SDC-ES Operations

Goals:

- Run the SIM code in other SDCs
- Distribute the SIM production

Motivations:

- Integration and optimization
- Debug
- Technical tests

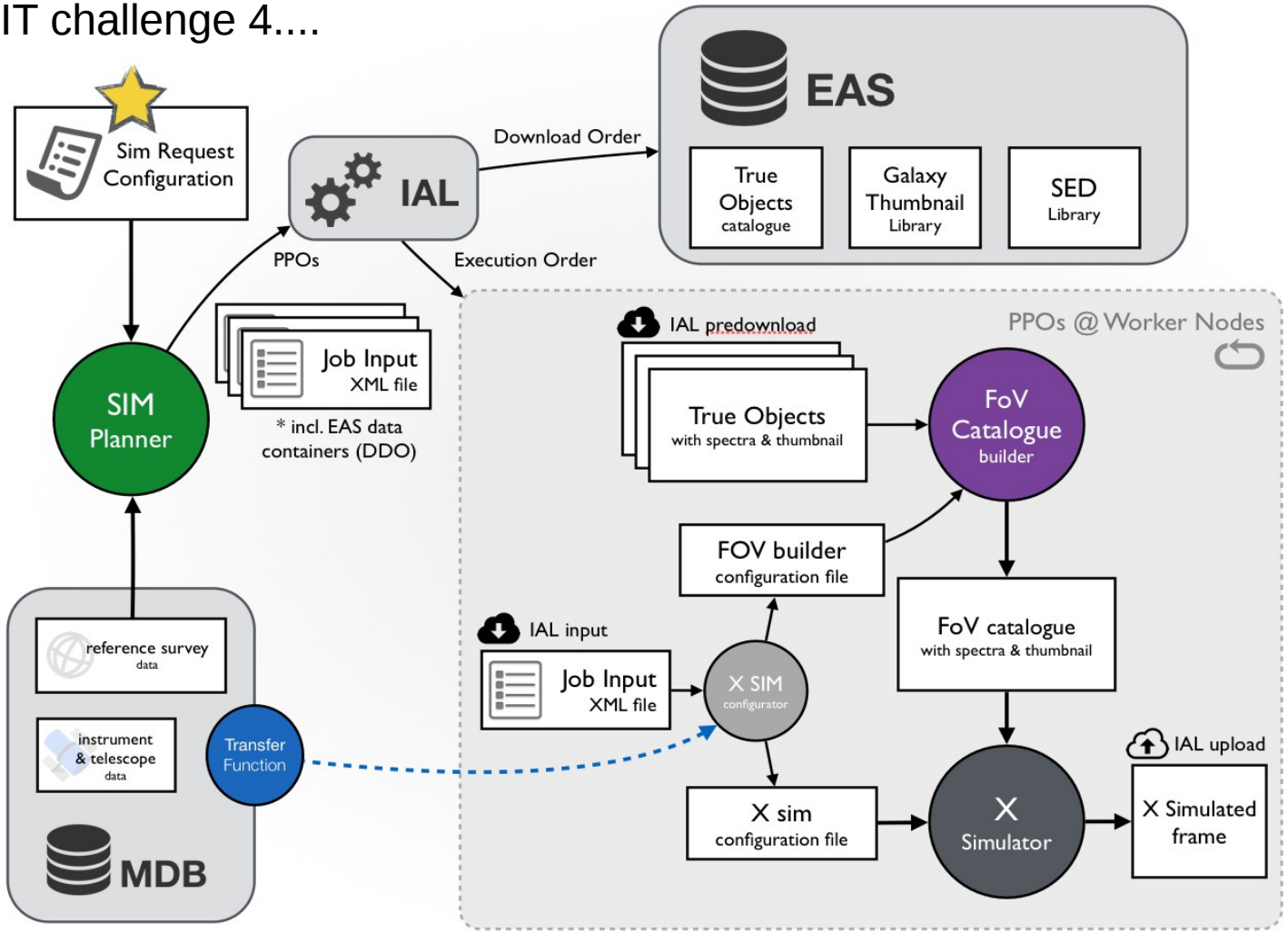
SDC-ES Operations

Outline:

SIM pipeline integration with Euclid tools
SDC-ES infrastructure and running setting
Production so far
How to run the SIM pipeline

SIM pipeline

From IT challenge 4....



Integration

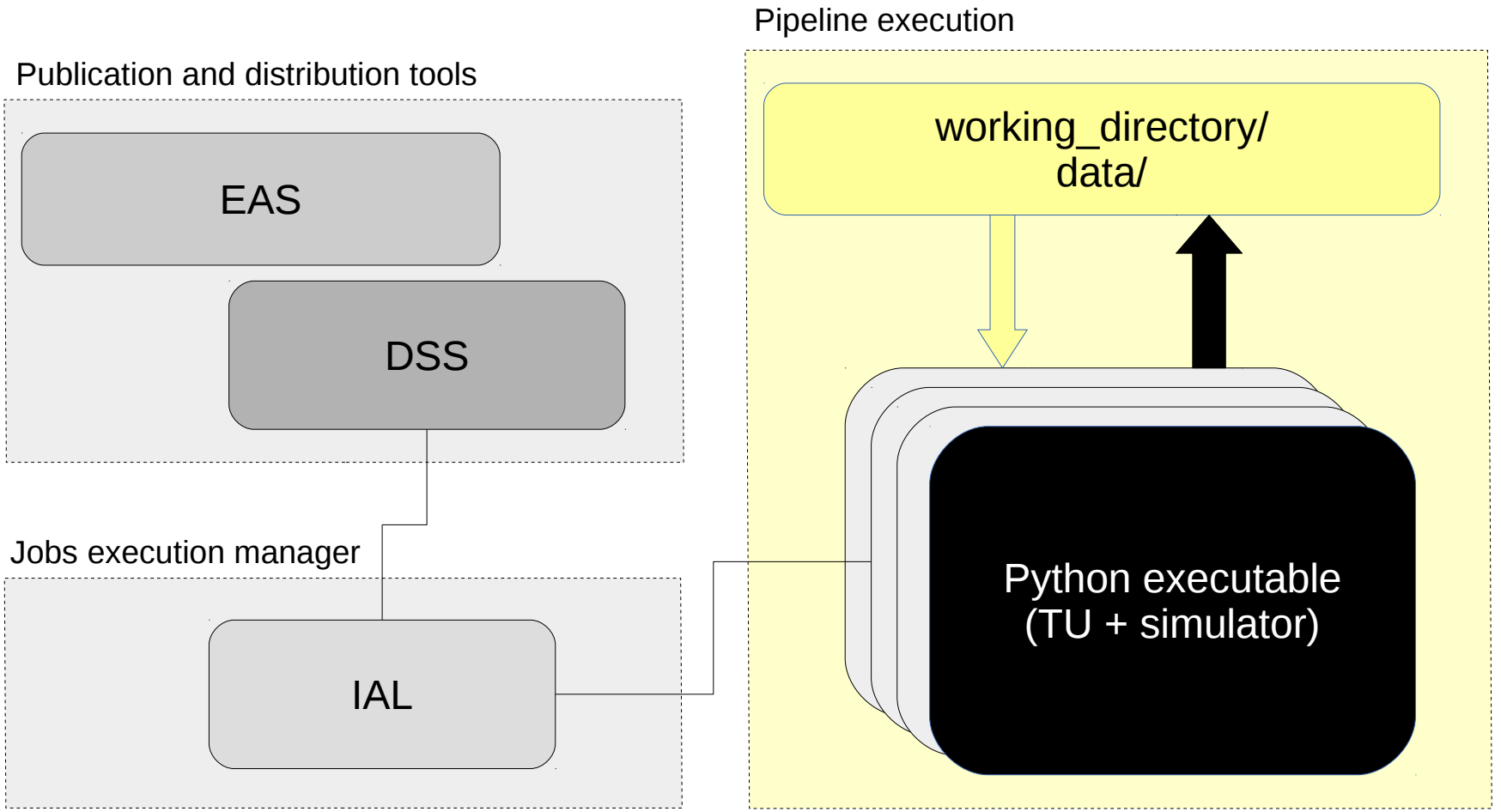
Integration completed for challenge 4

but

- **EAS - Data Model** in fast development, SIM execution often incompatible with EAS.
- Not many jobs (<=20) accepted to be run in parallel by IAL
- **IAL** in development, waiting for a “stable” version.
- **PPOs** needed “manual” intervention... solved.

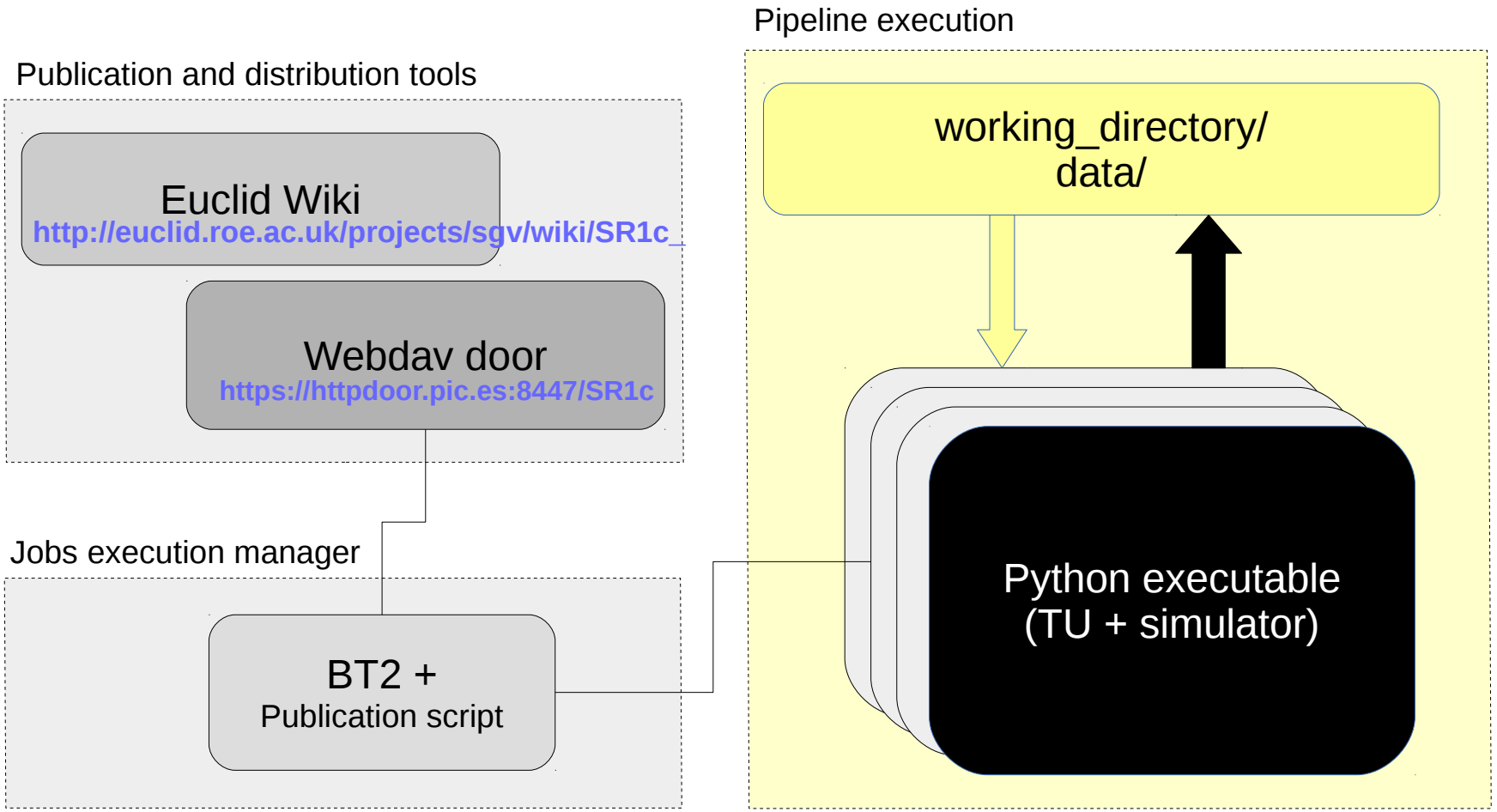
Challenge 4

From IT challenge 4....



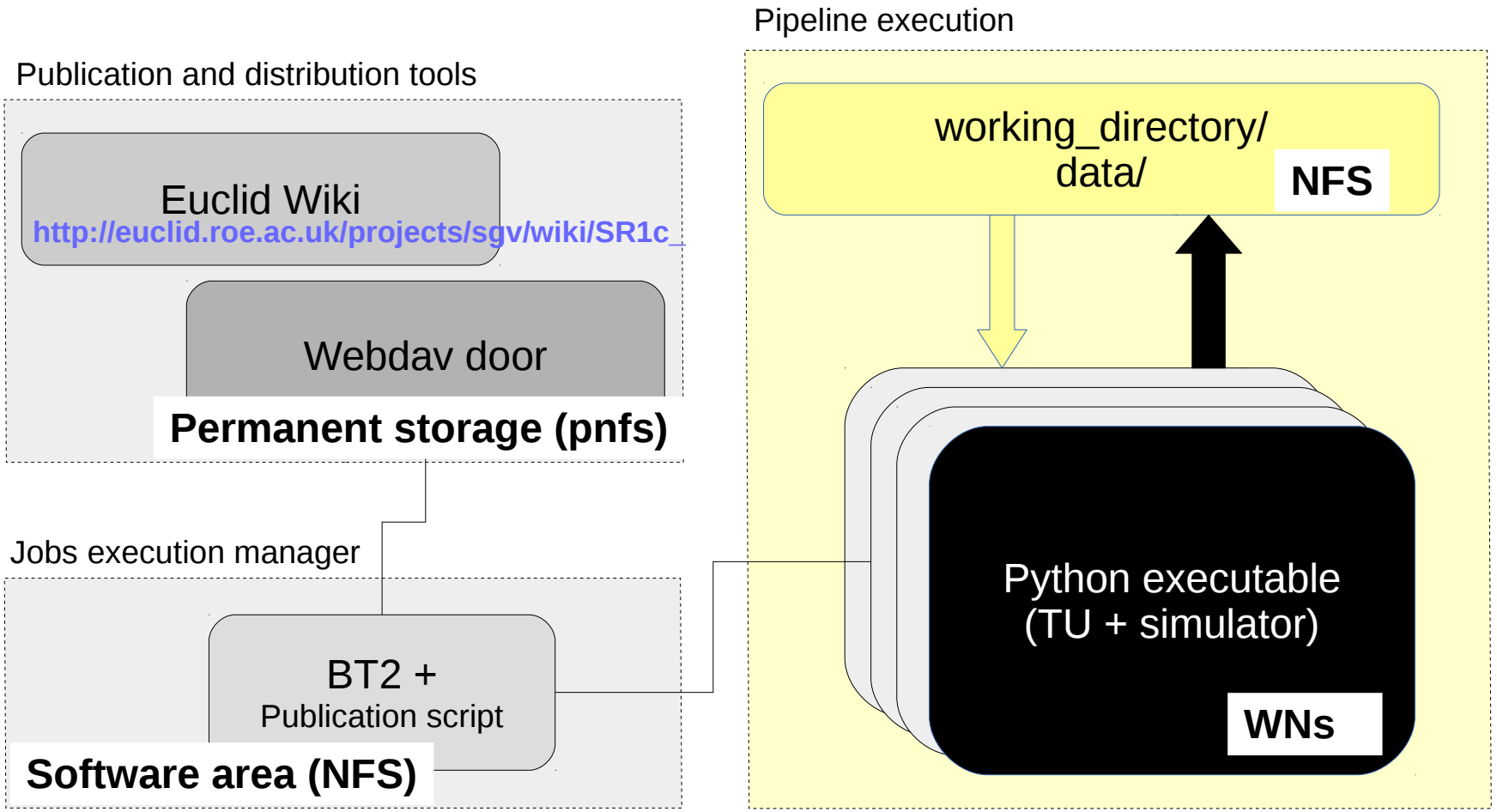
SR1b and 1c

SDC ES solution



SDC-ES infrastructure

SDC ES solution



SDC-ES SIM production

Data location: https://httpdoor.pic.es:8447/SR1b
https://httpdoor.pic.es:8447/SR1c

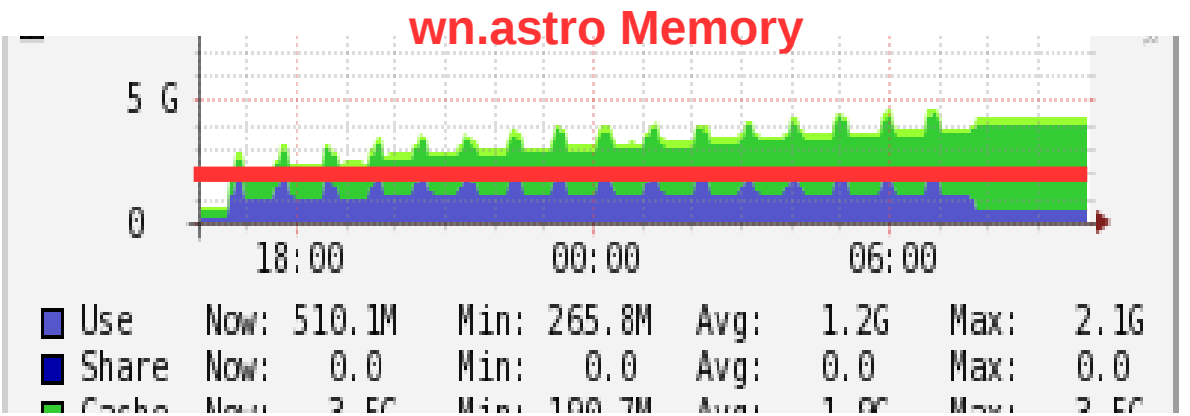
Wiki: http://euclid.roe.ac.uk/projects/sgv/wiki/SR1b_
http://euclid.roe.ac.uk/projects/sgv/wiki/SR1c_

SIM name	Date	Tot. n. images	Image size	Tot.space
SR1b-NIP SR1b-NIPb SR1b-NIPc SR1b-NIPd SR1b-NIPe SR1b-NIPf	2015-01-29 2015-04-13 2015-05-26 2015-07-13 2015-09-10 2015-09-21	300 x 16 detectors	8.3 MB x detector	
SR1b-NIS SR1b-NISb SR1b-NISc SR1b-NISd SR1b-NISe	2015-04-09 2015-04-24 2015-06-18 2015-07-13 2015-08-31	100 x 16 detectors	66.6 MB x detector	
SR1b-VISc SR1b-VISd SR1b-VISe	2015-06-25 2015-10-22 2015-12-09	100	2.4 GB	2.4 T
SR1c-NIPa... SR1c-NIPd	2015-11-20 2015-12-04	12 x 16 det.		
SR1c-NISa SR1c-NISb	2015-12-15 2015-12-18	4 x 16 det.		59 G

How to run it

- **SDC-ES WN specifications**
  - CPU: 1 core (single process, single thread)
  - Memory: 2-4 GB (4-8 GB RAM: limited resources, low availability)
  - Job killed by the system if too long (top limit: 48 hours)

- **Simulators checks before massive execution:**
  - Memory usage profile
    (really critical!!!!)
  - Execution time
    (no problem, just annoying...)
  - Room for optimization



How to run it

- **Environment**
  - LODEEN libraries (but not using LODEEN)
  - CODEEN Bindings
  - SIM svn code http://euclid.esac.esa.int/svn/EC/SGS/OU/SIM/
    - EuclidSIM
    - sim_planner
    - OU-SIM_NIP
    - EuclidVIS-SIM
  - Simulators code: http://svn.oamp.fr/repos/tips
  - Input **data/**

How to run it

- **Input files**
  - https://httpdoor.pic.es:8447/MDB/
    - Reference survey
    - Extintion law library
    - SED library
    - MDB file
    - ….
  - Star and Galaxy catalogs and Stamps

    https://httpdoor.pic.es:8447/TU/
  - Xml files produced by Simulation Planner

...all must be into the data/ directory

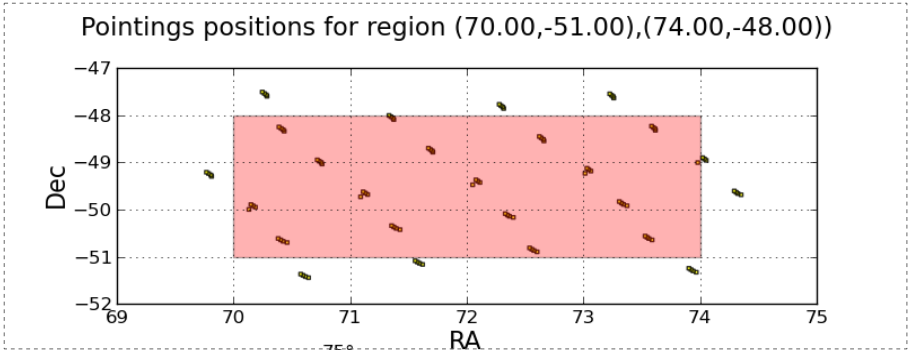
How to run sim_planner

python sim_planner/trunk/scripts/**Run_SIMPlanner.py**

- **--output-dir** /nfs/data/
- **--sim_request** sim_planner/trunk/tests/data/**sim_request.xml**

Simulation request (sim_planner/tests/data/) edited by hand:

```
<SimulationChannelList  NISP-S NISP-P VIS />
<SurveyConstraintArea>
        <RAMin>72.0</RAMin>
        <RAMax>72.0</RAMax>
        <DecMin>-49.5</DecMin>
        <DecMax>-49.5</DecMax>
</SurveyConstraintArea>
```



How to run sim_planner

python sim_planner/trunk/scripts/**Run_SIMPlanner.py**

- **--output-dir** /nfs/data/
- **--sim_request** sim_planner/trunk/tests/data/**sim_request.xml**

Many input files and options are hardcoded
Example:

```
starcatalog_release='StarRelease'   #'Empy' means 'do not simulate it'
galaxycatalog_release='GalaxyRelease'

query_EAS = False
starscat_filename = '/nfs/euclid/data/StarCatalogueCoverage.SC1b_r1.0.csv'
galscat_filename = '/nfs/euclid/data/GalCatalogueCoverage.SR1c_r1.01deuc.csv'

ref_filter_filename = 'EUC-TEST-FILJOHNV-2014-07-15T200000.000.fits'
ref_column_name = 'Mv'
survey_ref_filename = 'EUC-TEST-SURVEY-REF-2014-04-23.xml'
```

How to run a simulation

python /software/EuclidSIM/trunk/scripts/**EuclidNIS**

- **--input** /nfs/input/EUC-TEST-NISTASK-2015-12-15T091122.980.xml
- **--output** out.xml
- **--logdir** /nfs/log/ #optional
- **--workdir** ./ #optional

The EuclidSIM code:

- creates a tmp unique name dir where to run the job
- creates a link of data/ (config of the IAL-DRM) inside the working dir
- call the TU code and prepare the simulator configuration
- run the simulator
- copies the files listed in the output.xml to the **data/** dir

SDC-ES temporary changes:

- Wrapper modified to save the intermediate **TU files** in data/ (tmp solution).
- Script to recollect the output files from data/ and copy them in a structured shared dir.

SIM output dir example

```
job_rEgBHI                                  unique name dir
├── log                                     internal log dir
│   ├── EUC-TEST-SIMTASK.xml.err
│   └── EUC-TEST-SIMTASK.xml.out
├── wd                                      effective working directory
│   ...
│   ├── myTUGalaxyCatalog.fits              TU intermediate files
│   ├── myTUGalaxySpectra.fits
│   ├── myTUGalaxyThumbnailsPaths.dat
│   ├── config.fits                         simulator configuration file
│   └── outputSIM
│       ├── image_00.fits                   simulated images
│       └── image_01.fits
...
└── ws -> /nfs/euclid/                      where data/ is
```

Massive simulations in SDC-ES

0. Preliminary work

a. Update the code to test
b. Update input files in data/ if necessary
c. Configure simulation request and Run_SIMPlanner script accordingly
d. Run a single test job in a interactive wn and monitor RAM/CPU/IO...

1. Create input files

Configure and run sim_planner –-> /nfs/input_dir/EUC-TEST-SIMTASK-*.xml

2. Create a BT task and run the simulation

Run a job for each /nfs/input_dir/EUC-TEST-SIMTASK-*.xml

3. Publicate results

Run the script sim_planner/trunk/scripts/SIMOutput_separation.py to copy the output of the simulation in a shared directory and tgz it.

Massive simulations in SDC-ES: BT tasks

euclid_sim_challenge task

It creates an euclid_ simulation task for each EUC-TEST-SIMTASK-.xml file in tasks_path*

Example config:

```
logdir: /nfs/euclid/data/local_workspace/log_SR1c_NISb
tasks_path: /nfs/euclid/data/local_workspace/input_SR1c_NISb
workspace: /nfs/euclid/data/local_workspace
```

euclid_simulation task

Runs the Euclid simulation in the worker node with the specified parameters

Example config:

```
logdir: /nfs/euclid/data/local_workspace/log_SR1c_NISb
ports:
    input: data/EUC-TEST-NISTASK-2015-12-18T144901.842.xml
    output: out_EUC-TEST-NISTASK-2015-12-18T144901.842.xml
task: NIS
workspace: /nfs/euclid/data/local_workspace
```

Massive simulations in SDC-ES: Publication

Run the script sim_planner/trunk/scripts/**SIMOutput_separation.py**

- copy the output of the simulation from **data/** in a directory to be shared:
  - Parse the output.xml file
  - Copy all the files listed there, organize the output files for FPA
  - Copy additional TU files, if available in **data/**
- Tgz the result

https://httpdoor.pic.es:8447/

Documentation in the wiki:
http://euclid.roe.ac.uk/projects/sgv/wiki/SR1b_