

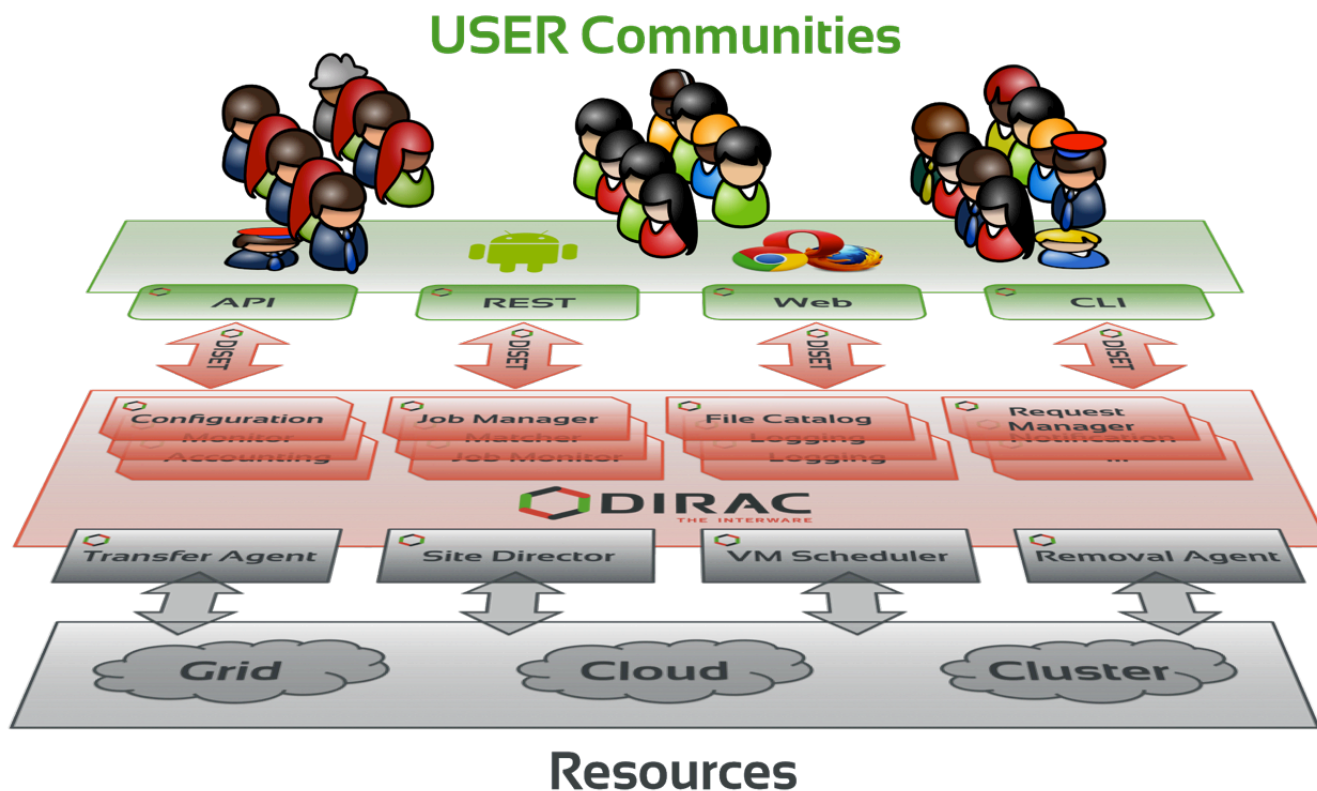
# DIRAC Services for Scientific Communities

*A. Tsaregorodtsev,  
CPPM-IN2P3-CNRS, Marseille,  
GRID'2018, JINR, Dubna, 10 September 2018*



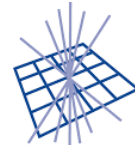
- ▶ Quick DIRAC reminder
- ▶ Multi-VO DIRAC services
- ▶ Developments necessary to support multiple VO's in a single installation
- ▶ Conclusions

- A software framework for distributed computing
- A **complete** solution to one (or more) user community
- Builds a layer between users and resources



- Started as an LHCb project, became experiment-agnostic in 2009
  - First users (after LHCb) end of 2009
- Developed by communities, for communities
  - Open source (GPL3+), [GitHub](#) hosted, python 2.7
  - No dedicated funding for the development of the “Vanilla” project
  - Publicly [documented](#), active [assistance forum](#), yearly [users workshops](#), open [developers meetings](#)
  - [diracgrid.org](#)
  - 4 FTE as core developers, a dozen contributing developers
- The DIRAC consortium as representing body
  - CNRS, CERN, University of Barcelona
  - IHEP, KEK, PNNL, University of Montpellier





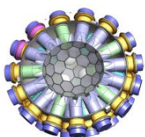
**GridPP**  
UK Computing for Particle Physics



A framework shared by multiple experiments/  
projects, both inside HEP, astronomy, and life  
science



Experiment agnostic  
Extensible  
Flexible



- ▶ 22 national infrastructures
- ▶ 2 international scientific organizations: CERN и EMBL

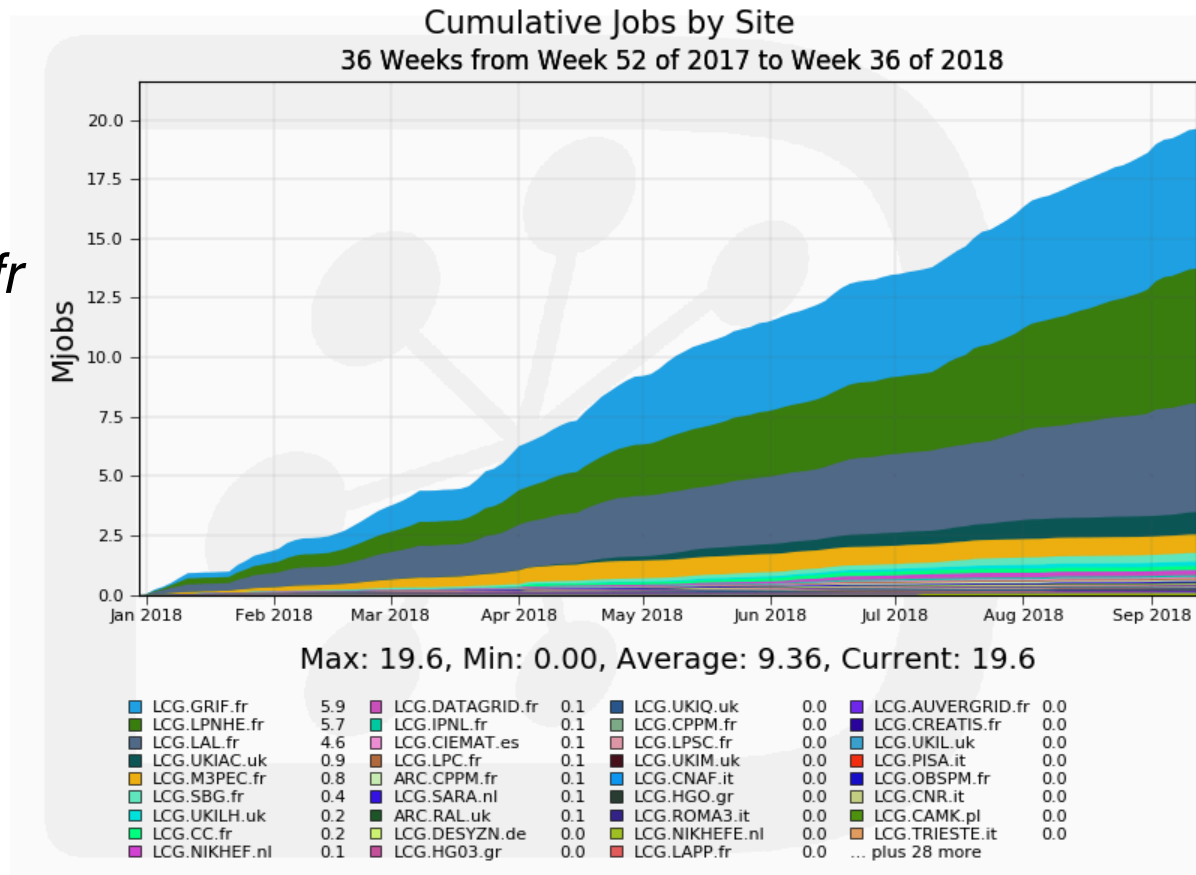




- ▶ Maintaining dedicated DIRAC services for small communities is not affordable
  - ▶ Requires expertise and effort to setup and run the system
- ▶ There was a clear need for services like DIRAC for an increasing number of communities with a low expertise in ( distributed ) computing and with high demands for computing resources
  - ▶ DIRAC framework was updated to support this kind of installations
- ▶ **DIRAC** services were provided by several National Grid Initiatives: France, Spain, Italy, UK, China, Romania, ...
  - ▶ Some of them did not survive (Spain, Romania)
  - ▶ Some are still in active production



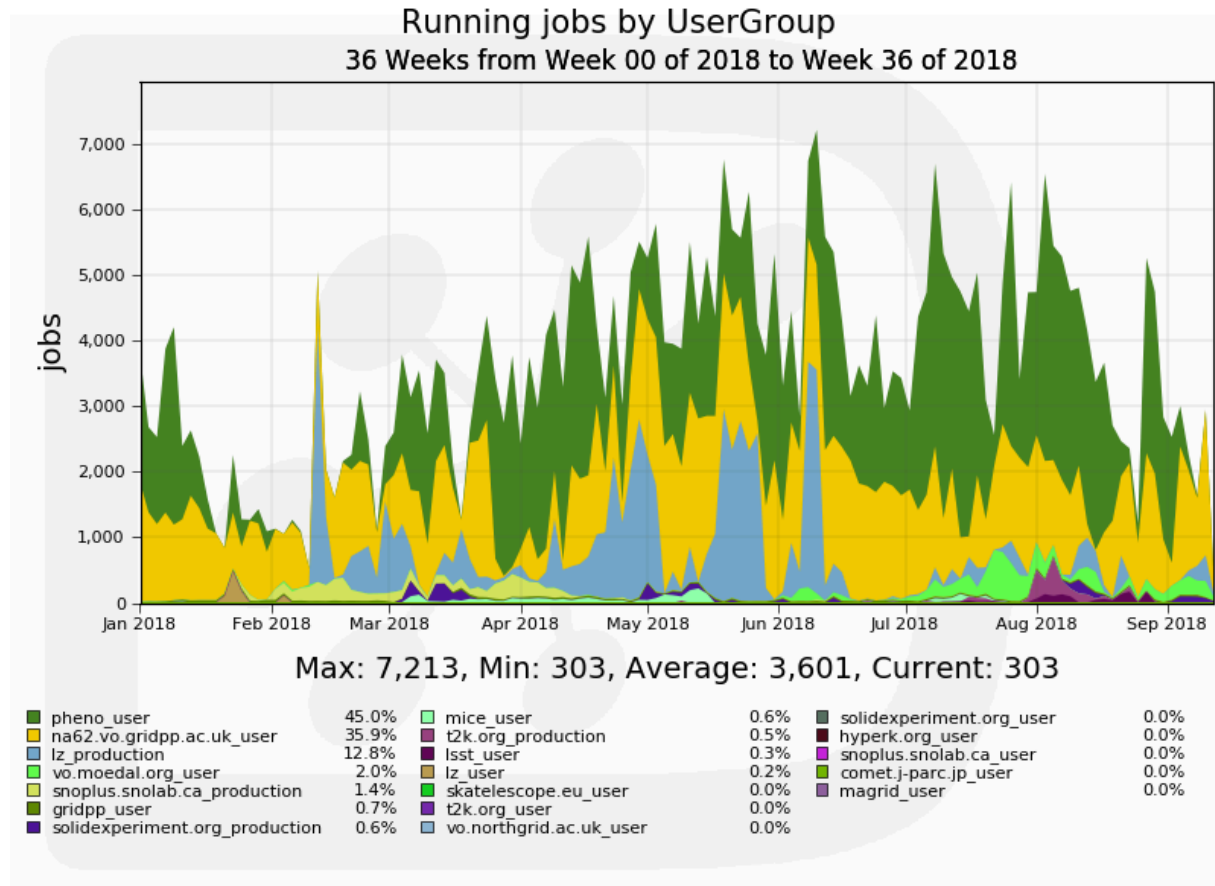
- ▶ Hosted by the CC/IN2P3, Lyon
  - ▶ *dirac.france-grilles.fr*
- ▶ Distributed administrator team
  - ▶ 5 participating universities
- ▶ In production since May 2012



Generated on 2018-09-12 19:20:08 UTC

- ▶ About 5 active communities – complex-systems, biomed, vo.france-grilles.fr, ...
- ▶ 9 ▶ > 20M jobs executed this year at 90 sites

- ▶ DIRAC service provided for the GridPP NGI
  - ▶ *dirac.gridpp.ac.uk*
  
- ▶ Hosted and operated by the Imperial College team



Generated on 2018-09-12 16:51:33 UTC

- ▶ About 10 active communities – LZ, NA62, Pheno, Snoplus...
- ▶ 10 ▶ > 5M jobs executed this year

## Partners

- ▶ Operated by EGI
- ▶ Hosted by CYFRONET, Krakow
- ▶ DIRAC Project providing software, consultancy
- ▶ Supported via the EOSC-Hub H'2020 grant
- ▶ [dirac.egi.eu](http://dirac.egi.eu)

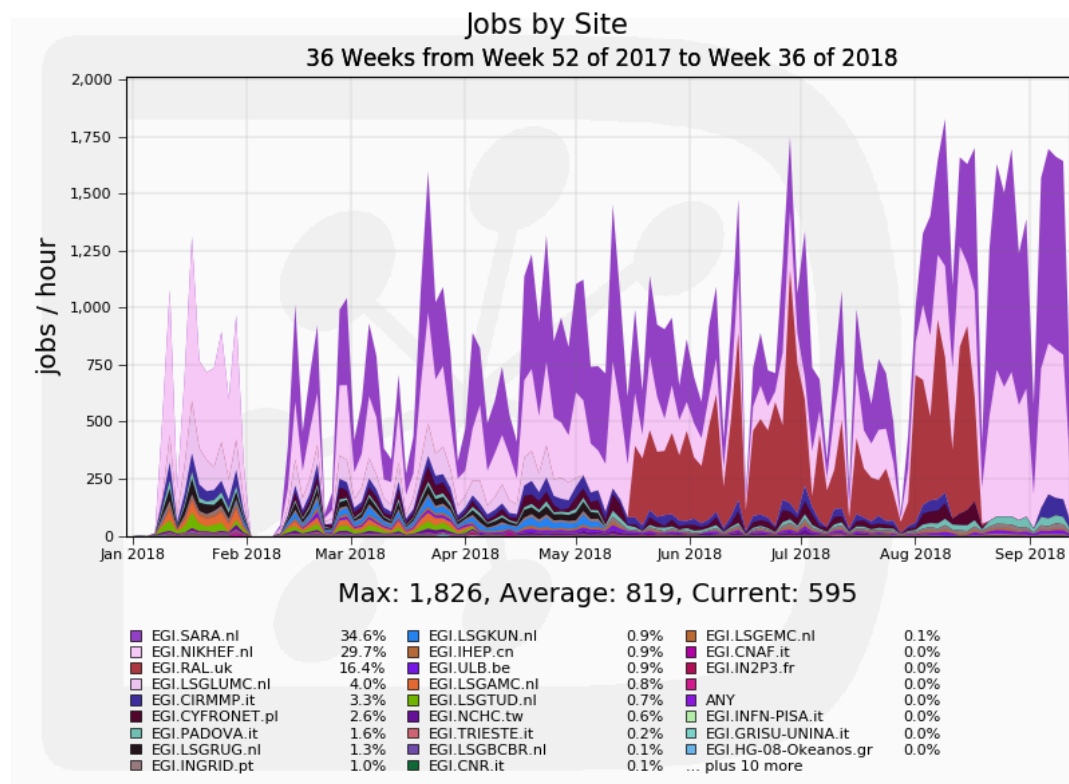
## 20 Virtual Organizations

- ▶ [enmr.eu](http://enmr.eu)
- ▶ [virgo](http://virgo)
- ▶ [eli-beams](http://eli-beams)
- ▶ [eiscat.se](http://eiscat.se)
- ▶ [fedcloud.egi.eu](http://fedcloud.egi.eu)
- ▶ ...

## Usage

- ▶ > 5 million jobs processed this year

## DIRAC4EGI activity snapshot



Generated on 2018-09-12 19:23:21 UTC

- ▶ The main goal is to provide a Workload Management service to access EGI computing resources
  - ▶ Replacement of the gLite/EMI WMS
  - ▶ Providing access also to cloud resources (VMDIRAC extension)
    - ▶ EGI FedCloud sites
    - ▶ France-Grilles Federated Cloud sites
    - ▶ GridPP relies on Vac/Vcycle project to access national cloud resources
- ▶ Getting requests for higher level services
  - ▶ Bulk job operations
  - ▶ Workflow management (DIRAC Transformations)
- ▶ Possibility to attach private computing resources
  - ▶ Without the need to install grid middleware, using SSH tunnels

- ▶ Data Management functionality is not in the formal mandate of the DIRAC4EGI service
  - ▶ However, there is a clear need for that
    - ▶ LFC decommissioning, access to data transfer services (FTS), etc
    - ▶ Managing user metadata
  
- ▶ DM components are provided by DIRAC
  - ▶ Configuration of the grid Storage Éléments
    - ▶ Possibility to add private storages by installing DIRAC SE service
  - ▶ General purpose File Catalog
  - ▶ Dedicated community File Catalogs
    - ▶ Help in importing replica data (from LFC, from physical storage dumps)
  - ▶ Specific developments are done for several Competence Centers
    - ▶ E.g. custom File Catalog service with specific data access rules for the Eiscat 3D community

- ▶ Providing training for the EGI users
  - ▶ General courses
  - ▶ Webinars
  - ▶ Training for specific communities (Virgo, Auger)
- ▶ Helping in porting the applications
- ▶ Developing specific Web Portal features
  - ▶ E.g. custom File Browser and Job Launcher for the Eiscat 3D Collaboration

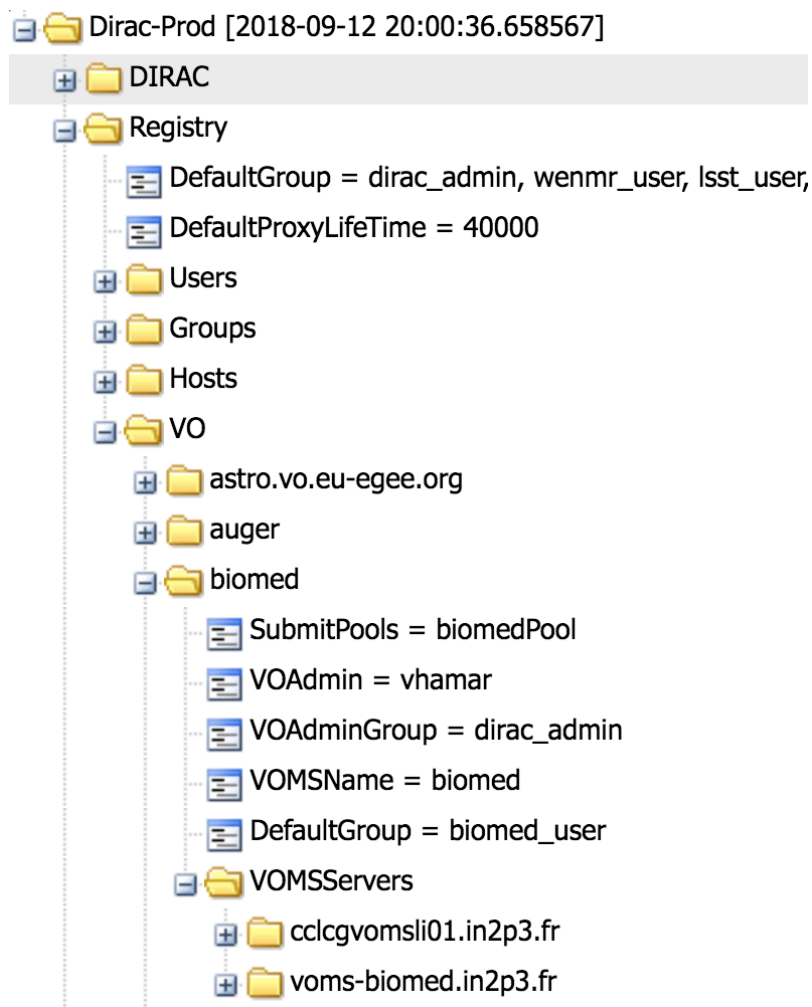
- ▶ Few other multi-VO service installations
  - ▶ CERN: ILC, CALICE
  - ▶ IHEP: BES III, Juno, CEPC
  - ▶ DIRAC@JINR (see *presentation by I.Pelevanyuk*)
- ▶ DIRAC service at CNAF
  - ▶ After initial installation and tests the service was attached to the DIRAC4EGI service
  - ▶ Providing several extra services
    - ▶ VIRGO File Catalog, DM services

- ▶ In order to minimize the overall effort there is a plan to merge DIRAC4EGI and FG services
  - ▶ Creating a single service administrator team
  - ▶ Keeping distinct web portals
  - ▶ Keeping dedicated subset of services (catalogs)
  - ▶ Sharing workload management system
  - ▶ Sharing configuration information (resources description)



- ▶ Providing a multi-VO service as opposed to single VO installations required several important developments of the basic DIRAC components
  - ▶ Enhanced per VO configuration
  - ▶ Easily available preconfigured clients
  - ▶ Language neutral interface
- ▶ Few more related developments are ongoing
  - ▶ Multi-VO pilot factories
  - ▶ SSO
  - ▶ VO specific resource management

- ▶ VO section in the Registry for specific parameters
- ▶ VO tags for all the configured resources (computing and storage éléments)
- ▶ VO specific operational parameters
  - ▶ A list of specific VO services



- ▶ Automatic synchronization of VO specific information with external services
  - ▶ User information – VOMS2CS agent
  - ▶ Resources information – BDII2CS agent
- ▶ Work in progress: introduction of the role of a VO administrator
  - ▶ Having limited write access to the CS in the parts related to a given VO

- ▶ DIRAC client installation is not very difficult
  - ▶ Usual part of tutorials
  - ▶ However, the client configuration can be complicated for some users
- ▶ Preinstalled clients on interactive nodes
  - ▶ CC/Lyon [cca.in2p3.fr](http://cca.in2p3.fr)
- ▶ CVMFS installation
  - ▶ It is enough to set up the environment to begin using DIRAC service, e.g.
    - ▶ Source `/cvmfs/dirac.egi.eu/dirac/bashrc_egi`
- ▶ Readily available Docker containers with DIRAC clients preconfigured for different installations

- ▶ Several VO using DIRAC services have dedicated application portals that can be interfaced to DIRAC
  - ▶ Need language neutral interface to the DIRAC services
- ▶ Some still use DIRAC command line client
  - ▶ Not good: needs installation and maintenance of the DIRAC client
- ▶ REST interface for a subset of the most common DIRAC services was introduced
  - ▶ Job management, file catalog
  - ▶ Example: vo.complex-systems.eu users are using DIRAC services via the OpenMOLE front-end written in *scala*

- ▶ Pilots are always submitted for a specific VO
  - ▶ Pilots can take jobs from different users but from the same VO
    - ▶ Security
    - ▶ Accounting
  - ▶ This might change in the future
  
- ▶ Now a separate Pilot Factory (Director) per VO
  - ▶ With dozens of VO's quickly becomes difficult to maintain
  - ▶ Work is ongoing on multithreaded multi-VO Directors
    - ▶ Other VO specific agents are already multi-threaded
      - VOMS2CS, BDII2CS, ...

- ▶ EGI provides the Check-in AAI
  - ▶ Federation of identity providers
  
- ▶ DIRAC Web portal is enhanced to allow login via the Check-in plugin
  
- ▶ Internally DIRAC still uses certificates/proxies for user identities
  - ▶ Generated automatically in case of SSO login. There are two possibilities:
    - ▶ User certificates provided by the DIRAC internal CA
      - If only communication with DIRAC services is needed
    - ▶ Use of Per User SubProxies (PUSP)
      - Community wide robot certificate with the user identity encoded as a CN extension
      - Several EGI services accept and understand PUSP proxies
  - ▶ This is the work in progress

- ▶ **Resource Management Service**
  - ▶ Resource status evaluation framework for organizing specific probes
- ▶ **Resource Status Service**
  - ▶ Serves status information to the clients
- ▶ **The current Resource Management does not allow separate probes and status per VO**
  - ▶ There is a clear need for VO specific probes and status
  - ▶ This is a work in progress



- ▶ DIRAC multi-VO services provide easy access to grid resources for multiple relatively small user communities
- ▶ There are several active multi-VO DIRAC services and there are more to come either independently or by joining existing installations
- ▶ Multi-VO support required several important developments specific for this mode of operation