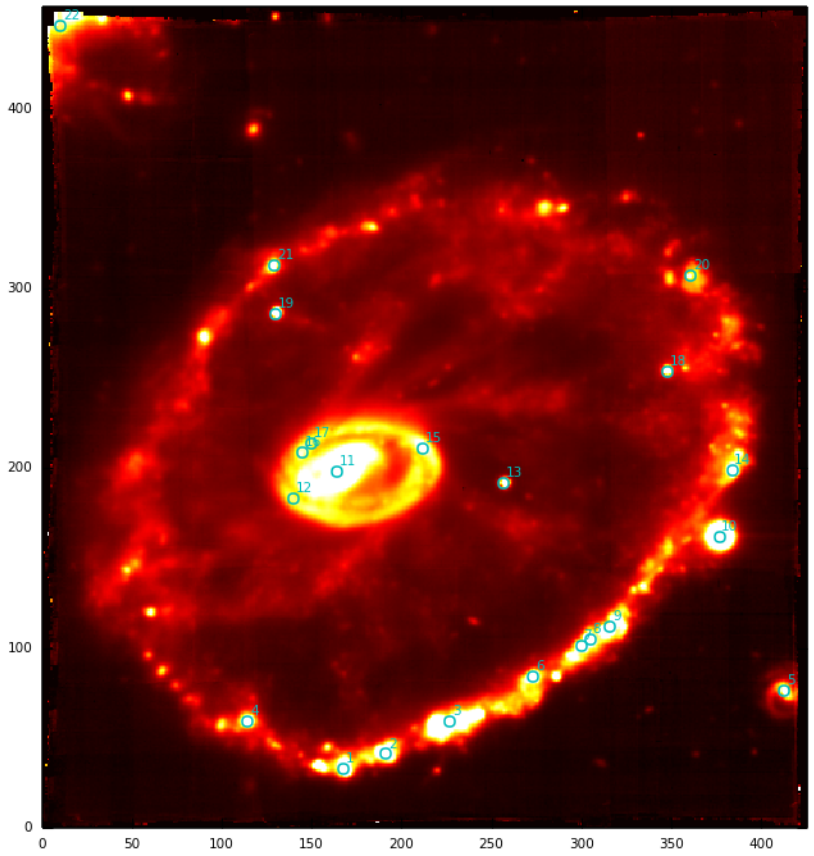
ESO-VLT science data processing

Reinhard Hanuschik

ESO

Data Processing and Quality Control Group

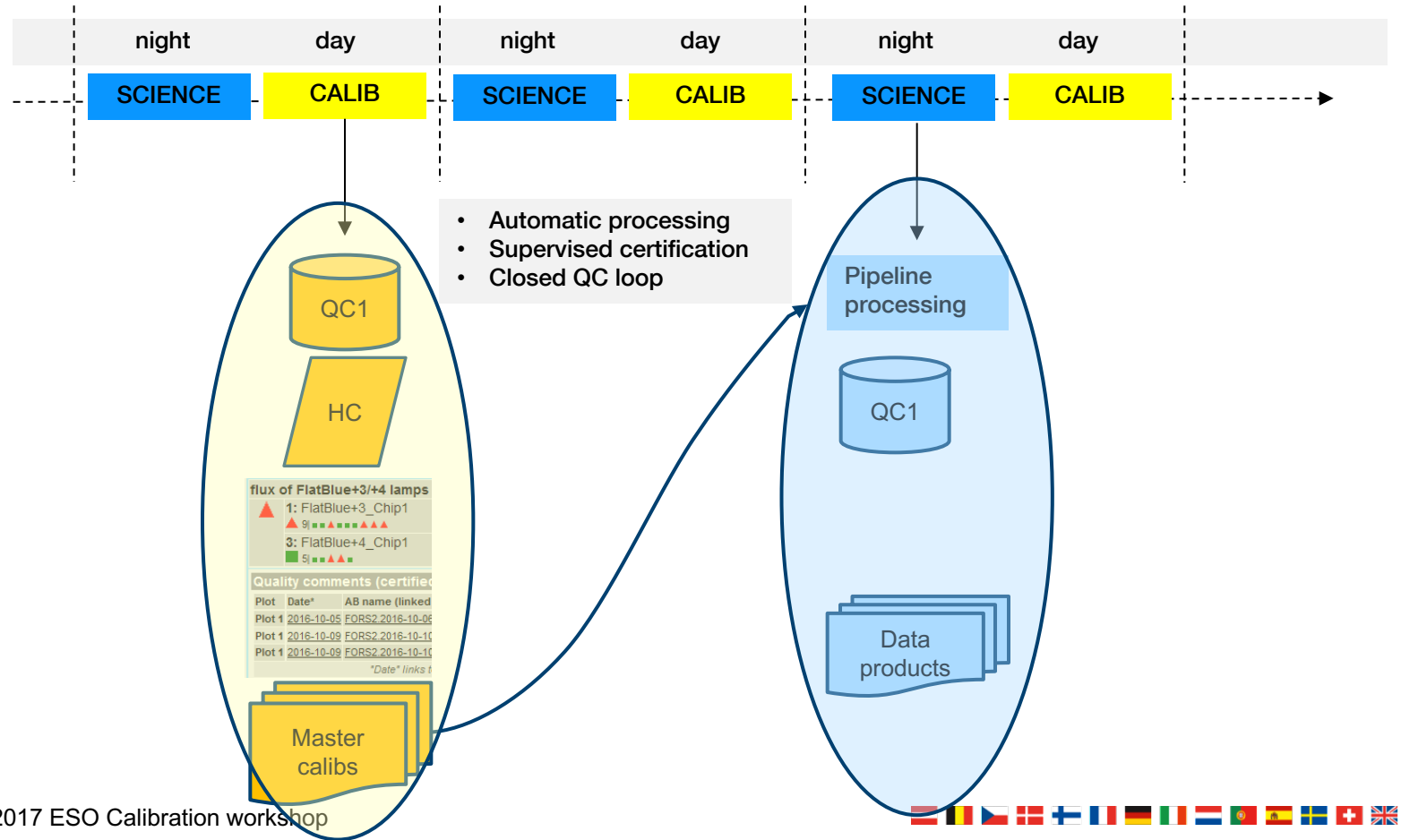
QCG: D. Dobrzycka, W. Hummel, I. Percheron, B. Wolff



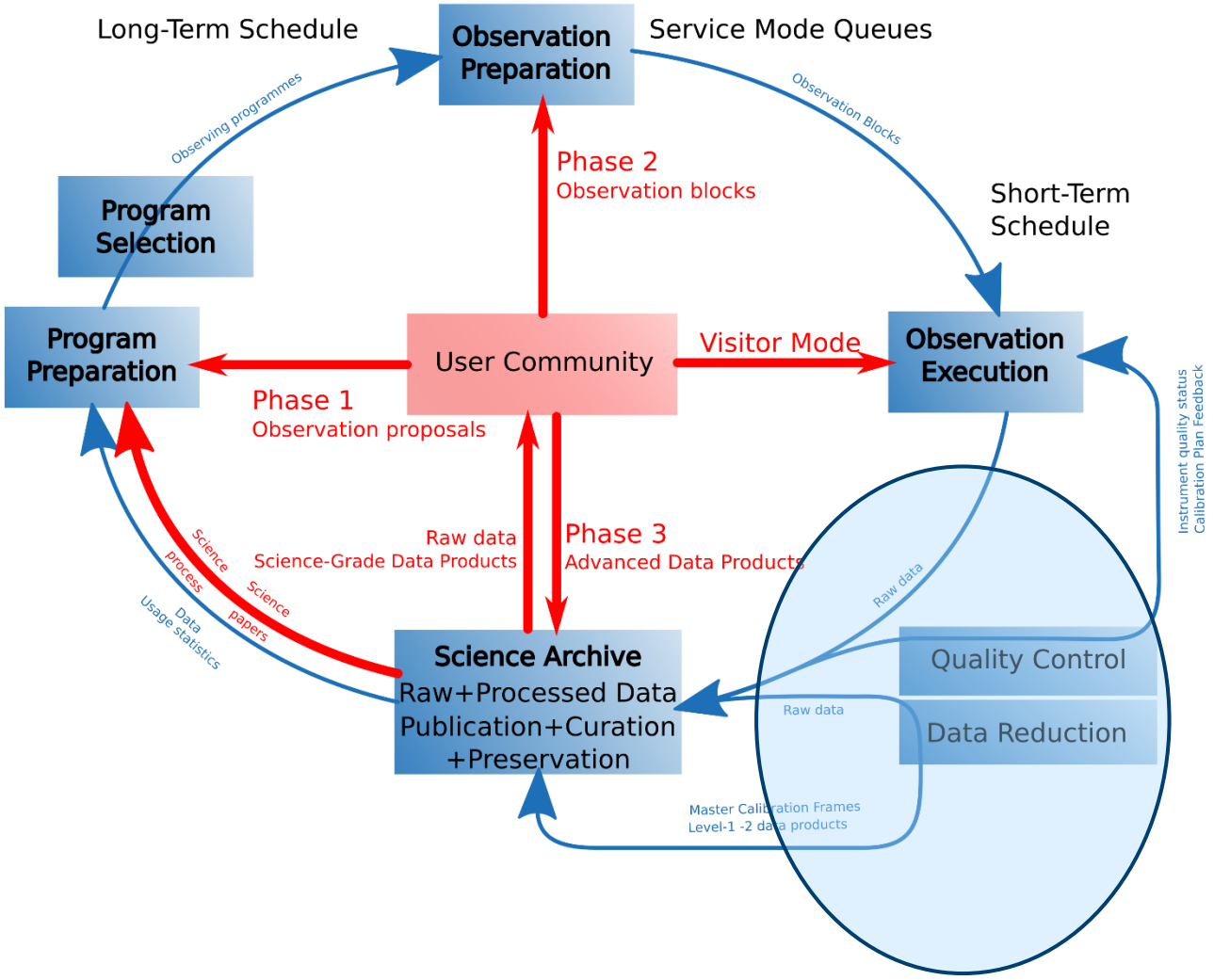
Data products: CALIB + SCIENCE

Two pillars of QC services at ESO:

- CALIB data processing, QC1 parameters, scores, trending
- SCIENCE data processing

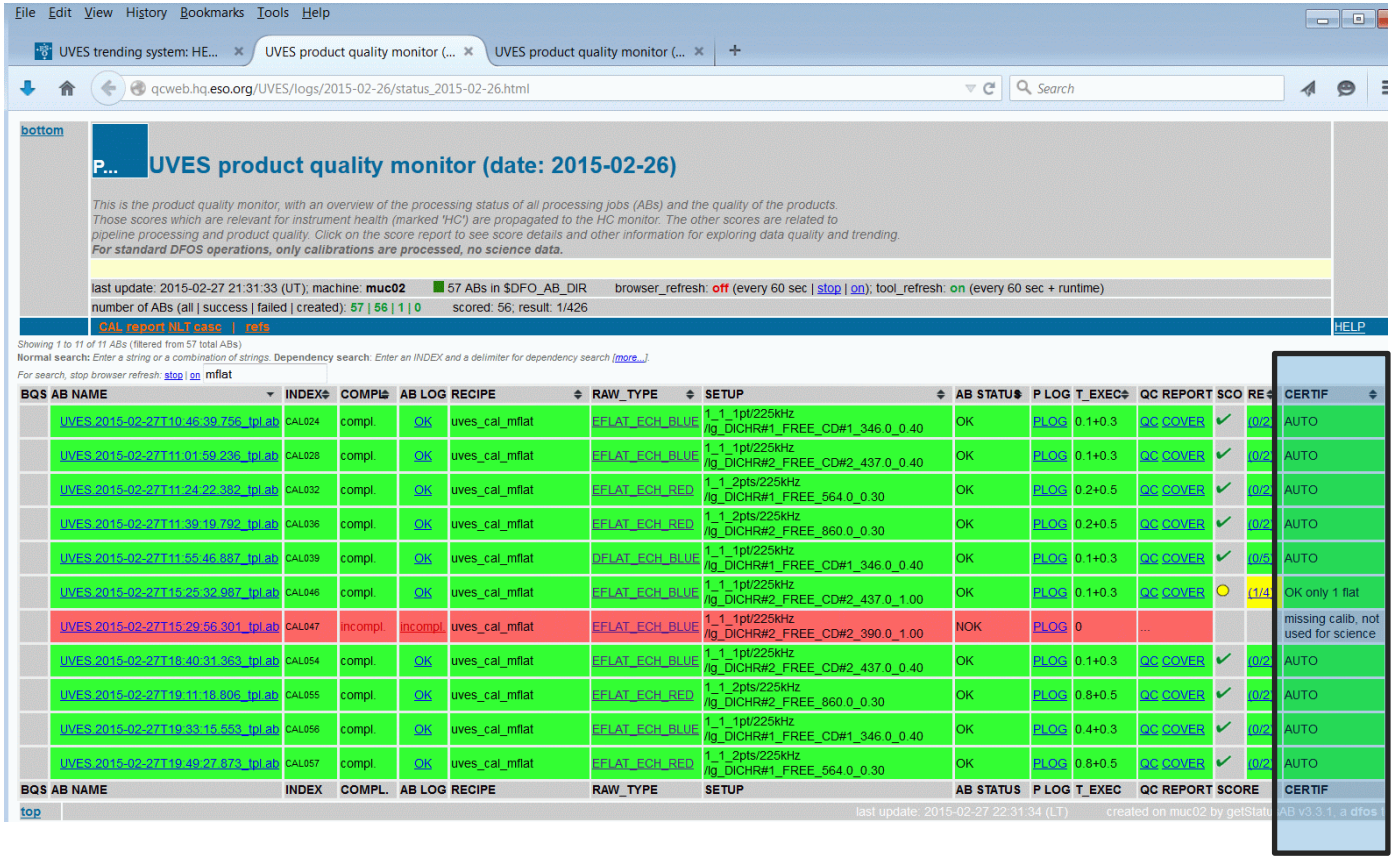


ESO Data flow model



Calibration data processing I

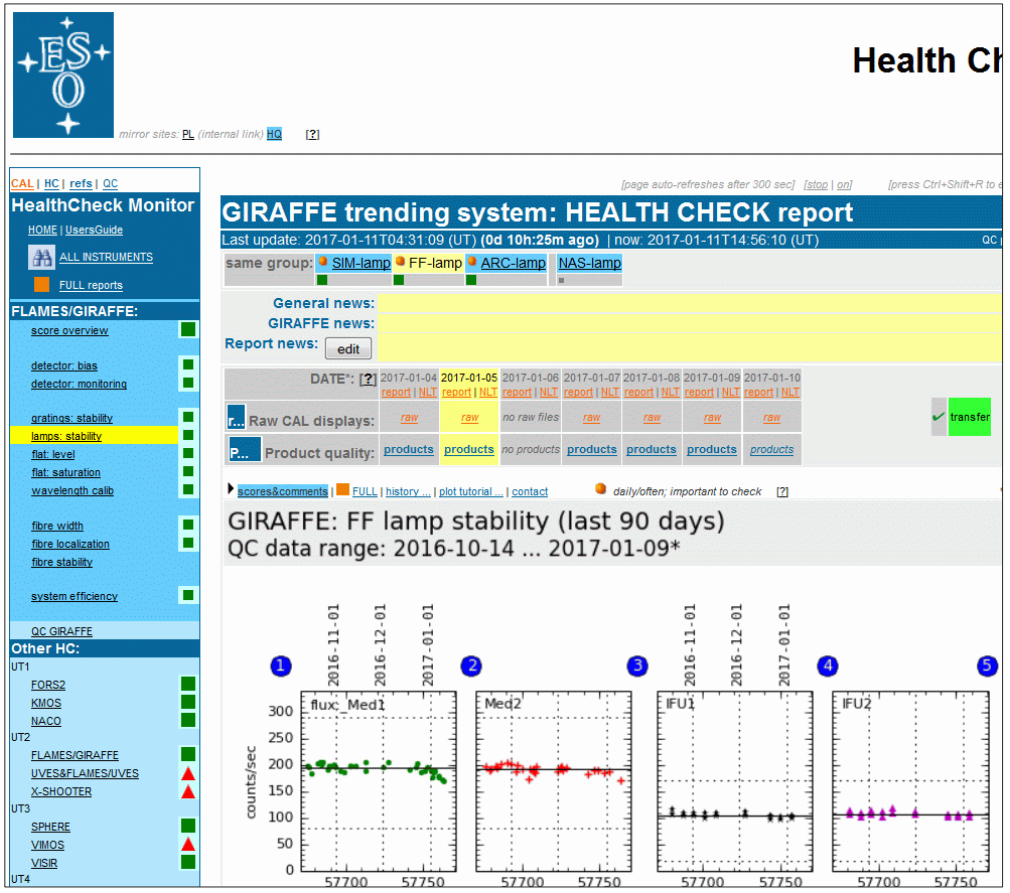
- Automatic pipelines for all 17 operational instruments
- QC parameters extracted & stored in database
- Scores help deciding about quality and certification
- certification



Calibration data processing II

- Instrument trending: Health Check Monitor
- Monitor performance in context
- Compare to specs, reference values

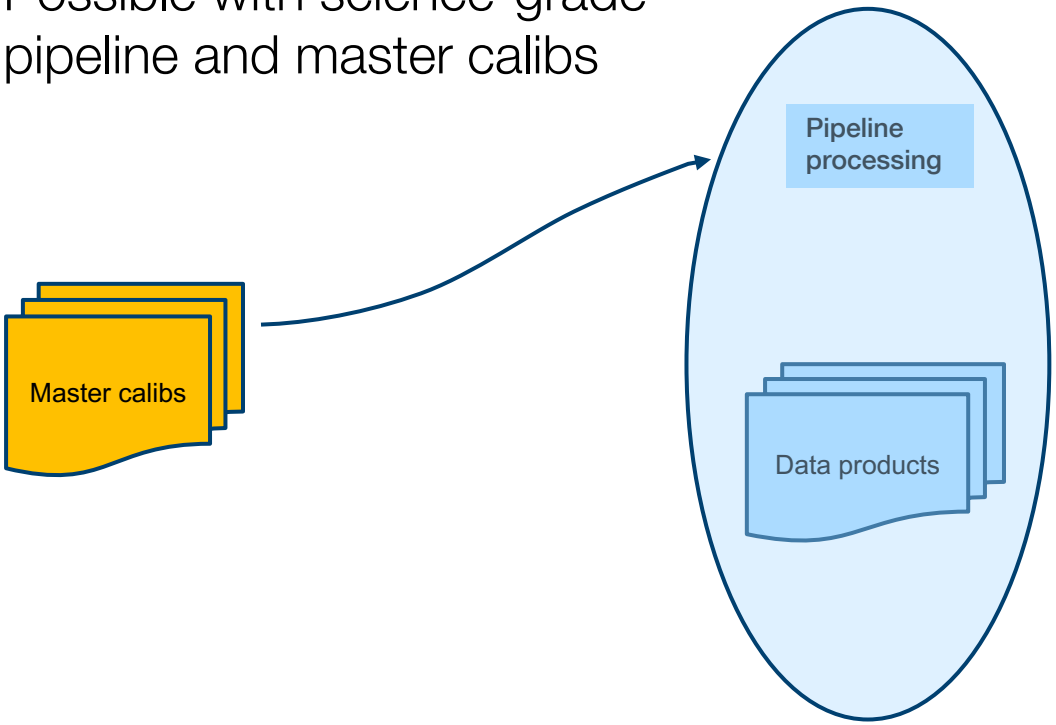
→ Complete knowledge about ins+atm effects stored in master calibration files



Data products: SCIENCE

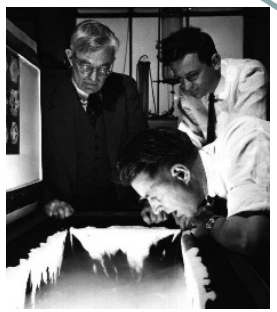
Two pillars of QC services at ESO:

- CALIB data processing, QC1 parameters, scores, trending
- SCIENCE data processing:
  - Possible with science-grade pipeline and master calibs



Traditional data reduction

- "Traditional" way towards data products in ground-based astronomy:
  - Get your raw data, reduce them by yourself, optimized for project
  - Cumbersome, long learning curve, "data tsunami"



1. Raw data

science photons
plus atmosphere
plus instrument

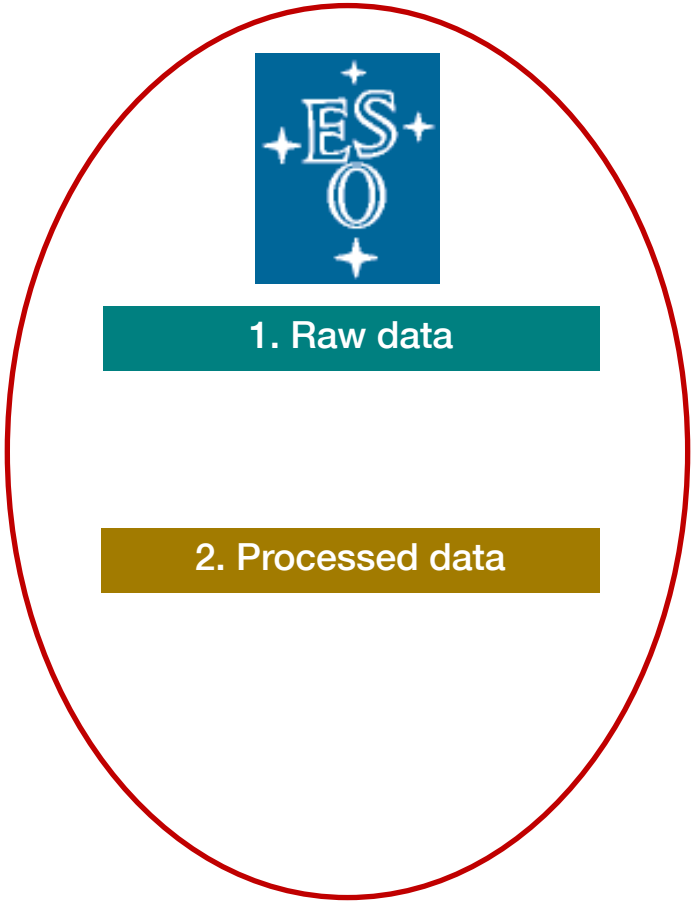
2. Processed data

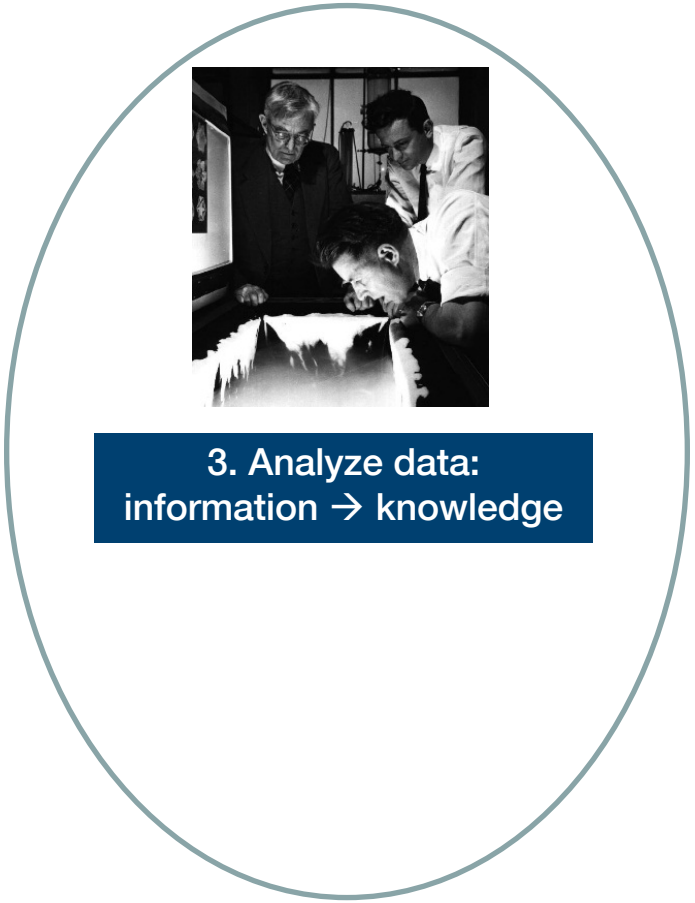
cleaning artefacts
extract signal

3. Analyze data:
information → knowledge

New ways: IDPs

- More efficient:
  - do it centrally
  - offer to whole community (PI, anyone)





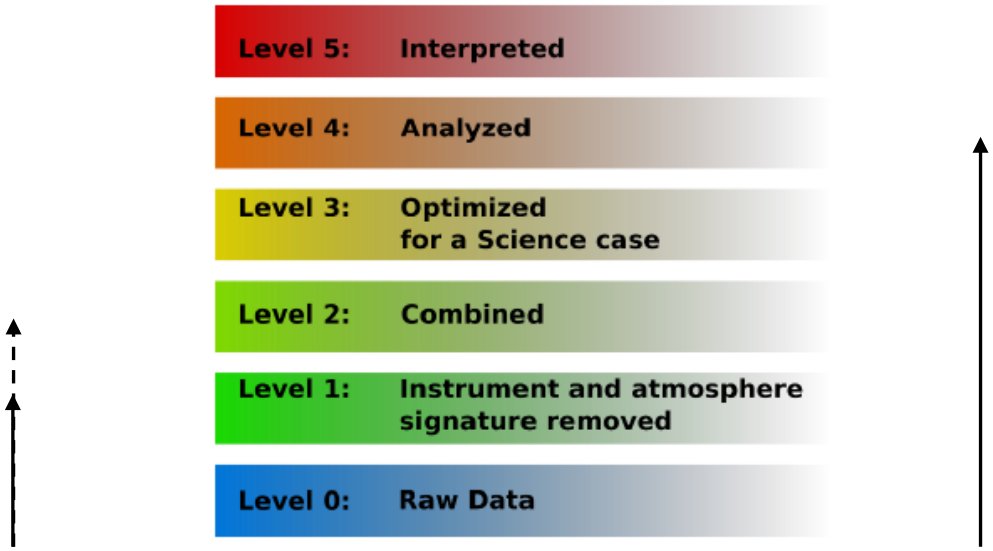
Science data products at ESO

- IDPs – Internal Data Products:
  science data processed in-house at ESO

- Up to science-grade level:
  - Instrumental signature removed, atm. effects removed if possible
  - Physical units (wavelength, flux)
  - Errors
  - Ready for analysis

- Reduction strategy:
  - Standard reduction (reproduceable and understood)
  - General strategy (not fine-tuned to specific case)
  - Reduction efforts done once, centralized, in mass production, with quality control

Science data products at ESO



IDPs:
- processed in-house
- complete by INS
- 100s of programmes per INS
- general-purpose

External data products:
- processed by PIs/data centres
- from surveys or large programmes
- specialized & optimized
- complementary to IDPs

IDP production

- Needed for the IDP process:
  - science-grade pipeline, review/certification
    (→ *Science Data Products Group)*
  - master calibrations in archive (covering all instrument aspects, calibration information complete)
  - QC scheme (QC1 database, scores)
  - preview plots
  - data product standard (for 1D spectra)
    (→ *Archive Science Group)*

Instruments covered

- IDPs:
  - not possible for all VLT instruments (pipelines)
  - since 2013: products from spectroscopic VLT instruments
  - So far 5 projects by QCG:
    - UVES (120,000 spectra);
    - XSHOOTER (45,000 spectra);
    - GIRAFFE (1,200,000 spectra);
    - MUSE (4500 data cubes)
    - coming soon: MUSE_DEEP (200 deep datacubes)
  - also (non-VLT): HARPS, FEROS
  - cover entire history of respective instrument
    (UVES: 16 years; GIRAFFE: 13 yr; XSHOOTER: 6 yr; MUSE: 2+ yr)
  - new data added in monthly batches

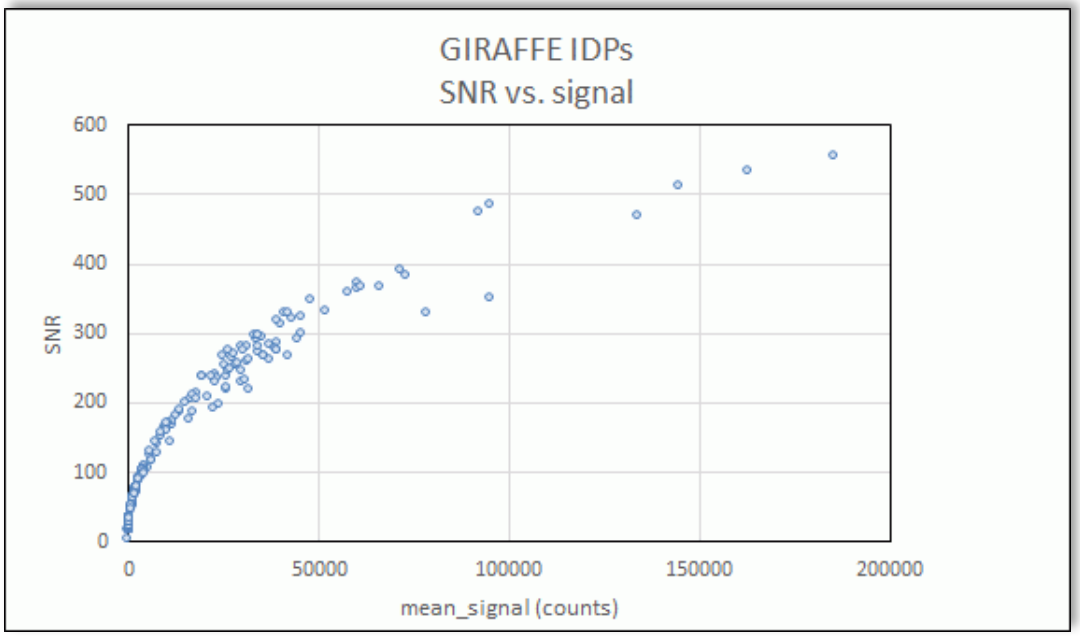
Phoenix: workflow tool

- Important:
  - Individual issues with calibrations can be ignored;
    reviewed at QC time → we build on the historical certification
  - Hence we can re-use the certified master calibrations blindly:
    Important condition for efficient mass-production
  - done with a workflow tool: ‘phoenix’

- Also:
  - need to check association quality (CAL <-> SCI)
  - QC checks for issues with SCI (e.g. saturation)

Process quality control

- Process QC (as opposed to data QC): we check statistically
  - SNR<->counts (photon noise dominates error budget)
  - Association quality (proximity to calibrations)
- Individually: saturation, number of cosmics
- Flags: negative counts, anomalies (like no SKY available)

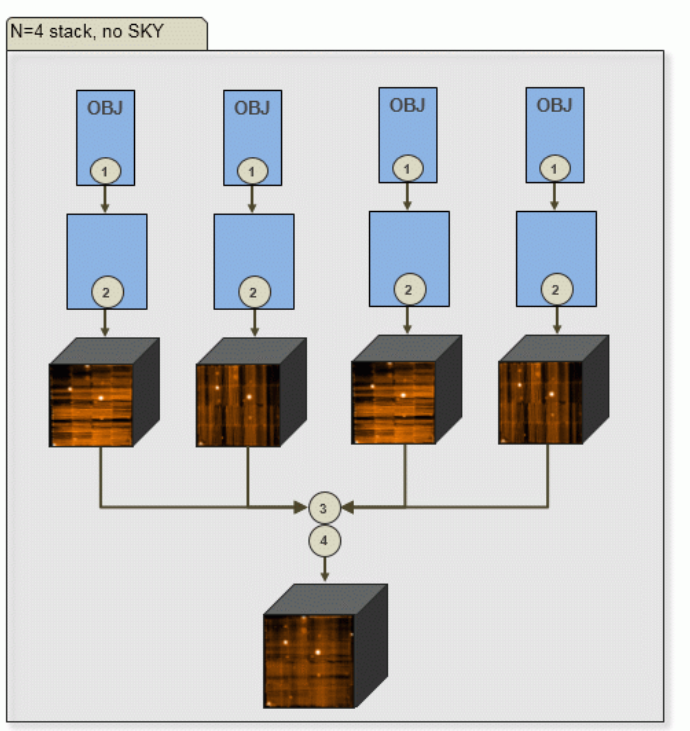


Example: MUSE

- IDP projects #4 and #5: biggest & most complex so far
- MUSE: 24 identical spectrographs (4Kx4K each)
  - Field of view: 1'x1' (WFM)
  - Spectral resolution: R=3000
  - Spatial resolution: 0.2"/pixel (WFM)
- MUSE data
  - Big files (1 GB, 384 mio px): ESO going "big data"
  - 90,000 spectra in one shot
  - Challenging in terms of data volume and complexity
  - Part of community not ready for reducing these data "by hand"

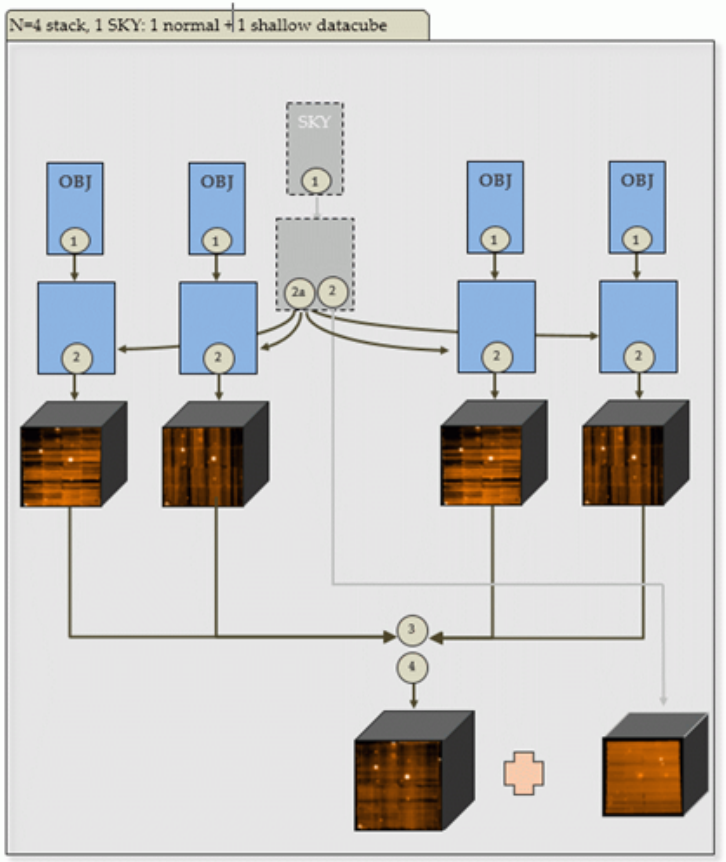
MUSE IDPs

- MUSE science data processing: requires careful orchestration
  - up to 5 pipeline recipes
  - up to 6 steps
- Typical dataset of 4 raw OBJ:
  - needs 24 cores, 100 GB memory
  - takes 2 hrs
  - delivers 1 OBJ datacube ≥ 3 GB size
  - flux-calibrated, sky-subtracted: science-ready (still some SKY residuals)



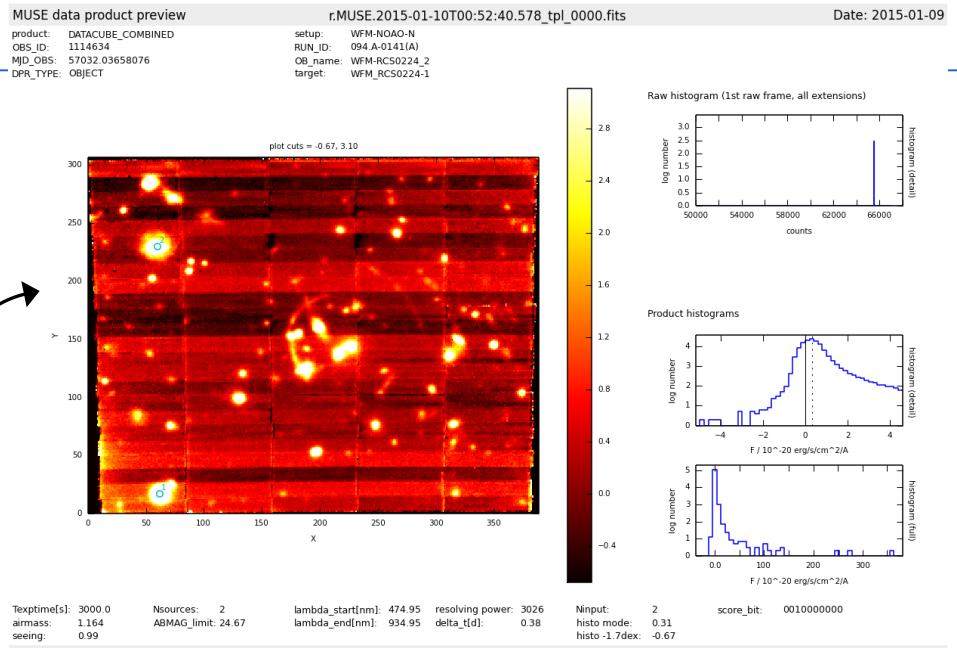
MUSE IDPs

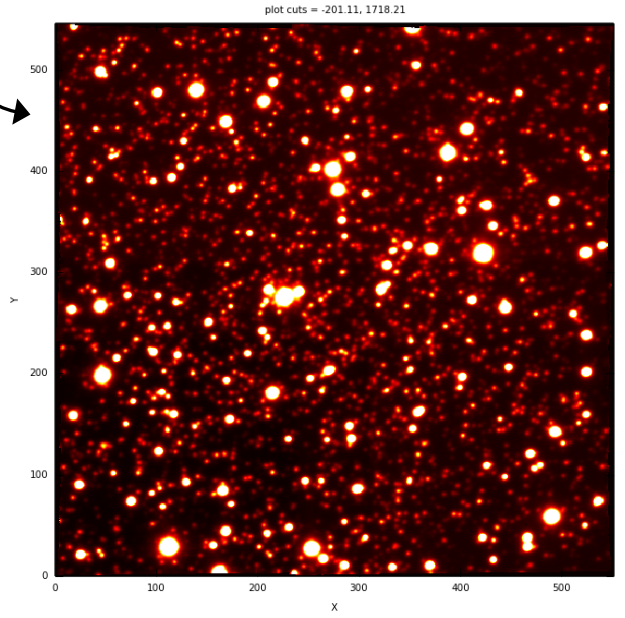
- OBJ with SKY:
  - delivers 1 OBJ datacube
  - plus 1 “shallow” datacube from SKY
- Archived, downloadable by any user
  (proprietary rights respected)



MUSE IDPs

- Quality control:
  - QC parameters on single and combined products
  - scores
  - QC reports
  - certification
  - visual inspection





Score report HELP

```
MUSE.2014-11-07T07:42:57.317_pst.ab
RAW_TYPE:       PIXTABLE_OBJECT
TPL_ID:         MUSE_wfm-noao_obs_genericoffset
TPL_NAME:               Observation in WFM with user defined offsets

setup:          WFM-NOAO-N
time range:     2014-05-11 ... 2015-05-06
```

AB | ALOG | PLOG | QC1 plotter | factsheet

Point your mouse on QC1 parameter name for short documentation.
Grey squares: unscored parameter.

NAXIS1	NAXIS2	NAXIS3	num_sat	Ishift_max	qc_deltatime	TPL_NEXP
■	■	■	■	■	■	■

explore >>

HC plot(s): non configured

QC report(s):

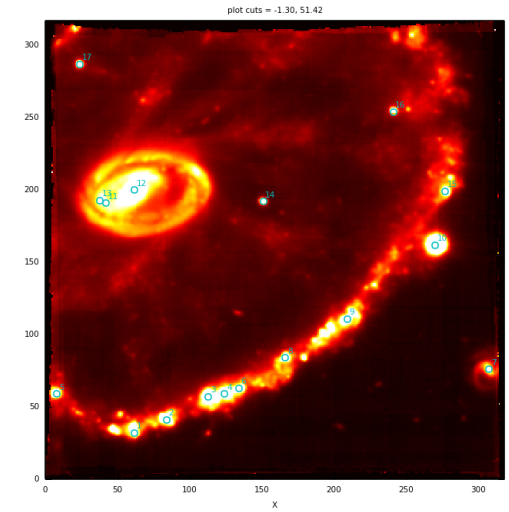
Score data: details ...

score result: 1/7 **best:** 0/7

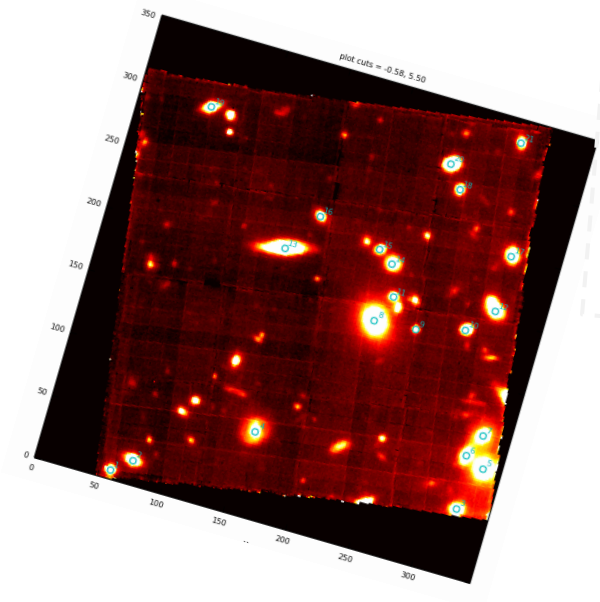
powered by QC [scoreQC v2.1.3]

MUSE

- Processed + ingested:
  - Science Verif. #1 and #2
  - Start of ops (2014-09) until now
- IDPs at the OB level (single pointings), like for all other IDPs



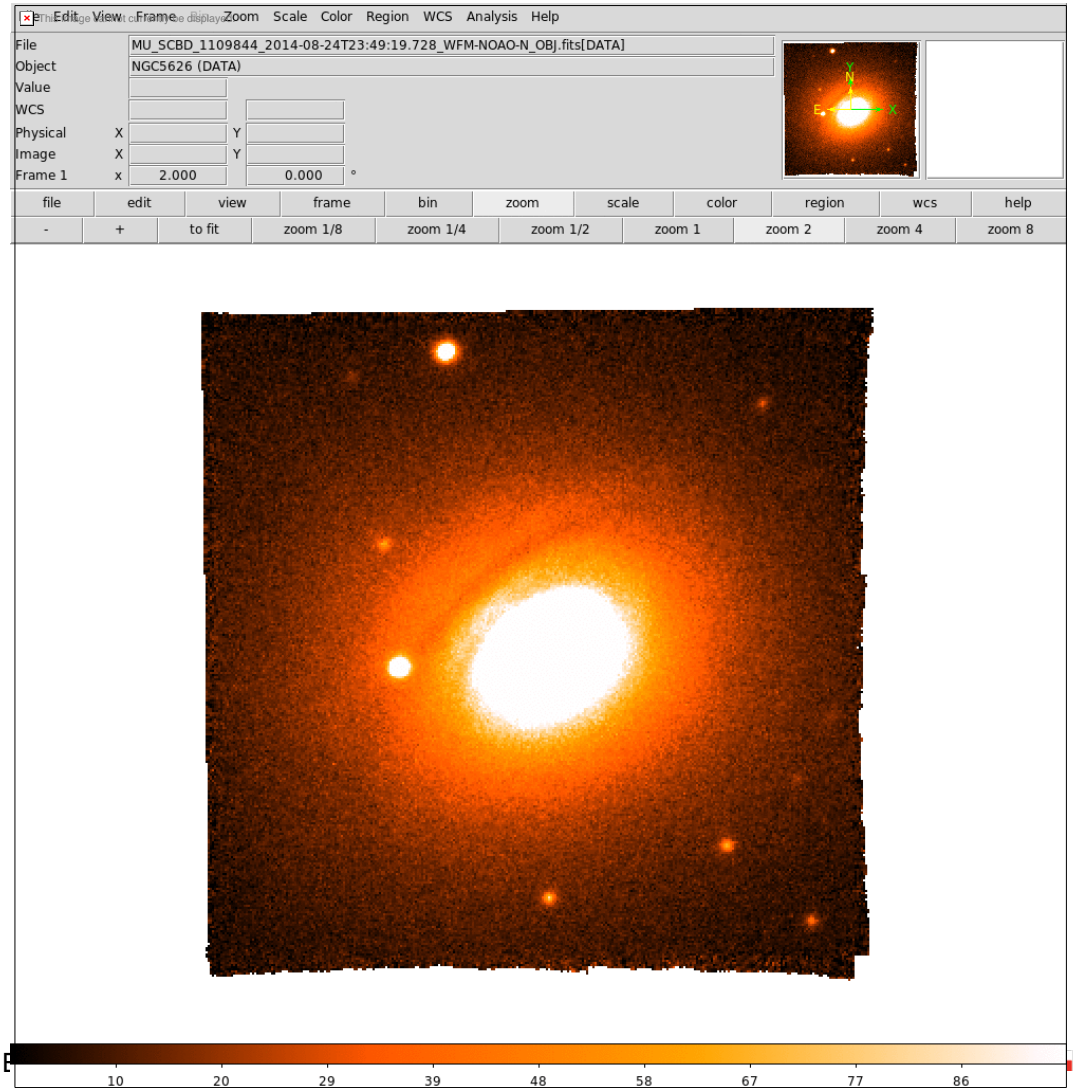
Cartwheel galaxy.
35 min, N=4



Abell 2744.
60 min, N=2

MUSE

- Data quality and content: amazing …

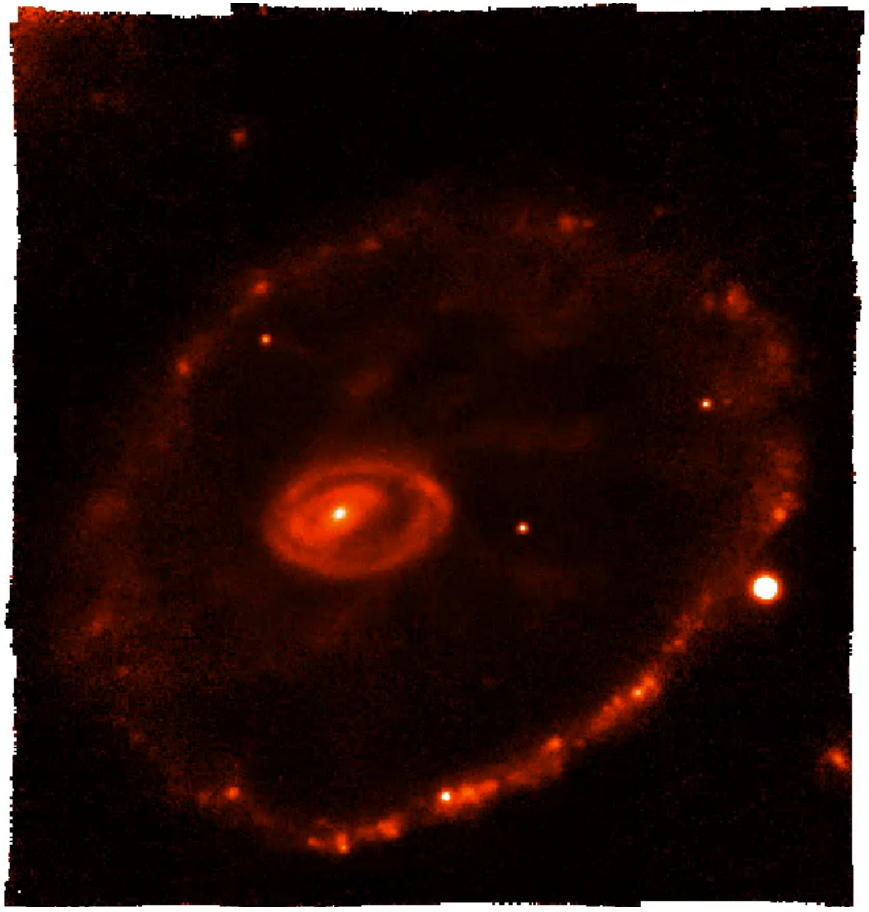


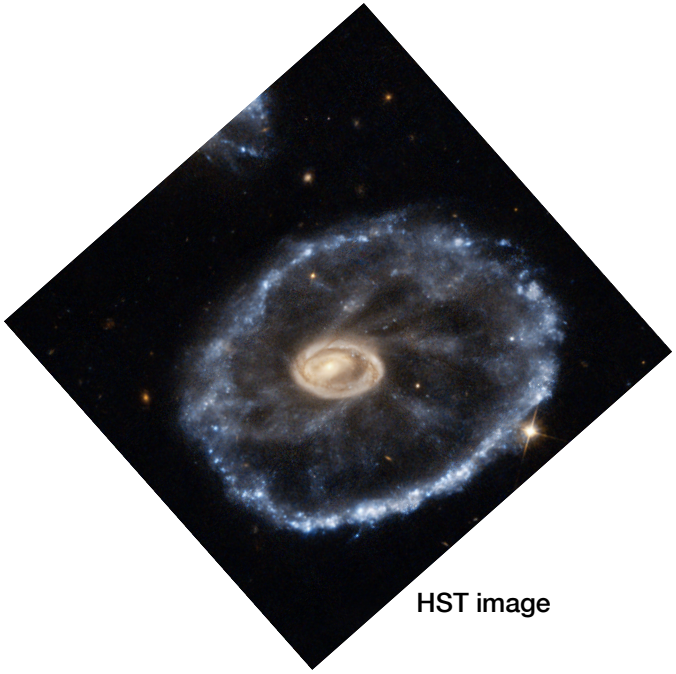
Science Verification run 60.A-9337A:

Velocity maps of emission-line regions in NGC 5626;

17 wavelength bins shown around 6740 A

MUSE





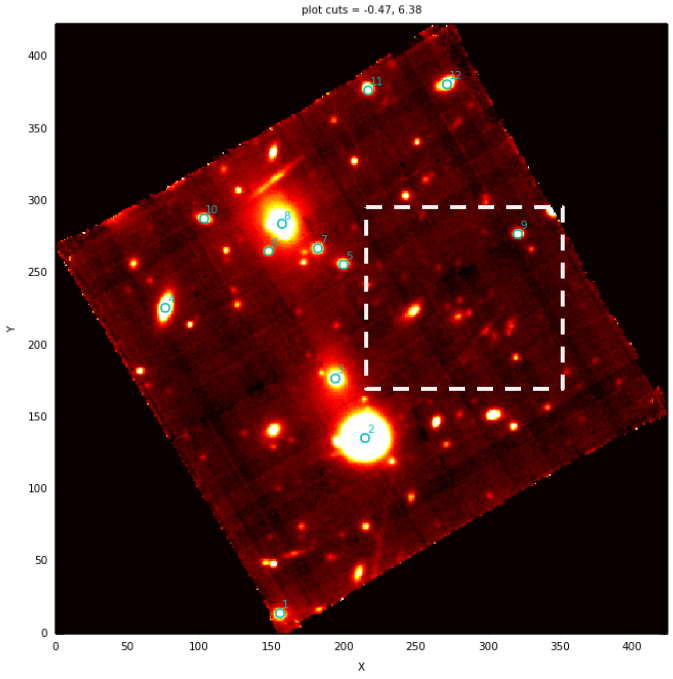
HST image

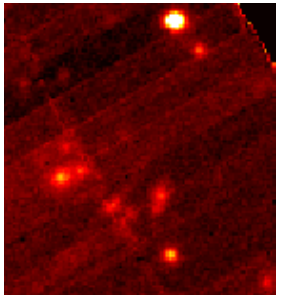
SV run 60.A-9333A

Emission-line regions in
Cartwheel galaxy
(6730-6800 A; 2 A/sec)

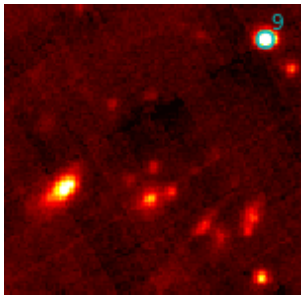
MUSE: going deep

- What we have: OB based datacubes
  - Typical stack of N=2-4, covering ~ 1 hour (operational limit)



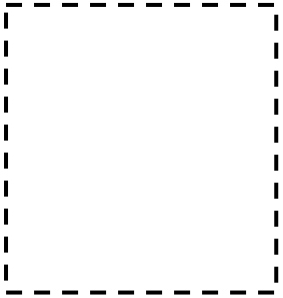


Single exposure, 10 min



One OB, 40 min, 4 exposures

MUSE OB-based IDPs



15 OBs, 11.2 hrs, 58 exposures

MUSE_DEEP IDPs

- for many programmes: not good enough
  - > 50% of all MUSE runs: want to go DEEP!
  - OBs designed with 10 hr or more, in different nights

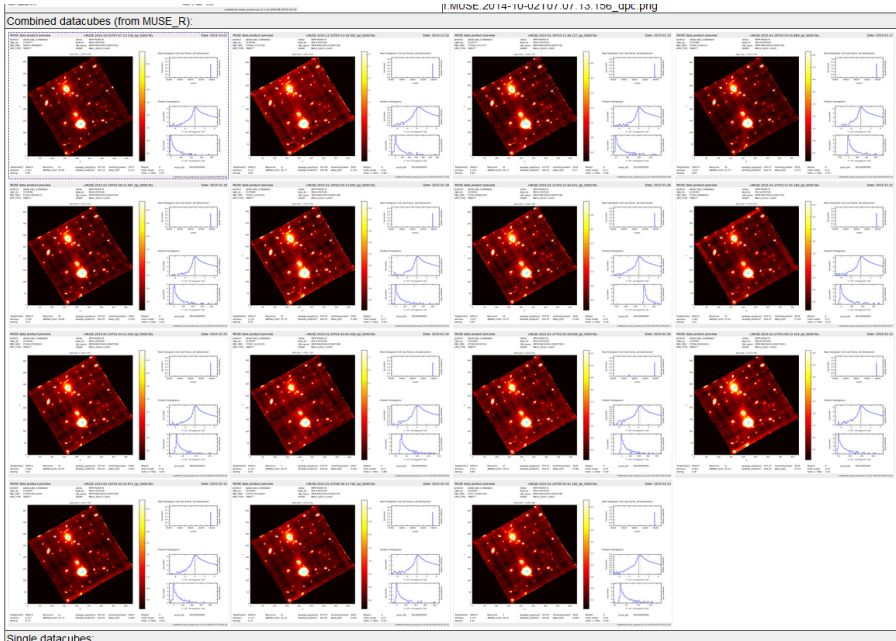
MUSE_DEEP

- going DEEP:
  - multiple visits of same target: multiple OBs
  - some spread across multiple run_IDs
  - challenge to find them

Abstract: *"... we propose a pilot study to spectrally image ... the HFF Cluster MACSJ0416.1-2403 with MUSE. The resulting datacube will yield robust redshifts for strong emission-line objects out to $z=6.6$, enable emission-line diagnostic, kinematics, and morphology studies ..."*

- stable calibration plan at its best!
- challenge to process them
  - 2 stages: OB processing
  - combine all products from all OBs to final DEEP product

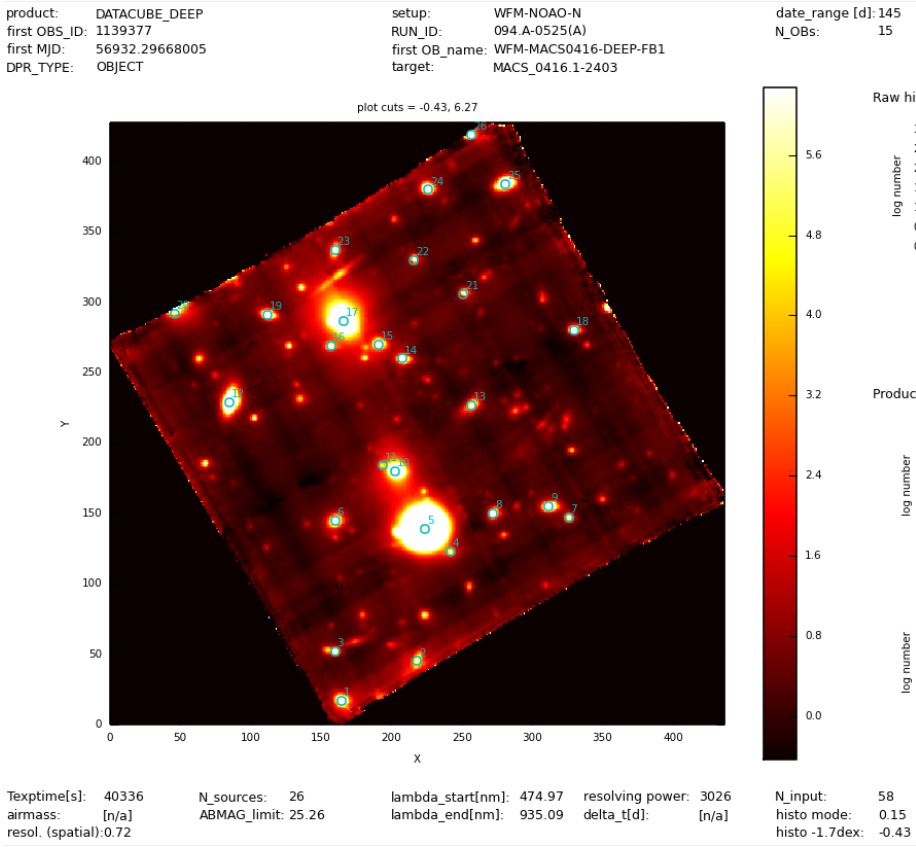
15 OBs spread across a semester; currently 15 individual datacubes; MUSE_DEEP: 1 deep cube

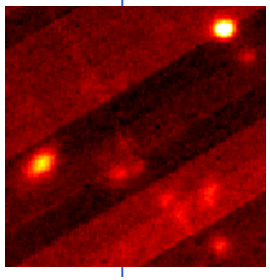


- full scientific signal available ONLY in deep cubes

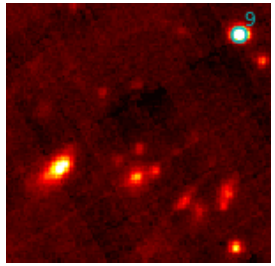
MUSE_DEEP

- Example: what DEEP can mean
  - 15 OBs combined
  - 58 input files
  - 145 days apart
  - 11.2 hr exposure time
  - 15.8 hr compute time on 48-core blade, 1 TB memory
  - Current high-end of IDPs
- 200 such deep cubes
  - from SV+P94+P95, to be available soon

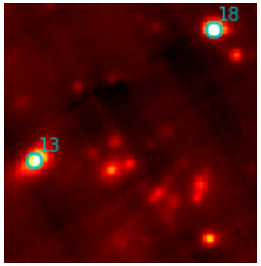




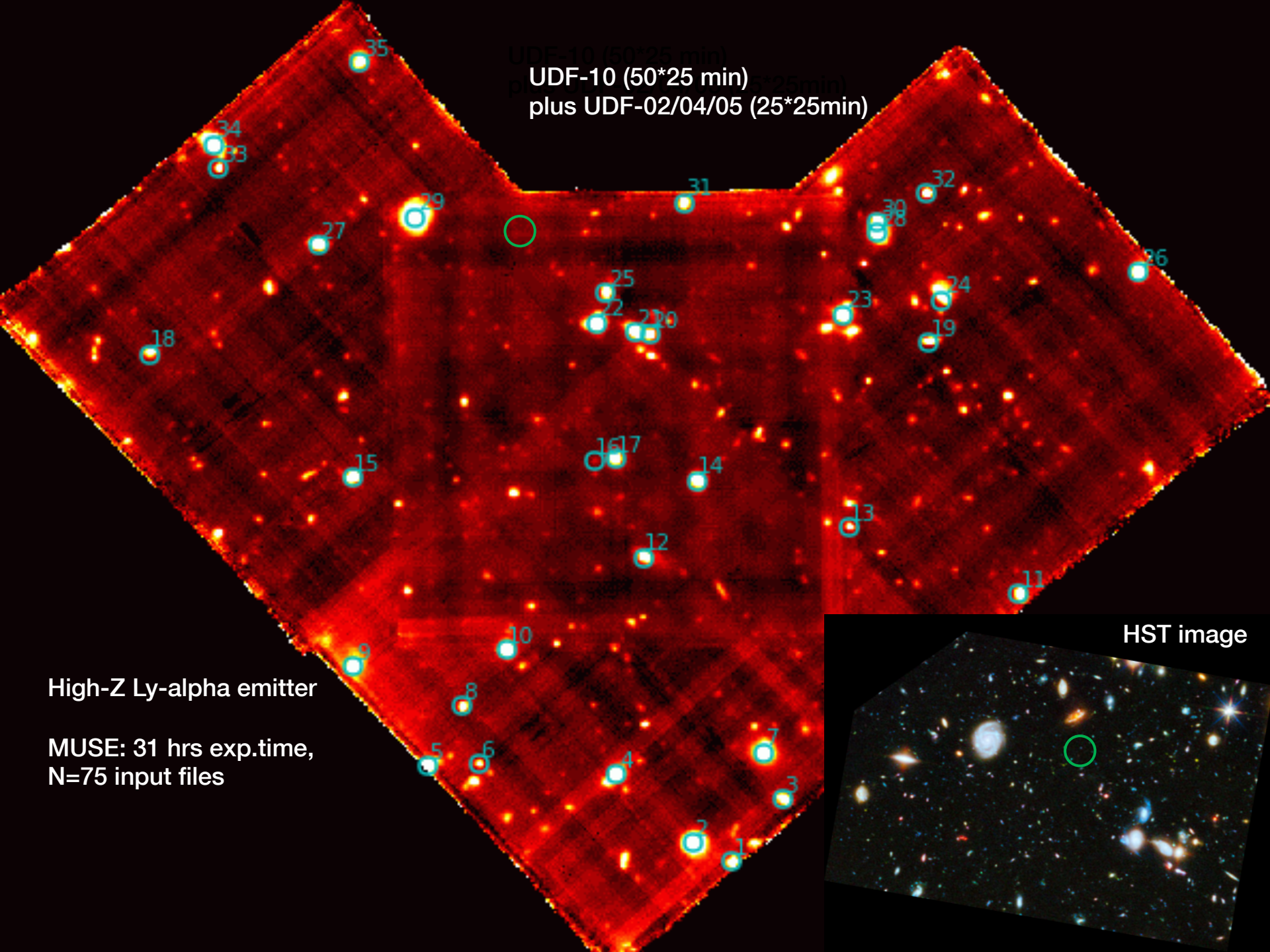
Single, 10 min



One OB, 40 min, 4 exposures



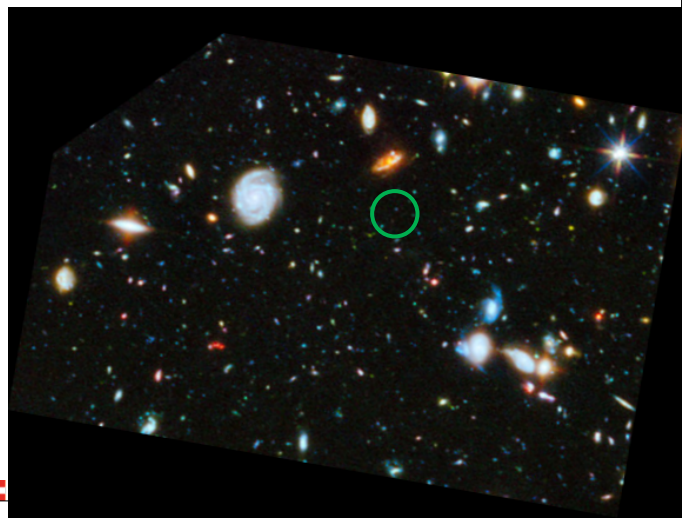
15 OBs, 11.2 hrs, 58 exposures


UDF-10 (50*25 min)
plus UDF-02/04/05 (25*25min)
HST image
High-Z Ly-alpha emitter
MUSE: 31 hrs exp.time,
N=75 input files

UDF-10 (50*25 min)
plus UDF-02/04/05 (25*25min)

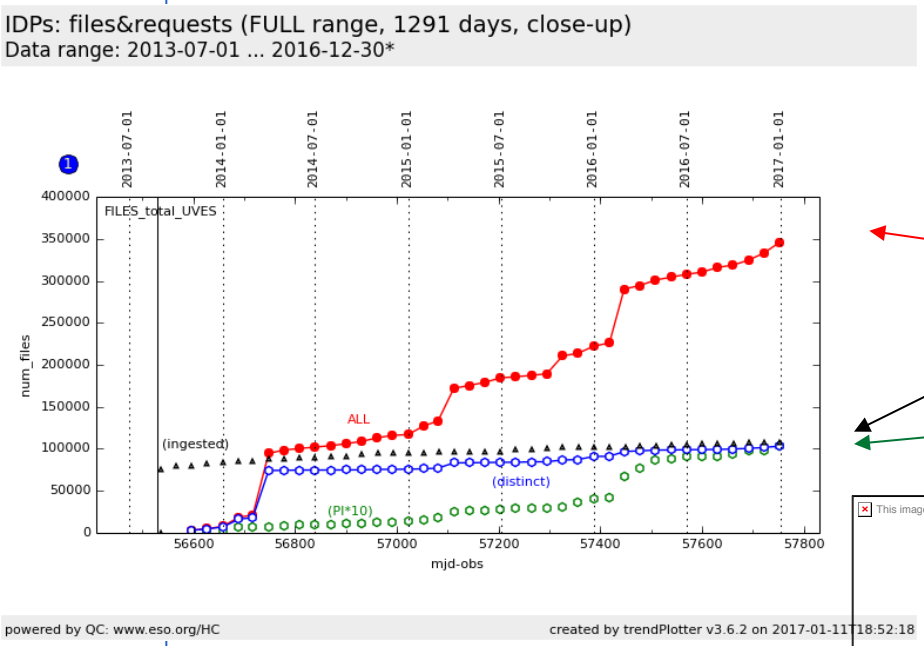
High-Z Ly-alpha emitter

Wavelength scan 669nm+



IDPs: downloads

- IDPs are used: most popular data products at ESO

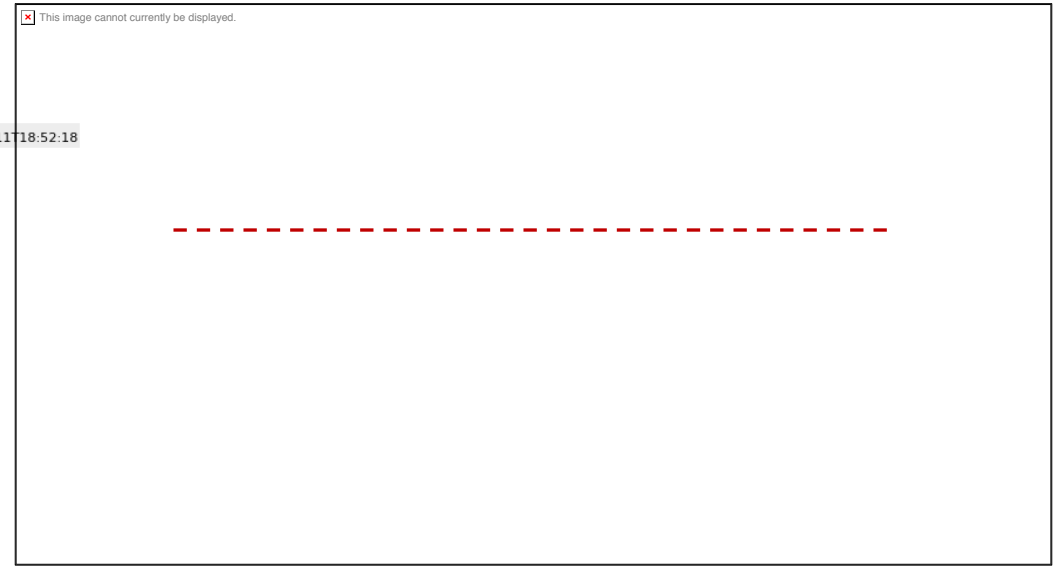


UVES IDP downloads

Total downloads (~350,000)

available IDPs (~105,000)

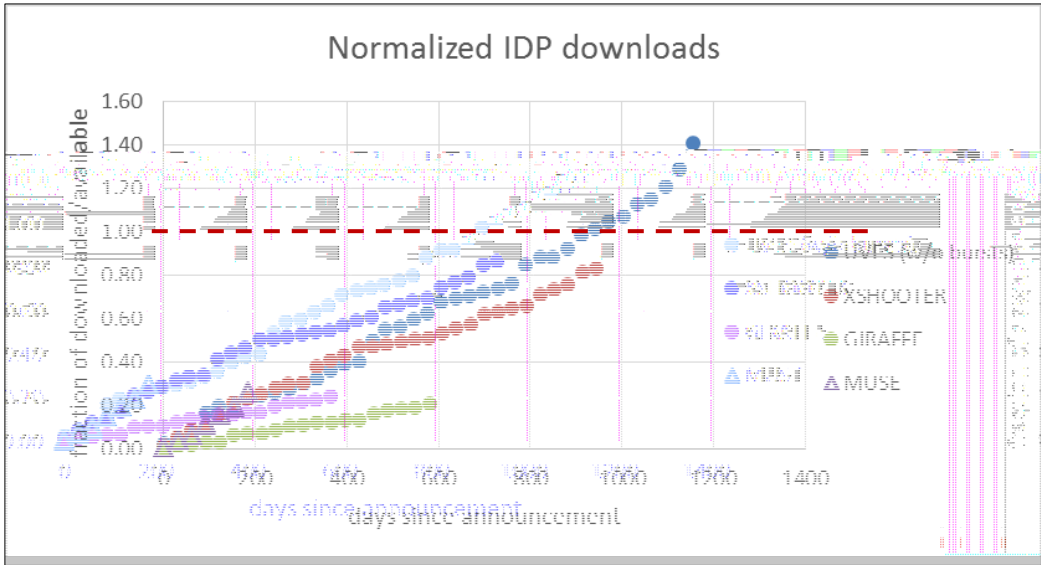
PI downloads *10 (~10,000)



IDP downloads normalized: fractional downloads vs. time since release

Additional value

- "magic" threshold of 1:
  - all available IDPs downloaded once (statistically)
  - usage *in addition* to traditional analysis by PI
  - looked at perhaps with other eyes or in broader context
  - may produce publications *in addition* to PI
- proves value of archive
- proves value of careful calibration and data processing efforts



Thanks!