VERA C. RUBIN OBSERVATORY

Brazilian IDAC

Julia Gschwend
on behalf of Carlos Adean and IDAC team

NSF AURA U.S. DEPARTMENT OF ENERGY SLAC CHARLES AND LISA SIMONYI FUND FOR ARTS AND SCIENCES LSST CORPORATION

Brazilian IDAC

BRA-LIN in-kind contribution program

- **Lite IDAC**
- Software + Data Products
  - DESC Pipe. Sci.
  - PZ Server
  - PZ Training Set Maker
  - PZ Compute
  - FTEs for commissioning process

LIneA = Inter-institutional Laboratory of e-Astronomy
(but in Portuguese)
www.linea.org.br

Carlos Adean
IDAC Contribution Lead

BRA-LIN key-people:

- **Program Lead: Luiz da Costa**
- Program Manager: Julia Gschwend

- **IDAC Cont. Lead: Carlos Adean**
- PZ Cont. Lead: Julia Gschwend
- DESC Pipeline Scientist: Sandro Vitenti

- In-kind Program Coordinator (from Rubin): Aprajita Verma

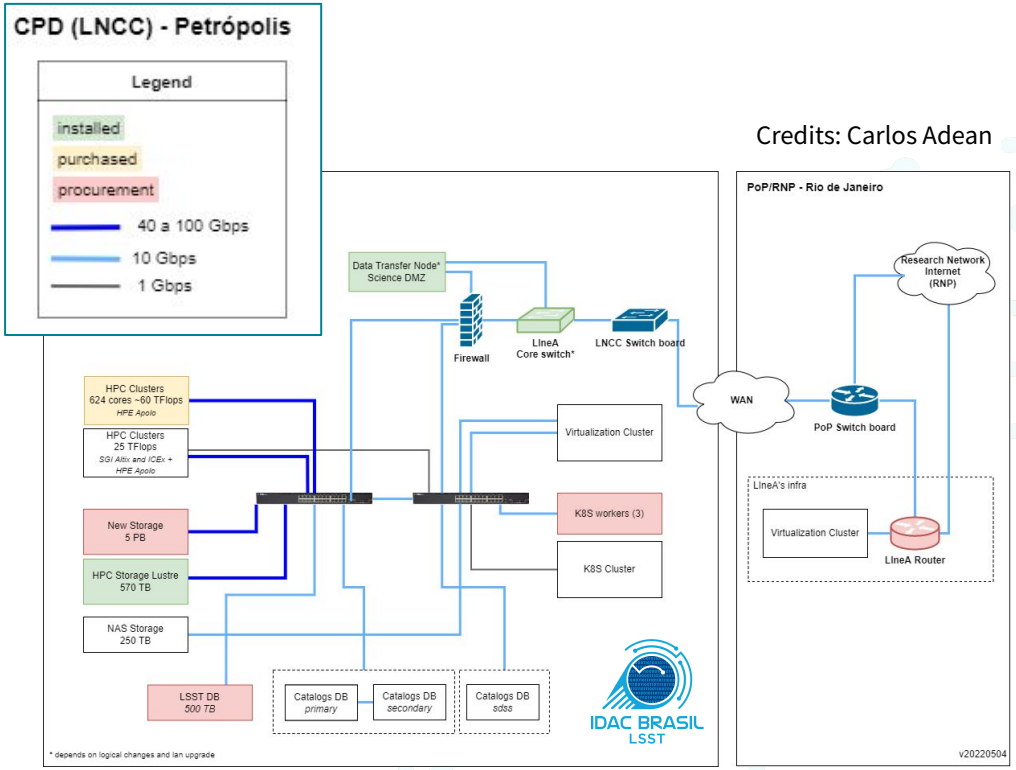
Brazilian IDAC - status and plans

Infrastructure layer

- ✓ Cluster 500 cores + access to supercomputer
- ✓ Lustre system (570 TB+)
- ✓ Definition of database (single server)
- ✓ Kubernetes
- ✓ CILogon (Satosa, COmanage, SIRTFI)
- ✓ Meetings with vendors
- ○ Analysis of ~20 proposals
- ○ Purchase and installation of new database server (postgreSQL)
- ○ Purchase Juniper router
- ○ Upgrade LAN
- ○ Development of data ingestion pipeline

✓ completed
○ ongoing/next

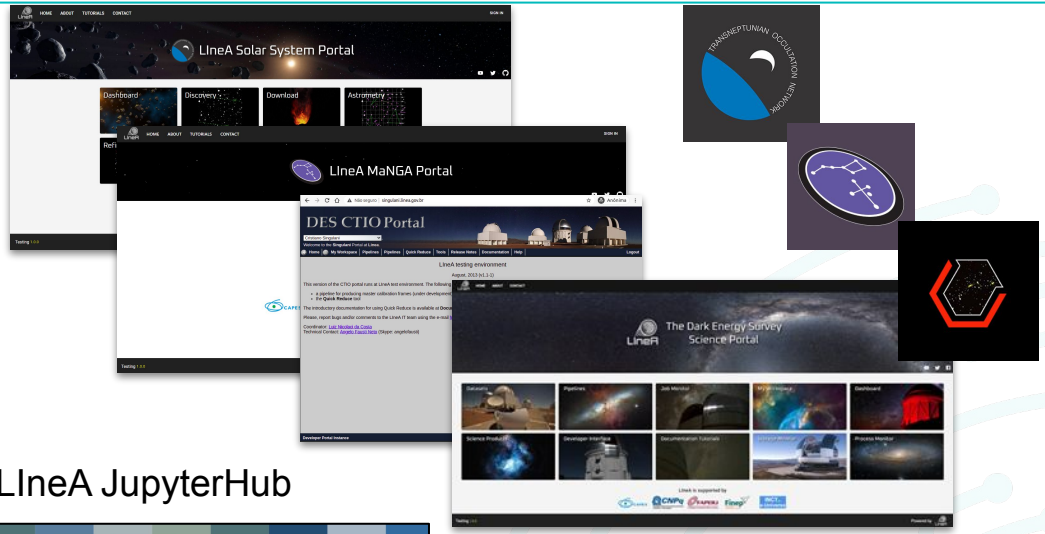
Credits: Carlos Adean



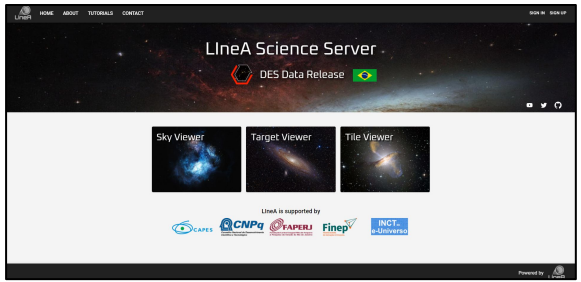
Brazilian IDAC - status and plans

Software layer

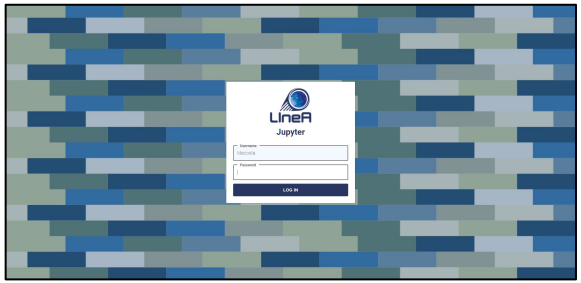
- LIneA Science Platform
  - ✓ JupyterHub (gold, silver, bronze)
  - ✓ Science Server
  - ✓ Daiquiri
  - ○ Integration of all services
  - ○ Single landing page (like RSP)
  - ○ Cross-match service* (AXS?)



LIneA Science Server



LIneA JupyterHub



Registered users

Science Server: 620
JupyterHub: 49

Brazilian IDAC - use case 1

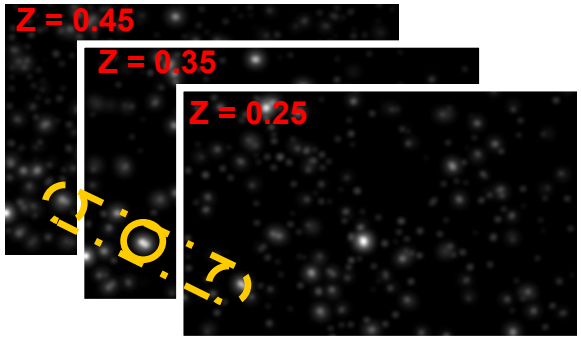
Galaxy Cluster (WaZP) & Stellar System (GAWA) finders

- What LSST and other data do you need? What information should an Object Lite data product contain?

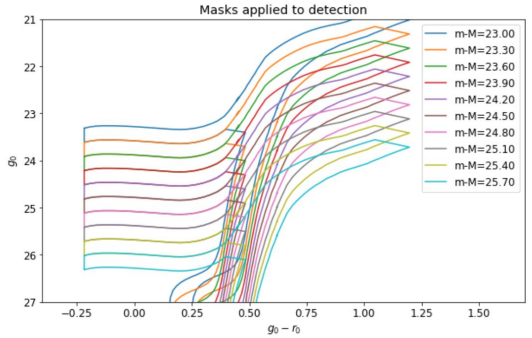
  Coordinates, photometry (6-band fluxes or magnitudes + errors) (photo-z related ancillary data), s/g classification flags, quality flags (SExtractor-like), depth maps (healpix/healsparse), X-Ray surveys (e.g. Chandra, XMM), SZ (e.g., ACT, SPT). Catalog of NGC objects.
- What software will you be running? Will you want an interactive environment or mainly be submitting batch computing jobs?

  WaZP Cluster Finder, Photo-z algorithms, cross-matching to other surveys (ClEvaR). Mostly batch computing jobs, but also interactive analysis of outputs.
- Does the computational workflow place particular demands on the hardware or software services?

  Massive parallel data processing, position-dependent, area chunks (w/ treatment of borders) plus 3rd dimension slicing.



Enabling Science program
Project info

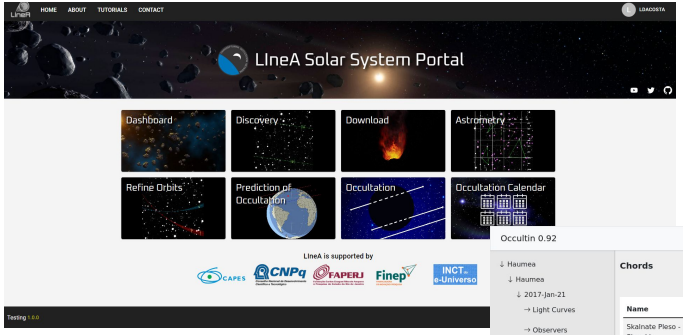


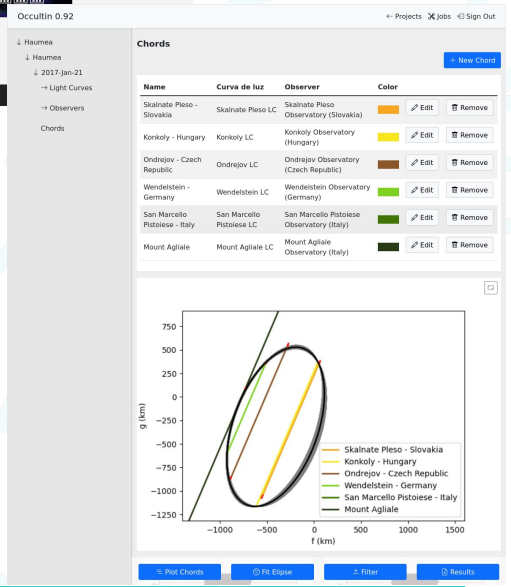
Kickstarter project
Project info
Documentation

Brazilian IDAC - use case 2

Solar System:

- ➢ Occultation Predictions (SS Portal)
- ➢ Analysis (SORA)



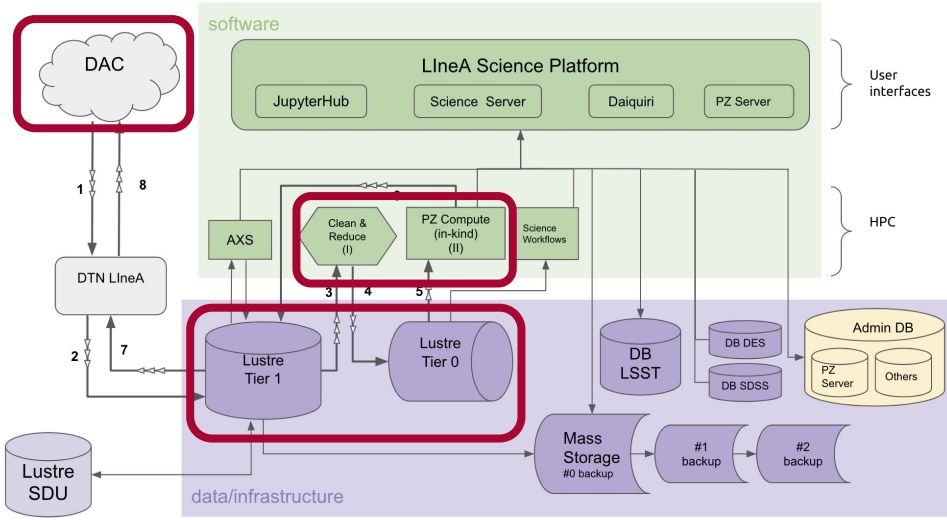


- What LSST and other data do you need?
  What information should an Object Lite data product contain?
  Astrometric positions, single-epoch photometry.
- What software will you be running? Will you want an interactive environment or mainly be submitting batch computing jobs?
  Occultations prediction, Orbit determination. Both interactive and batch computing are needed.
- Does the computational workflow place particular demands on the hardware or software services?
  Mainly on software and queries (bandwidth). Use of external data such as JPL ephemeris also place a different kind of demand.

Brazilian IDAC - use case 3

Photo-z Tables as federated datasets

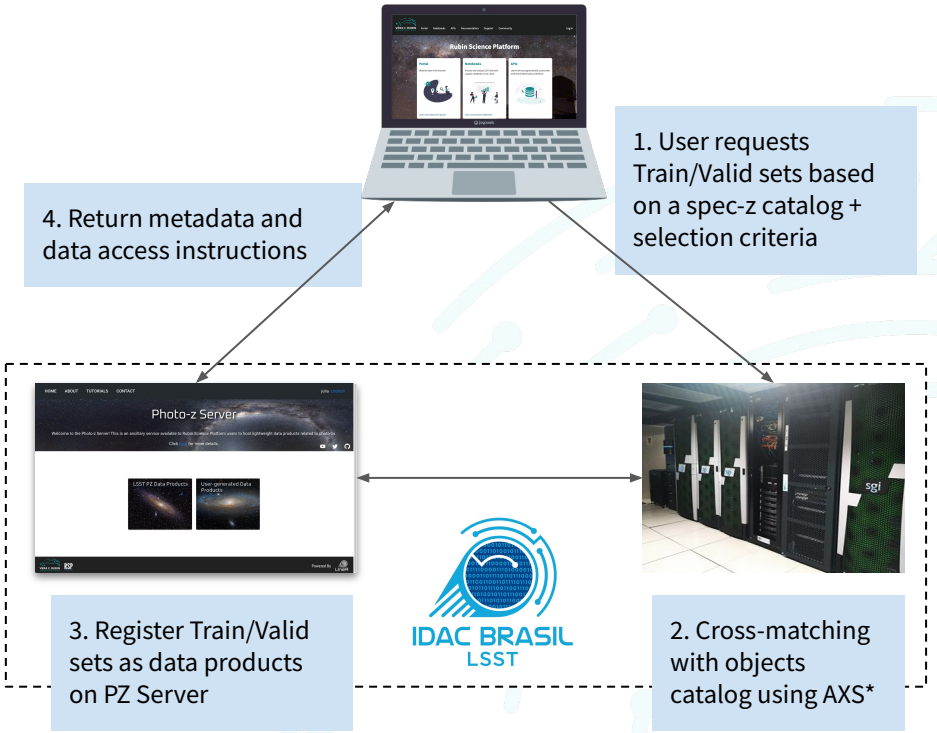
- What LSST and other data do you need? What information should an Object Lite data product contain?
  Photometry (6-band fluxes or magnitudes + errors), lightweight ancillary data (spec-z samples, priors, SED templates, etc)
- What software will you be running? Will you want an interactive environment or mainly be submitting batch computing jobs?
  Mostly batch computing jobs: Parsl workflow* (maybe use RAIL from DESC) with a photo-z code (yet to be defined by DM) plugged in as a blackbox. Afterburners for outputs characterization.
- Does the computational workflow place particular demands on the hardware or software services?
  Massive embarrassing parallel processing. No dependency on position (input easily balanced partitioning). Outputs potentially large (Photo-z PDFs, compressed?). Processing can be done partially as the data chunks arrive.



Brazilian IDAC - use case 4

Photo-z - Training Set Maker

- What LSST and other data do you need? What information should an Object Lite data product contain?
  Positions, Photometry (6-band fluxes or magnitudes + errors), other parameters, e.g. shape/size (less likely), spec-z samples (provided separate from the Objects catalog).
- What software will you be running? Will you want an interactive environment or mainly be submitting batch computing jobs?
  Positional cross-matching based on Spark (AXS). Register data and metadata. API for the users jobs submission and data manipulation.
- Does the computational workflow place particular demands on the hardware or software services?
  Spark cluster, data installed in AXS format (partitioned by zones). Authentication and authorization of users.



Brazilian IDAC

Burning questions (some remained from last year's IDAC session)

- (1) Will the data from DP1 and DP2 be available for the IDACs ?
- (3) From where will the IDACs get their data?
- (4) Which authentication/authorization system will be used?
- (5) Will the IDACs be able to use the DAC services for image display/cutouts?
- (8) Science-ready catalogs will also require some ancillary data, e.g. coverage maps - will these ancillary information be accessible and provided to the IDACs?

Some new ones:

- What is the recommended tool for international data transfer (e.g., Rucio, Globus, Aria2C)
- What is the data policies and governance? (list of related questions in an extra slide)
- Will IDACs provide HPC resource for users? Or just JupyterHub?
- How to provide image server, cutouts, mosaics? Produce ou download hips images? And ptiffs?

Brazilian IDAC

Extra slides

Brazilian IDAC

- What is the data policies and governance?
  - What is the size and time limit (query) for mydb ? Any special requirements?
  - What is the size and limit for user's files in directories?
  - How long user data should be available? Purge policies? Garbage collector?
  - How to protect the data?

Useful links:
Knut's document IDAC Knowledge Base and FAQ
RTN-003
RDO-13