SDC-CH:
Processing Infrastructure and Pipelines

Florian Dubath
Université de Genève

Swiss Euclid Days, 02-2020
• Data Processing Infrastructure

• SDC-CH Implementation

• Processing Functions & Data Model

• Pipelines

• SDC-CH Packages & PHZ Pipeline
Data Processing Infrastructure

- Too many data to move them around
- Data stored on SDCs, Meta-Data in a common DB
- Each software can be run on any SDC
SDC-CH Implementation

**ICINGA MONITORING**
http://ganglia-euclid.esc.rzg.mpg.de/ganglia

**CORRS**

**EAS-DPS**
http://eas-dps-cus.pip.euclid.astro.ru.nl (proxy, port 80)
http://application14.target.ru.nl (server, port 38550)

**EACS**

**EIS**

**SDC-XX DSS**

**ICINGA PROXY**
pieclmon00

**IAL**
piecldial00

**IAL-DRM**
piecldialdrm00 (Submission host)

**DSS**
piecldds00

**Local Storage**
/dss_storage

**SDC-CH**

**SQUID PROXY**
ecldprox01
/cvmfs/euclid.in2p3.fr

**CLUSTER**
isdc-cn20
isdc-cn21

**CernVM-FS Servers**
Ccernvmfss1Ii01.in2p3.fr
cernfs.roe.ac.uk

**Shared Storage**
/gpfs1/euclid/challenges/ial.drm/work_dir_host
PF & ECDM

- Data are processed in sequence

- Intermediate results: stored in DSS following pre-established format (Euclid Common Data Model)

- No loop

- Later PF’s have only access to upstream results

- Each PF build pipelines to be run in the processing infrastructure
Pipelines

- PF Pipelines: build by chaining software

- Input/output data format: ECDM

- Scientific software must be wrapped so they follow interfaces and understand ECDM

- Pipeline use fixed version of the software packages => tracability of products
SDC-CH Packages

Elements
(Framework used by SDCs for pipeline packages)

Alexandria
(common lib)

SourceExtractor++
(Source detection + measurement)

PhosphorosCore
(Template Fitting)

PHZ_NNPZ
(Nearest Neighbor)

PHZ_SomBiasCorrection
(Self Organized Map)

PHZ_Piepline
(Wrappers & simple tasks)

PHZ_IAL_Pipeline
(PHZ-Pipelines definition)

Phosphoros
(GUI over PhosphorosCore)

PHZ_Pipeline
(Wrappers & simple tasks)

Scientific community

Other SDCs' Packages

Euclid Data Processing
Example: Pipeline computing the Photometric Redshift (SC8 forecast)
Example: Pipeline computing the Photometric Redshift

**INPUT:**
MER Catalog containing photometries for Euclid and Ext filters
PHZ Pipeline

Example: Pipeline computing the Photometric Redshift

- ECDM Data
- Internal Data
- Executable

**Computation of the Redshift Probability Distribution (PDZ)**

(Phosphoros / NNPZ code are also available for use outside of Euclid)

**Computation of statistics on the PDZ**
PHZ Pipeline

Example: Pipeline computing the Photometric Redshift

Computation of the SED
(Needed by SHE for PSF determination)

Determination of the Tomographic Bin & bias correction
Example: Pipeline computing the Photometric Redshift

Wrapping PHZ results into ECDM