



ESO's role as data provider: Strategies and Challenges

- ESOs mandate
- address the challenge: Data Flow System
- provide quality content: Science Data Products
- future opportunities: ESO archive



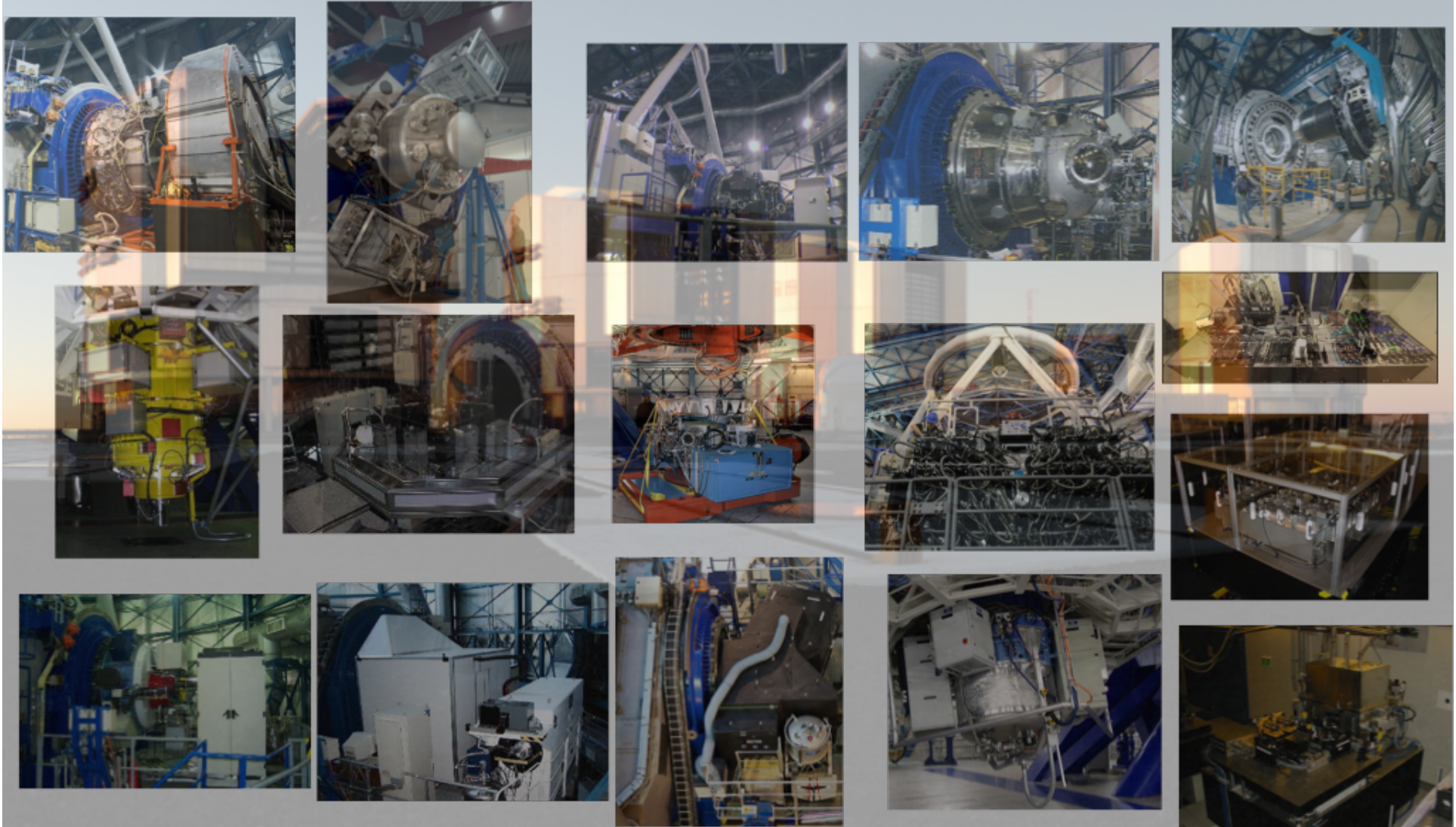
“Data” Mandate from the VLT/I Science Policy

- Monitor the long term evolution of instruments
 - instrument health
 - accuracy of calibrations
- Produce Data Products
 - remove instrumental signatures
 - calibrate in physical units
- Deliver
 - all raw, calibration and data products
 - proprietary and public data through the Science Archive Facility
 - pipelines and recipes (and increase their accuracy over time)
- Support the community
 - helpdesk
 - in the generation of Advanced Data Products

*provide scientifically useful,
trusted, high quality,
large quantity content*

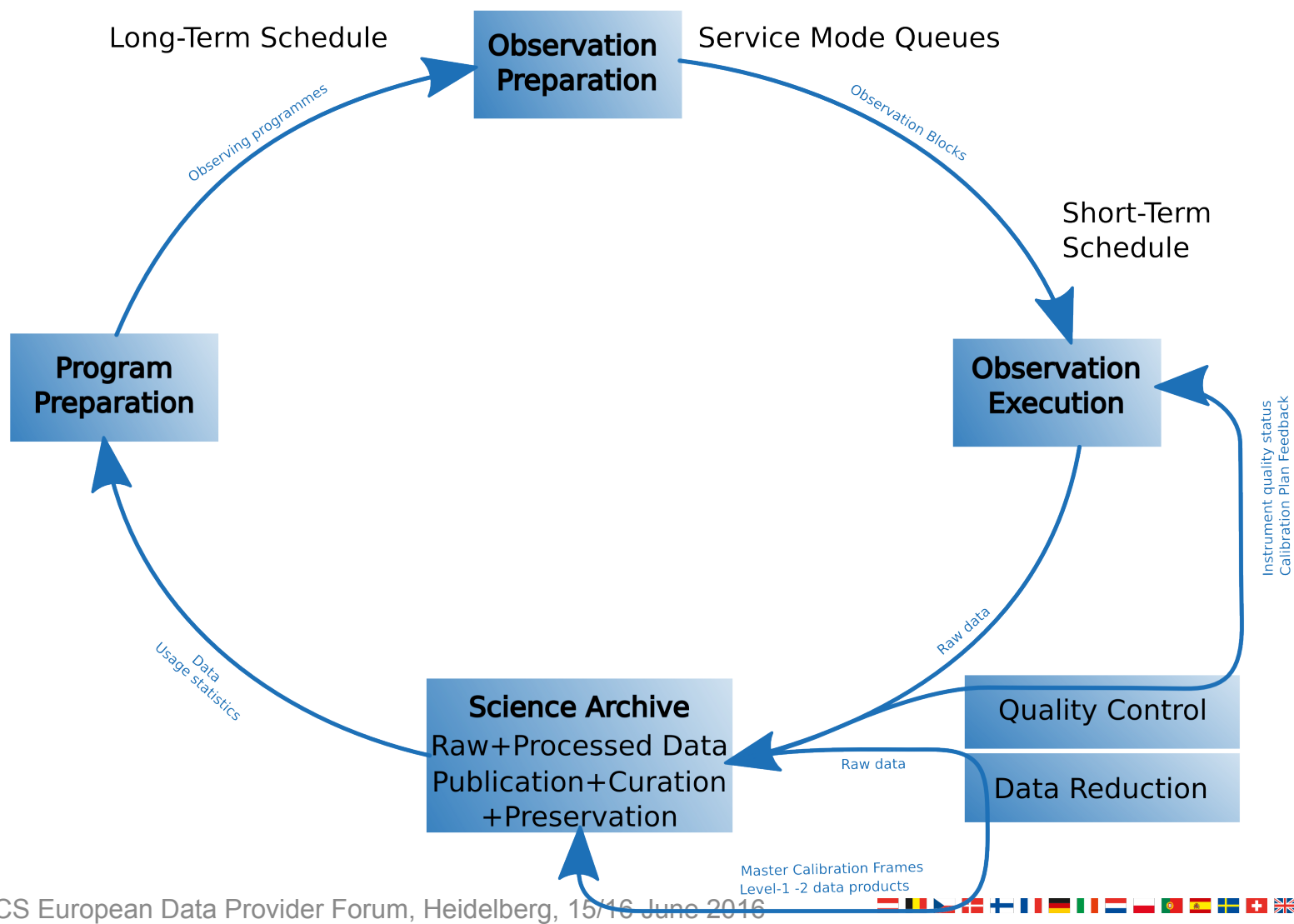


Some Challenges



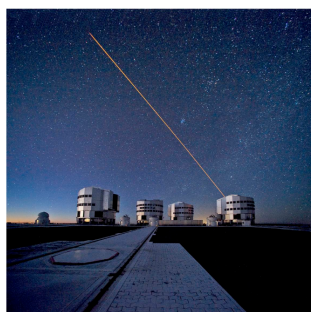


Mapping into Data Flow

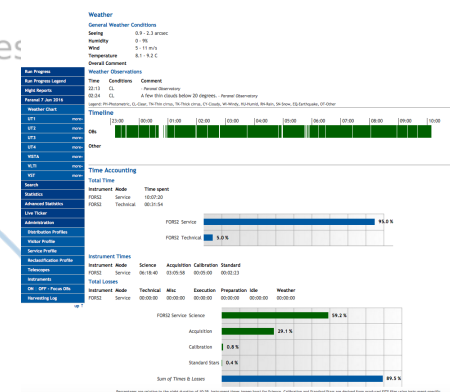




Mapping into Data Flow



Observation Preparation



ESO Call for Proposals
Proposal Deadline: 31 March 2016

Program Preparation

ESO User Portal Services

Phase 1
Download the proposal form
Submit an observing proposal
Check the time allocation information

Phase 2
Download P2PP
Submit a target or set-up change request
Check the status of your observing runs
Delegate Phase 2 tasks

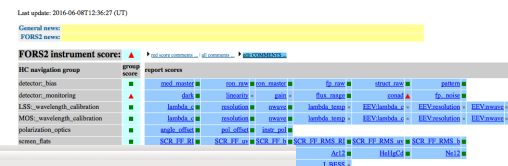
Phase 3
Download the Science Data Products Standard
Submit data
Check your Phase 3 submission status
Delegate Phase 3 tasks

Archive Services
Query the Archive for
La Silla Paranal raw data
La Silla Paranal reduced data
APEX reduced data
Phase 3 Catalogs
Delegate proprietary data access rights

Help
Ask for help
Find User Portal Information and FAQ
Check the Data Reduction Forum
Check the data reduction FAQ

Observation Execution

FORS2 trending system: overview of scores





Channels for SDP @ ESO

- In-house generation of Data Products (IDPs)
 - enabled through standardized acquisition and quality control processes
 - near-real time quality control process ensures certified master calibrations
 - un-attended processing through certified pipelines
 - goal: science grade data for all popular instrument modes
 - UVES, XSHOOTER, HAPRS, FLAMES/GIRAFFE
 - imminent: MUSE, HAWK-I, VIMOS (IMG), FEROS

- External Data Products (EDPs)
 - provided by public surveys and large programs (deliverables)
 - programs selected by their high legacy value
 - most use dedicated (non-ESO) user-pipes (eg CASU)
 - goal: advanced products (wide, deep, merged catalogs)
 - perspective: users *at large* contribute EDPs
 - quality assurance: published datasets only?
 - acknowledgement: DOIs?



SDPs, SDPS and Phase 3

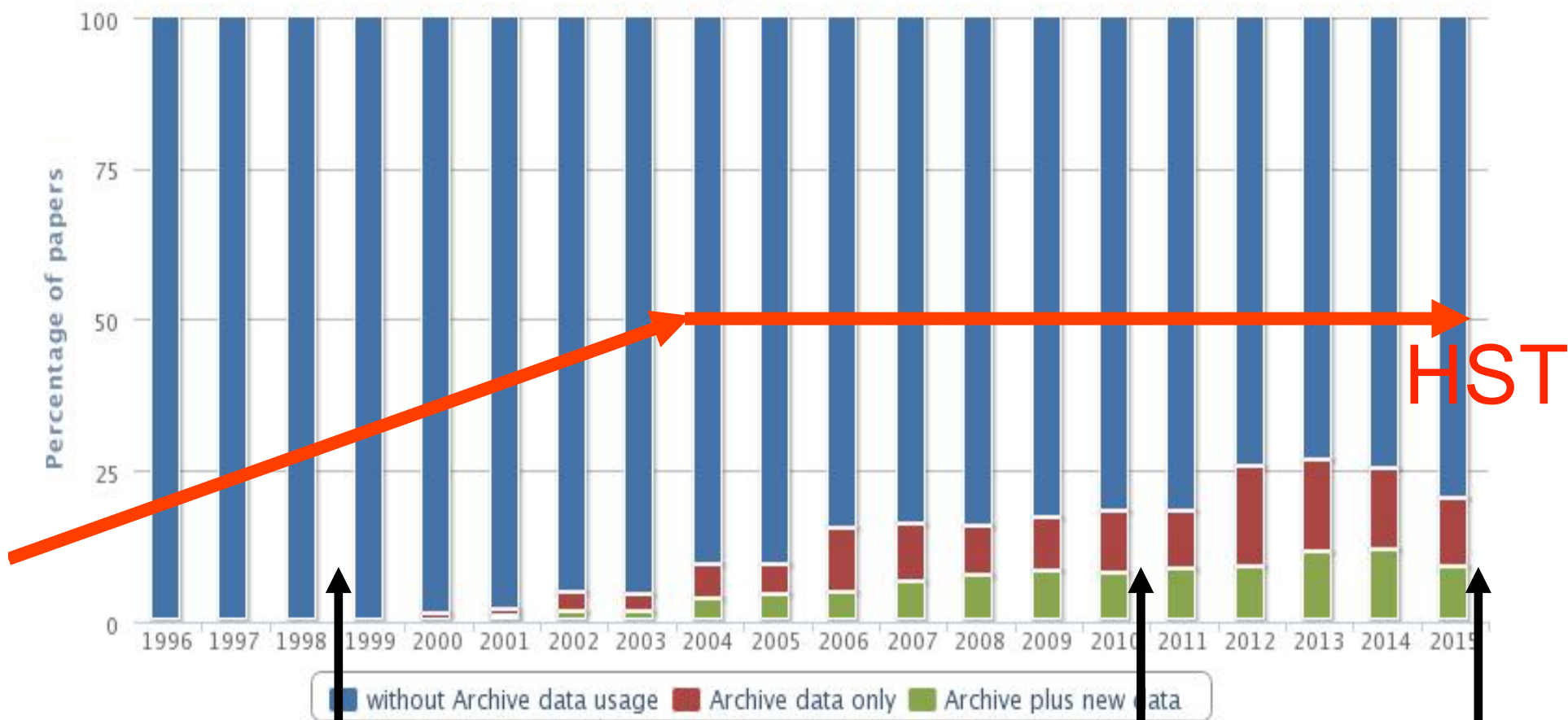
- ESO Phase 3 process enables
 - preparation, submission, validation and ingestion of science data products for storage in the ESO Science Archive Facility (SAF), and subsequent publication to the scientific community.
- ESO Science Data Product Standard is **required** for coherence of EDPs *and* IDPs in the SAF
 - defines format, meta-data, keywords, quality descriptors and processing provenance
 - generally derived from “VO” standards, when available
 - www.eso.org/sci/observing/phase3/p3sdpstd.pdf
- added-value through validated and curated content
- ESO SDPS sets pace
 - multi-epoch photometry (surveys, timeseries, NGTS)
 - processing provenance
 - 3D/IFU cubes (KMOS, MUSE!)
 - sub-mm/radio maps (APEX/ATLASGAL)



SAF as a science resource

U. Grothkopf et al., <http://www.eso.org/sci/libraries/edocs/ESO/ESOstats.pdf>

Archive Data Usage
Source: telbib, 1996 - 2012



start of facility operations

start archive population with DP

archive services interoperability





... and costs?

(fraction of total operation costs)

- data archive operations
 - archive infrastructure TCO (1PB, 3 safe copies) 0.3-1%
 - content management (production, curation) ~10%
- “systemic” data generation
 - facility (VLT) time for calibrations ~ 4%

- favorable cost-benefit relation
 - close monitoring, metrics...
 - effective use of resources (FTE and \$)



NEW ESO Archive Services: high level goals

- Build access services to the holdings of the ESO Science Archive Facility to maximize its scientific potential *within given resource constraints*
- The archive is a haystack of content, and users want to identify the needles they are interested in
 - make the two ends meet
- We build upon rich (curated!) metadata to enable complex queries based on the physical properties of the data
- Added-value services: previews, cutouts, solar system science, hierarchical file grouping, ...



NEW ESO Archive Services: project outline

■ Interactive access

- Query, display, interact, preview, retrieve

■ Programmatic interface

- incl. ADQL, TAP, ObsTAP/ObsCore, DataLink, AccessData...

■ Operational access

- Custom queries, full access

■ Underlying Infrastructure:

- Data storage, optimized for fast retrieval
- Databases, SQL and/or nonSQL (Solr/ElasticSearch etc)
- Full integration into Data Flow System



NEW ESO Archive Services: user interface

- New SAF user interface – key attributes:
 - **Graphical:** footprints, previews, aggregations, histograms, 2d distributions, next to the traditional tabular view
 - **Responsive:** Quick (in-browser) interaction with the data, while preserving their richness (images, cubes, spectra,...)
 - **Powerful:** Search by position, wavelength coverage, spatial/spectral resolution, limiting depth, SNR; programmatic access (VO protocols)
 - **Unifying:** unique entry point to all ESO science data
 - **Efficient:** fully integrated with ESO's Data Flow System



NEW ESO Archive Services: programmatic interface

- deploy VO services and protocols
 - incl. ADQL, TAP, ObsTAP/ObsCore, DataLink, AccessData (Simple Data Access)...
- Convergence to few stable VO protocols for data access
- Authenticated VO access
 - Access statistics are vital to understand our community, hence serve them better
 - Balance with ease of access and removal of access barriers
- VO accessibility of textual release descriptions
 - Vital information on global data quality, limitations and usability beyond mere file-by-file metadata



NEW ESO Archive Services: possible areas of collaborations

- assigning object categories to SAF assets to enable new ways of searching (e.g. *find spectra of $z>6$ QSO's*)
 - harvest meta-data?
 - distributed search?
- FITS serialization of new data models (e.g. optical interferometry, spectro-polarimetry)
- dynamic visualization of spectra/cubes in a web page
- incremental creation HiPS



NEW ESO Archive Services: implementation strategy

- We want to reuse existing components (Aladin Lite, VO libraries, etc.) as much as possible to build archive services tailored to ESO's requirements
- We maintain ownership of the application but not of the building blocks
- ASTERICS collaboration as opportunity to improve/further develop existing components
- Possible new developments @ ESO
 - usage of NoSQL search platform (Apache Solr, Elastic Search) to enable “real-time” exploration of archive contents (multi-dimensional aggregations/histograms)
 - Problem: different back-ends for programmatic/VO access and web/interactive access (data replication)



■ ASTERICS Project Contact

- Martino Romaniello
- Jörg Retzlaff
- Olivier Hainaut
- Stefano Zampieri
- Michael Sterzik

■ active exchange with CDS and ESA is ongoing