The European Hubble Space Telescope Archive is moving!

Jonas Haase (ESO/ESA), Christophe Arviset (ESA-ESAC), Pedro Osuna (ESA-ESAC) and Michael Rosa (ESA)

Abstract: The European Hubble Space Telescope Archive is moving from its home for the last 21 years at ESO, to ESA's European Space Astronomy Centre (ESAC) where it will join the archives of most other major ESA space science missions. The European HST archive stays in operation during the move, serving up-to-date files to the community. The main data production and transfer phase is currently underway and the final transition is expected to happen in the first half of 2012. This poster gives an overview of the archive move and discusses the challenges involved.

Background

The European Hubble Space Telescope Archive facility had been one of the primary formal deliverables of the Space Telescope - European Coordination Facility. The ST-ECF was instituted in 1984 at the European Southern Observatory as a common project of the European Space Agency (ESA) and ESO and was discontinued by the end of 2010.

At the time of the closure of the ST-ECF, it has been decided that the HST archive should be migrated from its original location at ESO to a final destination at ESAC. There it will take its place among ESA's other science archives and continue to serve very well calibrated and up-to-date science data to its community of more than 2000 users.

During the transition period the European archive stays fully operational at ESO, providing accessibility to all publicly available data from HST.

Migration Timeline

Early 2011 a timeline was put together to plan the move from one archive center to the other, while still guaranteeing normal maintenance and service at the old location. The move is predicted to last slightly more than a year:

- Preparatory work at ESO and ESAC (early 2011)
  - Early analysis and work plan. ✔
  - Build of a development and test archive system at ESAC. ✔
  - Provisioning of new hardware and infrastructure at ESAC. ✔
  - Adapting Cache to new infrastructure and expanding storage handling code. ✔
  - Processing tests on the ESAC Grid. ✔
- Migration to ESAC (late 2011)
  - Setup of Sybase servers. ✔
  - Transfer of all static telescope data from ESO to ESAC. ✔
  - Enable standard data and meta-data flow from STScI to ESAC. ✔
  - Reprocessing of all dynamic data on the ESAC Grid. IN PROGRESS
  - Transfer and ingestion of HLSP/HLA data. IN PROGRESS
  - Transfer of archive web pages and web applications.
- Commissioning phase at ESAC to demonstrate full equivalence in service and capabilities to the current state at ESO (early 2012)
  - Testing of subsystems.
  - Testing of front-end for the science users.
  - Testing of interfaces to STScI Archive and CADC.
  - Proving stable ops and capability for regular updates.
- Delivery Point (Late Spring 2012)
  - Switchover of all links to the European archive from ESO to ESAC.
  - ESAC assumes full responsibility of the European HST Archive and its operations.
- Final Switchover (Summer 2012)
  - Dismantling of the HST archive at ESO.

What is Changing? What is Not?

- The European HST Archive continues to serve up-to-date, highly processed science data to the community. Search interfaces and data delivery options remain the same as at ESO.
- The archive will continue to serve the newest available versions of The Hubble Legacy Archive data and other High Level Science Products.
- With the closure of ST-ECF, specialized science support for HST data is not available anymore. Support for the archive itself is available, and limited science help may be possible as well, but users may also get referred to the STScI help-desk.
- After the closure of the ST-ECF, the European Archive ceased to serve data in its proprietary period to PIs and COIs. It is not expected that this service is going to be resumed after the move either.
- The current VO offerings of the European HST archive are going to be replaced with new VO services in line with the currently available ones at the archives at ESAC.

Logistics

In the final years of the ST-ECF considerable effort went into making the HST archive lightweight and easy to maintain. In cooperation with the CADC, the HST Cache system was developed to automate the bulk of the data and metadata management tasks.

The Cache is an envelope around HST archive file production. It is a set of database tables and software agents that ensures that all Hubble Space Telescope pipeline products are locally available for rapid data retrieval and are in the best possible state at all times. The Cache includes mechanisms to discover new datasets to insert, and automatically triggers reprocessing of datasets which benefit from updates. It takes the data through the following processing steps:

Production of raw FITS files: HST telemetry data and metadata are run through the OTFR pipeline, provided by the STScI (Swam et al. 2001), to generate raw FITS files.

Calibration: Raw FITS are calibrated in the OTFC pipeline (Pirenne et al. 1998), using the newest available calibration software delivered by the STScI. Whenever possible a MultiDrizzled output product is provided as well.

Preview: Uniform preview products in FITS and JPEG/PNG formats are produced from calibrated data to be used for fast web and VO access.

Metadata: Metadata are extracted from the FITS headers and database tables and are used to generate improved metadata tables for direct use through a web interface, VO and as input for the CADC Common Archive Observation Model (Dowler et al. 2007). This includes footprints directly computed from the FITS images using the ST-ECF footprint finder (Stoehr 2008).

The HST Cache software system replaced an organically grown set of scripts and tables and did indeed minimize the amount of manual intervention needed in the archive work (Haase et al. 2010). Since the system was designed to be used both at the CADC and at ESO, it is a self-contained package with configurable interfaces to the local system infrastructure. As a positive side-effect the entire archive system now is not only easy to maintain, but also relatively easy to transfer to a new archive site.
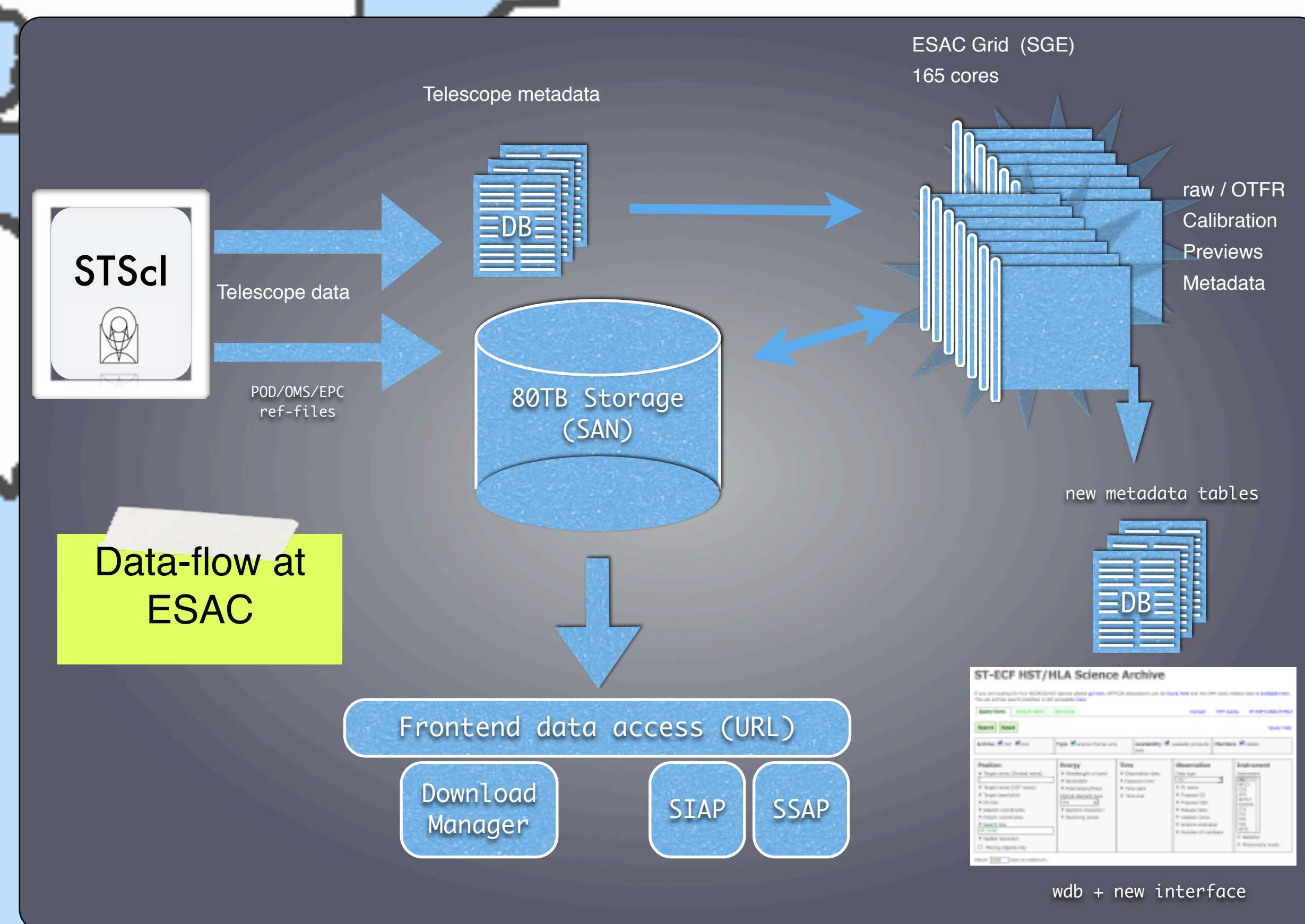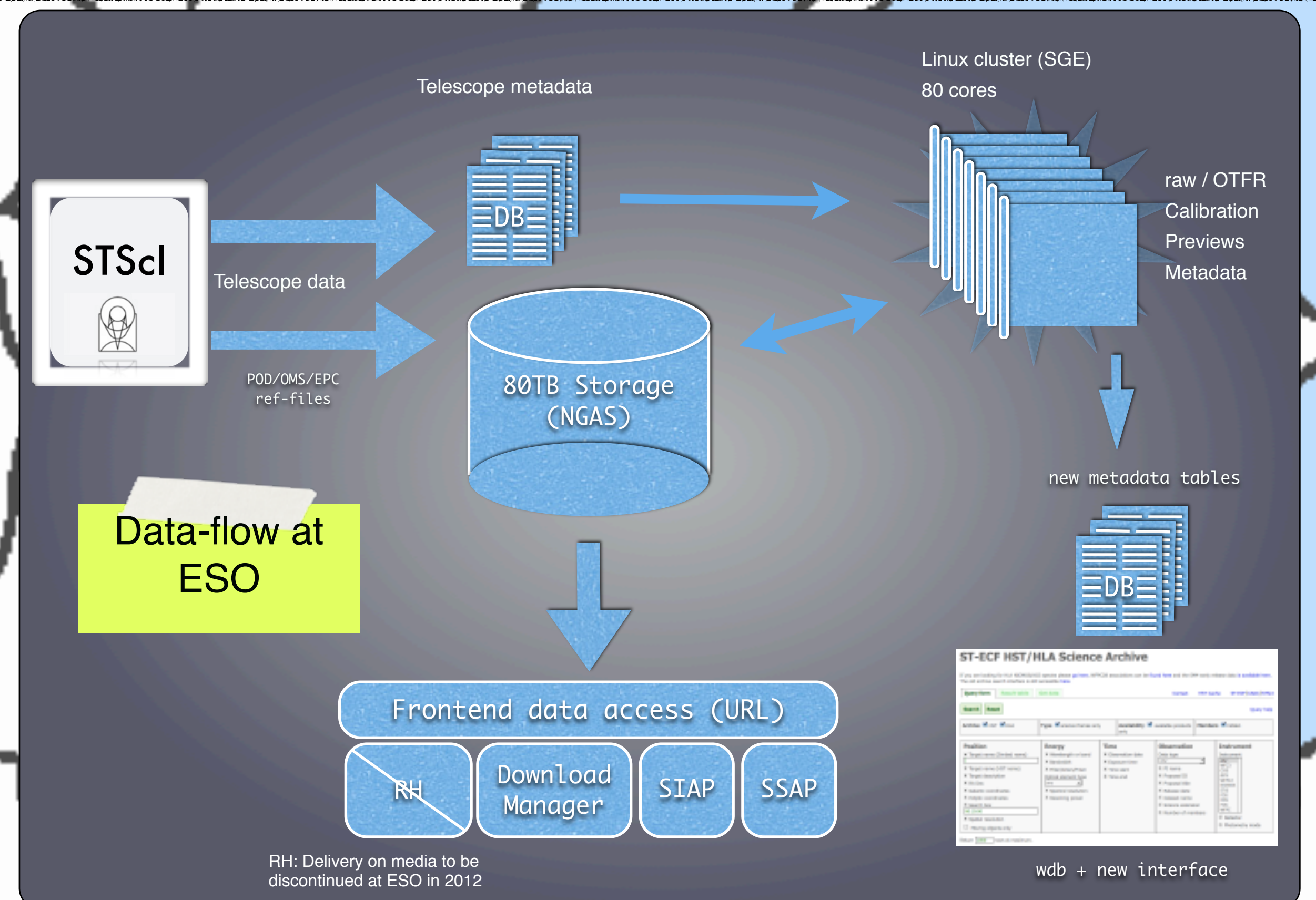
Since one of the goals of the move was to ensure a uninterrupted service for the user community it was decided to install the HST Cache software and surrounding systems on a new server infrastructure at ESAC and create a fully functional copy of the European HST Archive before the switchover was made. All software would be ported to the new environment with an absolute minimum of changes in order to remove all unnecessary risks. Most changes would purely be in the realm of configuration, the largest alteration was the adaption to the SAN storage system used at ESAC, which was a straightforward extension of the modular data storage access classes of the Cache system.

The current size of the HST archive exceeds 16 million files and a volume of 50TB (compressed). Luckily it is not necessary to transfer the vast bulk of the data, as all datasets from the newer instruments and previews for all data can be recreated from scratch by running the data through the three first production steps in the Cache system. This way it suffices to transfer the basic telescope data, reference files and the High Level Science Products, such as the Hubble Legacy Archive.

All transfer of static data is conveniently done with the data exchange methods built into the Cache software. Data processing happens on the ESAC science Grid which, like the currently operational processing nodes at ESO, is using the Sun Grid Engine for task management and can be used directly by the software.

Data-flow at ESO

Data-flow at ESAC

How much is transferred?

Data type	Number of Files	Volume / TB (compressed)
POD files	493,633	2.43
Observation logs	1,465,678	0.01
Legacy instruments (FOC,FOS,GHRS,WFPC)	679,612	0.13
HLA+HLSP	1,156,721	8.24
WFPC2B associations	252,212	0.37
reference files	90,056	0.74
Sum	4,137,912	11.92

How much is generated locally?

Data type	Number of Files	Volume / TB (compressed)
New instruments (WFPC2,Nicmos,STIS,ACS,COS,WFC3)	6,707,000	29.47
previews (all instr. and products)	5,542,572	10.26
Sum	12,249,572	39.73

References

Dowler, P. D., Gaudet, S., Durand, D., Redman, R., Hill, N., & Goliath, S. 2007, in ADASS XVI, edited by R. A. Shaw, F. Hill, & D. J. Bell, vol. 376 of ASP Conference Series, 347

Haase, J., Durand, D., Stoehr, F., & Micol, A. 2010, in ADASS XIX, edited by Y. Mizumoto, K.-I. Morita, & M. Ohishi, vol. 434 of ASP Conference Series, 275

Pirenne, B., Micol, A., Durand, D., & Gaudet, S. 1998, in ADASS VII, edited by R. Albrecht, R. N. Hook, & H. A. Bushouse, vol. 145 of ASP Conference Series, 341

Stoehr, F. 2008, Space Telescope European Coordinating Facility Newsletter, 45, 7

Swam, M. S., Hopkins, E., & Swade, D. A. 2001, in ADASS X, edited by F. R. Harnden Jr., F. A. Primini, & H. E. Payne, vol. 238 of ASP Conference Series, 291