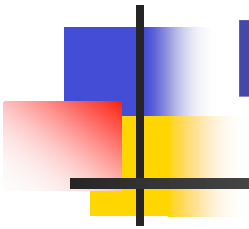
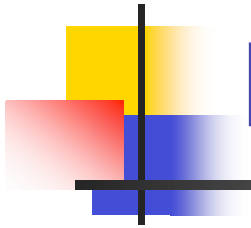
ISTITUTO NAZIONALE DI ASTROFISICA



Porting Astronomical Applications

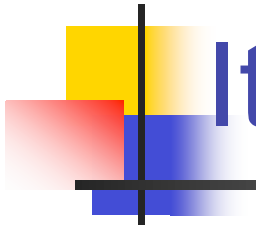
The INAF experience

Dr. Giuliano Taffoni (PhD)



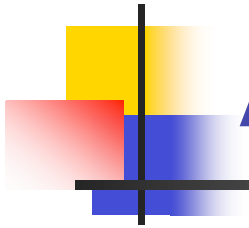
Plan of the talk

- Italian Institute for Astrophysics
- The Astronomical applications
  - Computational Intensive
  - Data Intensive
  - Parallel codes
  - DB related codes



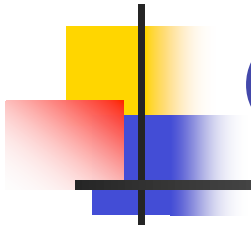
Italian Institute for Astrophysics

- Italian Observatories (19)
- Distributed in all the Country
- Heterogeneous community
  - Theoretical calculations
  - Observational analysis
  - Technology



Applications

- Pure Theoretical
- Techno
- Image analysis



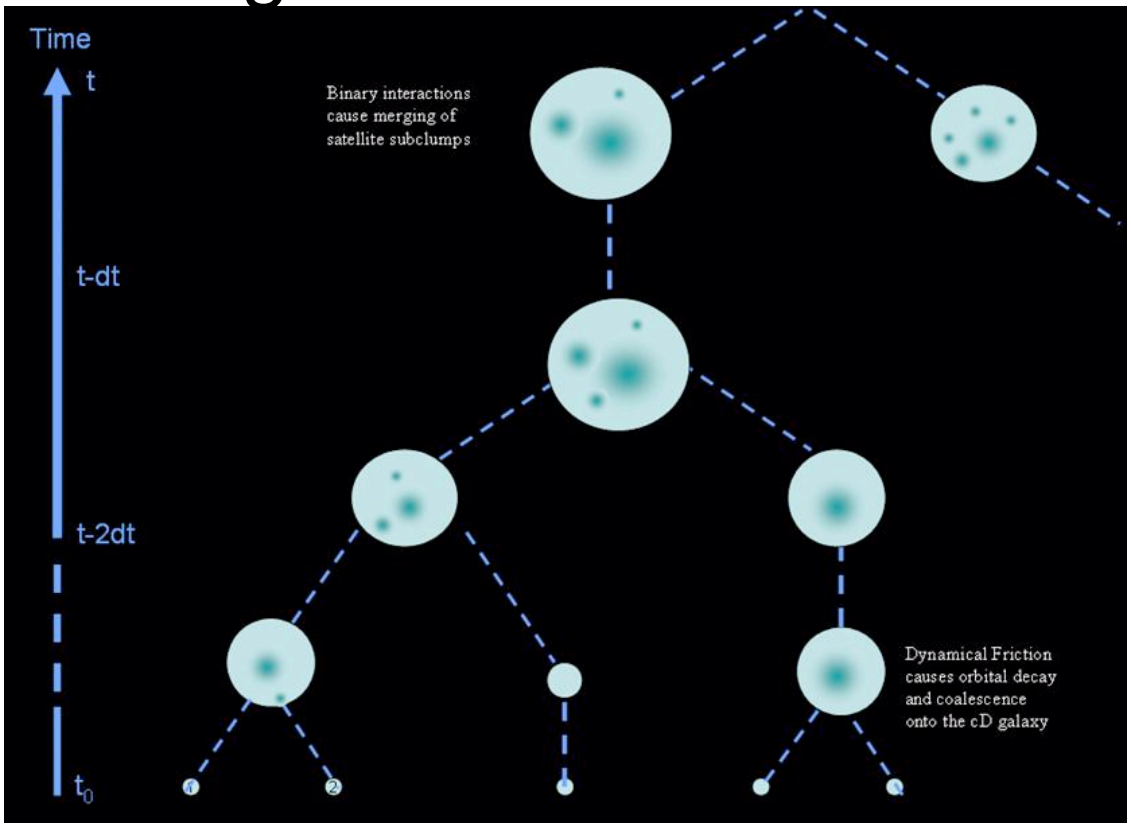
Galaxy formation codes

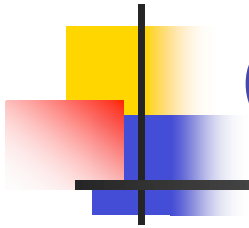
- Semi-analytical approach
- Uses Analytical recipes to:
  - Calculate merging history of halo
  - The substructure evolution of a halo
  - Barionic processes that forms stars and produces emissions
- Produces:
  - Astronomical observables related to galaxy formation history

Code Step 1

Code Step 2

- Following the evolution of substructures



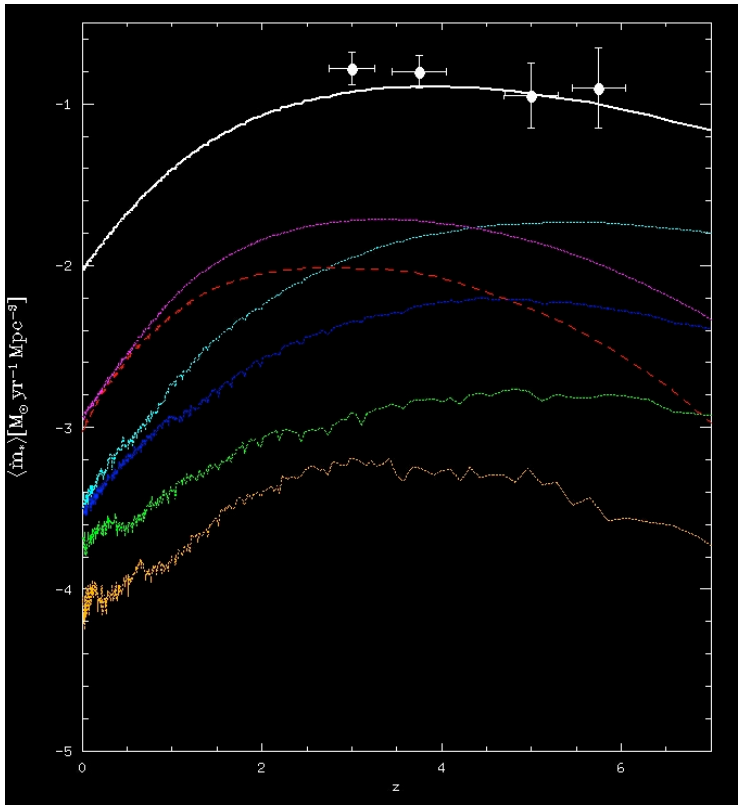


Code step 3

- Barionic evolution
  - Radiative cooling
  - Galactic disk formation
  - Star Formation
  - Interactions and star burst
  - Supernovae feedback
  - Evolution of stellar populations
  - Metallicity
  - Dust Absorption

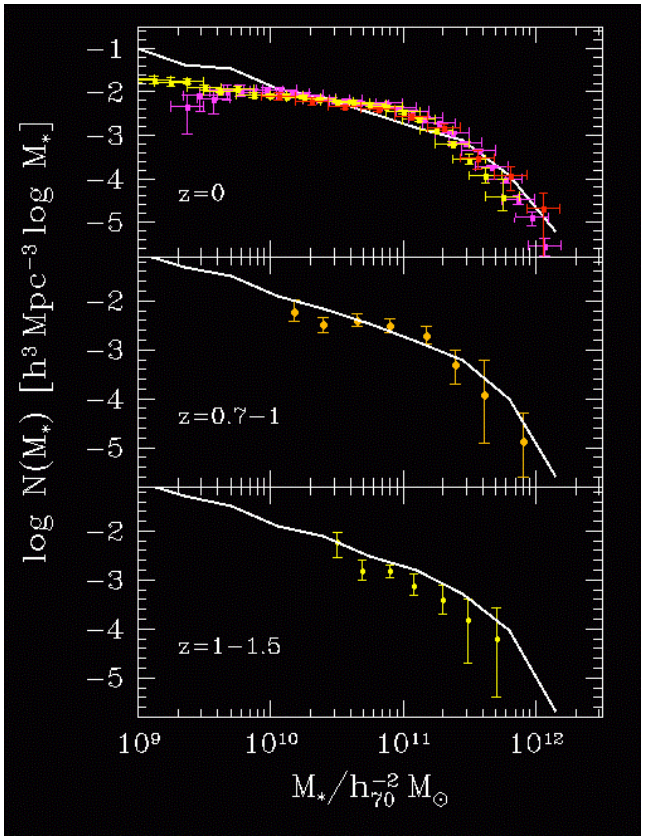
Observables

- Star formation Rate

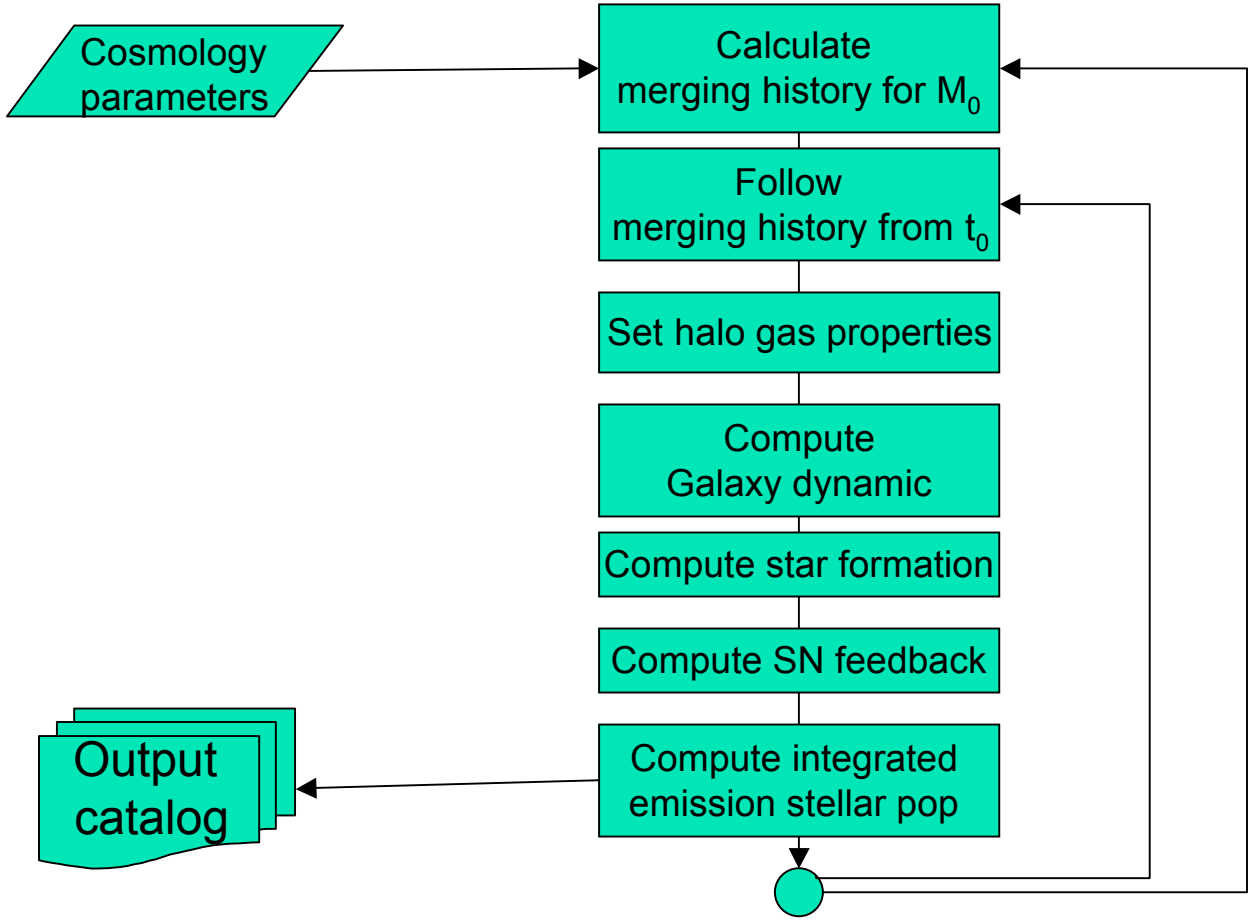


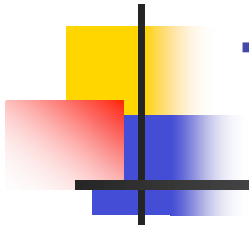
Observables

- Stellar Mass function



Code Structure

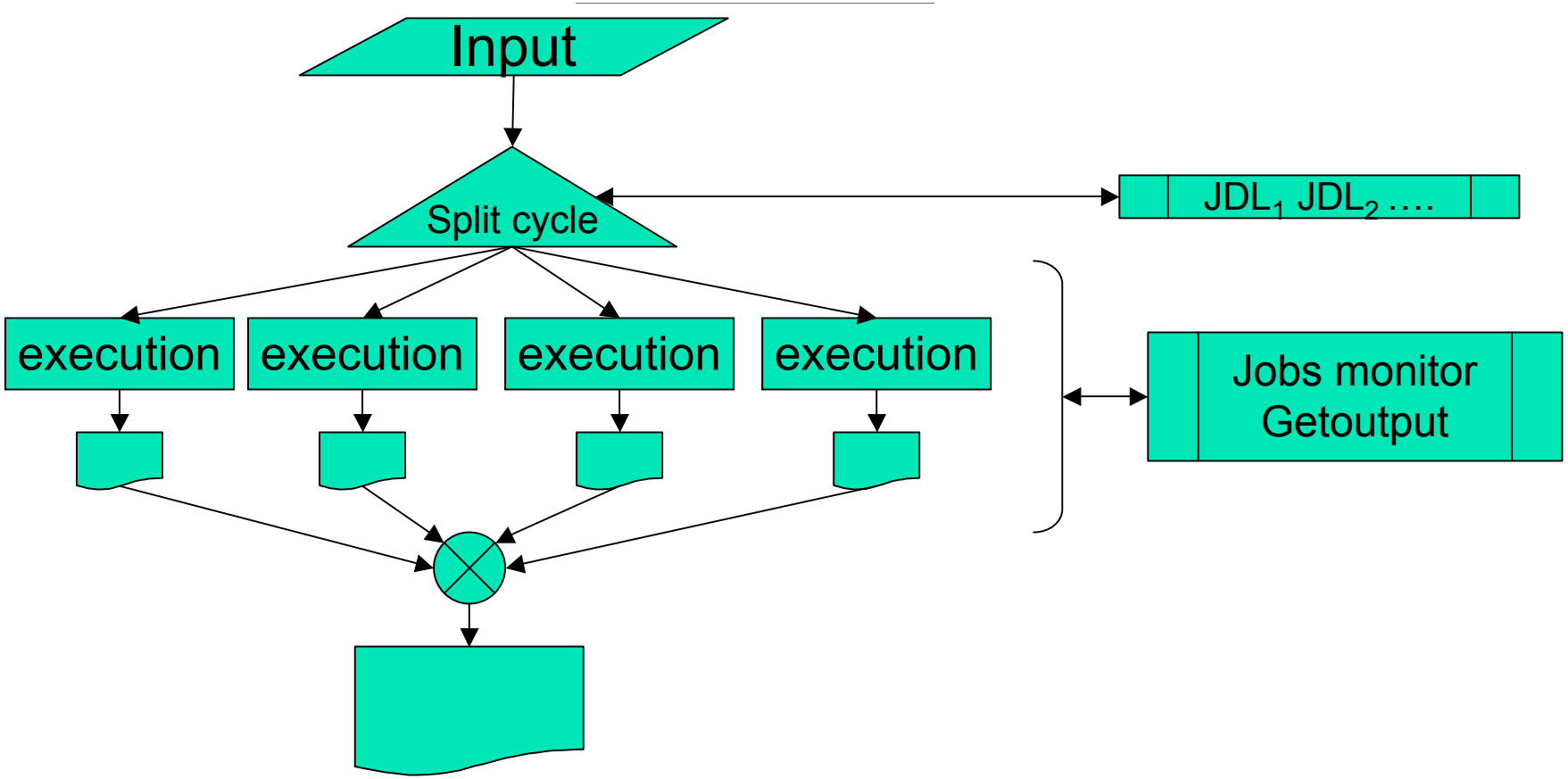


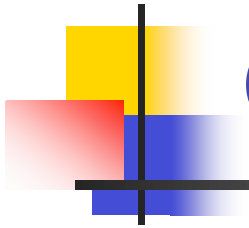


The Code

- 3 binary codes chained
- 18 parameters (Cosmological Parameters, etc.)
- 5 output files (< 5 MB compressed)
- Running time ~3 days for 1000 sampling

Galaxy formation on the Grid





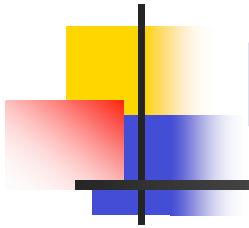
Galaxy formation on the Grid

- We do not need Grid FS (use sandbox)
- We do not need to modify code
- We develop a simple application env:
  - Create and submit a number of JDL one for each group of random generations
  - Monitor jobs status
  - Average on the results (locally on the UI)

Tips and Tricks

- MonteCarlo simulations are the optimum for grid computing
- **Scales ~ linearly**
- Split the job into a number of
- (latency of the WMS)
- Use as much as possible the S box

PARTITIONABLE JOBS?

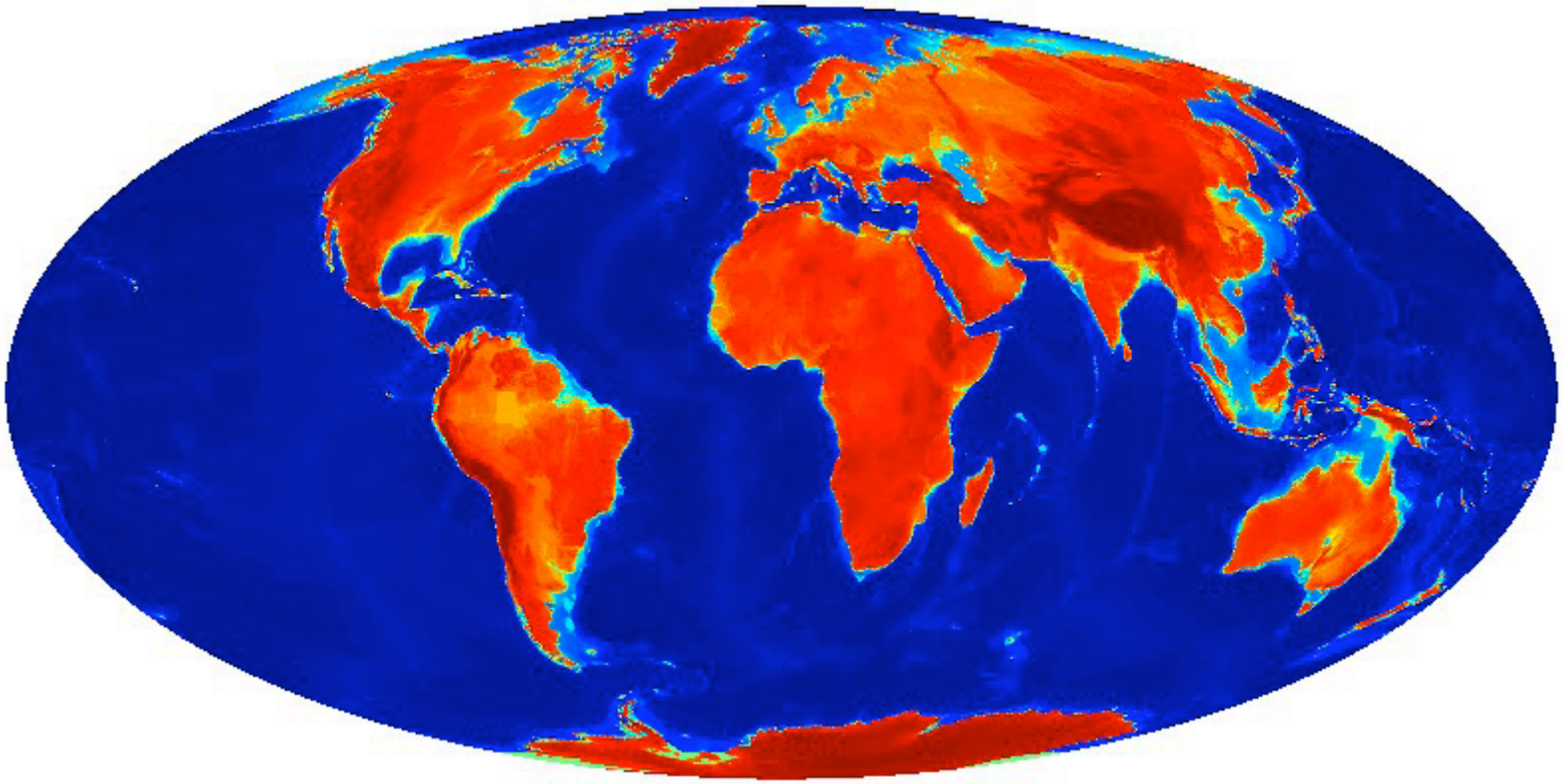


Planck Satellite Simulations

- Massive computing
- Massive data production (~PB)

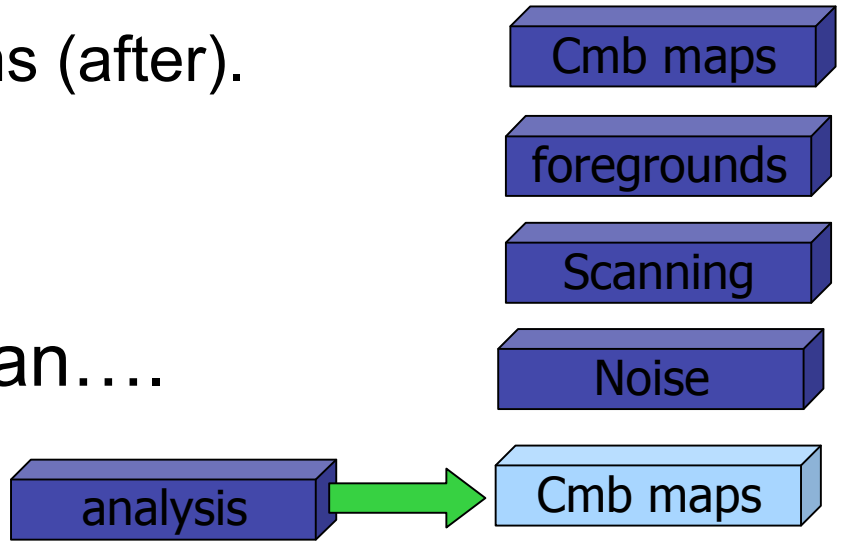
Planck Mission

- Measure cosmic microwa
  - succeeds COBE, Boomera
  - aims at even higher resolu
- Timeline
  - launch August 2007
  - start of observations 2008
  - duration >1 year
- Characteristics
  - continuous data stream (TOD
  - large datasets
  - changing calibration (parame
  - high-performance computing



Plank simulation is a set of 72 instances of the Pipeline.

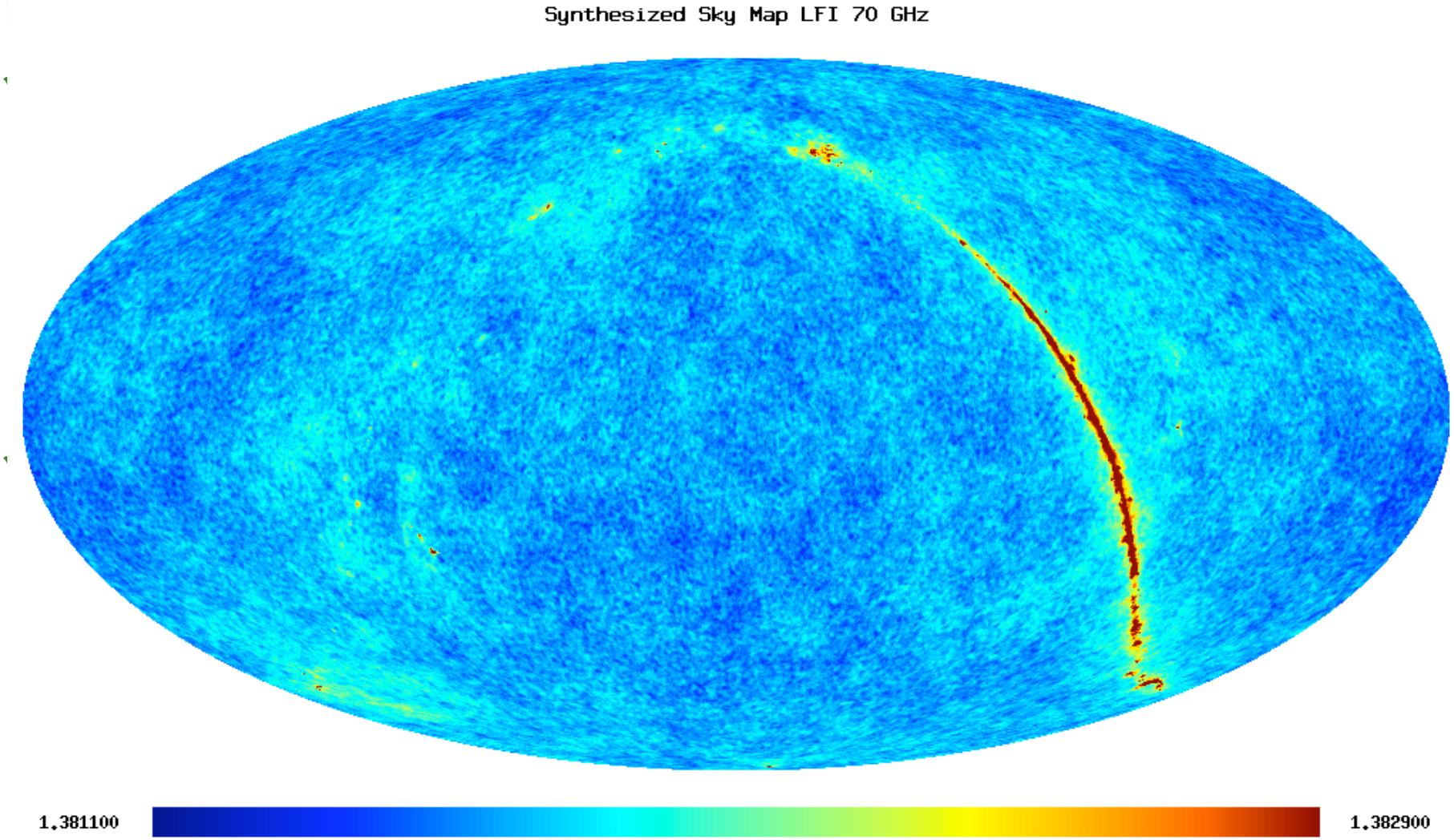
- Mission simulation Software:
  - ground checks (before);
  - control check & corrections (after).
- Pipeline:
  - Chained but not parallel;
- Stages are C/C++/Fortran….
- Shell/perl scripts;



Some Numbers

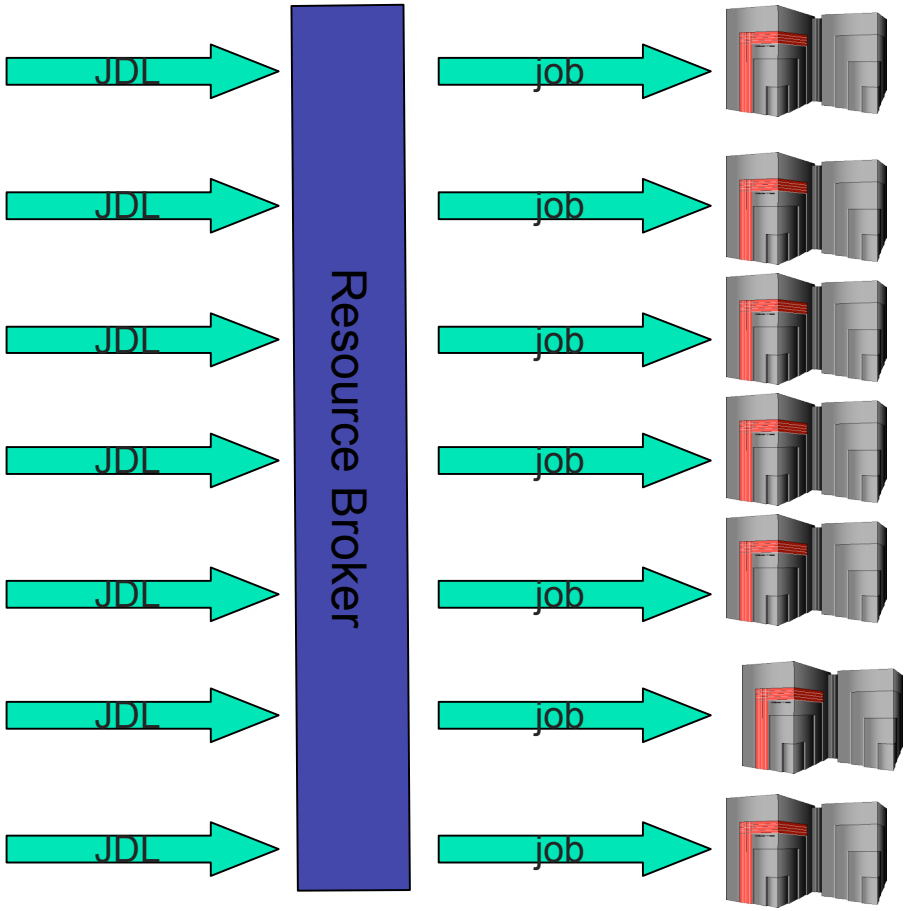
LFI 30 GHz (4)	LFI 44 GHz (6)	LFI 70 GHz (12)
+ HFI (50 channels) 100 GB each		
TOTAL (for LFI)		
255h	1.3 TB	

Planck on Grid



Application Environment

- Submission scripts:
  - Perl:
    - Coherent Configuration
    - JDL
    - Submission;
- WN environment:
  - Set up;
  - SE-WN interaction;
- GridFS tools:
  - grid-ls
  - grid-ll
  - grid-mv
  - grid-cp
  - grid-rm
- Metadata Management tools:
  - GDSE (IVOA)



Lesson learned

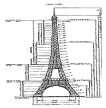
- **Massive data production on WN (> 40 GB):**
  - Big disks;
  - Complex site topology (parallel/distributed FS);
  - Compressing/RM-CR/removing file program;
  - FITSIO with fgal/gsiftp support;
- **CPUs not extremely important;**
- **Data handling:**
  - Complex data structure;
  - Complex data structure;

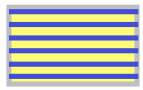
•METADATA!

- **1 GB RAM.**

10-15 terabytes
~20.000 CD-ROM

1 Eiffel Tower unit

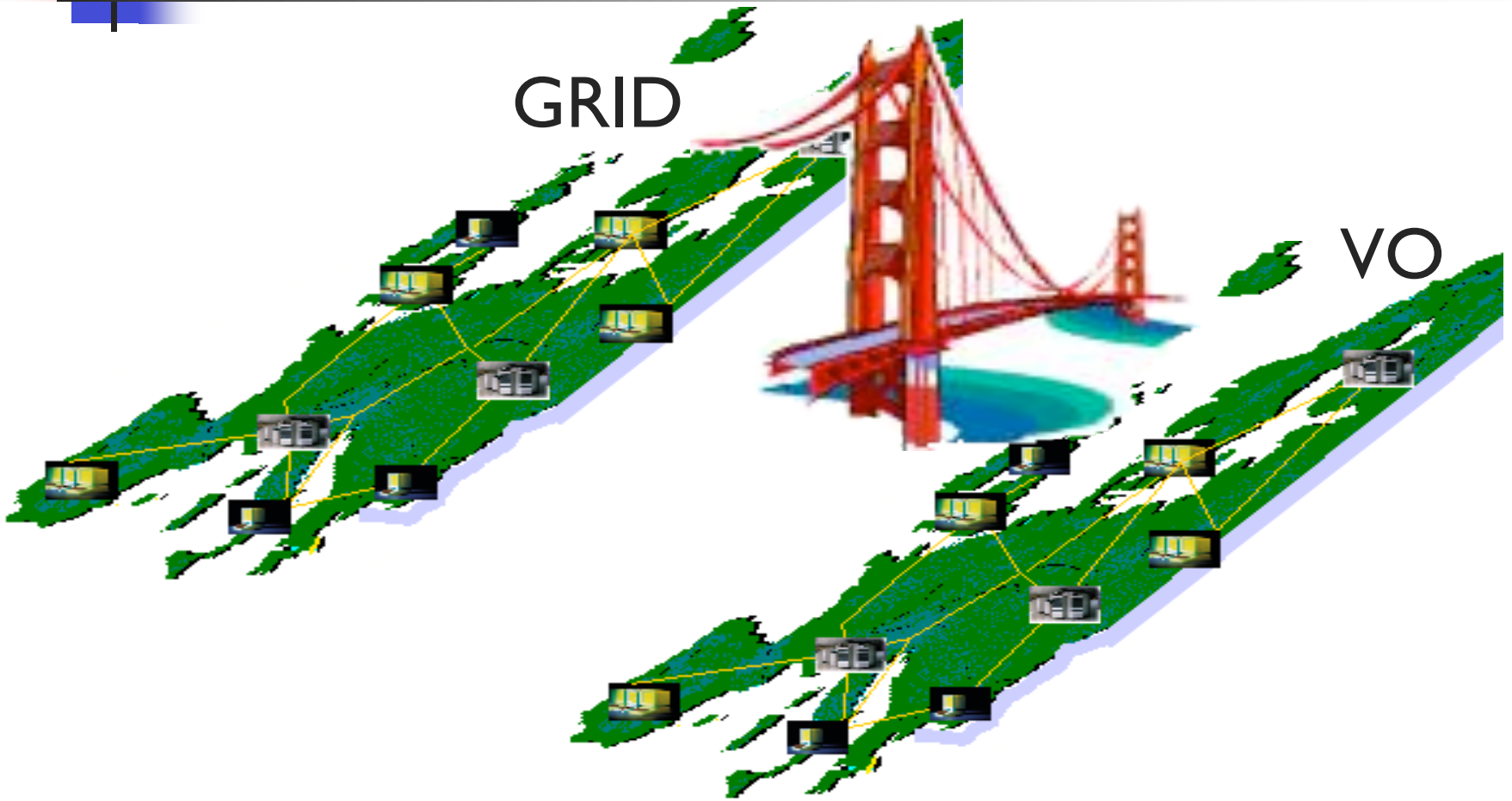


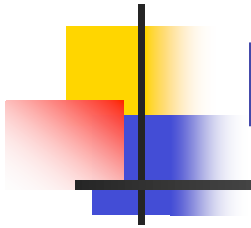


IVOA and DB in the Grid

The International Virtual Observatory Alliance (IVOA) was formed in June 2002 with a mission to *facilitate the international coordination and collaboration necessary for the development and deployment of the tools, systems and organizational structures necessary to enable the international utilization of astronomical archives as an integrated and interoperating virtual observatory*.

Two worlds



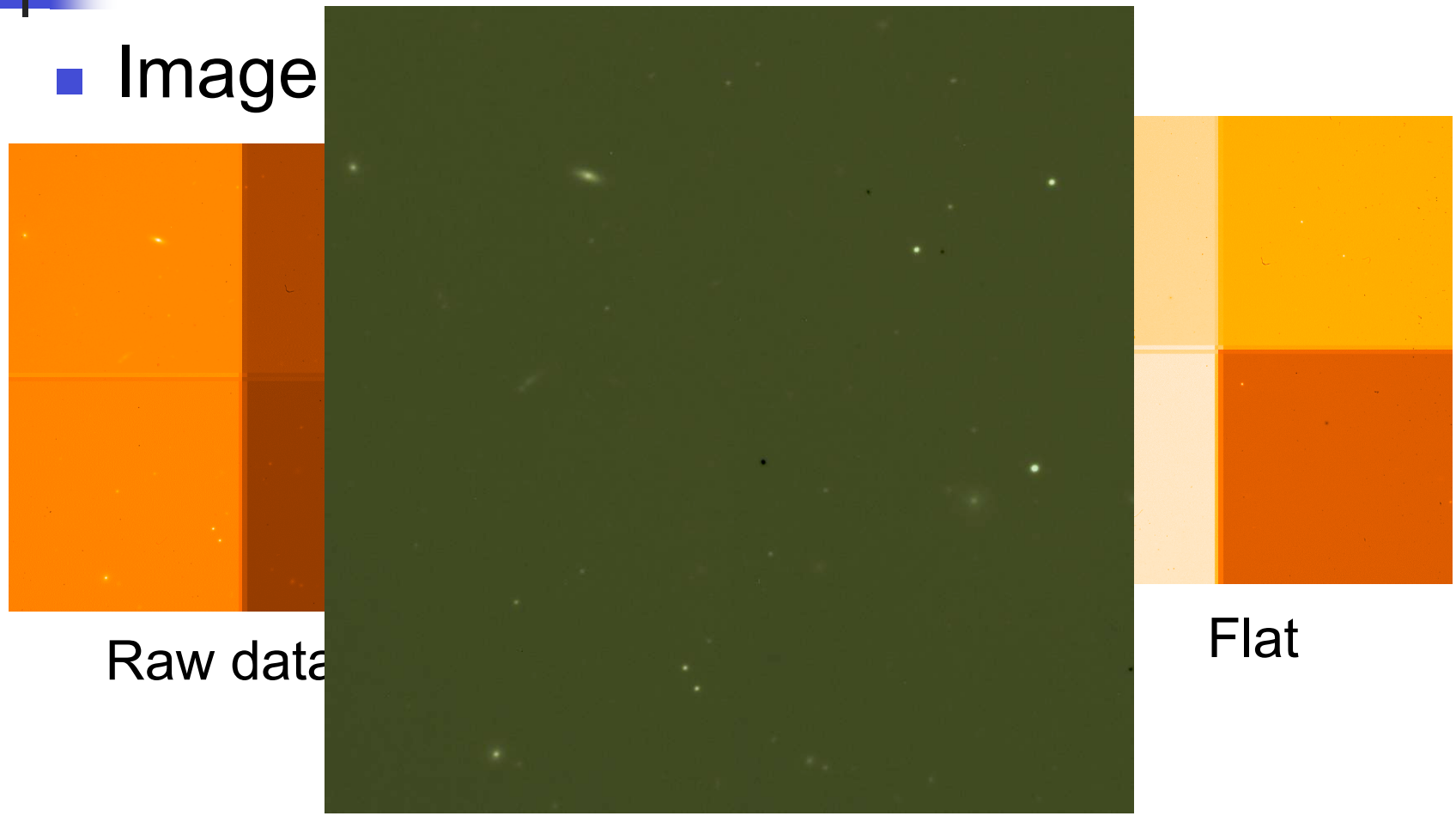


Requirements

- Allocate computational resources from VOBS (CEA)
- Locate Data from Grid
- Useful to compare theoretical data with observations
- Useful to share theoretical data to the community

Image Analysis

- Image

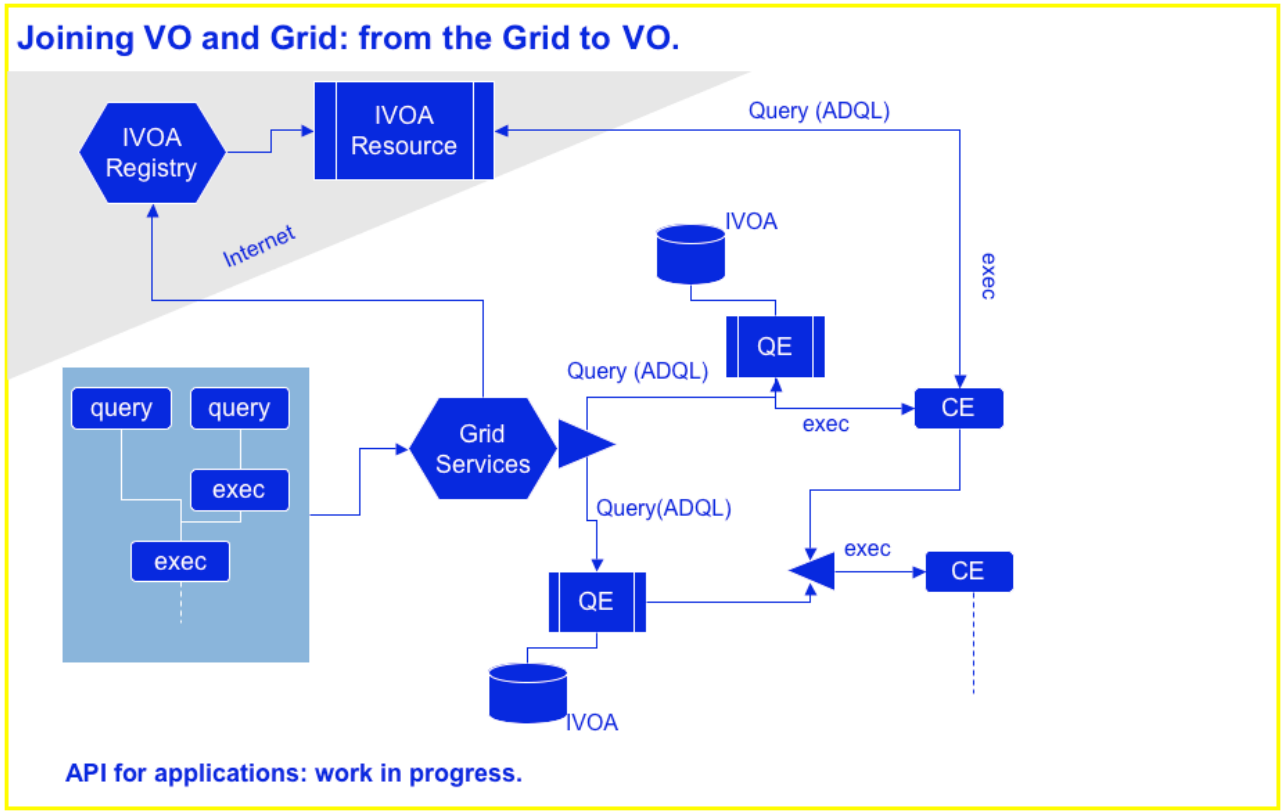


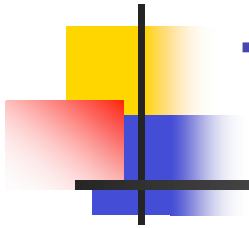
Raw data

Flat

Problem

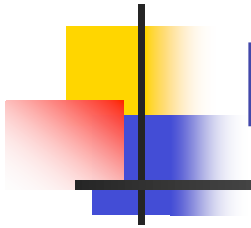
- Create complex workflows





The portal (the Bridge)

- Data saved in the Grid
- DB is an external entity
  - Connected to Grid (GDSE)
  - Connected to IVOA (Resource)
- Computations on the Grid



Lesson learned

- Moving images is too time consuming
  - A night of data > 30GB !!!!
- Distribute images in the Grid
- Metadata
- GDSE -> IVOA
- Theoretical jobs run in the Grid

Questions?