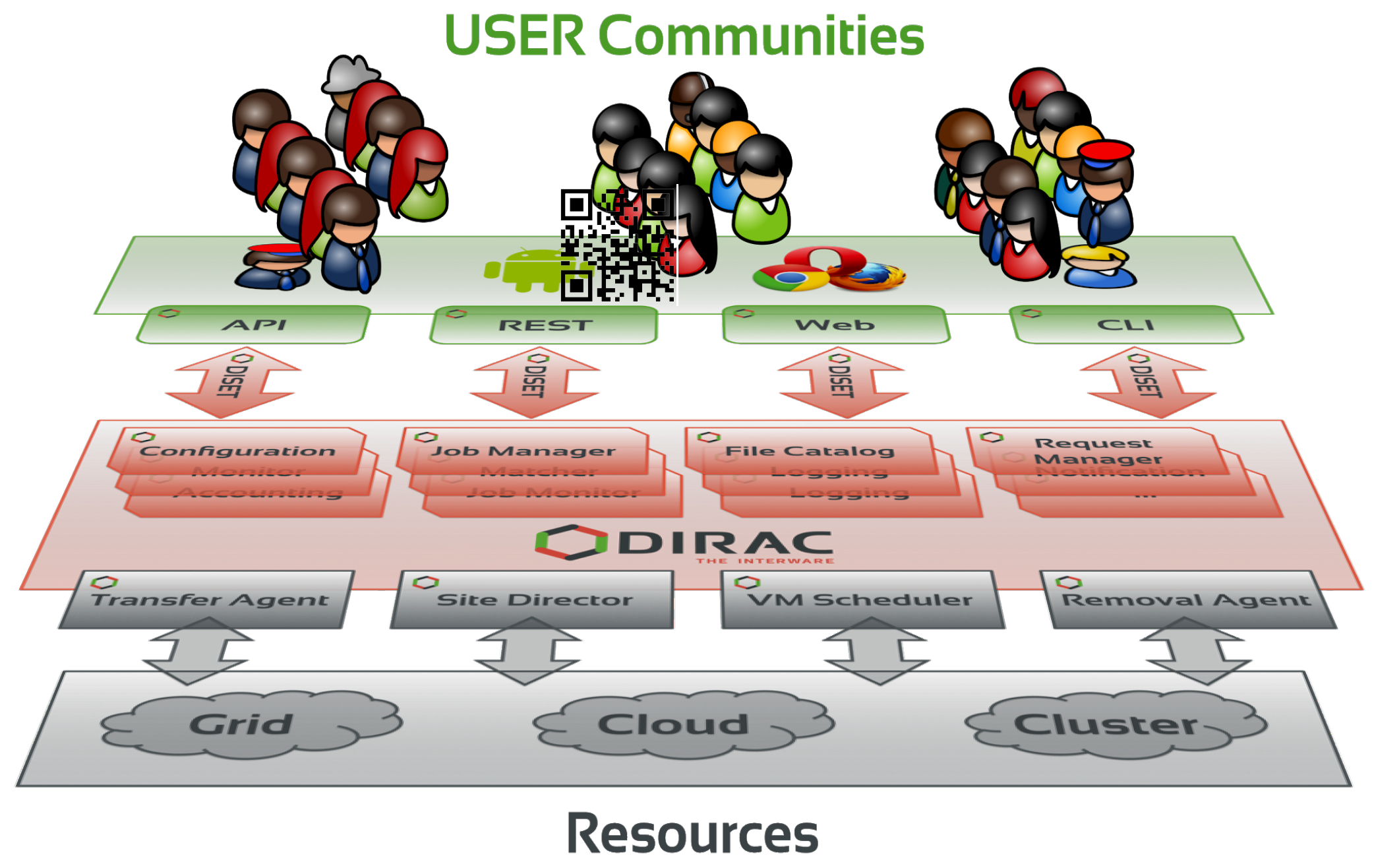
DIRAC
THE INTERWARE

Putting together heterogeneous resources for scientific applications

L. Arrabito[1], J. Bregeon[1], S. Camarasu-Pop[2], P. Gay[3], V. Hamar[4], F. Hernandez[4] and A. Tsaregorodtsev[5] for the DIRAC@IN2P3 project

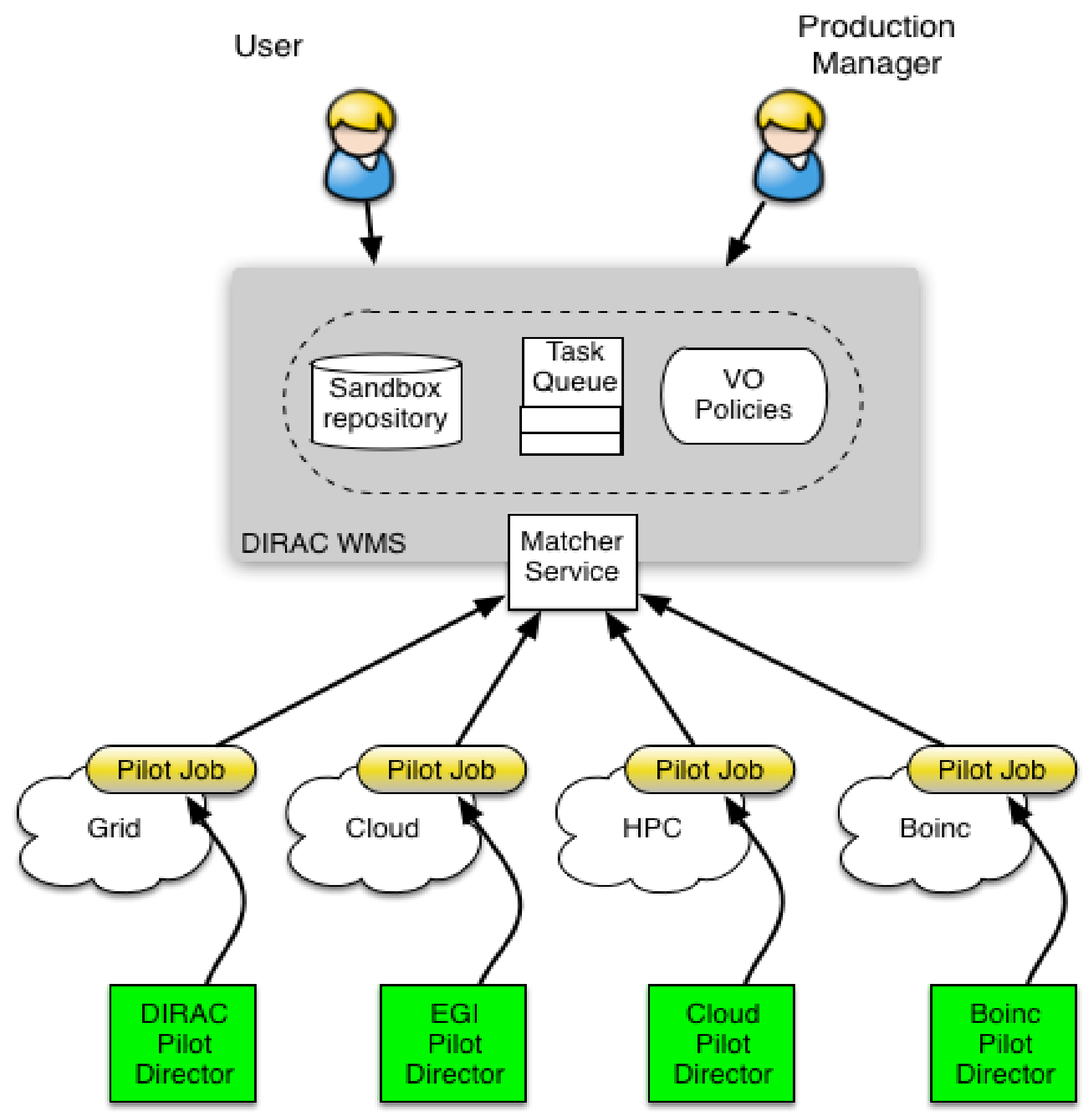


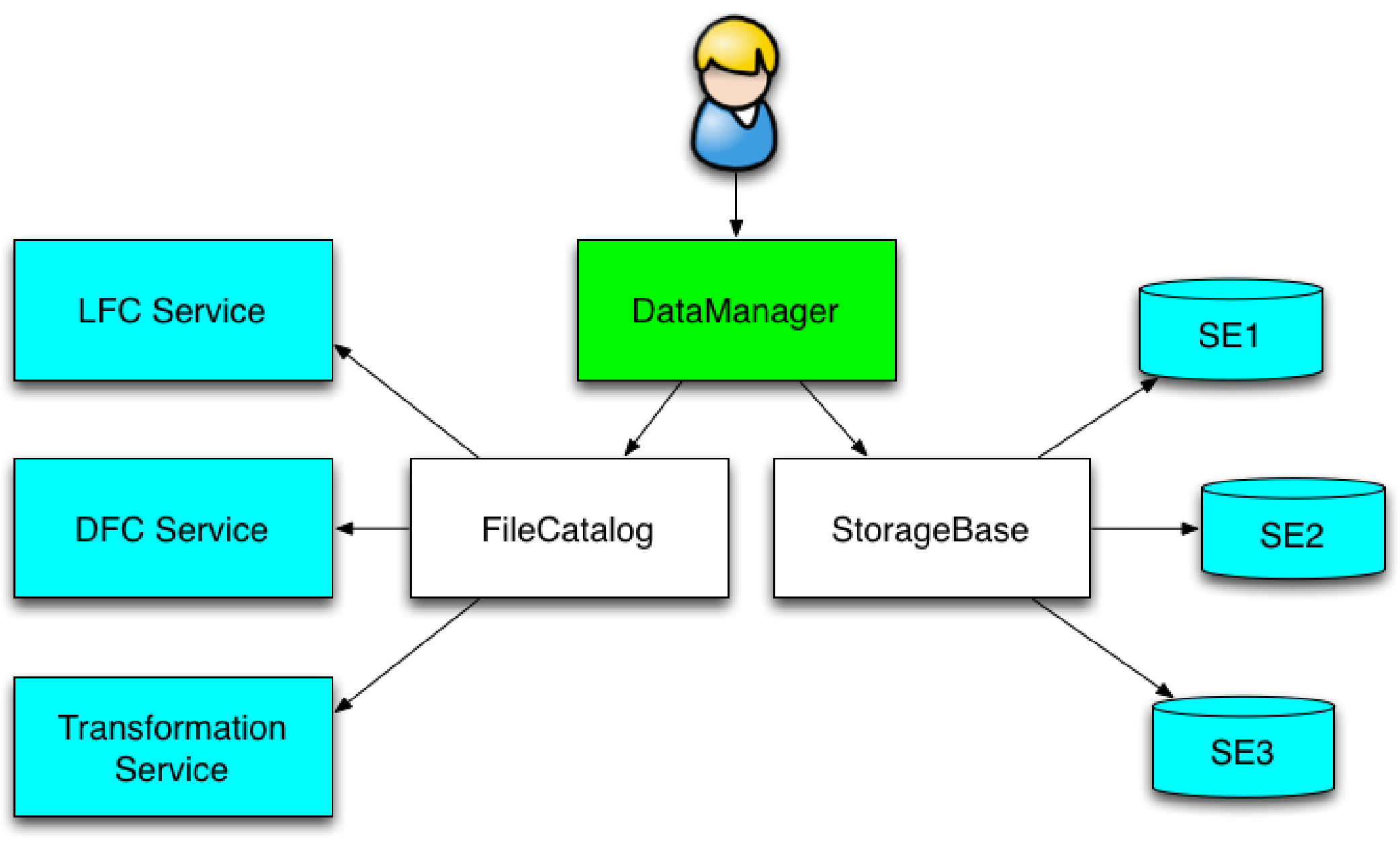
DIRAC, the interware

- A layer between users and computing resources
- A software framework for building distributed computing infrastructures
- A rich set of **Systems** providing a complete solution for user communities
- Seamless integration of heterogeneous computing and storage resources
- Friendly user interfaces hiding the complexity of a distributed infrastructure

Workload Management System

- Base on the pilot jobs paradigm for efficient user job execution and low failure rate
- Efficient application of resource usage policies for large communities
- Easy integration of heterogenous computing resources:
  - mecanisms of **plugins** for integration and job **tags** for allocation
  - grids operated with various middleware (CREAM, ARC, Globus)
  - clouds, private and commercial, from different providers (VMDIRAC)
  - HPC supercomputers and GPU nodes (through ssh, handles multicore jobs)
  - *ad-hoc* computing clusters (ssh, HTCondor...)
  - volunteer resources (BOINC)
  - containers on worker nodes (Singularity)



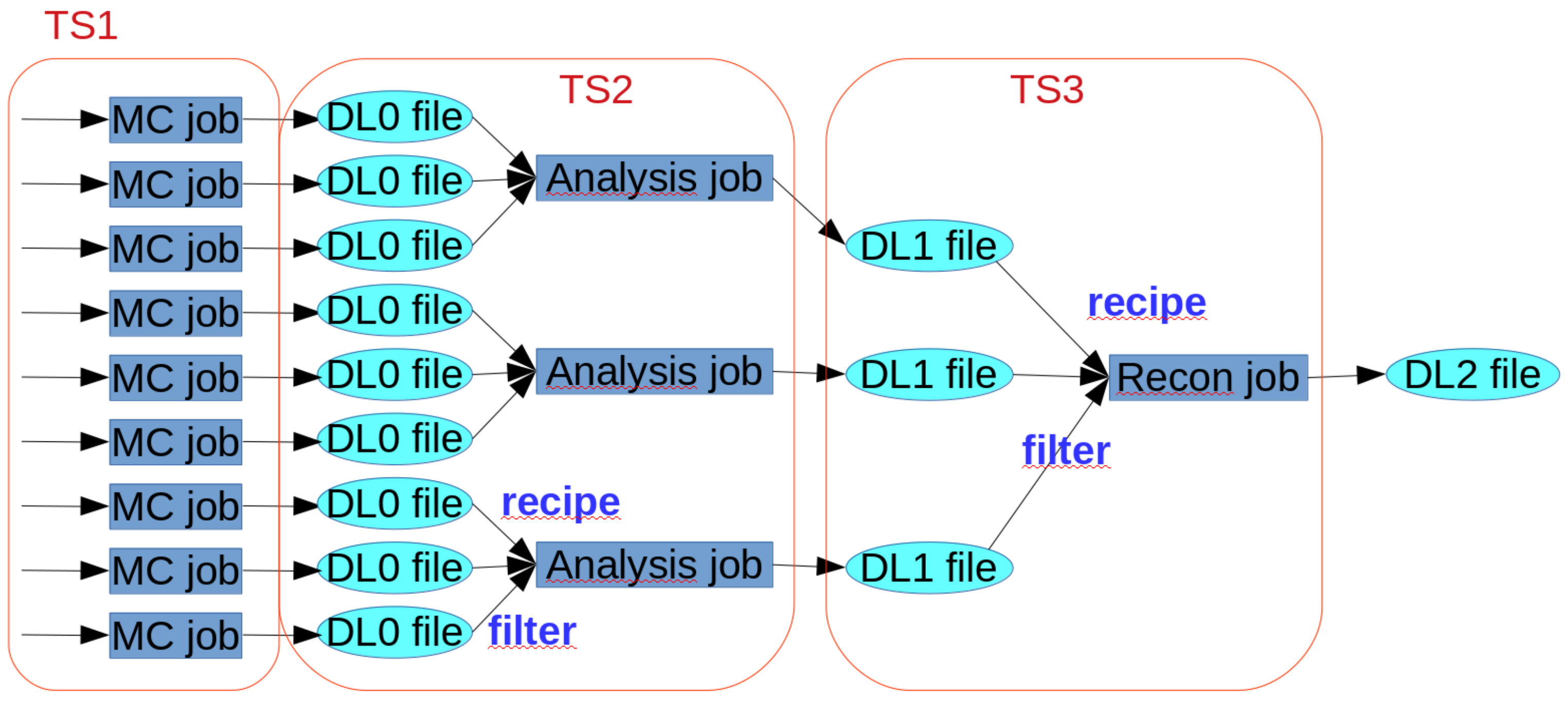


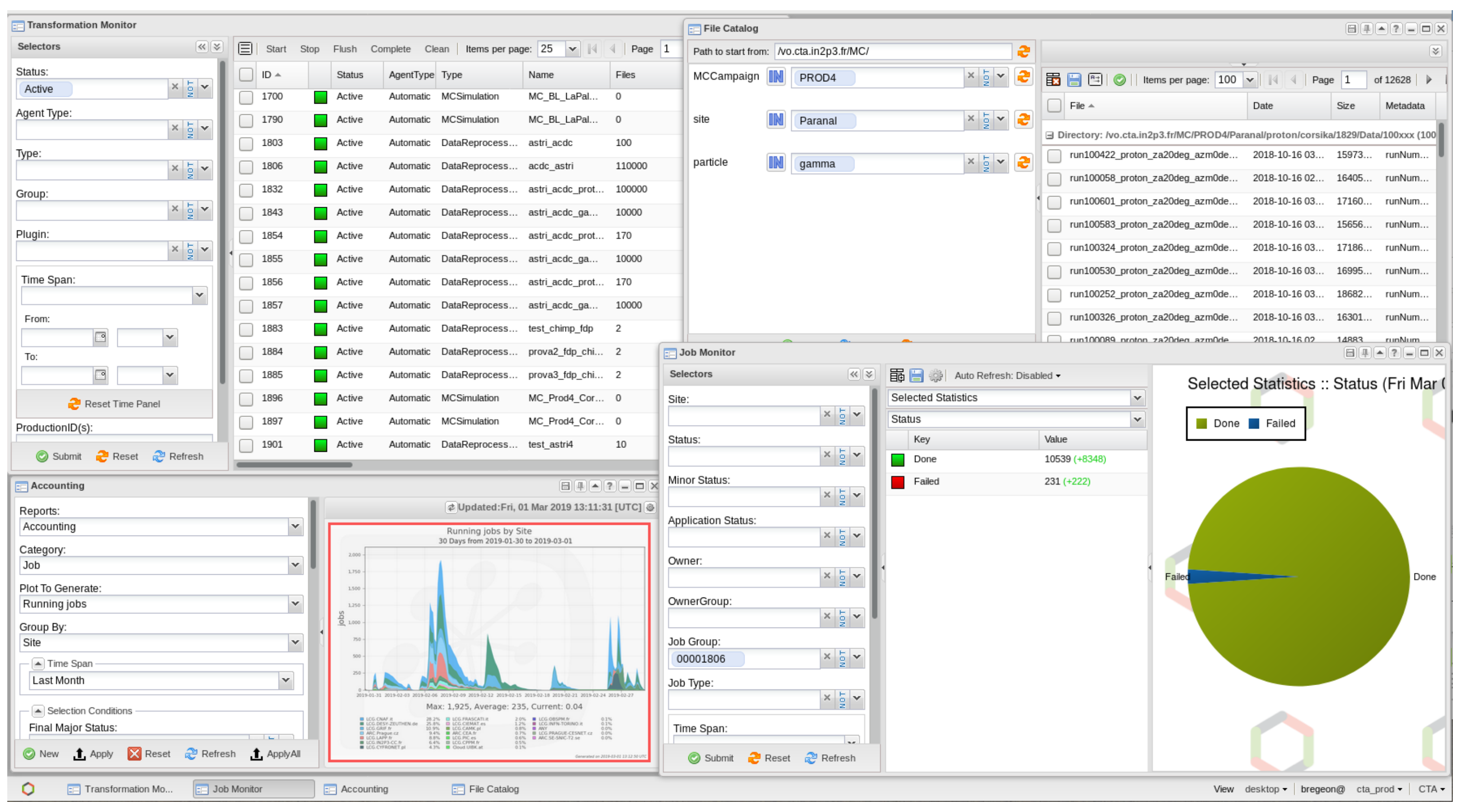
Data Management System

- Based on the abstraction of Storage Elements and File Catalogs
- Allows to present physically distributed user data as a single logical file system
- Implementation of abstract models available for most of the modern data storage technologies and several file catalog services
- The **DIRAC File Catalog** is both a replica and user metadata catalog that allows complex data models specific to user communities

Transformation (TS) and Production Systems

- Create complex workflows out of elementary data and workload management operations
- A **Transformation** is a **recipe** (task) applied automatically to data identified by a **filter** on their metadata
- Chaining transformations creates (meta) data driven workflows
- A **Production** is defined as a coherent set of **Transformations,** monitored and managed through a common interface



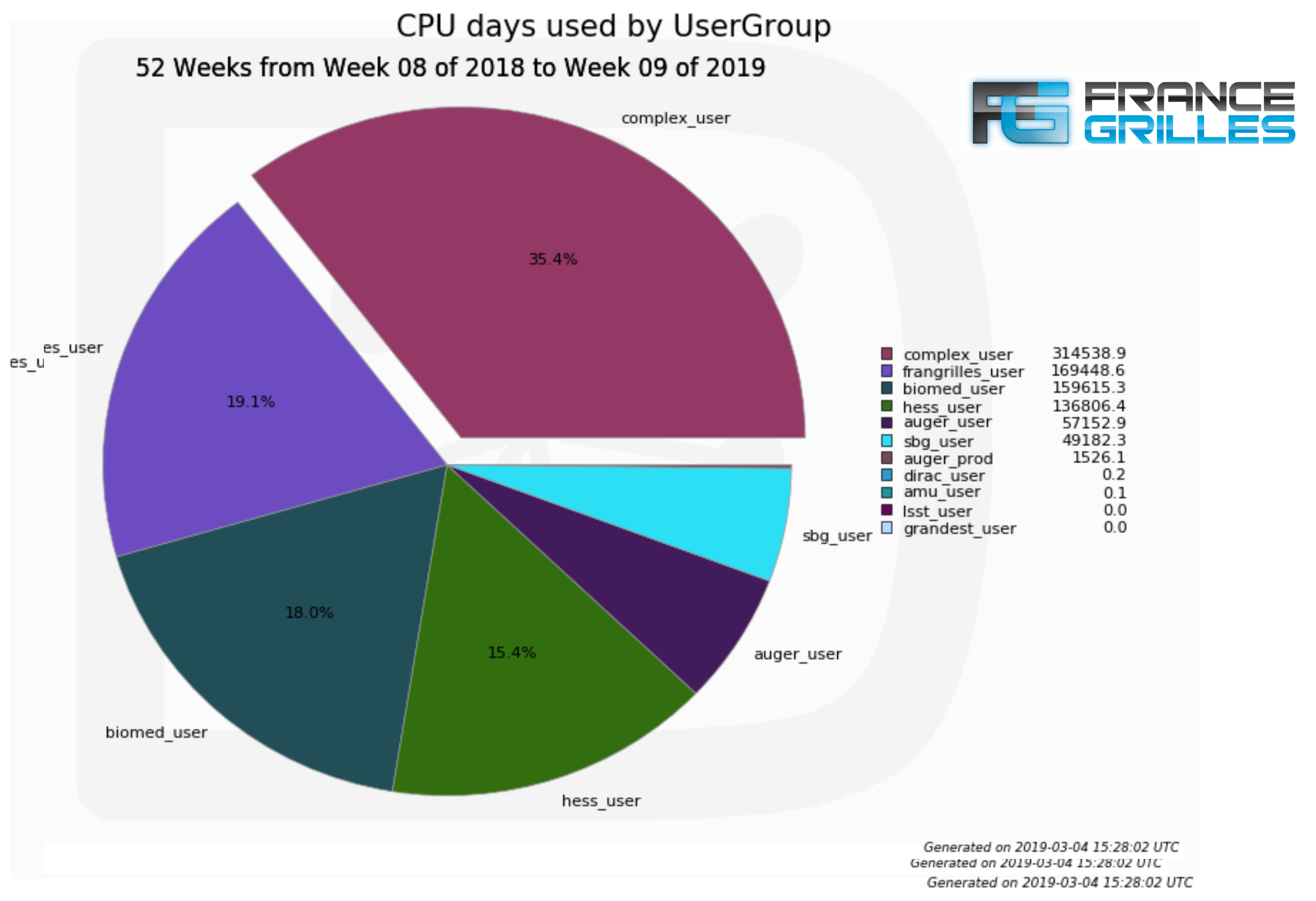


Standard user level and advanced programmatic interfaces

- Extensive command line tools: e.g. *dirac-wms-job-submit, …*
- Python API: e.g. *dirac.submit(job), …*
- REST interface for an essential subset of services
- Web portal providing views and interaction with the main systems
- User experience based on a paradigm of a single large scale computer:
  - Logical Computing and Storage elements (Hardware)
  - Global logical name space for data (File System)

Community based open source development on github/DIRACGrid

- Plugin mecanism, software modularity, new features upon user requests
- Software managed by the **DIRAC consortium** of users and developers
- Adopted by multiple collaborations in High Energy Physics, Astrophysics and other scientific domains for an ever increasing number of use cases



Multi-community DIRAC services available

- Supported by grid infrastructure projects: **France-Grilles**, **GridPP**, **EGI**, **JINR** ...
- Ideal for *small* user communities to benefit from the advantages of the DIRAC interware
- Example of the **France-Grilles DIRAC service**, hosted at CC-IN2P3:
  - First multi-community installation put in production in 2012
  - resources through ARC, CREAM, HTCondor and ssh access to farms with GPUs
  - distributed team of administrators from several universities and CNRS labs
  - multiple active communities: biomed, complex-systems, vo.france-grilles.fr, …
  - > 25M jobs executed in 2018 at 90 different sites

1) CNRS/IN2P3, LUPM, Univ. Montpellier, Montpellier, France
3) Univ. Bordeaux, Bordeaux, France
2) Univ Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, UJM-Saint Etienne, CNRS, Inserm, CREATIS UMR, France
4) CNRS, CC - IN2P3, Villeurbanne, France 5) Aix Marseille Univ., CNRS/IN2P3, CPPM, Marseille, France