

Ensuring the scientific quality of science products from several data providers

Laura Mascetti

M. Arnaboldi, N. Delmotte, A. Micol, J. Retzlaff

ESO Archive Science Group

Terma GmbH

Phase 3

Channel to submit and publish science-ready data for

- Distribution
- Long-term preservation

<http://www.eso.org/sci/observing/phase3>

It includes

- Data preparation
- Validation
- Ingestion
- publication

Phase 3

17 October 2017

- LIGO-VIRGO Gravitational Wave: Public Release of ESO Follow-up Observations [\[Read More\]](#)

29 September 2017

- First data release of spectra and catalogue resulting from the deep VIMOS survey of the CANDELS UDS and CDFS fields (VANDELS) [\[Read More\]](#)

21 September 2017

- Third data release of images, source lists and aperture matched multi-band catalogue resulting from the VISTA Kilo-degree Infrared Galaxy Survey (VIKING) [\[Read More\]](#)

[\[More News...\]](#)

- Pipeline-processed products (ESO)
- Public Surveys: VISTA (6+7), VST (3), SPS (4)
- Large Programmes
- Others

- Telescopes
 - VLT, VLTi, VISTA, VST, ESO-3.6,
MPG/ESO-2.2, NTT, APEX, GTC

- Types of data (single and stacked)
 - Images (monolithic, MEF)
 - Spectra
 - IFS (3D cubes)
 - Source tables
 - Catalogues
 - Submm flux maps
 - Interferometry

- Ensures integration into the Archive
- Defines data format, structure and metadata
- Defines how to encode
 - level of calibration
 - scientific quality
 - original files
 - ancillary data
- Growing document
- Allows access from a unique interface



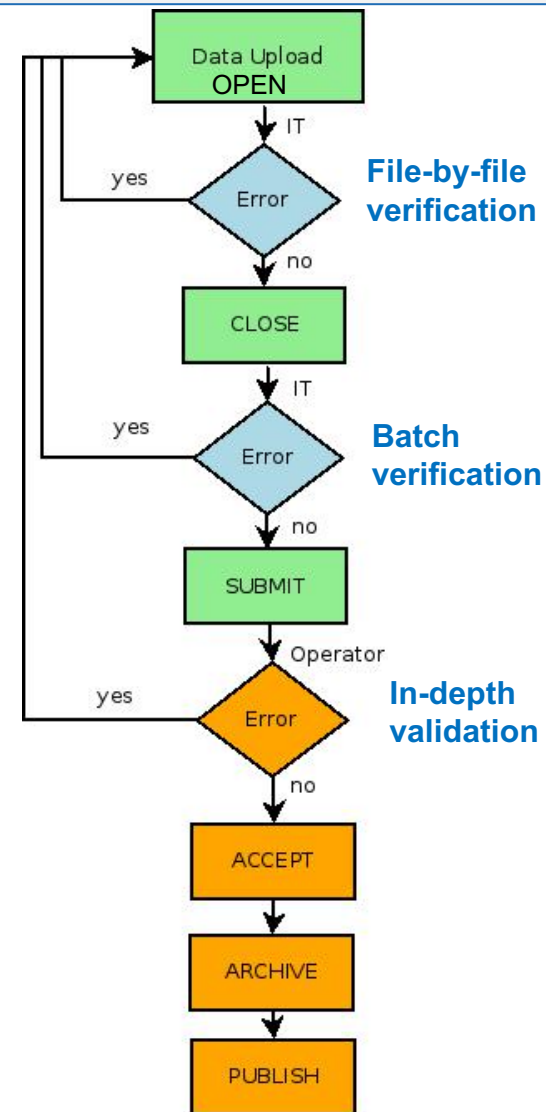
Audit process needed → **Data Validation**



- Support in generating compliant products
 - Conversion tools
 - Maintain up to date documentation (FAQs, templates)
 - Direct user support
 - Testing data sub-sets

- ESO provides Pipelines that generate Phase3 compliant products

- Governs the P3 workflow
- Written in Groovy
- Runs as a demon
- Checks the batches status
- Performs several operations





Data flow - Release Manager



European
Southern
Observatory

ESO — Reaching New Heights in Astronomy



[ESO Home](#)

[User Portal](#)

[Contact](#)

[Site Map](#)

Go

Laura Mascetti (Admin) | [Logout](#)

[Science User Information](#) / [ESO User Portal](#) / [Phase 3 Release Manager](#)

Phase 3 Release Manager

[New data collection](#)

[New batch](#)

[Report data problem](#)

[Change survey manager](#)

[Include stream batches](#)

[Time to publication report](#)

Regarding the data content of Phase 3 batch no. 4967 some issues have been found. Please fix the reported problems and replace the respective data within your ftp directory ftp://phase3ftp.eso.org/VPHASplus/batch_4967 Press CLOSE when you are done.

Batch	Update info	Data collection	Data release	Data collection type	Files	Volume	Date	State	
		All	All	All				All	
+ 5157	N	GW170817	DR1	Normal programme	18	31.76M	2017-10-16 14:00:28	PUBLISHED	
+ 5151	i	TESTSUB	Any kind of test	Normal programme	78 58	0.66G	2017-10-11 07:54:08	OPEN	
+ 4967	i	VPHASplus	DR3.2	ESO public survey	1180	0.53T	2017-09-14 12:38:06	OPEN	Close ⚙
+ 4578	i	185.A-0745	IN PROGRESS	Large programme	0	0.00B	2017-08-14 06:20:23	OPEN	Close ⚙
+ 3955	i	TESTAMI	IN PROGRESS	Normal programme	3	126.56K	2017-08-11 10:25:25	OPEN	Close ⚙
+ 3943	i	TESTSUB	Any kind of test	Normal programme	0	0.00B	2017-08-14 10:05:35	OPEN	Close ⚙
+ 3319	N	AMAZE	DR1	Large programme	52	6.73M	2017-08-08 08:48:20	PUBLISHED	- ⚙
+ 3314	i	TESTSUB	Any kind of test	Normal programme	78 58	0.66G	2017-10-10 14:29:05	OPEN	- ⚙

[Test your data](#)

[Content summary](#)

[Data documentation](#)

[Validation report](#)

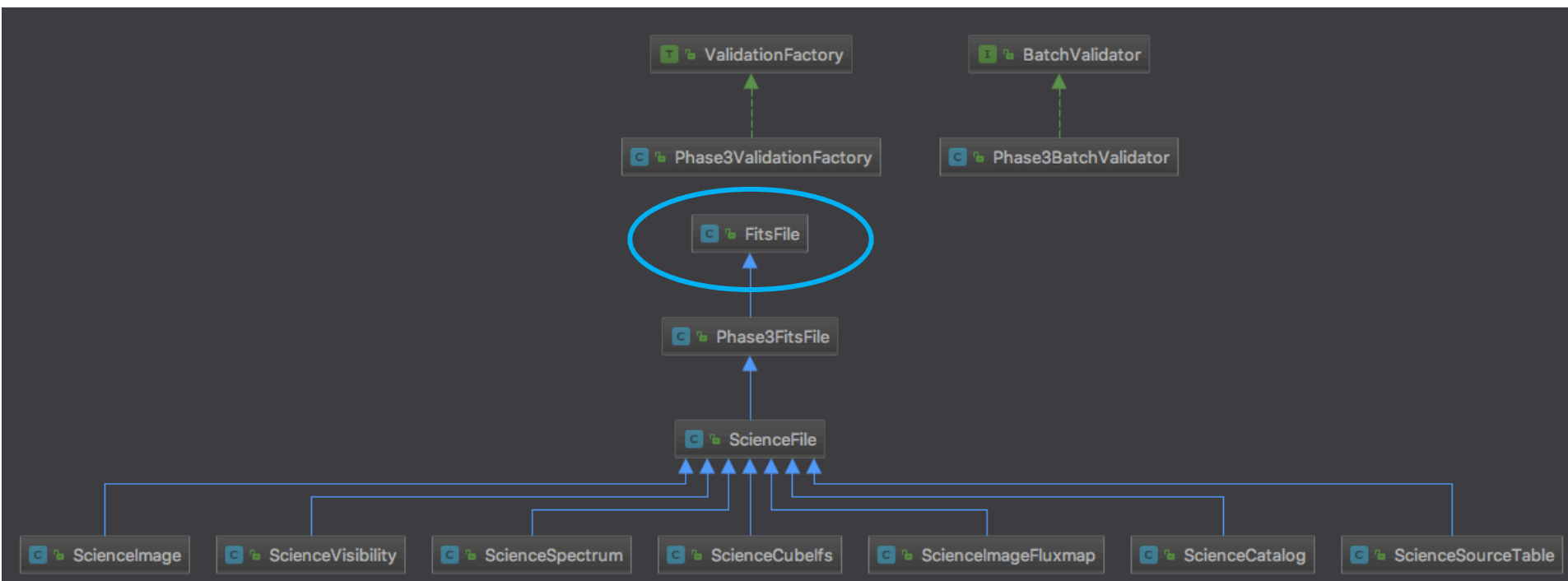
[Edit](#)



■ File-by-file verification

- Triggered upon files upload
- It checks:
 - *Fitsverify* / MD5 hash for non-FITS files
 - Mandatory keywords/keywords not allowed
 - Keyword type/values
 - Cross-header checks
 - Checks on the data, consistency data-metadata
 - File structure
 - Mandatory associations

■ Groovy hierarchical validation classes



■ Batch verification

- Triggered by CLOSING the batch
- Verifies the batch structure
- It checks:
 - PROVENANCE



- Catalogue tiles structure

■ In-depth content validation

Checks that cannot be performed automatically:

- Statistics (plausibility, outliers, null values)
- Checks depending on the type (UCDs)
- Spot checks
- Documentation review
 - Consistency with the data
 - Completeness
 - Clearness

Detailed validation report sent to the provider:

- Request to fix issues → batch re-OPENED
- Acceptance → batch ACCEPTED

Validation

- Ensures the science value of the data
- Guarantees quality of services offered
- Enables discoverability
- Enables scientific exploitation
 - ➡ legacy value
- Costs resources
 - ➡ minimize data curation

Thank you!

- Provide as much feedback as possible to the users at early stage
- Consolidate batch validation
- Validation report embedded in the Release Manager