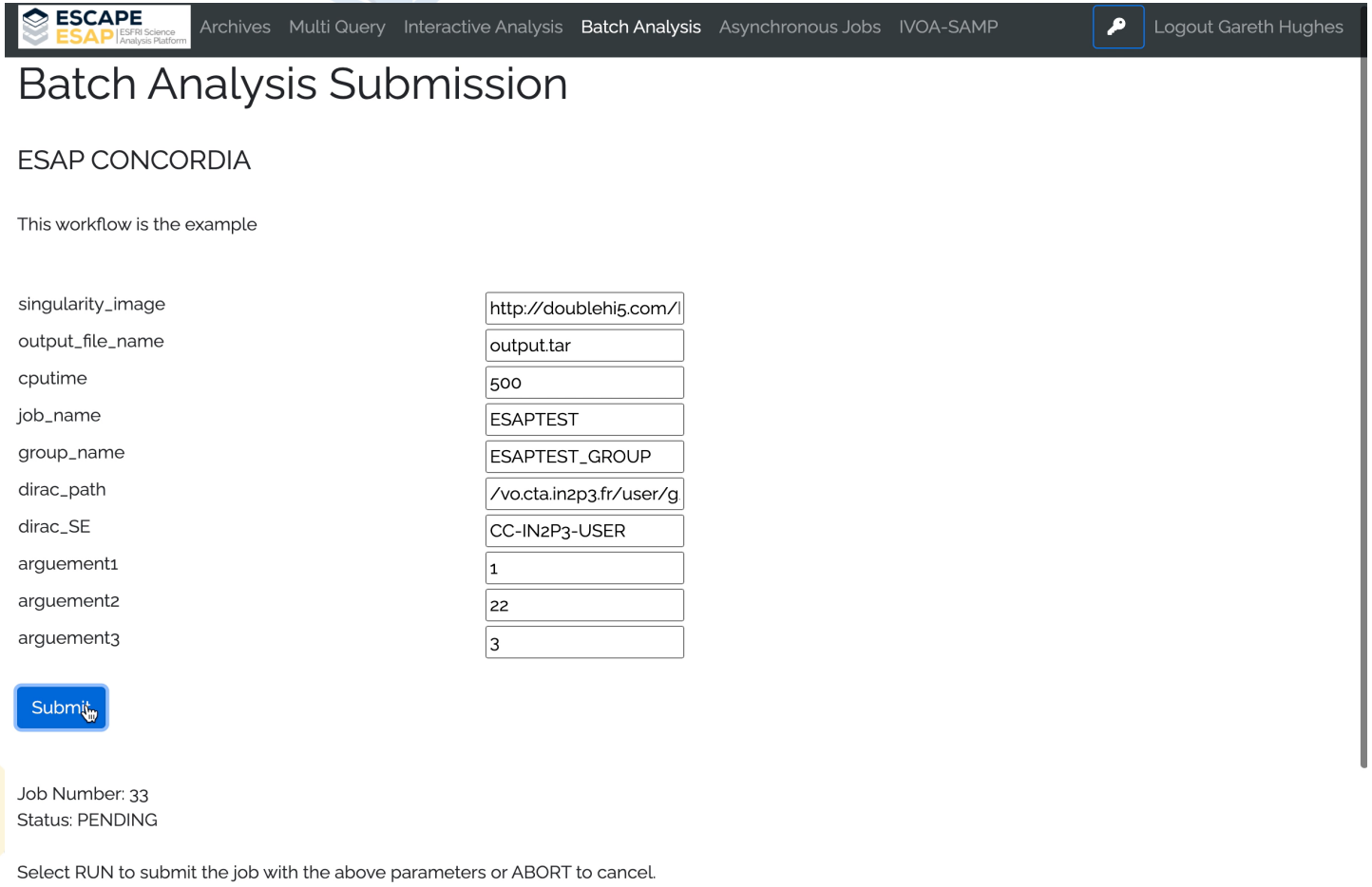
```
[[16:54:ghughes@esap]$ more esap_batch_config.yaml
- model: batch.facility
  pk: 1
  fields:
    name: CTADIRAC
    description: CTA-DIRAC instance
    url: https://pypi.org/project/CTADIRAC/
    facilitytype: DIRAC
    runtimeengine: Docker
- model: batch.facility
  pk: 2
  fields:
    name: Rosetta @ INAF OATS
    description: The Rosetta platform deployed at INAF OATS computing centre
    url:  https://esap-rosetta.oats.inaf.it/
    facilitytype: rosetta
    runtimeengine: Singularity
- model: batch.workflow
  pk: 1
  fields:
    name: CONCORDIA
    description: Containerized Corsika on DIRAC
    url: DIRACExampleWorkFlows/concordia.json
    ref: HEAD
    workflowtype: container
- model: batch.workflow
  pk: 2
  fields:
    name: Hello World
    description: This is the hello world example
    url: DIRACExampleWorkFlows/helloworld.json
    ref: HEAD
    workflowtype: container
[16:54:ghughes@esap]$
```



Funded by the European Union's Horizon 2020 - Grant N° 824064

Batch/DIRAC on ESAP

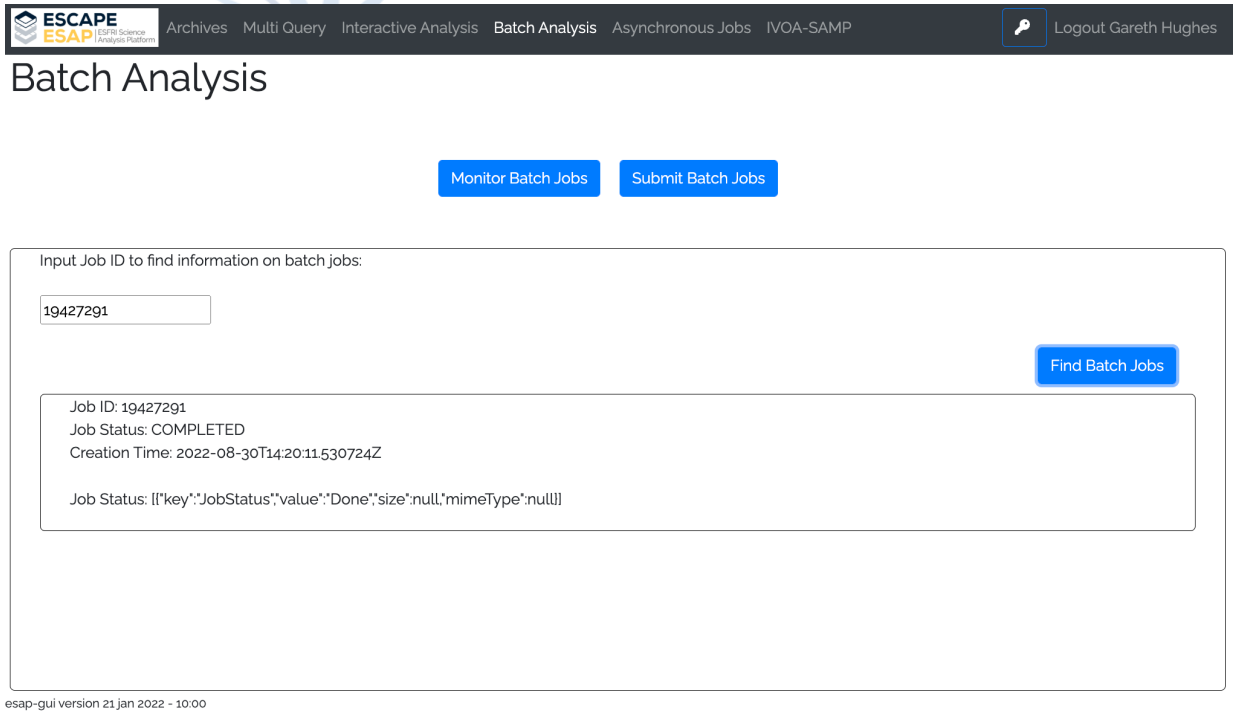
```
[16:56:ghughes@DIRACExampleWorkFlows]$ more concordia.json
{
   "uuid": "fa2b5ebc-ee25-10ec-7ea1-0247ac120006",
   "type": "URL",
   "title": "ESAP CONCORDIA",
   "description": "This workflow is the example",
   "workflow": "CONCORDIA",
   "Parameters": {
     "singularity_image": "http://doublehi5.com/helloworld.sif",
     "output_file_name ": "output.tar",
     "cputime": "500",
     "job_name": "ESAPTEST",
     "group_name": "ESAPTEST_GROUP",
     "dirac_path": "/vo.cta.in2p3.fr/user/g/ghughes/",
     "dirac_SE": "CC-IN2P3-USER",
     "arguement1": "1",
     "arguement2": "2",
     "arguement3": "3"
   }
}
```



Funded by the European Union's Horizon 2020 - Grant N° 824064

Batch/DIRAC on ESAP

Can check to confirm on the CLI

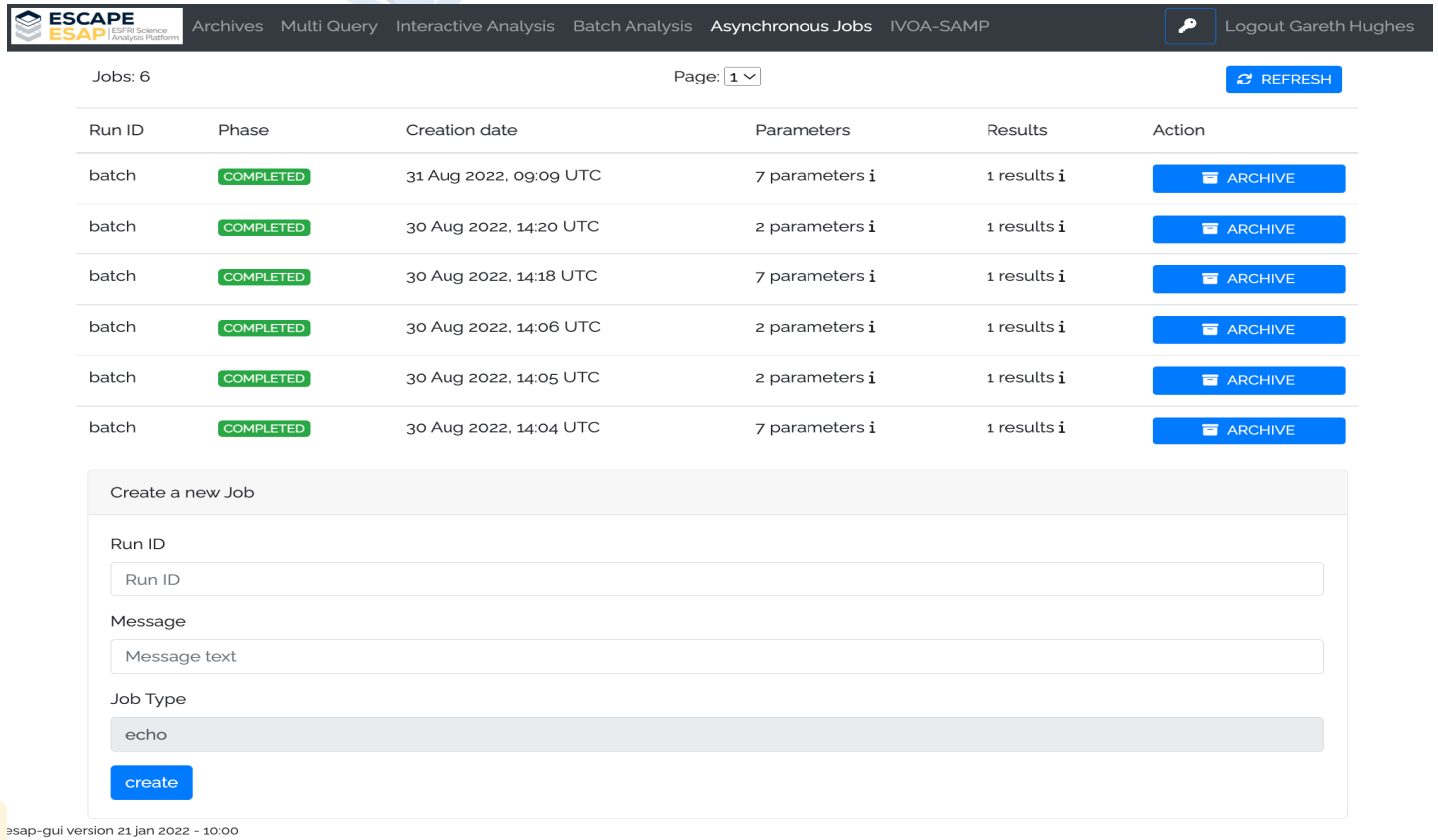



```
base) root@13cf83eb3f93:/ESAPWORKER/esap-worker# dirac-wms-job-status  19427291
obID=19427291 ApplicationStatus=echo "{"Parameter1": "999", "Parameter2": "2", "Parameter3": "33", "Parameter4": "GG", "Parameter5": "test", "Par
ter6": "default", "type": "batch.submit"}" successful; MinorStatus=Execution Complete; Status=Done; Site=LCG.IN2P3-CC.fr;
```


Funded by the European Union's Horizon 2020 - Grant N° 824064

Async UWS App

Able to monitor UWS jobs on the Async



Funded by the European Union's Horizon 2020 - Grant N° 824064

What is missing

- Authentication
  - x509
- Defining and finding the steering file
  - IVOA Execution Planner
- How to get your output
- More general framework for other WMS

Funded by the European Union's Horizon 2020 - Grant N° 824064

What next

- DIRAC-Rucio developments
  - WP2/DIOS CTA Rucio-DIRAC work
  - DIRAC-Rucio workshop!
- DIRAC developments
  - Tokens being brought into DIRAC
  - DIRAC REST API
  - DIRAC Jupyter-notebook extension
- CTA-ESAP deployment
  - CSCS & Observatoire de Paris working on deployments

Funded by the European Union's Horizon 2020 - Grant N° 824064

Things to note

- Things to note:
  - REANA
    - https://reanahub.io/
    - Reproducible research data analysis platform
  - DASK
    - Python library for parallel computing
    - Could be used within an interactive analysis
  - Provenance
    - Observatoire de Paris
    - UWS IVOA platform
      - able to perform local batch operations
      - captures provenance information

Funded by the European Union's Horizon 2020 - Grant N° 824064

Conclusion

- Framework for adding batch computing exists in ESAP
- Authentication is hard
  - Future developments will likely make this much easier
- Some work to finish up in the next few weeks
  - Documentation to write
  - Then think about long term possibilities

Funded by the European Union's Horizon 2020 - Grant N° 824064