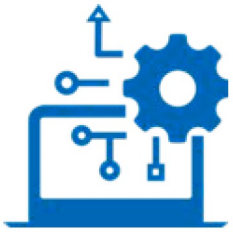
EUROPEAN OPEN SCIENCE CLOUD

DIRAC EGI service

A.Tsaregorodtsev
CPPM-IN2P3-CNRS, France,
DIRAC User's Workshop, 19 June 2024,
CC/IN2P3, Doua, France

- EGI Workload Manager – official service in the EGI portfolio
- Service hosting at CC/IN2P3
  - Openstack VMs
  - 2 development servers
    - DIRAC v9.0.0a29
  - 1 Alma Linux VM for the DIRAC certification tests
  - MariaDB server
  - Elasticsearch server
- 1 server at CPPM, Marseille
  - CS, SE
- DIRAC 8.0.24
  - DIPS services
  - Tornado TokenManager

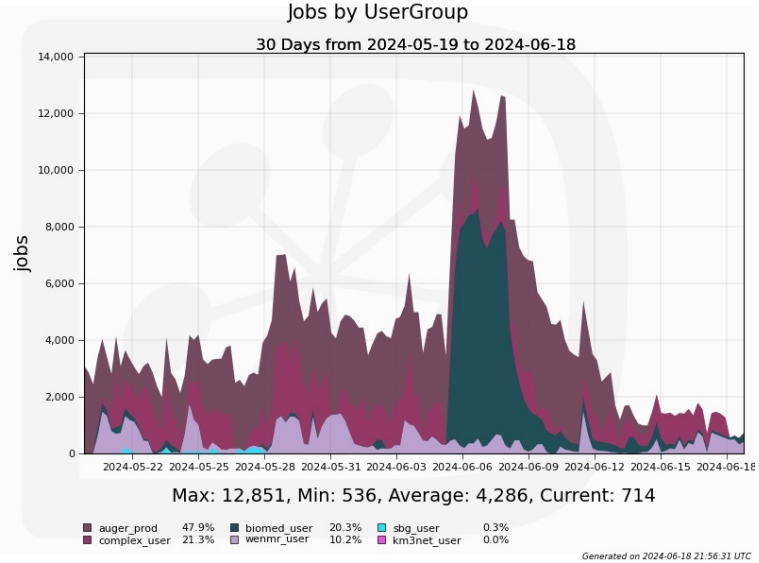


EGI DIRAC service

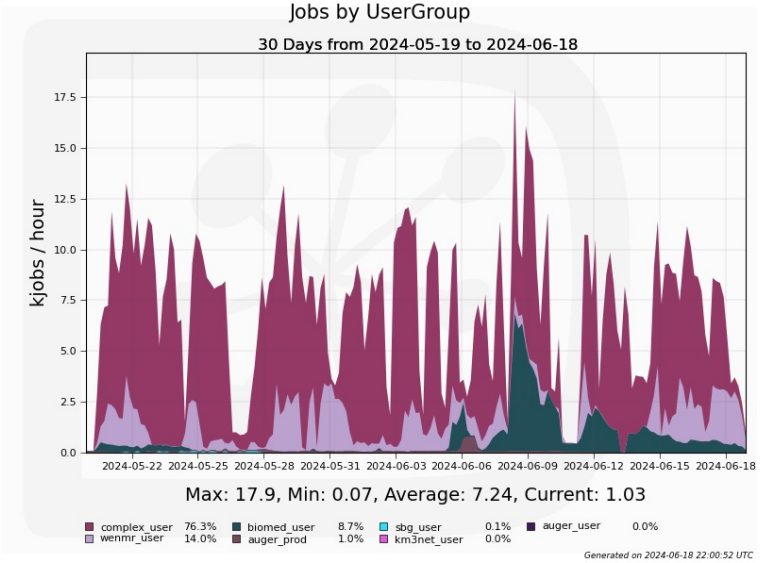
- The EGI DIRAC services are maintained by the members of the DIRAC@IN2P3 Project
  - CPPM/Marseille, LUPM/Montpellier, CC/IN2P3/Lyon
- DIRAC services provided
  - WMS services
    - Transformation service is enabled but not much used (demos, training)
  - DMS services
    - Multiple FileCatalogs (General, Biomed, Eiscat, HESS, Auger, …)
  - RMS
  - Accounting
  - Monitoring
- ElasticSearch problems to be still investigated
  - OpenSearch client incompatibilities ?

Operations: Job execution rates

Running jobs

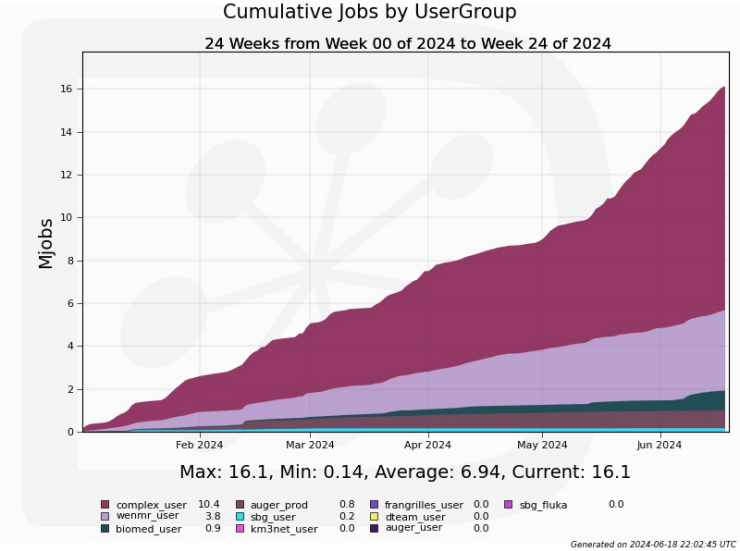


Job Execution rate

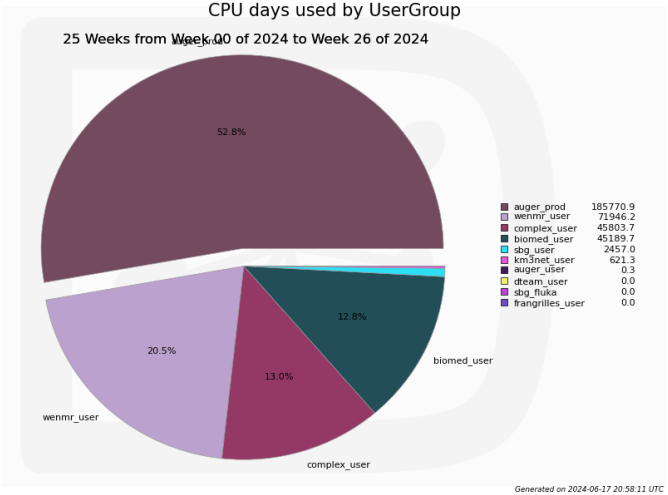


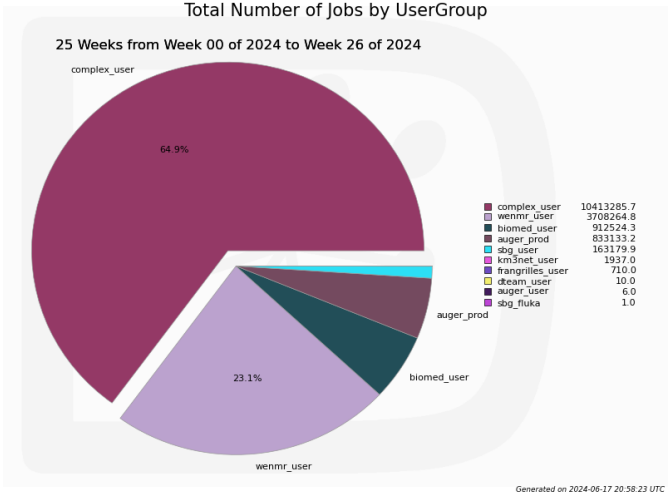
- Up to 5Hz of executed jobs
  - Up to 200K jobs per day
- Up to 13K running jobs

Operations: Workloads executed



- 16M jobs since 1 Jan 2024
  - 16m jobs in the whole 2023
- Very different job patterns
  - Lots of short jobs vs few long jobs

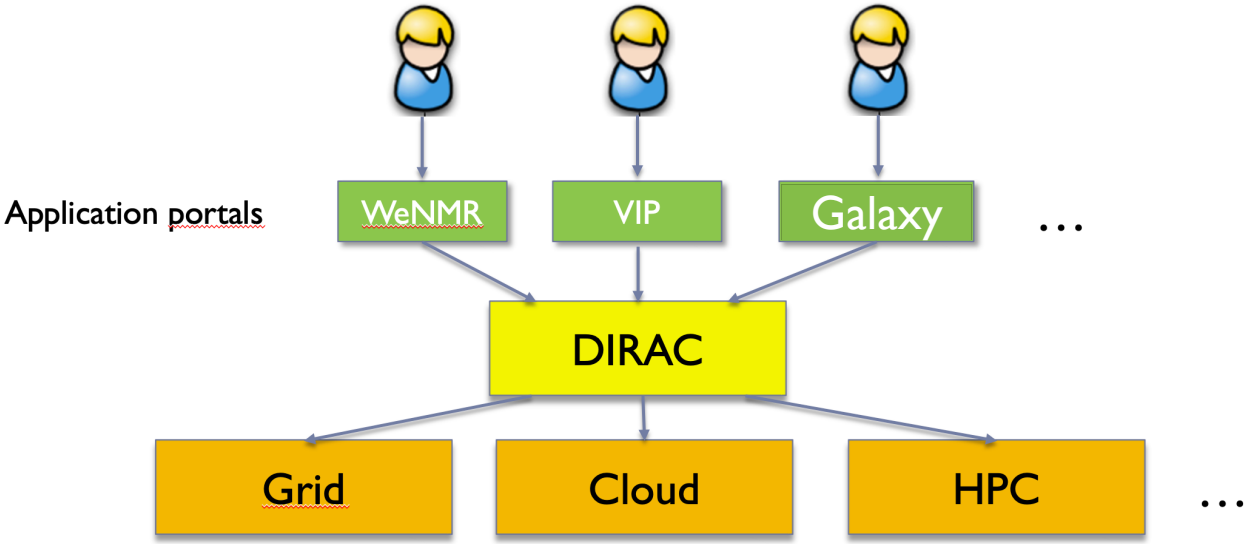




Communities

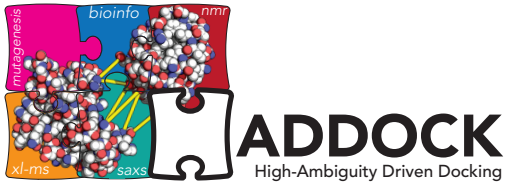
- ~20 communities with ~700 registered users
- Two distinct types of communities
  - Use of DIRAC by individual users
    - Auger, HESS, SBG, KM3NeT, biomed, OpenMOLE, …
    - Needs more effort for the user support, individual user registration, etc
  - Application portals between users and the DIRAC service
    - Users are managed by the portals
      - Registration, profiles, accounting
    - Portals are represented as a single user to DIRAC
      - Easier case for DIRAC, better synergy

Application Portals



- WeNMR (https://wenmr.science.uu.nl/), Haddock application, prothein docking
  - >44K registered users
- VIP, medical imaging (https://vip.creatis.insa-lyon.fr)
  - ~1500 users
- Galaxy, an open source, web-based platform for data intensive biomedical research (https://usegalaxy.org/)
  - Potentially many
  - Work in progress on the DIRAC job submission (runners)

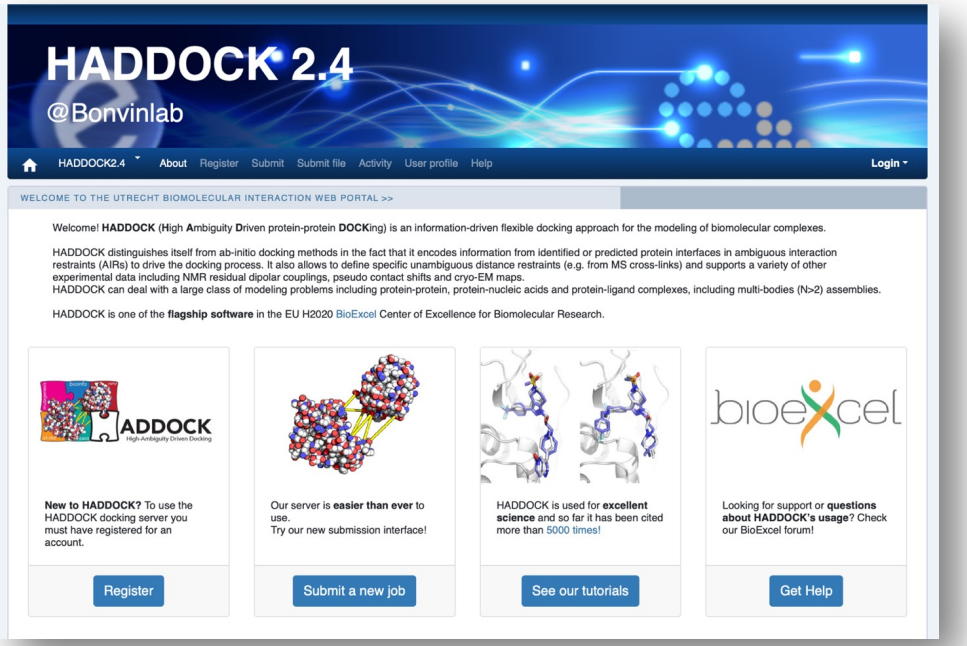
HADDOCK web portal (WeNMR)



- **> 44600 registered users**
- **> 650000 served runs since June 2008**
- **65% on the EOSC HTC resources (>85% for 2.4) using the EGI workload manager (DIRAC)**
- **Integrated in the EOSC marketplace**

De Vries *et al.* Nature Prot. 2010

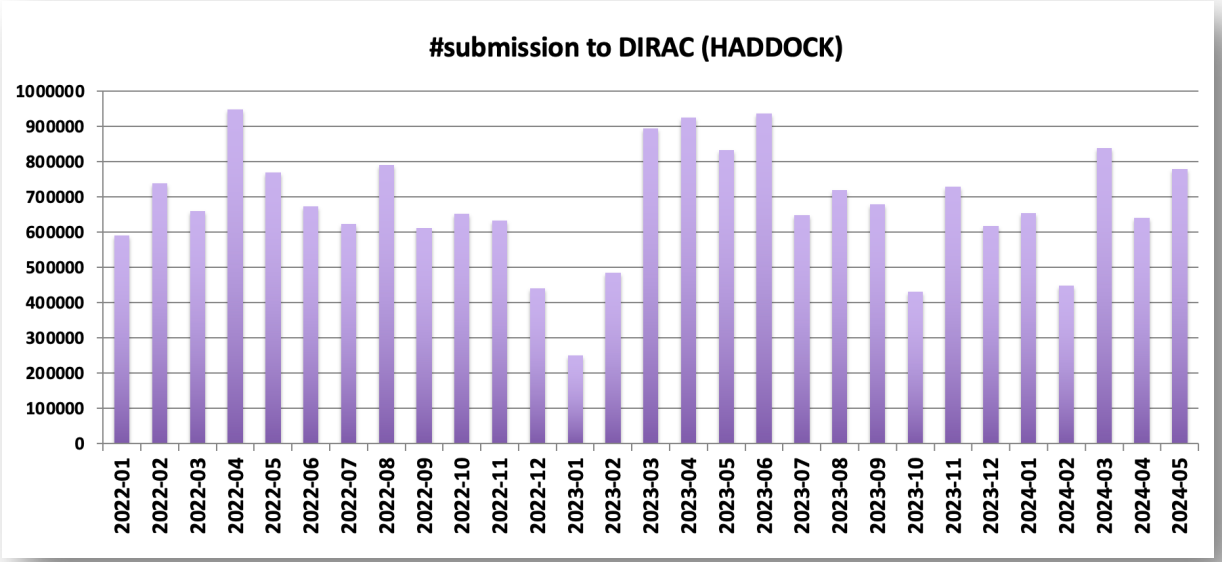
Van Zundert *et al.* J.Mol.Biol. 2016



https://wenmr.science.uu.nl

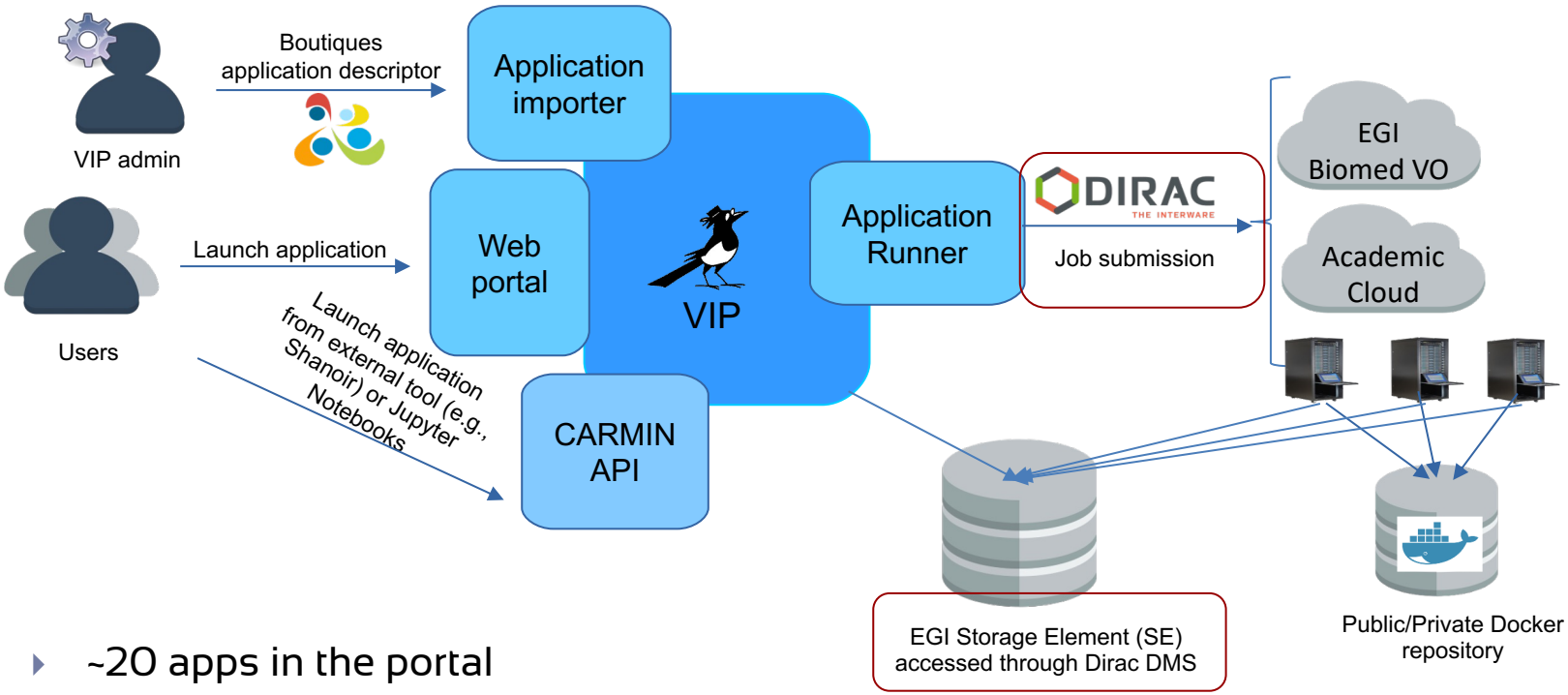
HADDOCK DIRAC submissions

Average of ~675000 job submissions to DIRAC per month by the HADDOCK server

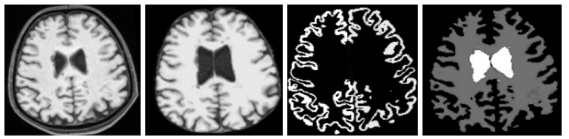


- Using HTC and Cloud sites
  - No tokens
- Smooth running
  - Few problems with VOMS configuration changes non properly propagated to sites
- App software from CVMFS
- Future version of Haddock
  - will execute complex workflows (MPI) on worker nodes – will require more **cloud** resources
  - will require different data transfer mechanisms (in/out)

Biomed: VIP Portal



- ~20 apps in the portal
- Using HTC and Cloud sites
  - Using EGI Check-In tokens and VOMS certificates
  - Dedicated FC and custom services
- Containerized applications (Docker)
- Using custom cloud VMs with SSHCE
  - Accessing GPU's
- Request for definition of CloudCE queues for VMs with GPU access configured



OpenMOLE

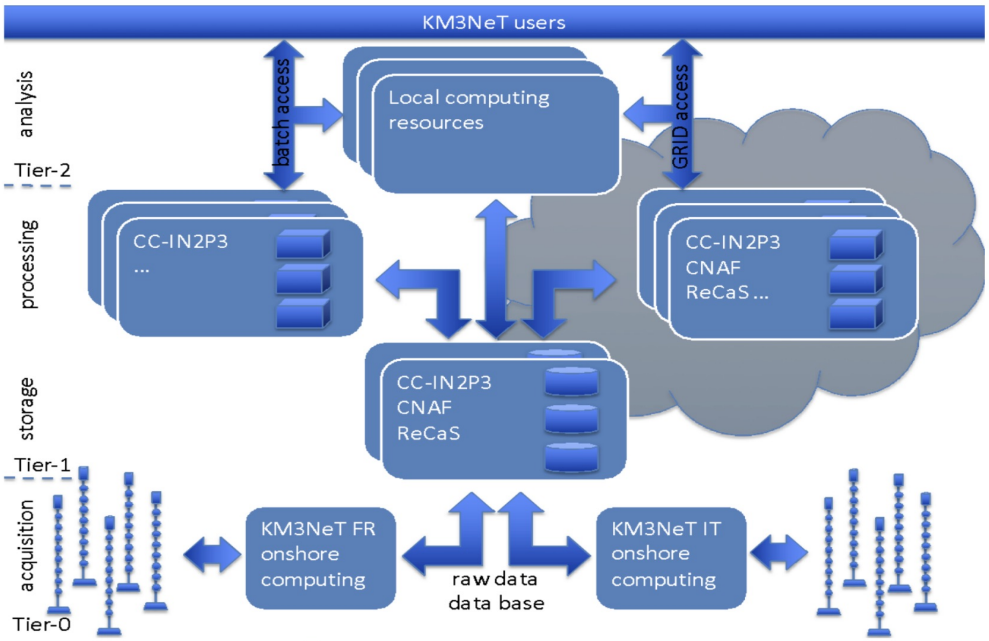
- Complex systems modeling applications (https://iscpif.fr/projects/openmole/)
- thousands evolutionary algorithms running in parallel (as jobs) orchestrated by a single global algorithm
  - Resulting in many short jobs in DIRAC
- Modeling software in different languages (java, scala, …)
  - Using REST (ancient !) DIRAC interface
- Requests:
  - REST interface for both jobs and data
    - Most impassionate potential DiracX users !
    - Both data bookkeeping and data access (webdav)
  - Moving to use tokens – to be quickly done for jobs. For data – to be seen
    - User management with Check-In is to be set up

Astrophysics communities: Auger

- Pierre Auger Observatory
  - Running heavy MC data productions with the EGI DIRAC services on HTC resources
  - Peculiarities:
    - Dedicated FC managed and hosted on the community server *dfc-auger.grid.cesnet.cz*
    - Using ***Perun*** user management service
  - *See Jiri's presentation*

Astrophysics communities: KM3NeT

- The KM3NeT is a large Research Infrastructure (RI) that will consist of a network of deep-sea neutrino detectors in the Mediterranean Sea
- Tier structure of the Computing Model



Astrophysics communities: KM3NeT

- Starting with the EGI DIRAC service
  - MC production
  - Mass Data Processing
- Currently all workflows are implemented in the Snakemake workflow engine https://snakemake.github.io/
  - Job submission to DIRAC
- User and community management in an Indigo IAM service
  - Should benefit from multi-IdP support in DIRAC
- Setting up Rucio service for data management
  - DIRAC-Rucio bundle is becoming more and more popular
- Looking at organizing mass productions with the DIRAC TS
- Problems encountered:
  - The DiracOS2 not well implemented on arm architecture when running on local osx client.
  - Need Admin access to see Accounting for jobs of all the VO users – need VOAdmin level privileges.
- Wish list:
  - KM3NeT Dirac/Rucio container standalone
  - Integration between Dirac Transformation System and Rucio
  - Better documentation ☺

CVMFS dirac.egi.eu repository

- The CVMFS dirac.egi.eu repository is maintained
  - Admins: Andrei, Daniela
- The repository is updated automatically nightly:
  - DIRACOS2 installers
  - Pilot files
- The DIRAC clients for 8.* and 9.* releases are installed automatically (github action)
  - Backed up by a cronjob
  - Intel and ARM versions
- Unstable behavior
  - Problems with Ceph mounted FS:
    - Failures in releases installations
    - Slow publishing to CVMFS clients
  - Managers are trying to cope but with a limited success so far

Conclusions

- Stable EGI DIRAC service operation with increasing numbers of payloads
- New communities are showing interest in the service and need guidance
- Points of user's interest
  - Token based user management (Check-In, IAM)
  - More versatile use of Cloud resources
  - Looking for DIRAC-Rucio combined services