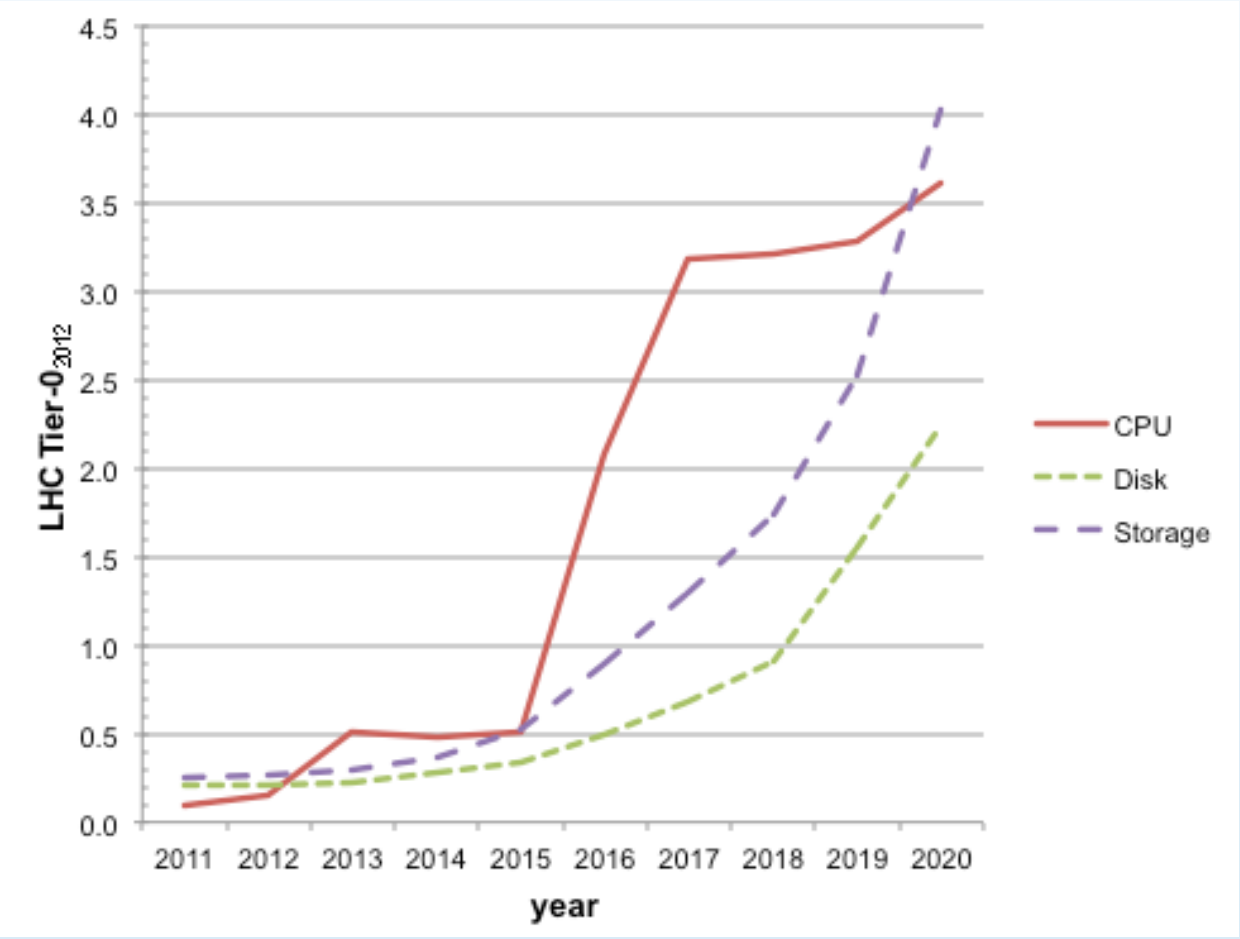
The François Arago Centre (FACe):
Challenges in Data Analysis

Volker Beckmann
François Arago Centre
APC, Université Paris Diderot

Challenges in Data Analysis

- After particle physics (LHC), also more and more *astro*particle physics projects face huge processing challenges
- Pbyte scale computing (e.g. LSST: 10-100 Pbyte per year)



Current use and future requirements (data collected in 2013) in CPU usage and short term disk and long-term (tape) storage space of astroparticle physics projects in Europe.
ApPEC white paper, 2014

ApPEC: Astroparticle Physics European Consortium

Solutions

François Arago Centre

Different processing options:

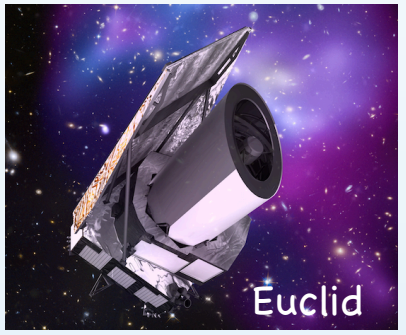
- Local cluster

  e.g. processing services for INTEGRAL

- Several independent processing centres

  e.g. GAIA (20-50 Gbyte per day)

  Euclid (~Pbyte per year), LSST …

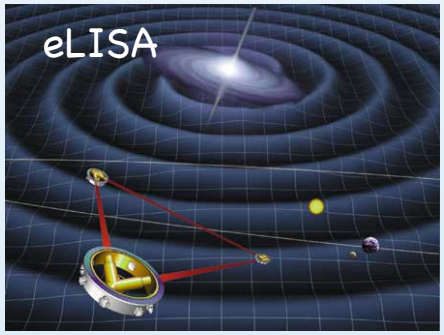


- GRID

  e.g. LHC (37 TByte/day → 11 Tier 1, 10-15 PByte/year), CTA

- Cloud infrastructure

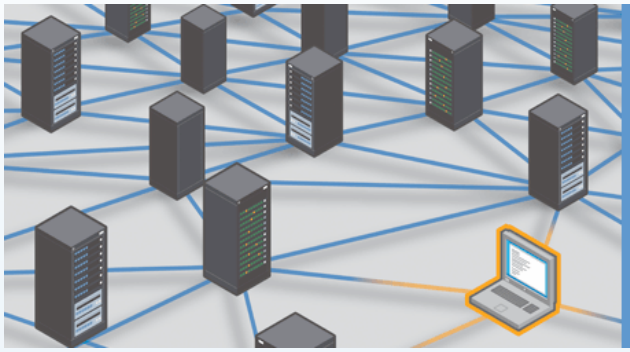
  e.g. eLISA development (and SDC)



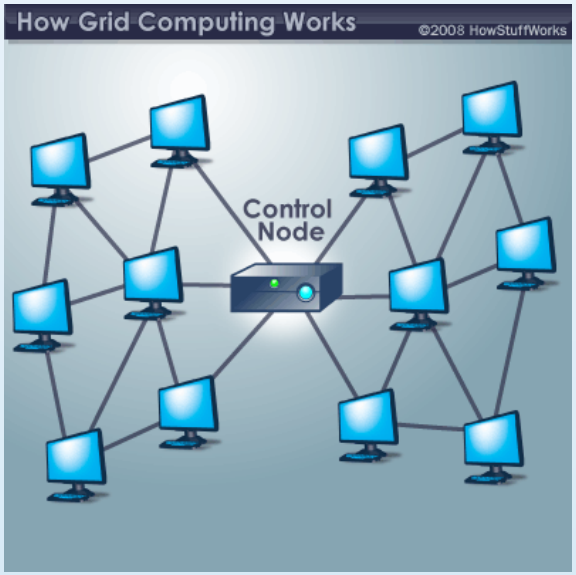
- Cloud infrastructure (commercial)

  e.g. LHC for Higgs discovery, space?

GRID



- Several participating centres
- Same installation
- Middleware (e.g. gLite, EMI)
- high entry level
- EGI: Heavily supported by FP-6, FP-7, FP-8 (>100 M€ for EGEE and EGEE-II)
- Data intensive processing: needs dedicated infrastructure (e.g. LHC)
- Few, well connected large centres
- Long term projects (>> 10 years)



EGI: European Grid Infrastructure
EGEE: Enabling Grids for E-Science in Europe (FP-6 / FP-7)

Science Clouds

Cloud computing skepticism *(Shane Canon, Lawrence Berkeley National Lab)*:

- *Overhead to convert to Cloud environments*

Not more heavy than training colleagues on clusters (depends also on application)

- *Virtual instances underperform bare-metal systems*

True (especially for heavy i/o), but portability, collaboration in consortium, service, long-term possibility to process data

- *Less cost effective than most large centers*

Might be true when considering commercial clouds (again depends on application). Science cloud: in comparison with clusters, probably less costs for IT. Better in optimising load of computing system.

Best solutions

François Arago Centre

Best solution depends on task + politics

GRID approach for heavy + long term + well financed tasks

Cloud environments can be a flexible solution for space projects

But: “The more communication, the worse the performance becomes” (Jackson et al. 2010, IEEE 2nd International Conf. on Cloud Comp. (Cloud Com))

Hybrid cloud solutions appear to satisfy many of the demands of space missions

Commercial cloud for temporary needs only

Example: recommendation of CNES to use federated clouds for eLISA (ESA’s large mission L3 in the Cosmic Vision program)

How to find for each project the best way to do their processing?

Interface center, providing some processing power, connect and train for use of HPC and GRID, train users on the use of cloud systems.

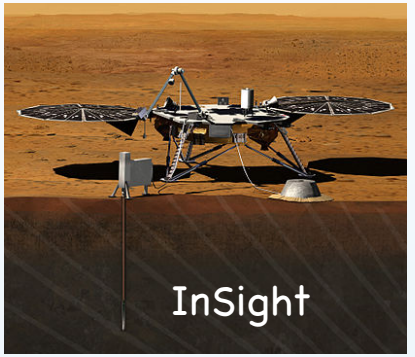
François Arago Centre

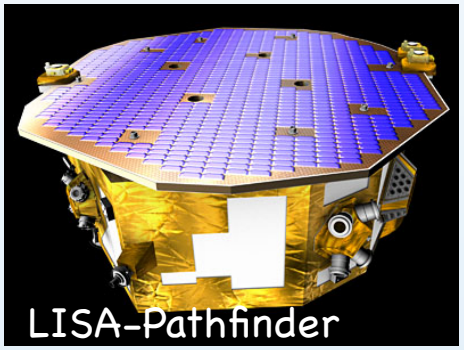
- A data treatment centre for space based projects
- Heavy support by CNRS/IN2P3 & CNES
- Many partners
- strong link to CC-IN2P3
- Project at APC, opened October 2010

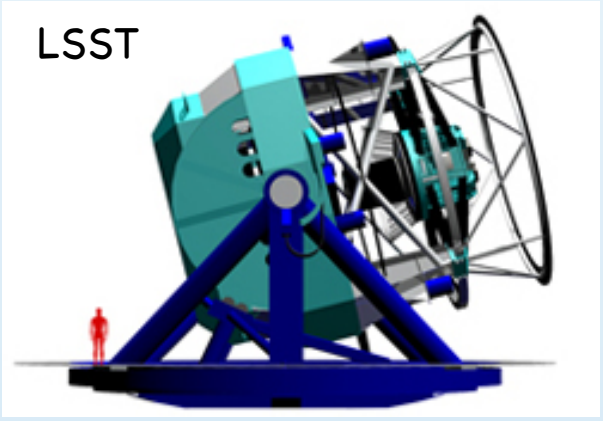
IN2P3: *Institut national de physique nucléaire et de physique des particules*

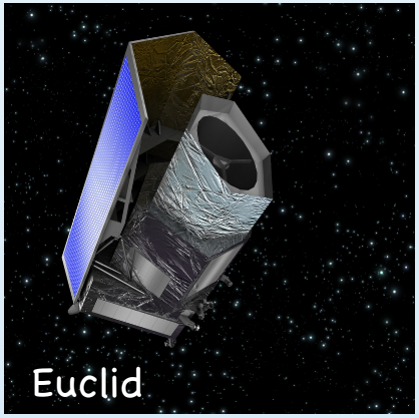
Projects

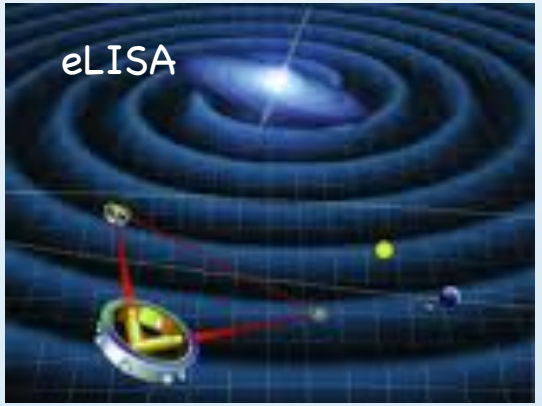
- INTEGRAL (X-/gamma-ray)
- Euclid (optical/IR satellite)
- LISA-Pathfinder, eLISA
- HESS, CTA (VHE)
- LSST (optical; US)
- SVOM (Chinese-French GRB)
- Astro-H (X-ray; Japanese)
- InSight (Mars mission; NASA)



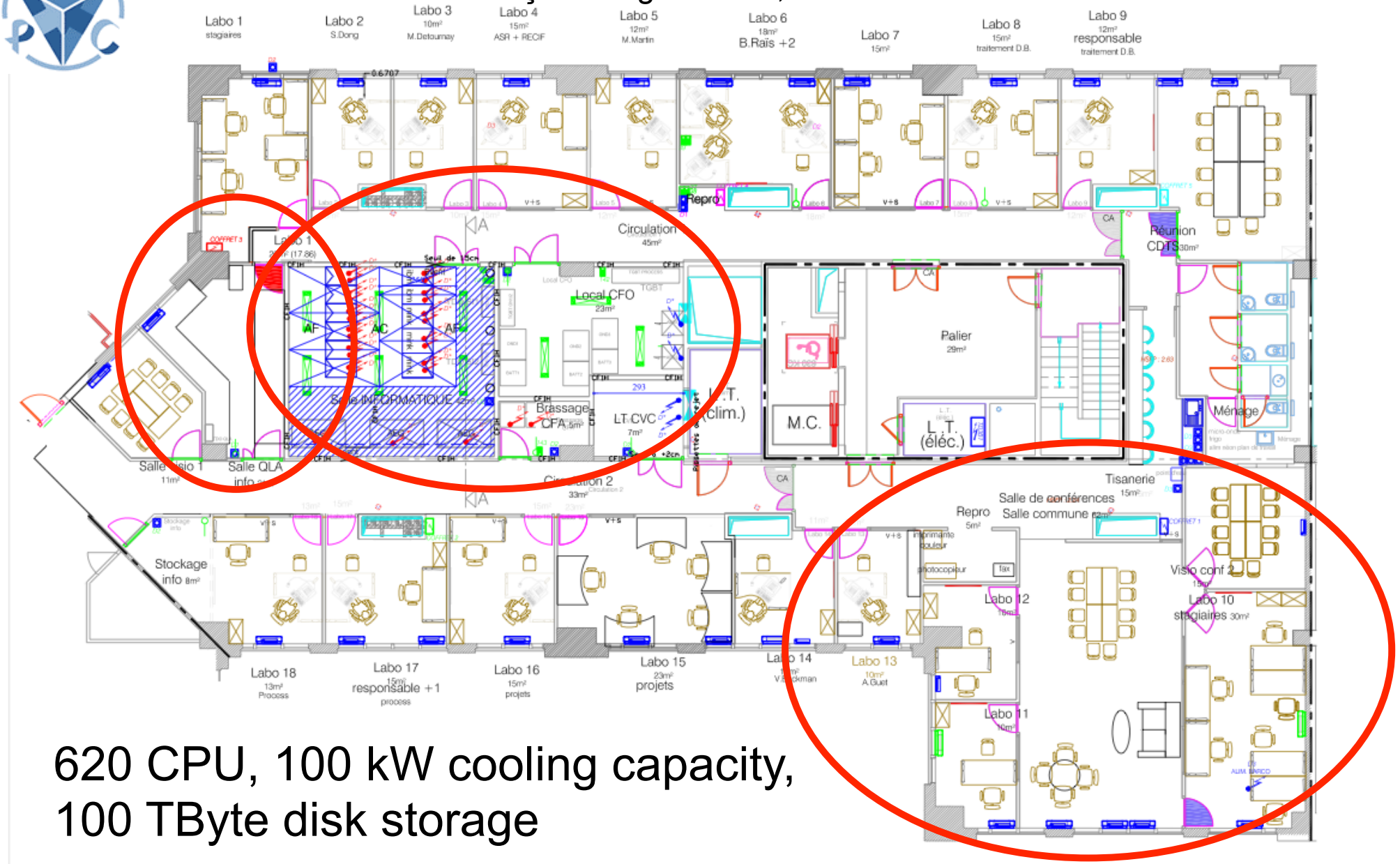








François Arago Centre, 13 Rue Watt



620 CPU, 100 kW cooling capacity,
100 TByte disk storage

Operations room, Concurrent Design Facility

“Quick Look” room

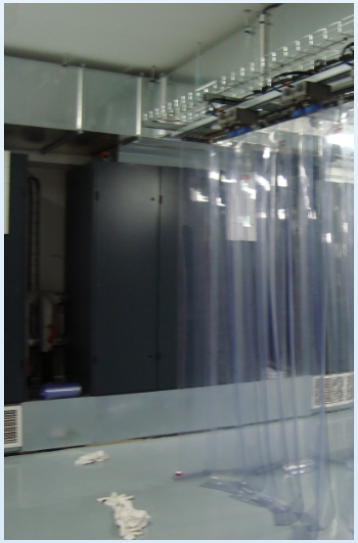
François Arago Centre

François Arago Centre

Operations room

Quick-look room

Battery backup



Computer room

Infrastructure

- Cluster, 640 CPU, 100 kW cooling, 100 TByte disk
- 10 Gbit/s connection
- 2 video conference rooms
- Concurrent Design Facility
- 2 meeting rooms
- Offices on demand
- Hard- and software support

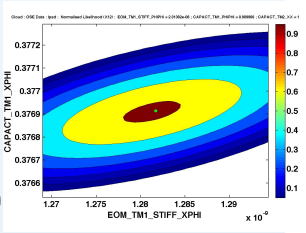
Concurrent Design Facility

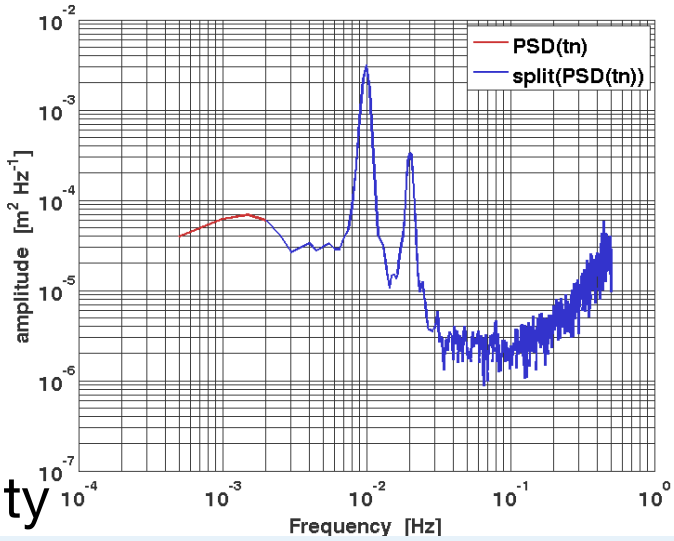
François Arago Centre

- R&D activities
- coordination and collaboration with CNES, ESA, NASA, and other large international partners
- Computing network, videoconference hard- and software
- Installed in 2011 and used e.g. in the re-definition of eLISA and for the *LISA Pathfinder data challenges*.

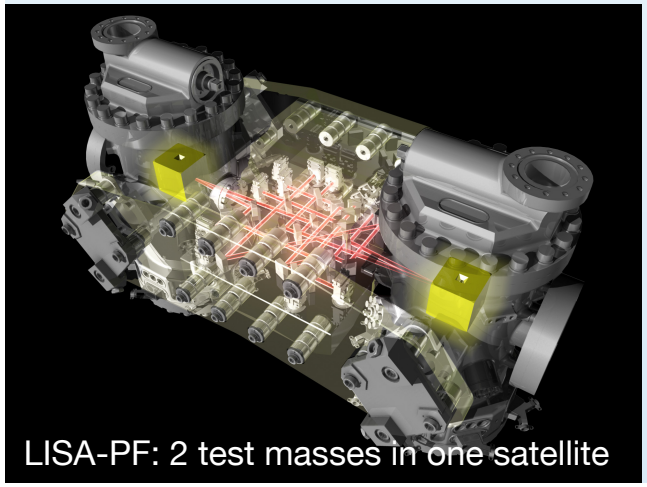
Accomplishments LISA-PF

- Demonstrate gravitational wave measurement techniques in space
- FACe is the Complementary Data Centre for the LISA-Pathfinder mission (STOC @ ESAC, MOC @ ESOC)
- s/w development
- Improve processing speed by parallelisation
- LISA-PF processing in the cloud





- Data challenges (2-3 times a year)
- FACe is the centre to be for the LPF community
- Launch summer 2015, 9 month main activity, post-mission

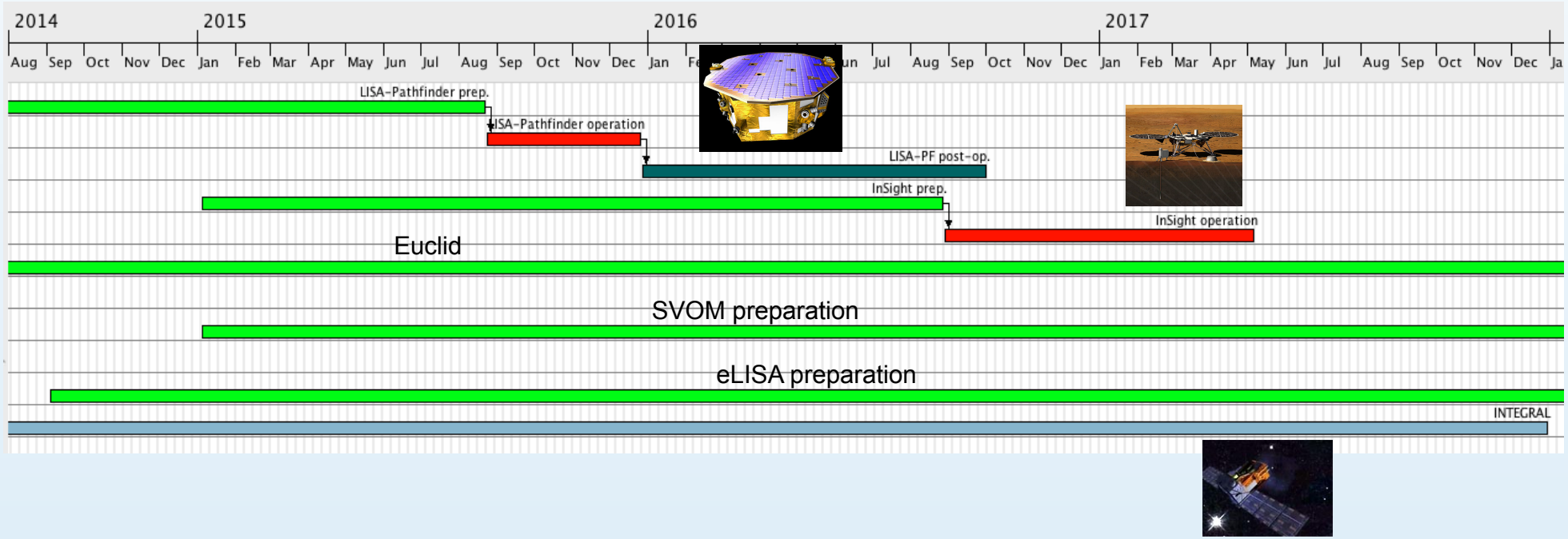

LISA-PF: 2 test masses in one satellite


LISA-PF data challenge 2012


Meeting with ESA in the CDF

Perspective

- FACe established as data analysis and operations centre for astroparticle physics
- near future (until 2020) dominated by INTEGRAL, LISA-Pathfinder, eLISA preparation, InSight, Euclid, SVOM, and LSST



Conclusions

- Best processing solution project-dependent
- The François Arago Centre is a multi-project centre
- Expertise in space projects (e.g. INTEGRAL, LISA-PF)
- Ready to provide services to upcoming missions
- The centre is up and running since autumn 2010:
  http://www.apc.univ-paris7.fr/FACe/