EUCLID
Data Handling System and Ground Segment

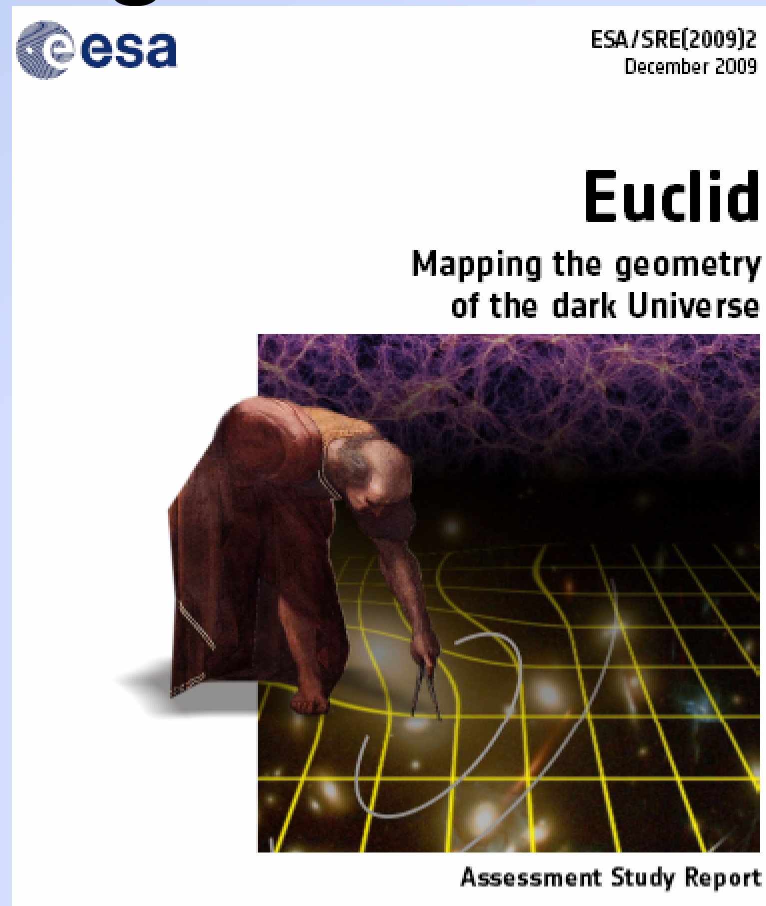
Edwin A. Valentijn
ESST
ESST Data handling working group

OmegaCEN - University Groningen

Observing the Dark Universe with EUCLID conference
ESTEC 18 November 2009

CAT: Merge DUNE SPACE -> data handling issue

- Feasibility study
  - Showstoppers ?
- How to
  - Data handling system



6 The Scientific Information Track – Operations and Dataflow

Euclid will deliver an unprecedented large volume of data for astronomical space missions: e.g. about 4

Dataflow WG

- WG members
  - Valentijn – Chair – OmegaCEN- Univ Groningen – NL

  - Niranjan Thatte – Oxford- UK
  - Matthias Steinmetz - AIP- D
  - Rene Laureijs – ESA
  - Anthony Brown – Leiden – GAIA
  - Fabio Pasian – INAF – Trieste - I
  - Herve Aussel – CEA - Fr
  - Andrey Belikov – OmegaCEN - NL
  - Marco Scodeggio – Milano – INAF -I

Raw data budget

- **20 000 deg²** of extragalactic sky |b|>30 degrees during the mission lifetime of **5 years**
- A ***field*** **~0.5 deg²** is dithered - **2400 sec**
- On a daily basis, Euclid will observe ***strips*** 20 degrees long = **36 fields**
- On approximately monthly basis a ***patch*** will be completed, 400 deg²
- **deep** field **40 deg²**, 2 mag deeper - no additional requirements
- covering of 20000 + 10% overlap deg² with leads to 44000 fields **->** **3. 4 years** of observing time. Add time for slewing, deep fields and calibrations plus maintenance periods **+1 more year** √

Instrument (frame size)	Collecting	Number of frames	Volume (Gbit/frame)	Compress Factor	Compressed Volume (Gbit/frame)
VIS (36 CCDs of 4x4k)	1 visible band of 0.5x1.0 degrees, 4 dithers to fill gaps	4	38.81	2.8	13.86
NIP (18 detectors of 2x2k)	4 NIR photometric bands, 4 dithers each, 24 2x2k detectors	12	14.5	2.5	5.81
NIS (8 detectors of 2x2k)	1 spectroscopic field, 4 integrations	5	2.69	1.5	1.80
Totals			56		21.47

36 fields * 22 Gb = 792 Gbit /day <-> new K band 850 Gbit / ~4 hours slot √

Data rates

	EUCLID	Planck	GAIA	Astro-WISE (KIDS)
Data Storage	5 PB	TB-scale (2 TB/year raw)	~ 1 PB (200 TB raw)	363 TB (2009) ~ PB (2010)
Data Processing	10^{20} Flop	5 Tflops (?)	10^{21} Flop	

40 TB raw compressed integer /year - > 270 TB/year floats -> 1 Peta / year-> 5 PB √

- All this in a distributed Mission Archive – MA √
- GAIA type process peak ~ 1 Tflop √
- extensive existing processing packages √

Single Mission Archive / science information system enabling:

- **quality control at all levels**
  - → feedback report to SOC/MOC
  - → monitoring satellite and instruments
  - → all aspects propagating as systematic errors ← **IOCs SDCs**
- **Connect various instrument and science teams**
  - → exchange and verification of results
  - → connect Ground based observations
  - → connect simulated data
- **EUCLID Legacy Archive**
  - → science ready data -> VO
  - → data processing from raw data to ELA – additional studies

Key Components GS

Satellite → Ground Station → MOC → SOC

ESA responsibility

Instrument Operation Centers (IOCs) →

Science datacenters (SDCs)

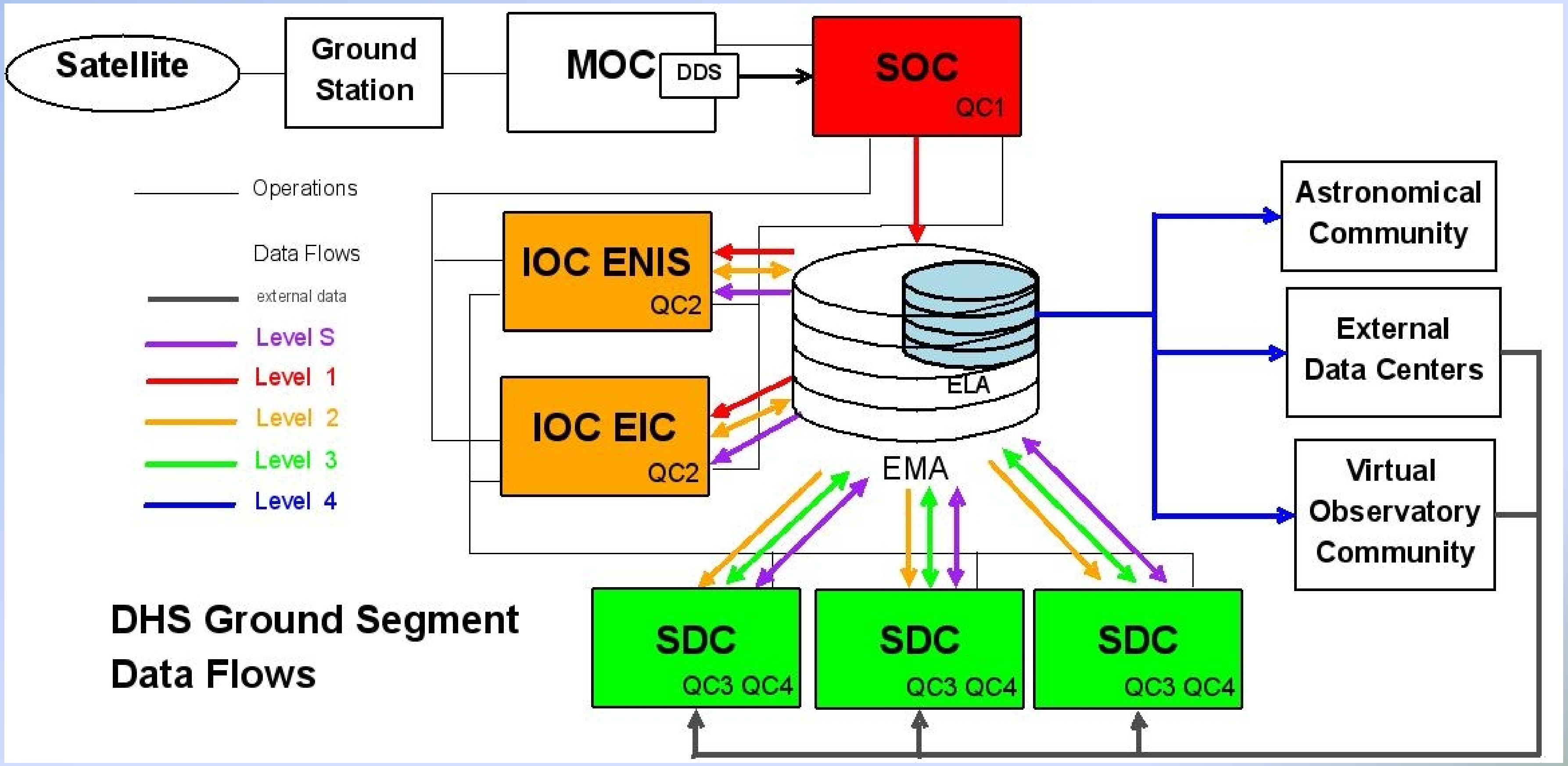
Consortia responsibility

Mission Archive → EUCLID Legacy Archive

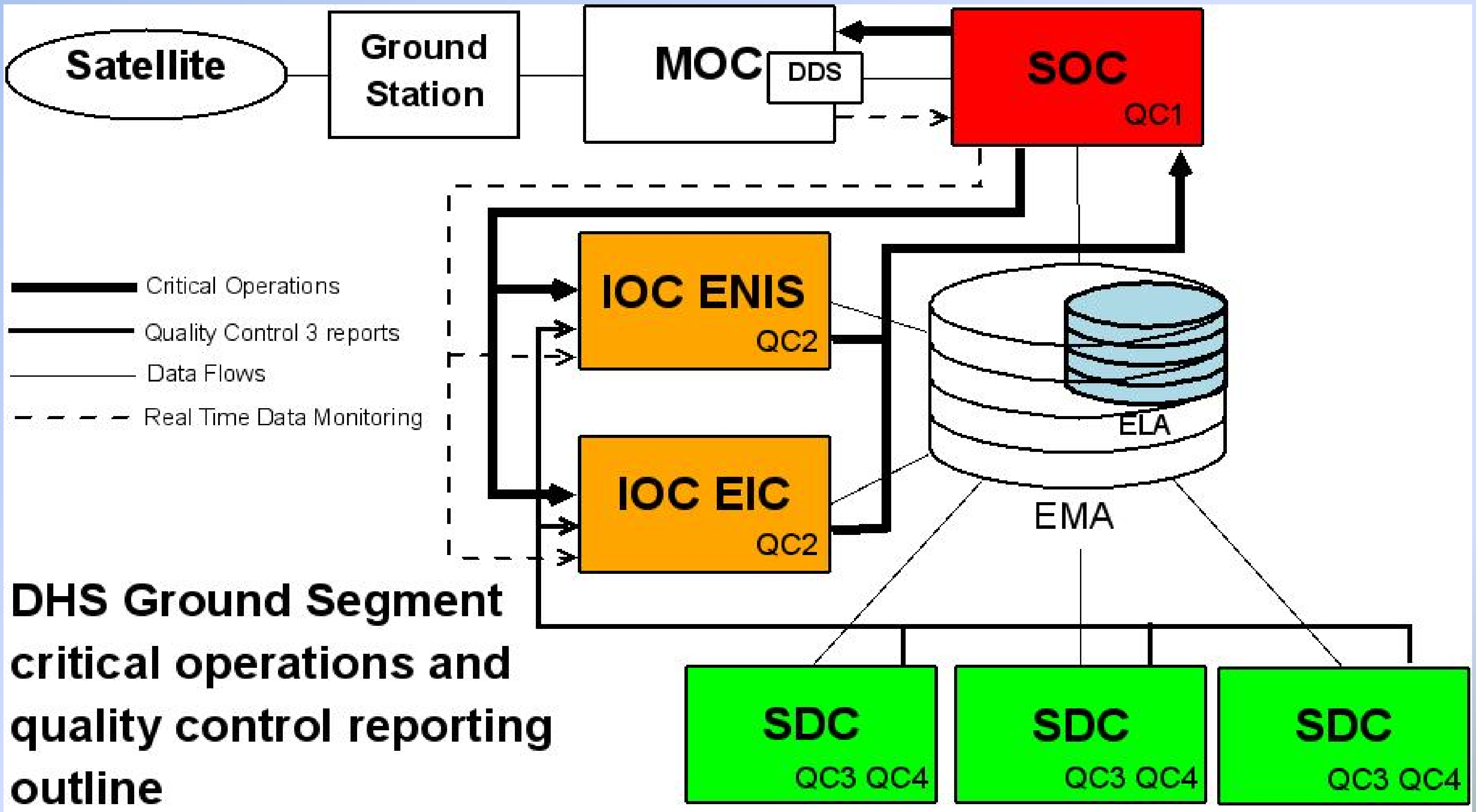
joint responsibility of ESA and Consortia

The GS in turn will be propelled by a data handling system (DHS), which amongst other things will maintain all administration of data at various stages of pro-cessing, data products and quality controls

GS Components
Data Flows

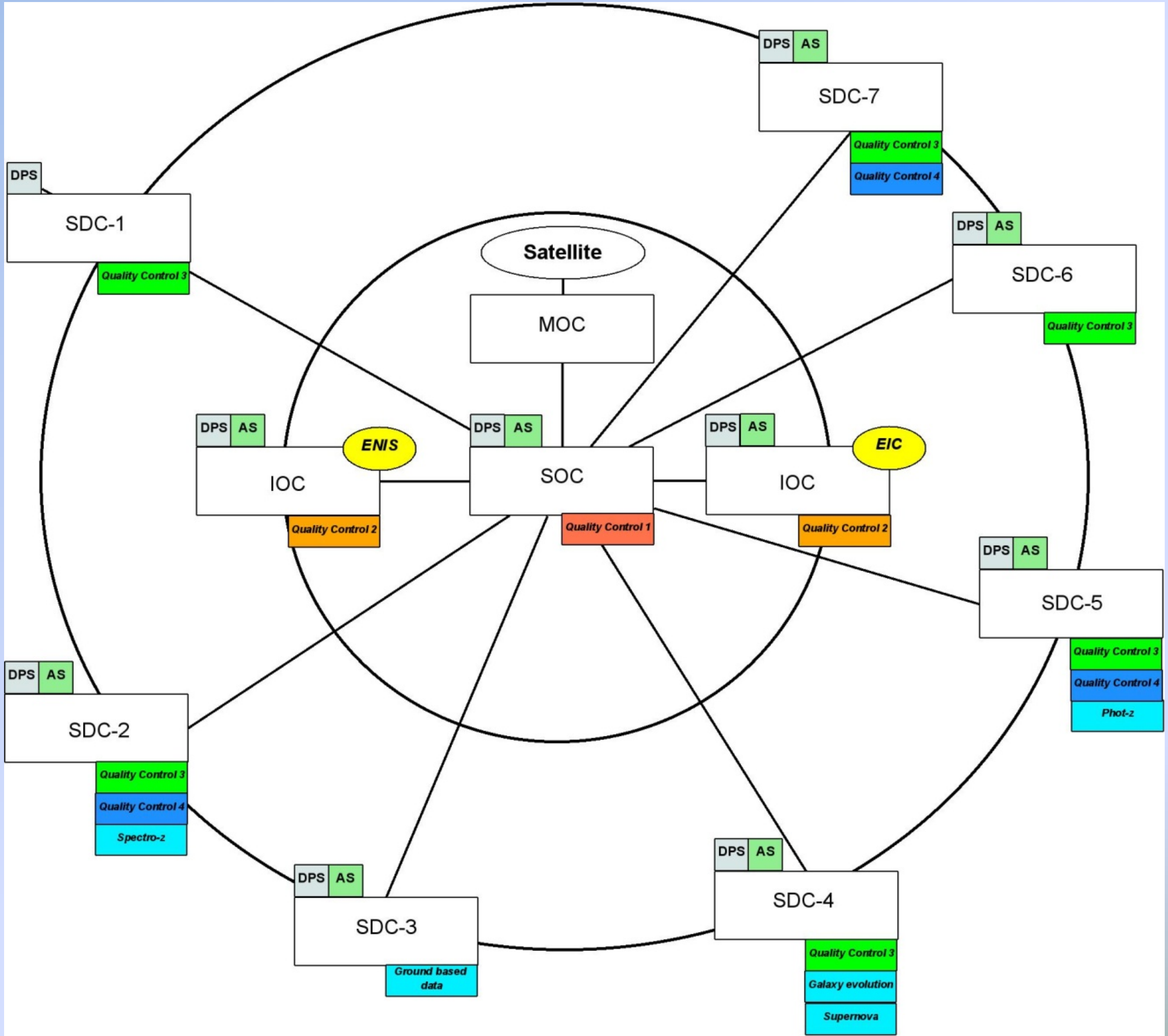


GS Components
Mission Support Operations



DHS Ground Segment critical operations and quality control reporting outline

GS Components



Data processing challenges critical for systematic errors

all requiring interoperability along 2 axis:

Vertical: Along data flow (feed-back)

Horizontal: Between teams (cross check)

- Photometric redshifts
  - vs Euclid Spectrogr. redshifts
  - vs ground based
- PSF modelling- CTI - glitches
- E.g. Illumination corrections- photometry
- Simulations

DHS
single distributed infrastructure
the dark energy of EUCLID DHS

- Component Based Software Engineering (CBSE) – modular approach
- Common Object-Oriented Data Model
- Persistence of Data Model Objects
- each processing step and lineage is saved in the system- full backward chaining (to ...)

Conclusions -DHS

- Data rates, data volumes √
- Including external data, Ground based, simulations √
- QC reporting and sharing √
- Cross verifications - redundant and competitive data processing systematic errors √

No duplications

- Save on data processing
- Save on staff
- Save on interfaces (one single interface for data access)
- Save on storage space (no unnecessary duplication of data items)
- Save on processing facilities (shared by participants)