



News from ESO Archive services: Next Generation Request Handler and Data Access Delegation

N. Fourniol, J. Lockhart, D. Suchar, L. E. Tacconi-Garman, C. Moins, T. Bierwirth, P. Eglitis,
M. Vuong, A. Micol, N. Delmotte, I. Vera, A. Dobrzycki, V. Forchi, U. Lange, F. Sogni

European Southern Observatory, Garching bei München, Germany



Abstract

Up until March 2011 electronic delivery of data from the ESO Science Archive was conducted through the use of ftp from an user dedicated staging area. Users would have access to publicly available data while access to proprietary data was only possible for the programme Principal Investigators. This poster presents the new ESO Archive services which improve the electronic data access via the Download Manager and also provide PIs with the option to delegate data access to their collaborators via the Data Access Control.

ESO Archive new services

1) The Next Generation Request Handler

From the User's point of view: Fig. 1 presents a schematic view of data request submission and download. Starting from the Archive query forms [2, 3], Users have an option to choose the former "FTP or Media delivery" service (using the old RH) or the "Instant Download" as shown in Fig. 2. The historical Request Handler service (and its FTP support) is meant to be decommissioned in the near future. The "Instant Download" button allows direct http download.

ESO Next Generation Request Handler



Fig. 1: Schematic view of the new download capability

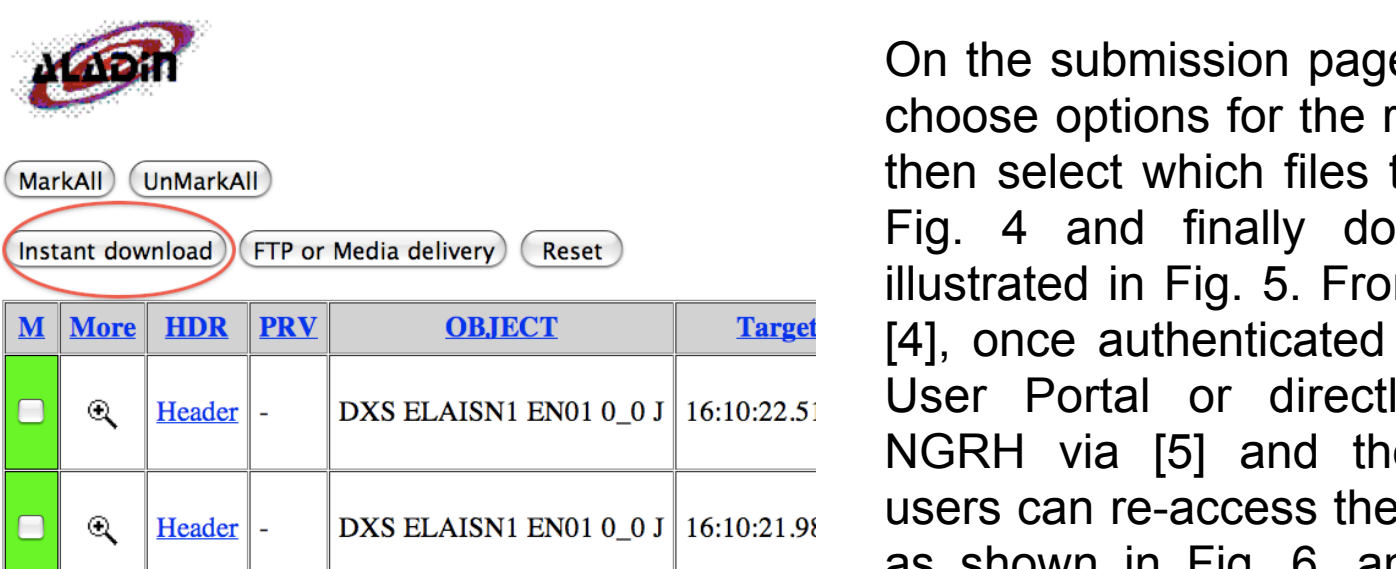


Fig. 2: Instant download new button

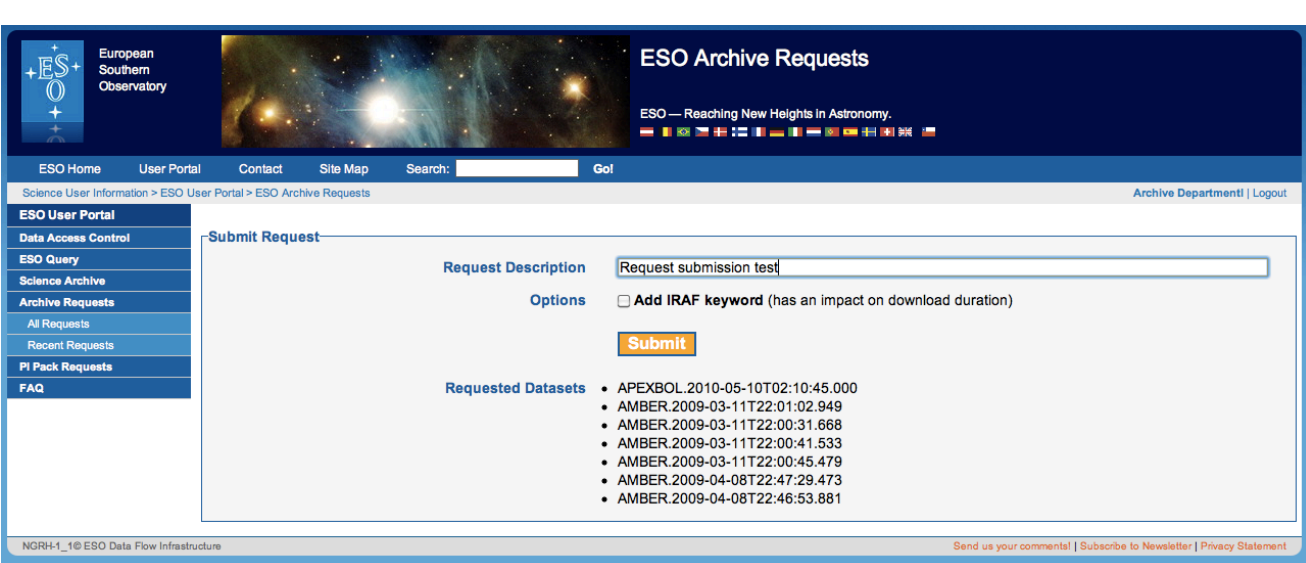


Fig. 3: Request submission page

A lot of efforts have been put to make the NGRH a very user-friendly tool for the Archive Users. It allows data download directly from the Archive but also download monitoring. The Download Manager itself allows one to retry a failed download, pause or resume the download. More Information in [6] and [7].

From the Operators point of view: the NGRH is also used by the ESO Operators to search, monitor and troubleshoot requests in a very efficient way. As shown in Fig. 7 one can view the state of all the requests in a range of dates, for a specific user, a type of request or a request state. For instance, a request in a failed status (▲) can be further troubleshooted by clicking on the failed icon. This will lead to a page which allows one to check the error message (Fig. 8).

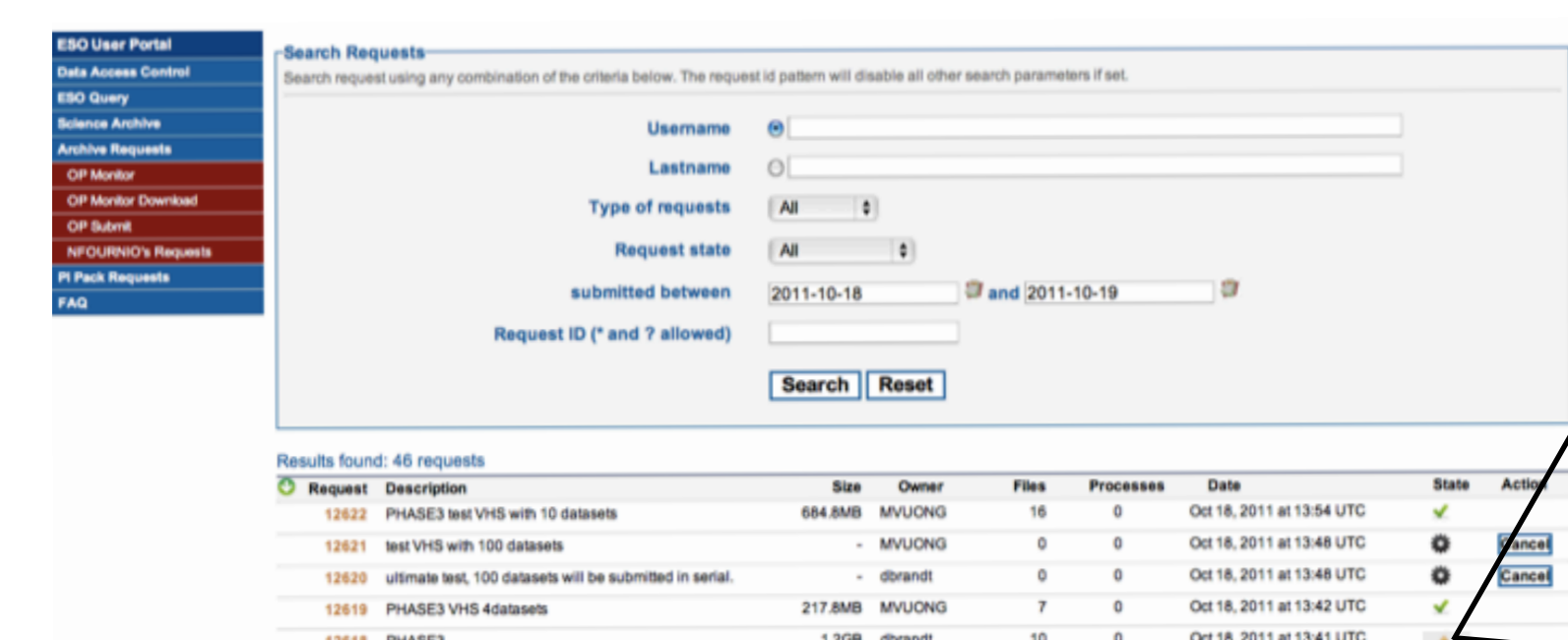
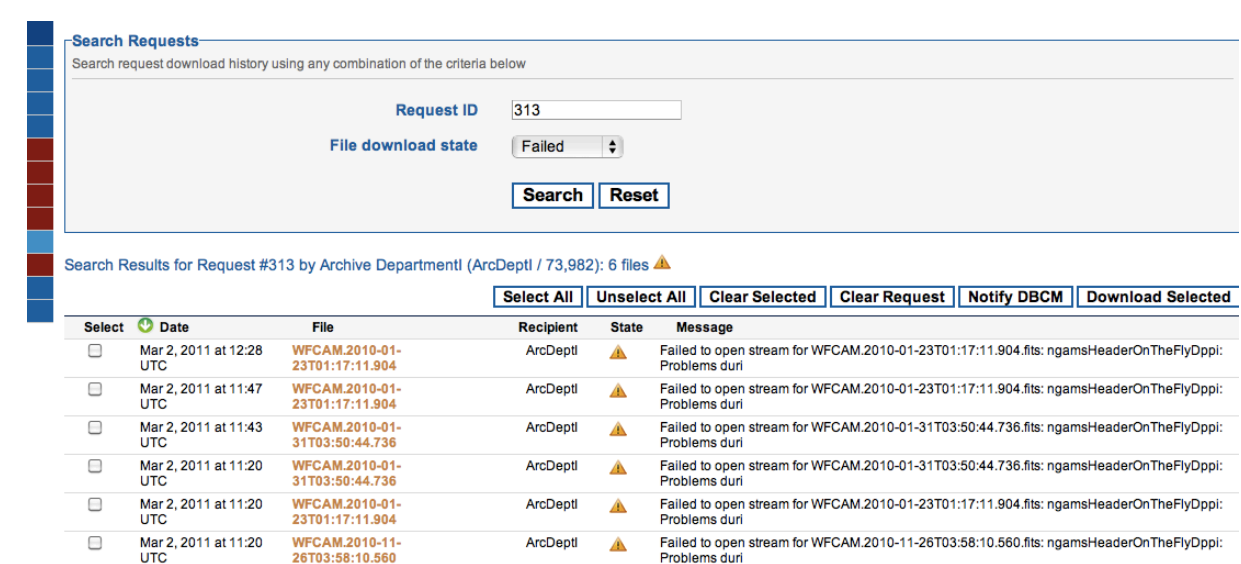


Fig. 7: Operator Monitor tool



Case of a failed download (Fig. 8): after checking the files, the operator can change the status of the failed files back to "OK" (✓). When all the files are OK, the status of the request itself is back to "OK" and it can again be downloaded by the user.
Case of a failed process (Fig. 9): in this case the operator can "Resubmit" the process

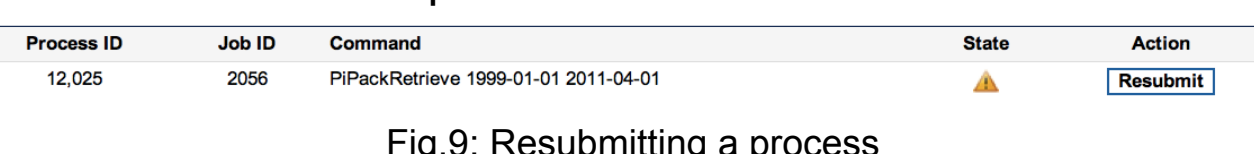


Fig. 9: Resubmitting a process

5) The S/W Architecture for the NGRH and Data Access Control

The NGRH and the Access Control software has been developed by the Software Development Division at ESO but part of the credit must go to CADC due the original software that the NGRH is based on was developed by CADC for the ALMA project. Fig. 18 shows the architecture of the whole system: it consists of two web applications, the actual NGRH and the Data Portal. As already mentioned, the NGRH allows a User to submit requests to retrieve data products, to see a request's content and to view the history of all requests previously submitted. A scalable processing infrastructure allows us to run external, asynchronous processes such as the CalibrationSelector to associate calibration frames to the requested data sets and to add them to a request via a REST interface. The ESO DownloadManager Java applet used to download data from the Data Portal uses parallel streams in order to maximize bandwidth utilization over high-latency network. Users who prefer command line usage are provided with a shell script to download files with curl/wget. The Data Portal supports file download of both files archived in the ESO Primary Archive as well as non-archive files on a staging area. Access permissions are verified against a central access control service (ACE) which is also used by Principal Investigators when they grant access to their proprietary data to their collaborators. Both applications support user authentication through the CAS single-sign on service.

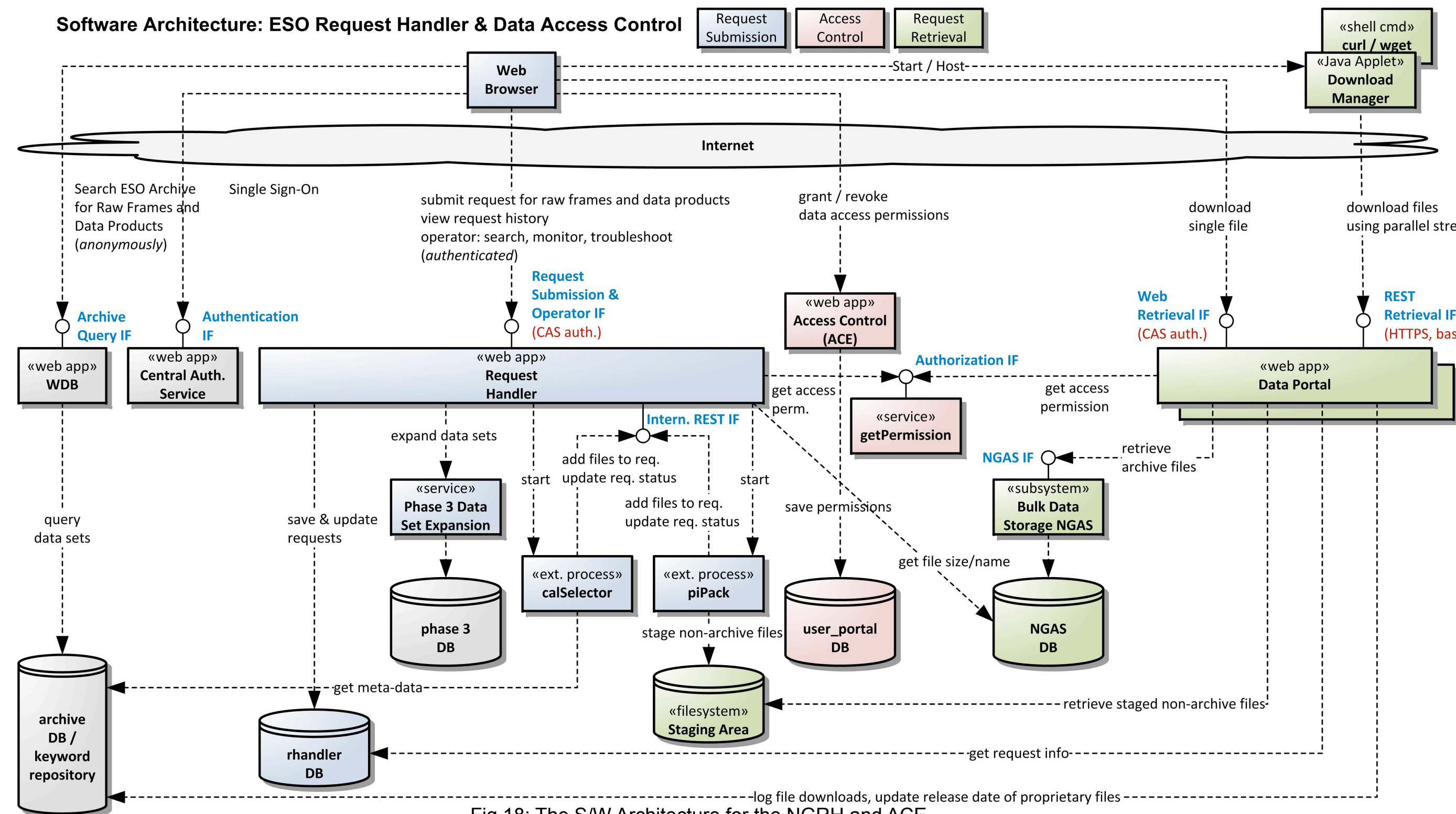


Fig. 18: The S/W Architecture for the NGRH and ACE

References

- [1] N. Rossat et al., "ESO Archive Services", ASPC 2005
- [2] http://archive.eso.org/eso_archive_main.html
- [3] <http://archive.eso.org/cms/esodata/instrument-specific-query-forms>
- [4] <http://archive.eso.org/cms/data-portal>
- [5] <http://dataportal.eso.org/rh>
- [6] <http://archive.eso.org/cms/faq>
- [7] N. Fourniol, "Archive Users Manual: Retrieval of data with the Download Manager and Data Access Delegation", ESO document
- [8] <http://www.eso.org/ace>

Background

The ESO Archive contains over 200 TB of scientific data originating from the La Silla Paranal Observatory, in-kind contributions (WFCAM and soon GTC data) as well as internal and external products. Until March 2011 the delivery of these data has been handled uniquely by a so-called Request Handler (RH, [1]), monitored internally by an operators team. A new tool called the Next Generation Request Handler (NGRH) has since then been introduced, currently running in parallel with the old one but which will soon fully replace it. We present here 1) the NGRH new tool from the User and Operator points of view, 2) some numbers reflecting its use within the community, 3) a future functionality of the NGRH, 4) information on an additional service: the data delegation service, 5) a S/W architecture diagram of the NGRH and Data Access control and 6) a H/W infrastructure diagram.

2) A few numbers for the NGRH use

The Archive delivers about half the volume it contains to the astronomical Users (see Fig. 10). With the arrival of the Survey Telescopes (VISTA and just operational VST) the amount of additional data has doubled in year 2010. Up until the NGRH was released to the public in March 2011, the average number of requests per day, all handled via the old request handler (RH), was about 75 requests. Now the ratio is 1/3 for the old RH and 2/3 for the NGRH, in over 600,000 http downloads in total. That there are still some requests to the old RH may be explained by the fact that there is not yet an option for media in the NGRH, and that products are not all accessible via the NGRH. Both services soon to come.

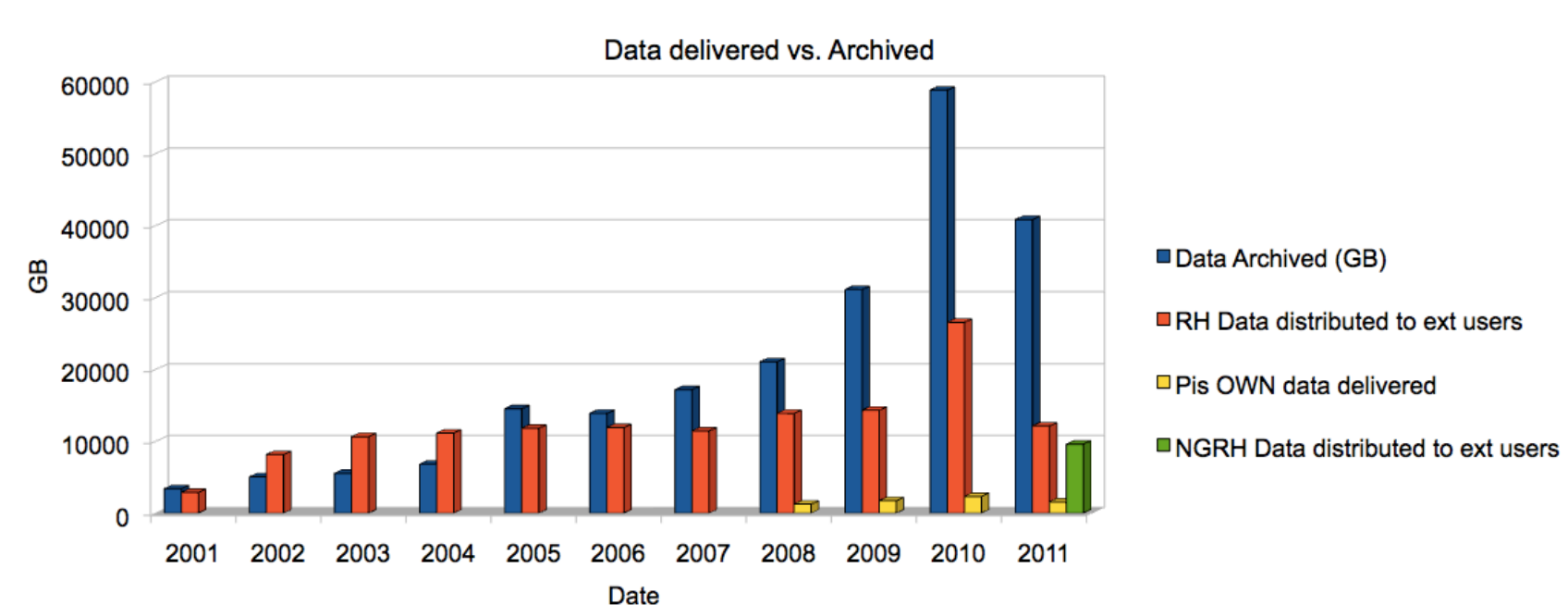


Fig. 10: Delivery vs. Archiving

3) A future functionality of the NGRH

The Calibration Selector, Cal Selector for short, is a new functionality soon to be released to the archive Users via the NGRH. This will allow the Users to retrieve all the calibration raw files as well as the master calibration frames associated to the science files requested. As visible on Fig. 11 a new "Cal selector" button will soon appear on the request submission page as an option. The result will be an expandable dataset (see Fig. 12 and 13) that can be downloaded as a whole or by part.

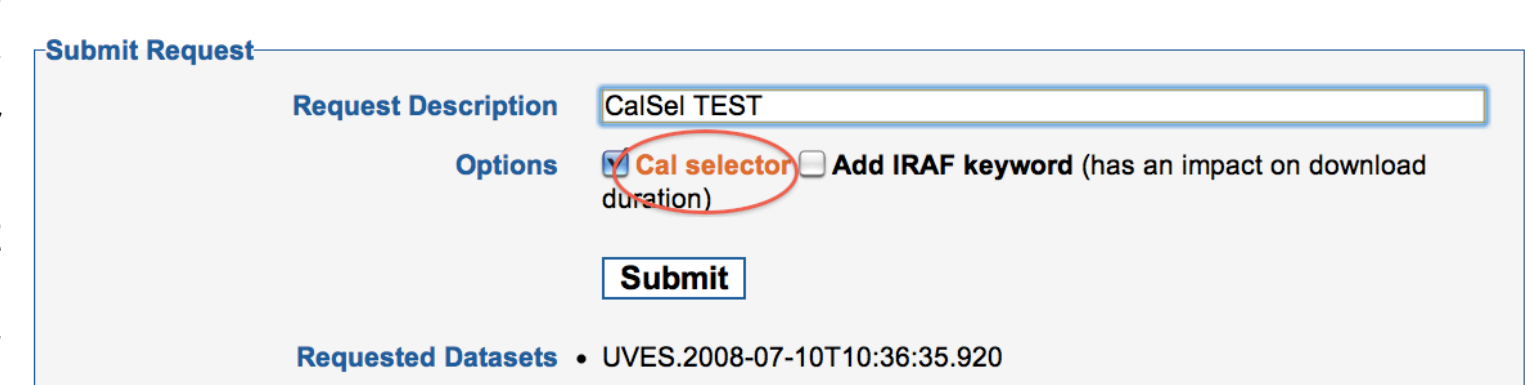


Fig. 11: The Cal Selector button

Requested Datasets
UNIX/Linux Shell script if you prefer command line download of complete request: [downloadRequest21130script.sh](#)

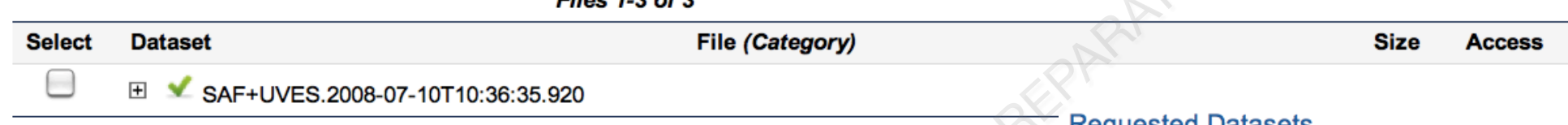


Fig. 12: One CalSel dataset

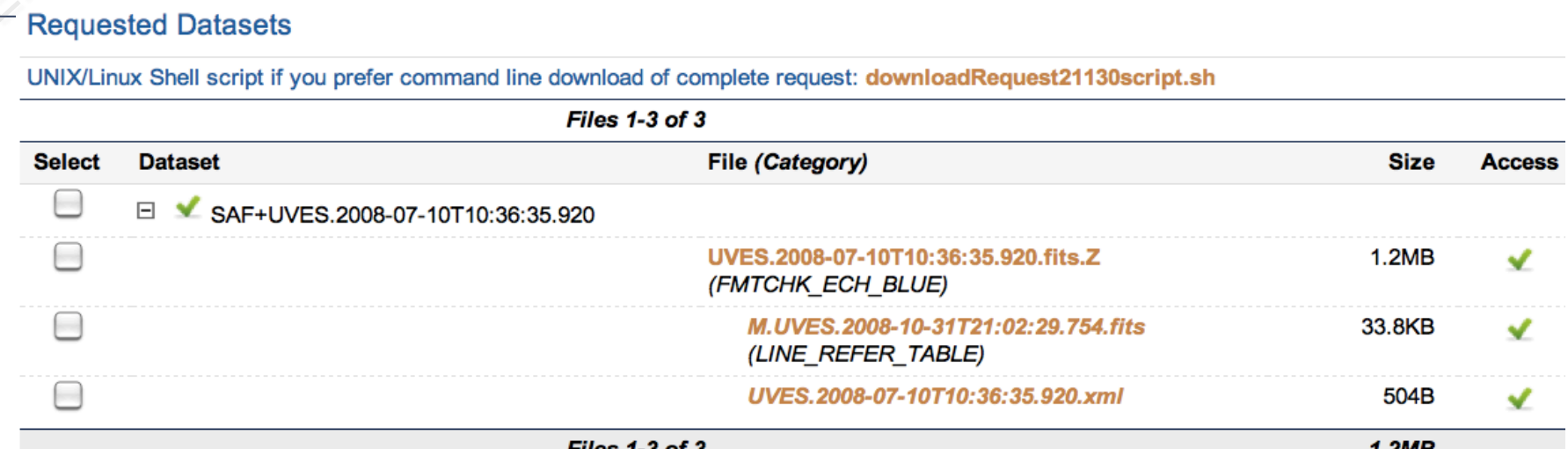


Fig. 13: Expansion of a dataset

4) The Data Delegation service

From the ESO Archive Request interface [4] one can access the "Data Access Control" service (Fig. 15), which allows a PI to delegate access to his/her data to his/her collaborators assuming the latter are registered in the ESO User Portal (Fig. 14). Runs can be delegated as long as they contain proprietary data, or have the potential to acquire proprietary data (that is, the run is still viable). These are visible on the first page corresponding to the menu item "Admin Data Access" (Fig. 16 and [8]). There is also a possibility for the PI to revoke the access delegation at any time. The uptake of the service is illustrated in Fig. 17. After an initial peak, there is now an average number of 15 delegations per month.

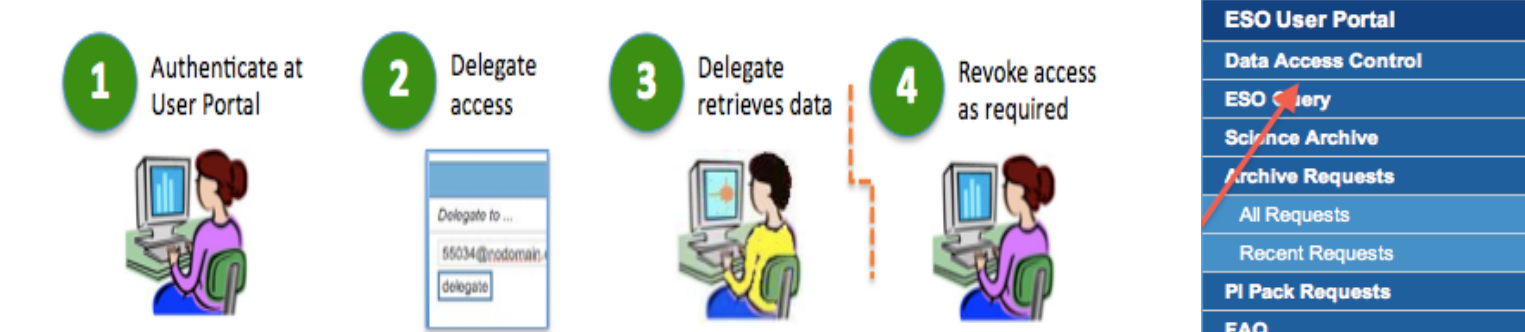


Fig. 14: Schematic illustration of data access delegation

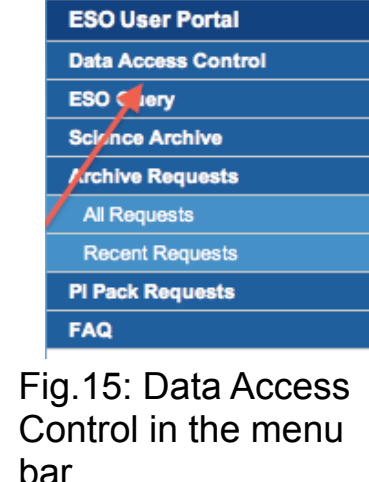


Fig. 15: Data Access Control in the menu bar

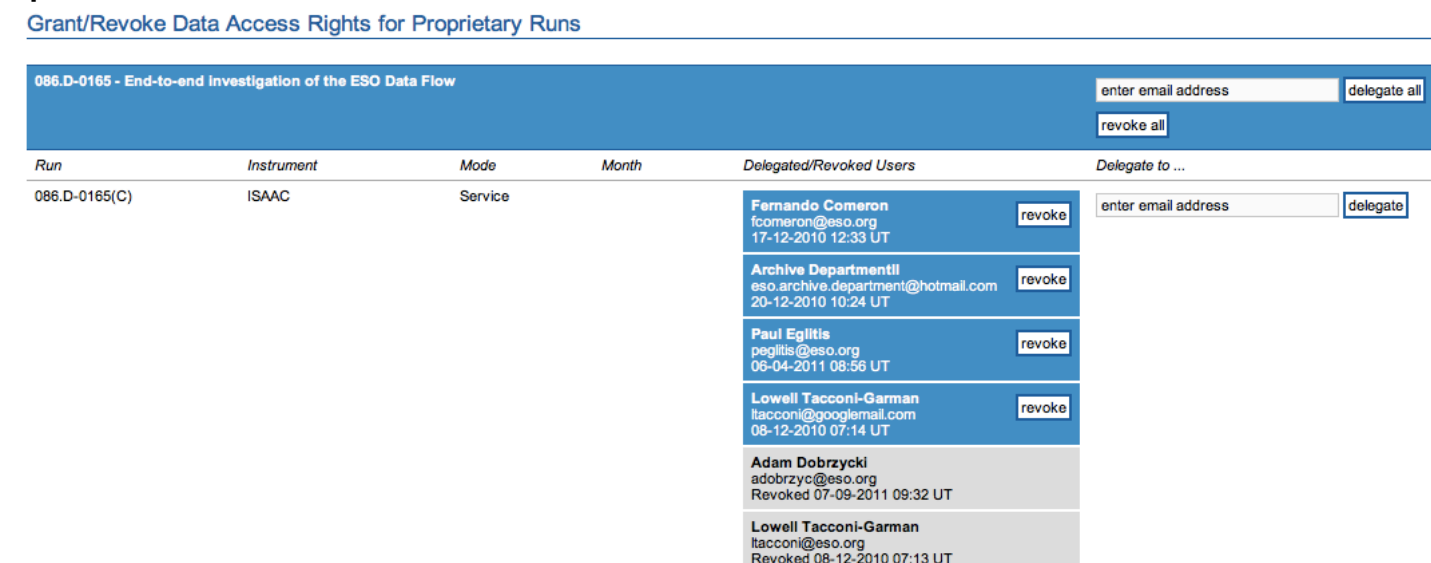


Fig. 16: Data Access rights Administration page

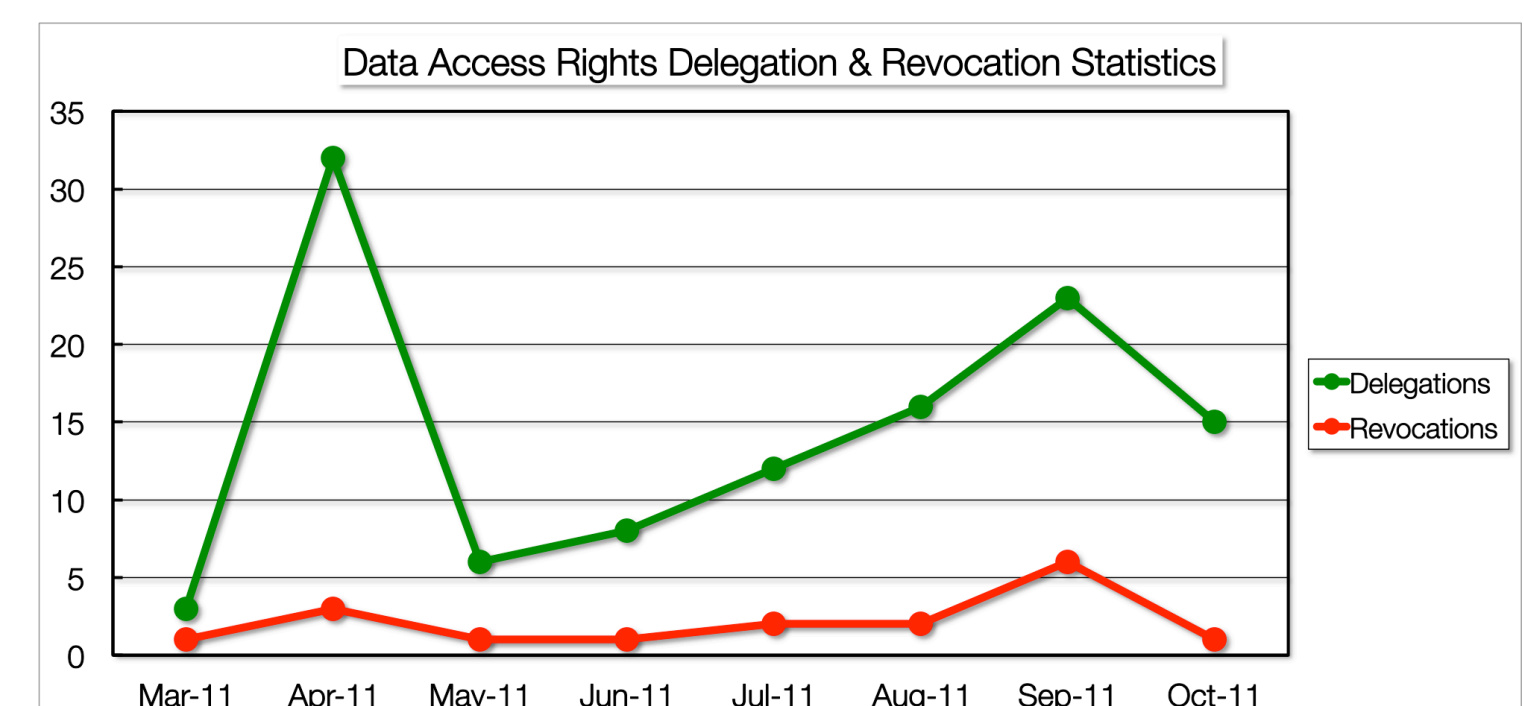


Fig. 17: Data Access Delegation Statistics

6) The H/W infrastructure for the NGRH

The ESO Request Handler is deployed in a distributed manner within the main ESO network infrastructure as visible in Fig. 19: the NGRH and the DataPortal applications are deployed in tomcat containers on a number of virtual machines within the Data Management and Operations Division portion of the ESO DMZ. The internal archive operation processes are not visible to the public and are therefore deployed in a secure area behind an internal firewall, however certain file-systems are being made available in read-only mode. Also behind the firewall are the actual data holdings stored in a Primary Archive (on Next Generation Archive System nodes) and the supporting meta-data stored in both Sybase Adaptive Server Enterprise and Sybase IQ Relational Database Management Systems.

Legend

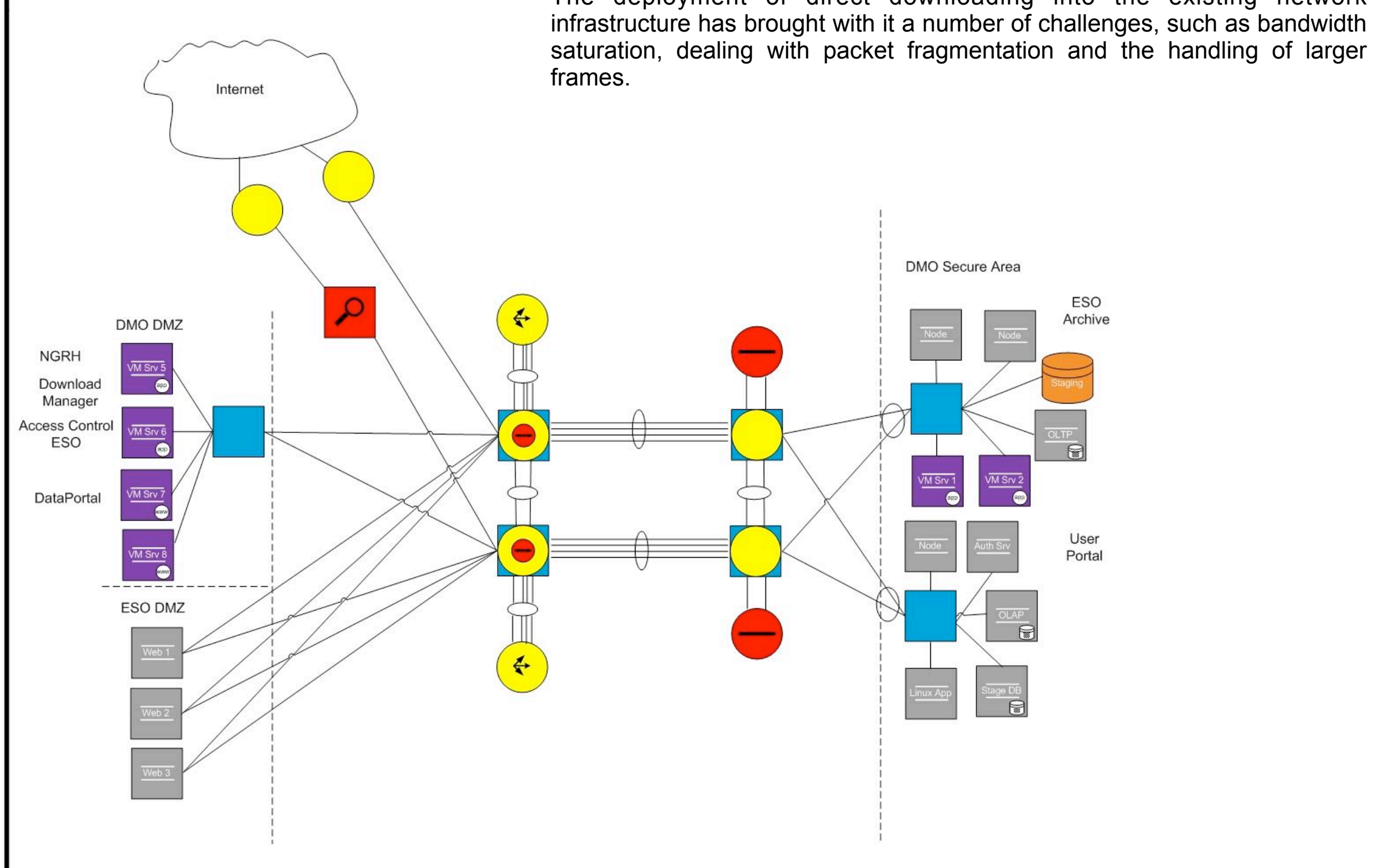


Fig. 19: The H/W infrastructure for the NGRH and ACE

Acknowledgments

We would like to thank: Felix Reinel for his dedication when deploying the system in the operations, Dmytro Muravov for thoroughly testing the software, Felix Stoehr for providing very useful statistics on the download performances, Nicolas Rosse and Laura Mascetti for helping out with the testing.