Quality assessment of environments and procedures

1h discussion session

contacts: K. Rygl, E. Liuzzo, M. Massardi

aeneas

Advanced European Network of E-infrastructures
for Astronomy with the SKA

Why do we need QA?

- verification of the instrument and data analysis procedures (Science Data Processor) sanity

- delivering science-ready, consistent data products to the PI and the community (archive)

- legacy value of archival data

What will the QA procedures be?

- QA done on all primary products in the archive

- QA done on all (advanced) products in the archive?

  - more straightforward: moment maps, light curves

  - problematic: source catalogs

- Require consolidated QA procedures per product

- How could these be best implemented?

  - separate & a-posteriori or in the same step as the product creation?

  - QA done only on the products (images,tables) or also on the ms (visibility analysis)?

Who is doing the assessment

- done by professionals (analysts) based on generated diagnostic plots/graphs/tables?

- done by AI (SDP, primary data products?) which just sends reports when failed?

- QA0: tune SDP pipelines on a small amount of data
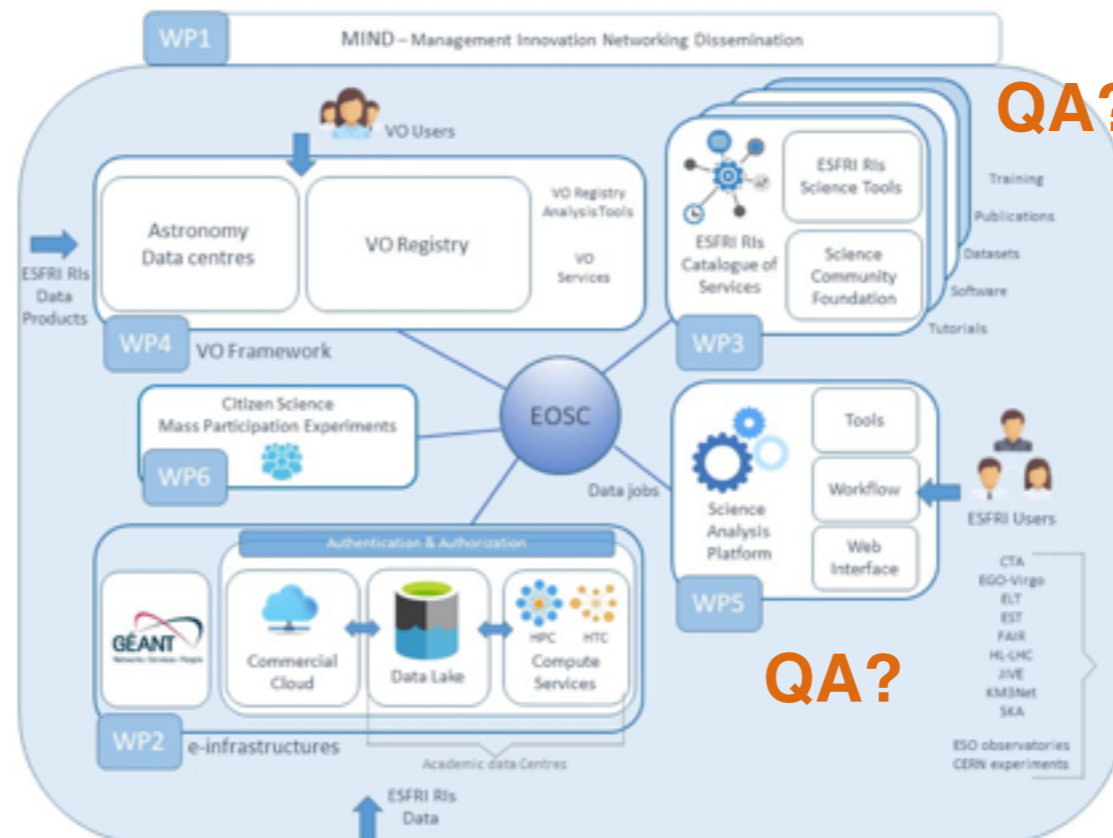
Traceable assessment

- Assessment should be accessible/useful to the community given a "QA manual" which implies QA report in archive

- QA report: well chosen plots/graphs/tables, clear definitions of quantities and units

Extra slides

T3.6   Validation, Verification & Proof of concept activities utilizing SKA pathfinder and pre-cursor facilities

Table 3: Summary table of working environments used within WP3.

Env.	Name	Processor	Memory	Cores	Represents
1	MacBook Pro	3.5 GHz Intel Core i7	16 GB DDR3	4	Basic User
2	Linux Box	Intel Xeon E5-2640 v4	256 GB DDR4	40(*)	Advanced User / HPC
3	GridPP1		16 GB	8	Grid (Standard PP)
4	GridPP2	Tesla V100 GPU	16 GB HBM2	640	Grid (Accelerated)
5	SurfSARA	Intel Xeon Gold 6148	32 GB	4	Grid (Fat)

* 2×10 hyper-threaded

Table 2: Summary table of processing Use Cases used within WP3.

No.	Name	Input Data	SWGs
1	Calibration & Imaging	Calibrated Visibilities	1, 3
2	Pulsar Re-folding	Pulsar Candidates	11
3	Rotation Measure Synthesis	Image Cube [4]	5
4	Object Detection and Classification	Image Cube [1]	1, 3, 4, 5, 6, 7, 9, 10
5	Automated Object Classification	LSM Catalogue	1, 3, 4, 5, 6, 7, 9, 10

Table 6: Summary table of reference data sets used within WP3.

Dataset No.	Dataset Name	Equivalent SDP Product	SWG	Size (% SDP)
1	LOFAR1	Calibrated Visibilities	8	10.5 TB (15%)
2	LOFAR2	Image Cube [4]	1,9	36 GB (7%)*
3	MWA	Image Cube [1]	9	2.2 GB
4	SciServer	LSM Catalogue	6,9	0.34 GB (0.03%)
5	GBT	Pulsar Candidates	11	1 MB

* Image cube parameters for HPSO-27.

Table 6: Summary table of Use Case instances.

Use Case No.	Use Case Name	Data Set	Environment				
			1	2	3	4	5
1	Calibration & Imaging	LOFAR1	N	Y	Y	—	Y
2	Pulsar Re-folding	Pulsar	Y	Y	Y	—	N
3	Rotation Measure Synthesis	LOFAR2	Y	Y	Y	—	N
4	Object Detection and Classification	MWA	Y	Y	Y	—	N
5	Automated Object Classification	SciServer	Y	Y	Y	—	Y

WP3 Computing Req. Ashdown

Where is QA done?