O programa de contribuição *in-kind* BRA-LIN

Uma parceria entre o LIneA e o Rubin Observatory

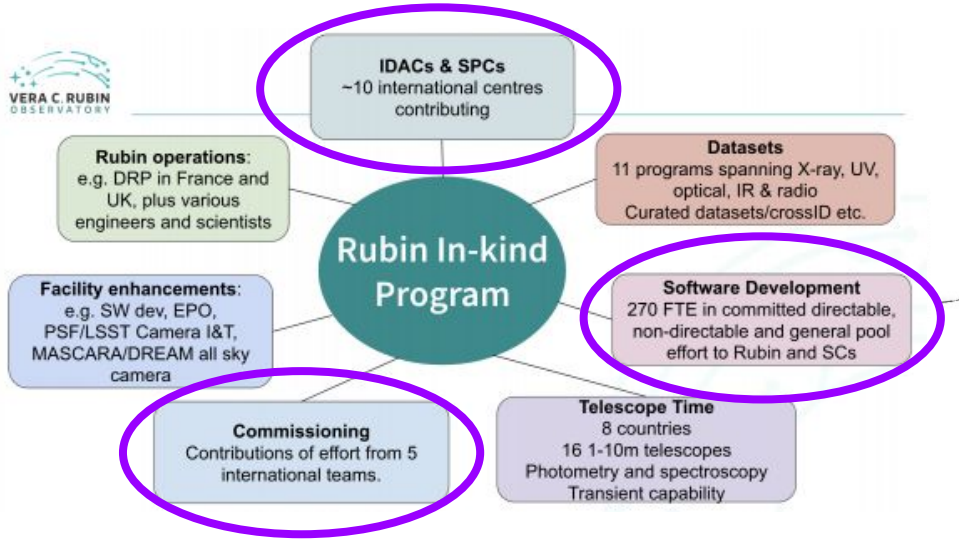
Workshop
Workflows e Plataformas Científicas: preparação para o LSST DR1

Julia Gschwend

25.03.2025

The Rubin LSST In-kind Program

https://www.lsst.org/scientists/in-kind-program



153 in-kind contributions to Rubin and the LSST science community

43 individual international teams from **28** countries + US/Chile

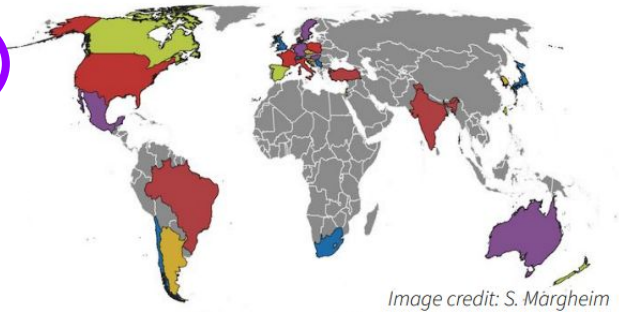


Image credit: S. Margheim

BRA-LIN in-kind contribution program

country institution

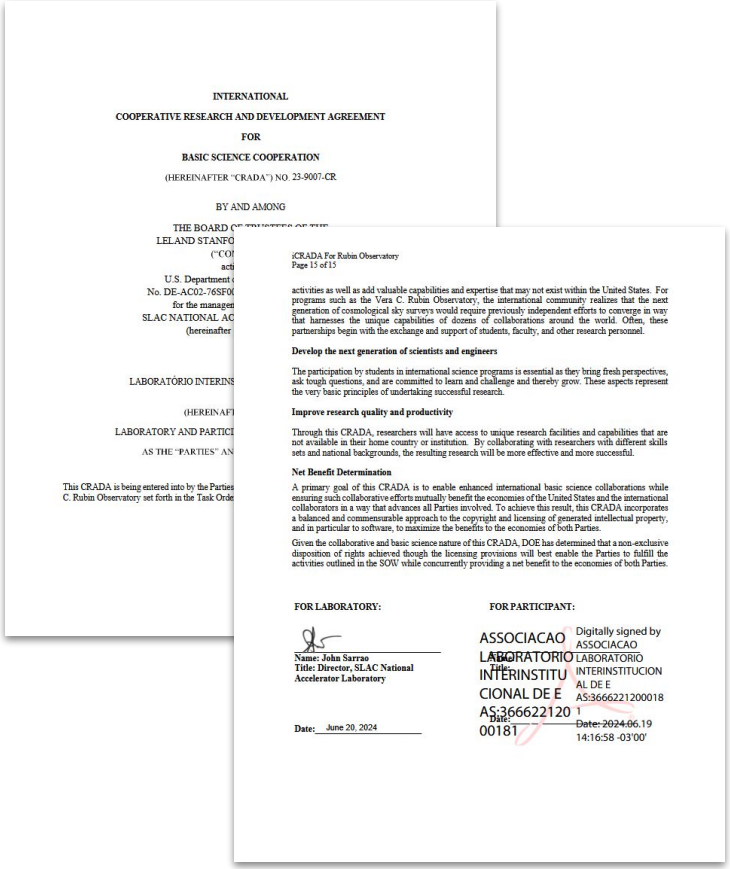
slide credits: Aprajita Verma - LSST PCW 2023

BRA-LIN in-kind contribution program

Computing infrastructure, software products, human resources

Office of Science

Data rights for 80 Brazilian scientists (16 PIs and 64 JAs)



INTERNATIONAL

COOPERATIVE RESEARCH AND DEVELOPMENT AGREEMENT

FOR

BASIC SCIENCE COOPERATION

(HEREINAFTER "CRADA") NO. 23-9007-CR

BY AND AMONG

iCRADA For Rubin Observatory
Page 15 of 15

activities as well as add valuable capabilities and expertise that may not exist within the United States. For programs such as the Vera C. Rubin Observatory, the international community realizes that the next generation of cosmological sky surveys would require previously independent efforts to converge in way that harnesses the unique capabilities of dozens of collaborations around the world. Often, these partnerships begin with the exchange and support of students, faculty, and other research personnel.

Develop the next generation of scientists and engineers

The participation by students in international science programs is essential as they bring fresh perspectives, ask tough questions, and are committed to learn and challenge and thereby grow. These aspects represent the very basic principles of undertaking successful research.

Improve research quality and productivity

Through this CRADA, researchers will have access to unique research facilities and capabilities that are not available in their home country or institution. By collaborating with researchers with different skills sets and national backgrounds, the resulting research will be more effective and more successful.

Net Benefit Determination

A primary goal of this CRADA is to enable enhanced international basic science collaborations while ensuring such collaborative efforts mutually benefit the economies of the United States and the international collaborators in a way that advances all Parties involved. To achieve this result, this CRADA incorporates a balanced and commensurable approach to the copyright and licensing of generated intellectual property, and in particular to software, to maximize the benefits to the economies of both Parties.

Given the collaborative and basic science nature of this CRADA, DOE has determined that a non-exclusive disposition of rights achieved though the licensing provisions will best enable the Parties to fulfill the activities outlined in the SOW while concurrently providing a net benefit to the economies of both Parties.

FOR LABORATORY:

Name: John Sarrao
Title: Director, SLAC National Accelerator Laboratory

Date: June 20, 2024

FOR PARTICIPANT:

ASSOCIACAO LABORATORIO INTERINSTITUCIONAL DE E AS:36662212000181

Digitally signed by ASSOCIACAO LABORATORIO INTERINSTITUCIONAL DE E AS:36662212000181
Date: 2024.06.19 14:16:58 -03'00'

Program lead: Luiz da Costa
Program manager: Julia Gschwend

In-kind contributions:

- **S1: Lite IDAC - access to lite catalogs for the whole collaboration**
- **S2 and S3: Pipeline Scientists - software development efforts**
- **S4: Photo-z infrastructure - software and advanced data products periodically delivered**

BRA-LIN - cronograma de entregas

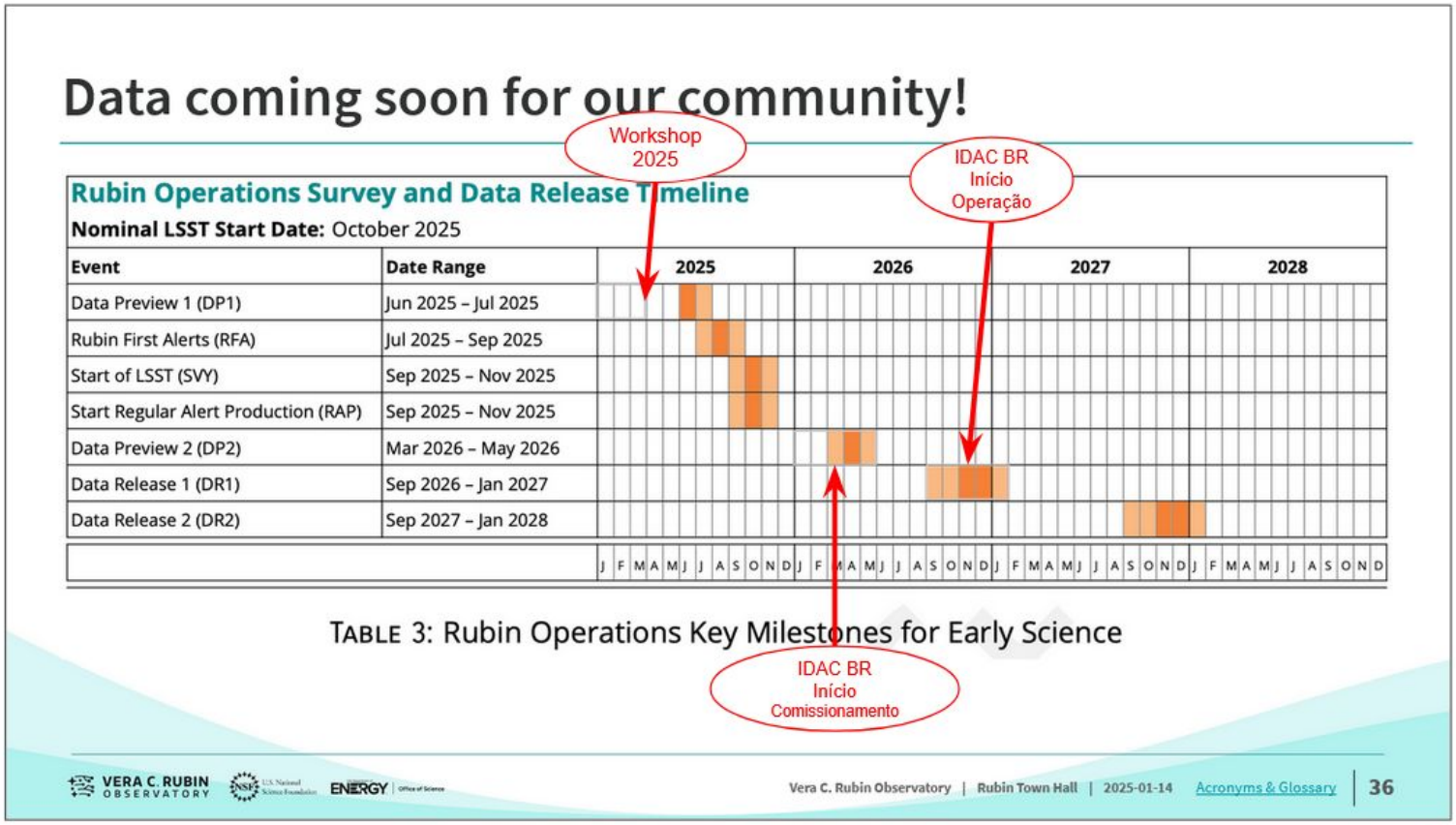
- **S1: Lite IDAC (5 PIs)**
  - 2021 - 2025: Implementation and tests
  - 2026/1: Commissioning
  - 2026/2 - 2039: Operations
- **S2: 1 pipeline scientist for DESC PZ**
  - contribuição retirada da proposta
- **S3: 2 pipeline scientists for DESC TJP (2 PIs)**
  - 2021-2024*: 0.25 FTE each (total 2.0 FTE)
- **S4: Photo-z infrastructure for Rubin DM (8 PIs)**
  - 2021-2024: Implementation and tests
  - 2025-2039: Operations (PZ Operations starts with LSST Commissioning)
- Program Manager **(1 PI)**
  - 2021 - 2039

Acordo de 2015 (Infraestrutura de rede - contribuição da RNP) (8 "historical" PIs)

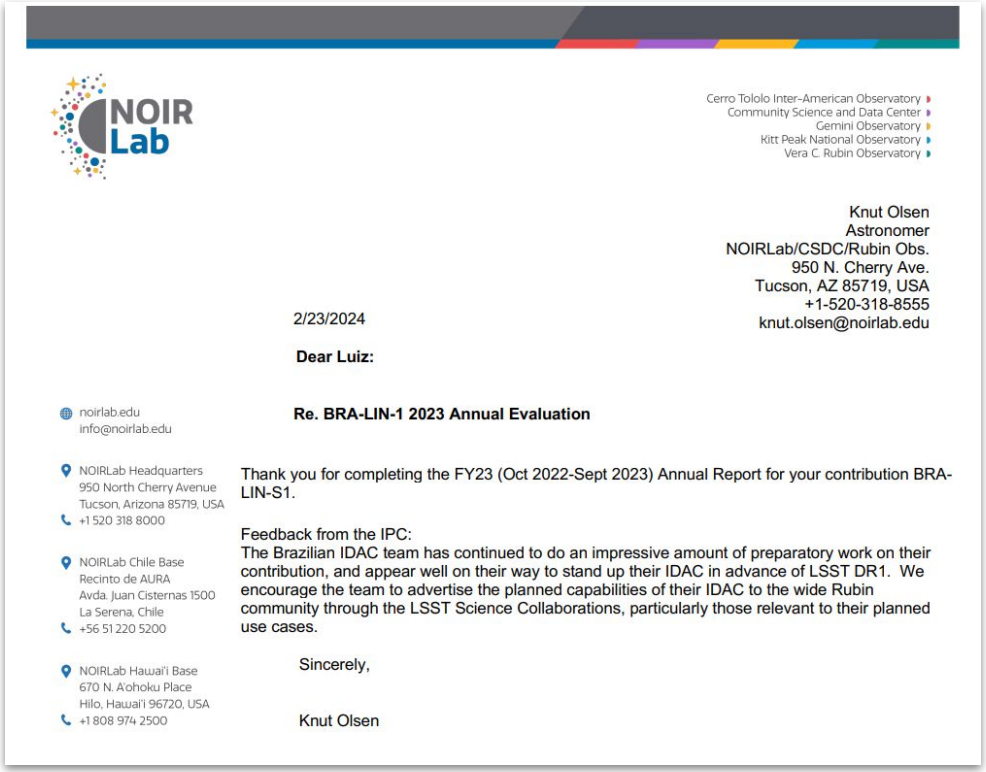
Total: 120 posições para pesquisadores brasileiros (24 PIs + 96 JAs) → Brazilian Participation Group (BPG)

Cronograma oficial de DRs apresentado em janeiro no AAS Town Hall

em vermelho: anotações Carlos Adean



Relatórios avaliações e feedback



NOIR Lab

Cerro Tololo Inter-American Observatory
Community Science and Data Center
Gemini Observatory
Kitt Peak National Observatory
Vera C. Rubin Observatory

Knut Olsen
Astronomer
NOIRLab/CSDC/Rubin Obs.
950 N. Cherry Ave.
Tucson, AZ 85719, USA
+1-520-318-8555
knut.olsen@noirlab.edu

2/23/2024

Dear Luiz:

Re. BRA-LIN-1 2023 Annual Evaluation

noirlab.edu
info@noirlab.edu

NOIRLab Headquarters
950 North Cherry Avenue
Tucson, Arizona 85719, USA
+1 520 318 8000

NOIRLab Chile Base
Recinto de AURA
Avda. Juan Cisternas 1500
La Serena, Chile
+56 51 220 5200

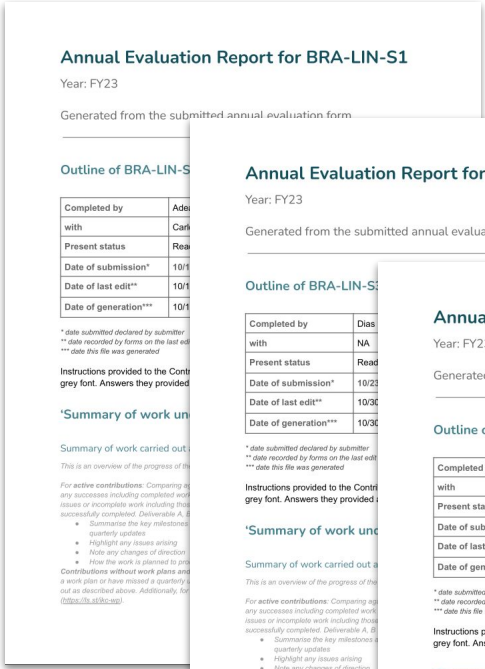
NOIRLab Hawai'i Base
670 N. A'ohoku Place
Hilo, Hawai'i 96720, USA
+1 808 974 2500

Thank you for completing the FY23 (Oct 2022-Sept 2023) Annual Report for your contribution BRA-LIN-S1.

Feedback from the IPC:
The Brazilian IDAC team has continued to do an impressive amount of preparatory work on their contribution, and appear well on their way to stand up their IDAC in advance of LSST DR1. We encourage the team to advertise the planned capabilities of their IDAC to the wide Rubin community through the LSST Science Collaborations, particularly those relevant to their planned use cases.

Sincerely,

Knut Olsen



Annual Evaluation Report for BRA-LIN-S1

Year: FY23

Generated from the submitted annual evaluation form

Outline of BRA-LIN-S

Completed by	Ade
with	Carl
Present status	Rea
Date of submission*	10/1
Date of last edit**	10/1
Date of generation***	10/1

* date submitted declared by submitter
** date recorded by forms on the last edi
*** date this file was generated

Instructions provided to the Contr
grey font. Answers they provided

'Summary of work un

Summary of work carried out

Annual Evaluation Report for BRA-LIN-S3

Year: FY23

Generated from the submitted annual evaluation form

Outline of BRA-LIN-S

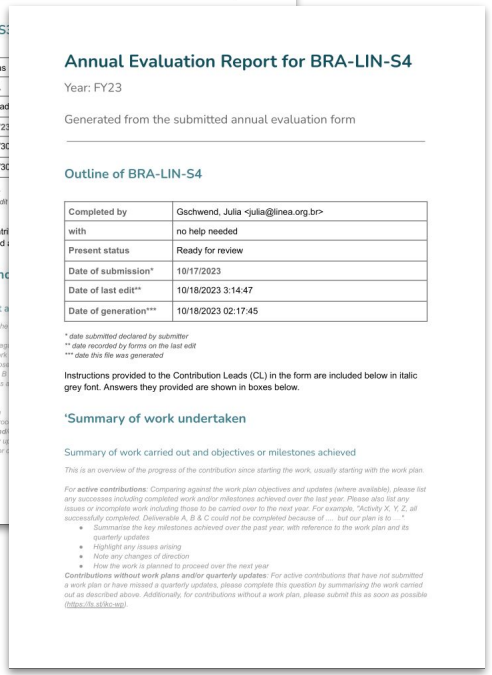
Completed by	Dias
with	NA
Present status	Read
Date of submission*	10/23
Date of last edit**	10/30
Date of generation***	10/30

* date submitted declared by submitter
** date recorded by forms on the last edit
*** date this file was generated

Instructions provided to the Contr
grey font. Answers they provided

'Summary of work und

Summary of work carried out



Annual Evaluation Report for BRA-LIN-S4

Year: FY23

Generated from the submitted annual evaluation form

Outline of BRA-LIN-S4

Completed by	Gschwend, Julia <julia@linea.org.br>
with	no help needed
Present status	Ready for review
Date of submission*	10/17/2023
Date of last edit**	10/18/2023 3:14:47
Date of generation***	10/18/2023 02:17:45

* date submitted declared by submitter
** date recorded by forms on the last edit
*** date this file was generated

Instructions provided to the Contribution Leads (CL) in the form are included below in italic grey font. Answers they provided are shown in boxes below.

'Summary of work undertaken

Summary of work carried out and objectives or milestones achieved

This is an overview of the progress of the contribution since starting the work, usually starting with the work plan.

For ***active contributions****: Comparing against the work plan objectives and updates (where available), please list any successes including completed work and/or milestones achieved over the last year. Please also list any issues or incomplete work including those to be carried over to the next year. For example, "Activity X, Y, Z, all successfully completed. Deliverable A, B & C could not be completed because of …. but our plan is to …".*

- *Summarise the key milestones achieved over the past year, with reference to the work plan and its quarterly updates*
- *Highlight any issues arising*
- *Note any changes of direction*
- *How the work is planned to proceed over the next year*

Contributions without work plans and/or quarterly updates*: For active contributions that have not submitted a work plan or have missed a quarterly updates, please complete this question by summarising the work carried out as described above. Additionally, for contributions without a work plan, please submit this as soon as possible (https://ls.st/ikc-wp).*

The IDAC

BRA-LIN-S1: The Brazilian IDAC

Contribution Lead: Carlos Adean

IDAC Lite Resources

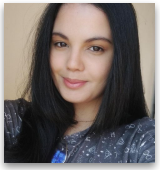
- 5 PB storage
- 500 TB database PostgreSQL
- 500 cores HPC (~25 TFlops)
- 20 Gbps network
- Kubernetes cluster (JupyterHub)

(2+1) annual data releases

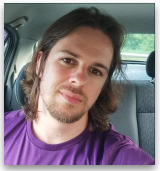
https://www.linea.org.br/idac

Current IDAC Infrastructure Team

Carlos Adean
System Analyst
IDAC Contribution Lead



Nubia Garcia
System Analyst



Eloir Troyack
System Analyst

IDAC BRASIL

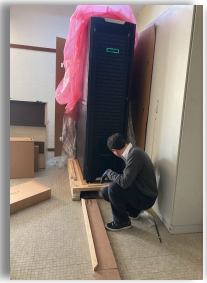
Financiado por:

Finep
INOVAÇÃO E PESQUISA

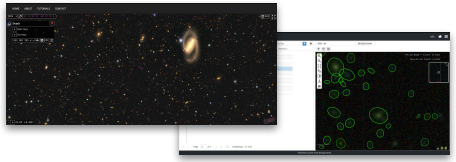
INCT do e-Universo

LineA

O IDAC Brasil já é realidade! (fotos de 2024)



Brazilian IDAC User Interfaces

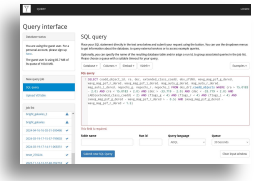


Sky Viewer / Target Viewer
Image visualization
and catalog overlay

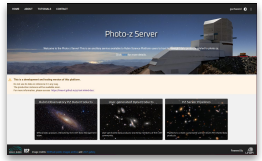
jupyterhub

OPEN OnDemand + jupyter
interactive app JupyterLab

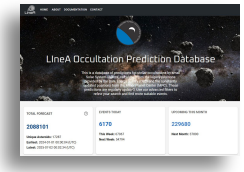
IDAC BRASIL
LSST



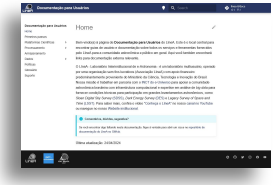
User Query Tool
SQL/ADQL
MyDB



PZ Server



Solar System Portal



Documentation Site
user manuals
tutorials

LIneA

Não perca! Amanhã:

15:35	Plataformas científicas para exploração dos dados do LSST & *LSST Data Products - Definições e características dos dados do LSST*	Julia Gschwend

Tópicos abordados:

- Quais tipos de dados o Rubin Observatory vai disponibilizar?
  - para os membros do projeto X para o público geral
  - via DAC X via IDACs
- Onde e como acessar os dados?
  - do LSST em geral
  - hospedados no IDAC-Brasil
- Onde obter informações?
- Como obter ajuda?

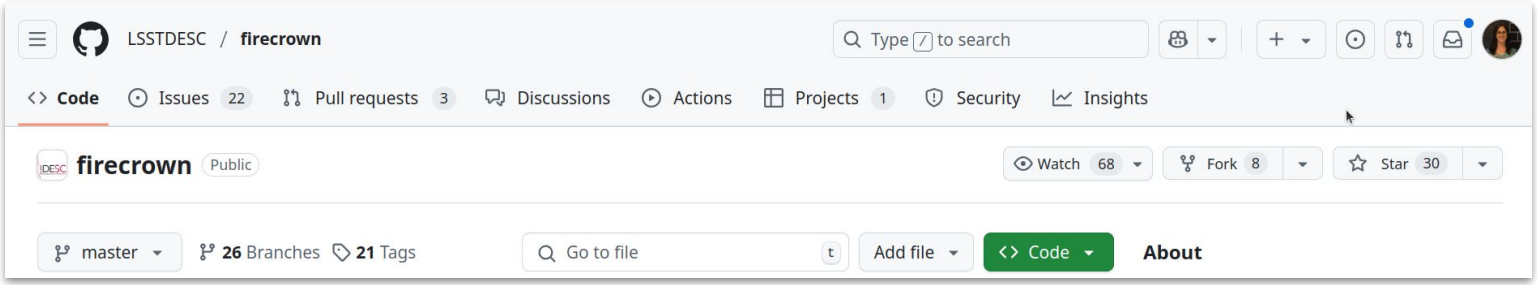
DESC Pipeline Scientists

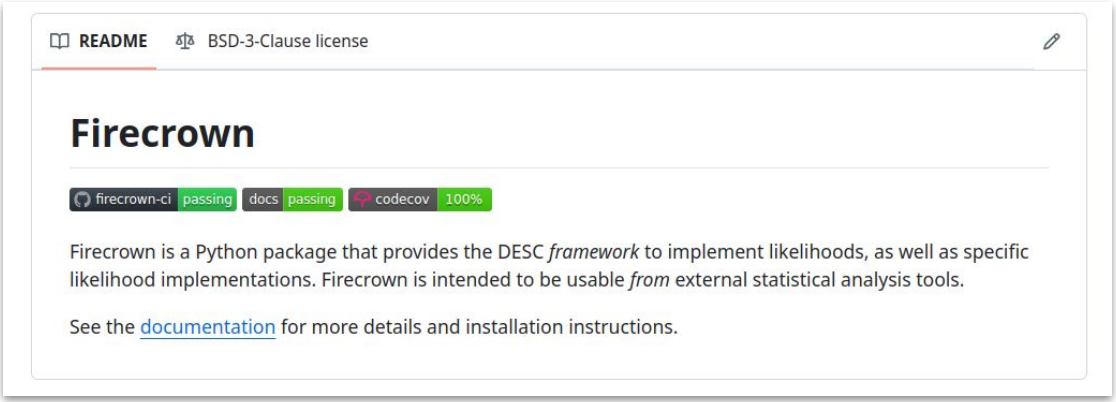
Firecrown

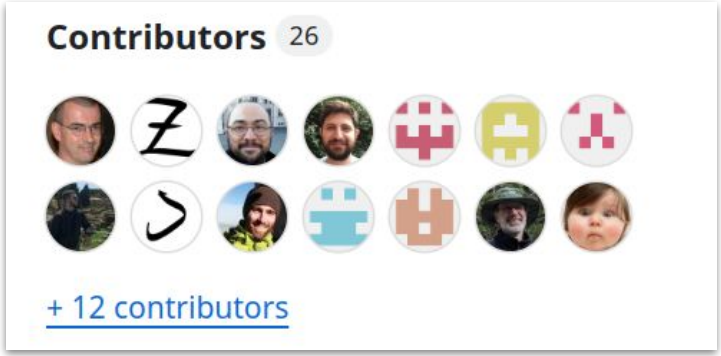
Contribution Lead: Dr. Sandro Vitenti

Recipient: LSST DESC TJP Working Group

Entregue! ✅







ML applications for WL

Contribution Lead: Dr. Clecio de Bom

Recipient: LSST DESC Time Domain WG

Primary goals:

Construct a pipeline to find strongly lensed supernovae, and support infrastructure for strong lens finders.

- Tentativa de suprir as 0.85 FTEs que ficaram pendentes com a saída do Felipe Oliveira.
- Ainda em fase de negociação com o recipient group. As entregas começam a contar a partir dos relatórios trimestrais.
- Detalhes na apresentação do Clécio de Bom.

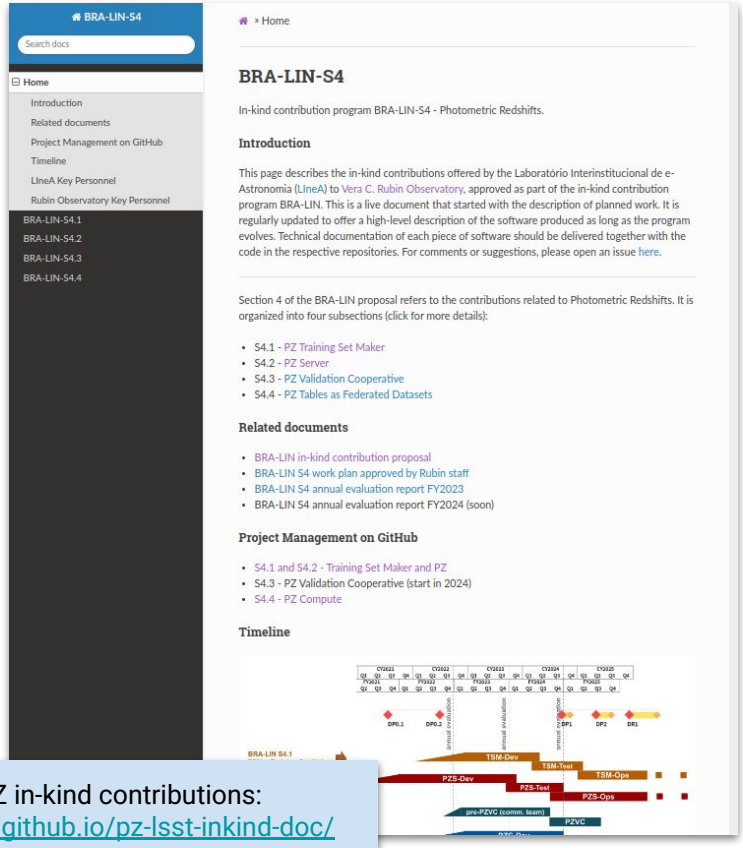
PZ Services

BRA-LIN-S4: Photo-z Services

Contribution Lead: Julia Gschwend

Four in-kind contributions

- S4.1: Training Set Maker
- S4.2: PZ Server
- S4.4: PZ Compute
- S4.3: PZ Validation Cooperative



Overview of PZ in-kind contributions:
https://linea-it.github.io/pz-lsst-inkind-doc/

Motivation:

Data Management Roadmap to photo-z for LSST Objects (ls.st/dmtn-049)

Previous experience with Photo-z Pipelines and production of value-added catalogs in the DES Science Portal

Astronomy and Computing 25 (2018) 58–80

Contents lists available at ScienceDirect

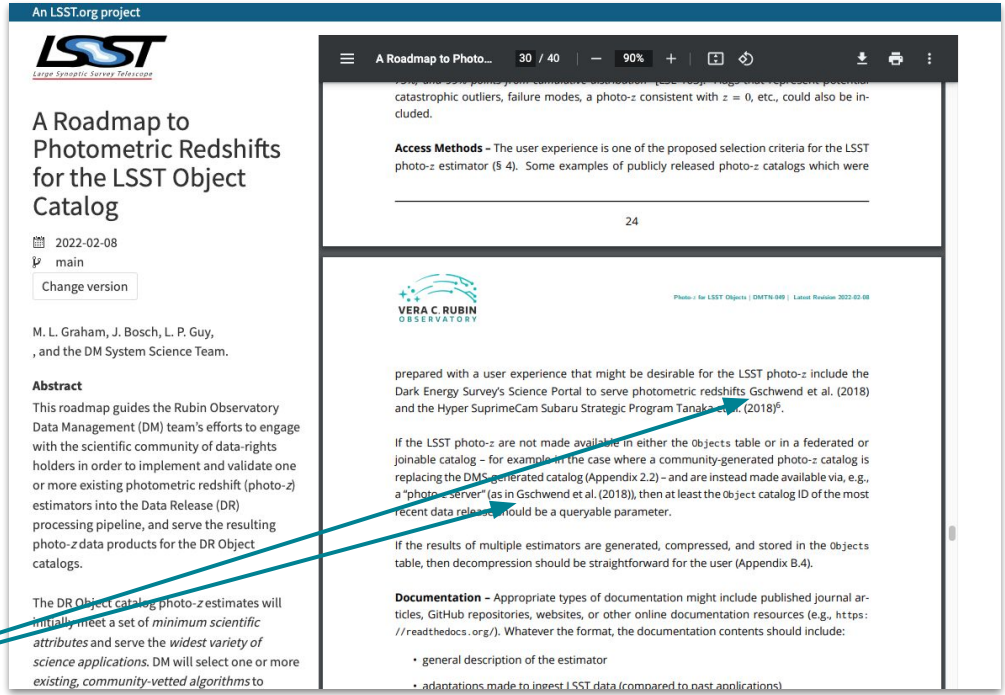
Astronomy and Computing

journal homepage: www.elsevier.com/locate/ascom

Full length article

DES science portal: Computing photometric redshifts

J. Gschwend [1,2,*], A.C. Rossel [1,2], R.L.C. Ogando [1,2], A.F. Neto [3,2], M.A.G. Maia [1,2], I.N. da Costa [1,2], M. Lima [4,2], P. Pellegrini [1,2], R. Campisano [5,2], C. Singulani [1,2], C. Adean [2], C. Benoist [6,2], M. Aguena [1,2], M. Carrasco Kind [7,8], T.M. Davis [9,10], J. de Vicente [11], W.G. Hartley [12], B. Hoyle [13], A. Palmese [12,14], I. Sadeh [15], T.M.C. Abbott [16], F.B. Abdalla [12,17], S. Allam [18], J. Annis [18], J. Asorey [9,19,20], D. Brooks [12], J. Calcino [21], D. Carollo [22], F.J. Castander [20,23], C.B. D'Andrea [24], S. Desai [25], A.E. Evrard [24,26], P. Fosalba [20,23], J. Frieman [18,26], J. García-Bellido [27], K. Glazebrook [28,10], D.W. Gerdes [24,26], R.A. Gruendl [7,8], G. Gutierrez [18], S. Hinton [9,10], D.L. Hollowood [29], K. Honscheid [30,31], J.K. Hoormann [9], D.J. James [32], K. Kuehn [33], N. Kuropatkin [18], O. Lahav [12], G. Lewis [34], C. Lidman [35], H. Lin [18], E. Macaulay [36], J. Marshall [37], P. Melchior [38], R. Miquel [17,39], A. Möller [28,10], A.A. Plazas [40,41], E. Sanchez [11], B. Santiago [12,42], V. Scarpine [18], R.H. Schindler [43], I. Sevilla-Noarbe [11], M. Smith [44], F. Sobreira [45,2], N.E. Sommer [46,10], E. Suchyta [47], M.E.C. Swanson [7], G. Tarle [24], B.E. Tucker [35,10], D.L. Tucker [18], S. Uddin [41], A.R. Walker [16]



An LSST.org project

LSST Large Synoptic Survey Telescope

A Roadmap to Photometric Redshifts for the LSST Object Catalog

2022-02-08

main

Change version

M. L. Graham, J. Bosch, L. P. Guy, , and the DM System Science Team.

Abstract

This roadmap guides the Rubin Observatory Data Management (DM) team's efforts to engage with the scientific community of data-rights holders in order to implement and validate one or more existing photometric redshift (photo-*z*) estimators into the Data Release (DR) processing pipeline, and serve the resulting photo-*z* data products for the DR Object catalogs.

The DR Object catalog photo-*z* estimates will initially meet a set of *minimum scientific attributes* and serve the *widest variety of science applications*. DM will select one or more *existing, community-vetted algorithms* to

A Roadmap to Photo... 30 / 40 90%

catastrophic outliers, failure modes, a photo-z consistent with $z = 0$, etc., could also be included.

Access Methods – The user experience is one of the proposed selection criteria for the LSST photo-z estimator (§ 4). Some examples of publicly released photo-z catalogs which were

24

VERA C. RUBIN OBSERVATORY

Photo-z for LSST Objects | DMTN-049 | Latest Revision 2022-02-08

prepared with a user experience that might be desirable for the LSST photo-z include the Dark Energy Survey's Science Portal to serve photometric redshifts Gschwend et al. (2018) and the Hyper SuprimeCam Subaru Strategic Program Tanaka et al. (2018)[6].

If the LSST photo-z are not made available in either the `Objects` table or in a federated or joinable catalog – for example in the case where a community-generated photo-z catalog is replacing the DMS-generated catalog (Appendix 2.2) – and are instead made available via, e.g., a "photo-z server" (as in Gschwend et al. (2018)), then at least the `Object` catalog ID of the most recent data release should be a queryable parameter.

If the results of multiple estimators are generated, compressed, and stored in the `Objects` table, then decompression should be straightforward for the user (Appendix B.4).

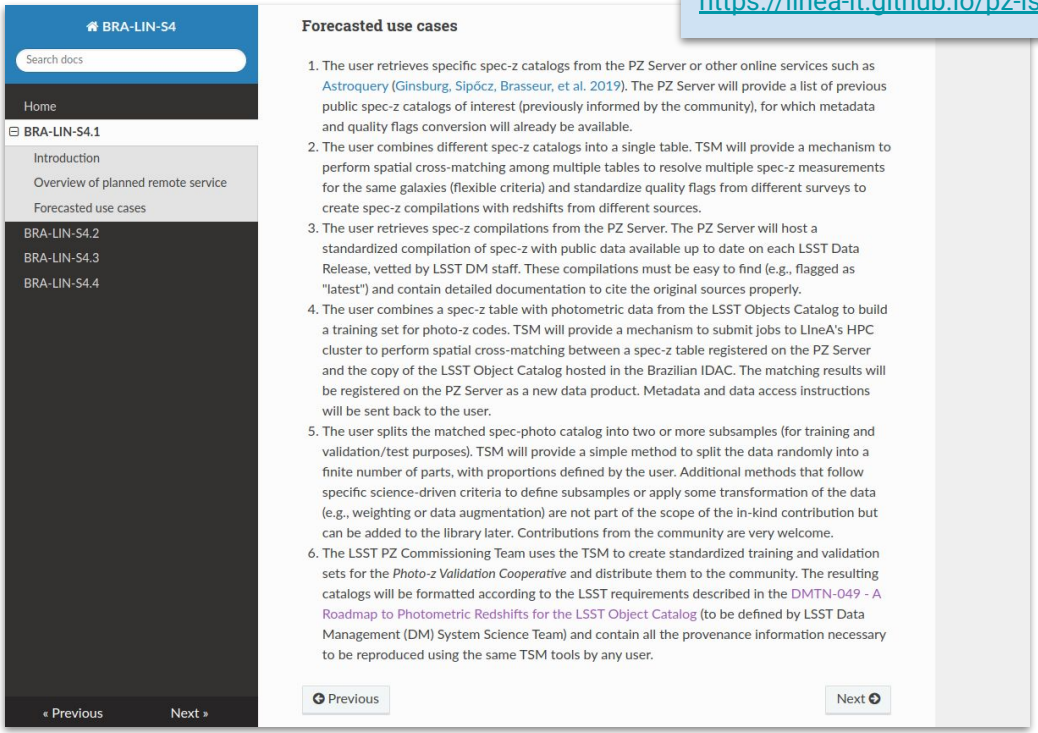
Documentation – Appropriate types of documentation might include published journal articles, GitHub repositories, websites, or other online documentation resources (e.g., `https://readthedocs.org/`). Whatever the format, the documentation contents should include:

- general description of the estimator
- adaptations made to ingest LSST data (compared to past applications)

LineA

Motivation: forecasted use cases

Overview of PZ in-kind contributions:
https://linea-it.github.io/pz-lsst-inkind-doc/



Summary of in-kind contribution **BRA-LIN S4**: Photo-z (PZ) Services

Contribution Lead: Julia Gschwend

Recipient: Rubin Observatory DM team (PZ Coord. Group)

Contact: Melissa Graham

4 Deliverables in Section S4:

S4.1 Standardized training/validation sets

- Data products (recurrent): spec-z compilation, matched spec+photo set
- Software: Python package **Training Set Maker**
- Cross-matching as a service using IDAC's infrastructure

S4.2 PZ Server

- Data management tool and user interface to facilitate sharing results among RSP users (100% software, development/maintenance).
- Web-based platform and API (Python library)

S4.3 Photo-z Validation Cooperative

- FTEs to help on the upcoming Photo-z Validation Cooperative: organization of validation results, write reports, prepare notebooks.

S4.4 PZ tables

- Data products (recurrent): PZ tables as federated datasets available in the Brazilian IDAC.
- Software: pipeline PZ Compute (wrapping of PZ codes and run for massive datasets using the IDAC's computing resources).

The PZ Server

PZ Server overview

A specialized service for creating and hosting pz-related derived data products

Glauber Vila-Verde
Front-end developer

Jandson Vitorino
Front-end developer

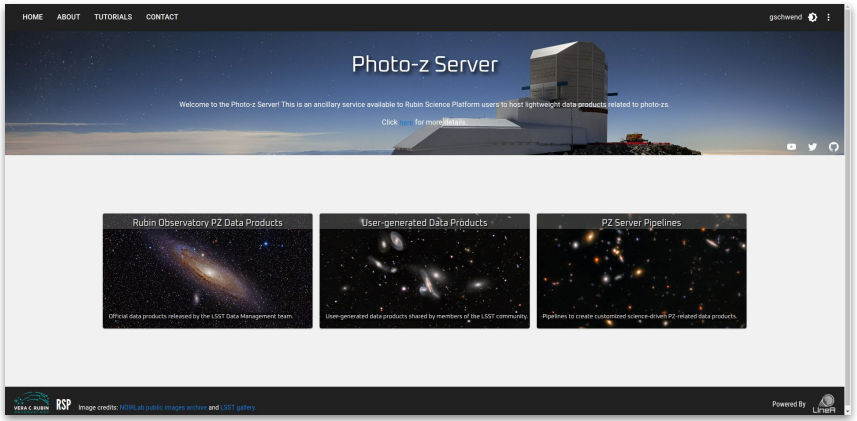
Cristiano Singulani
Back-end developer

Luigi Silva
Scientist

Julia Gschwend
Scientist
Contribution Lead

https://pzserver.linea.org.br

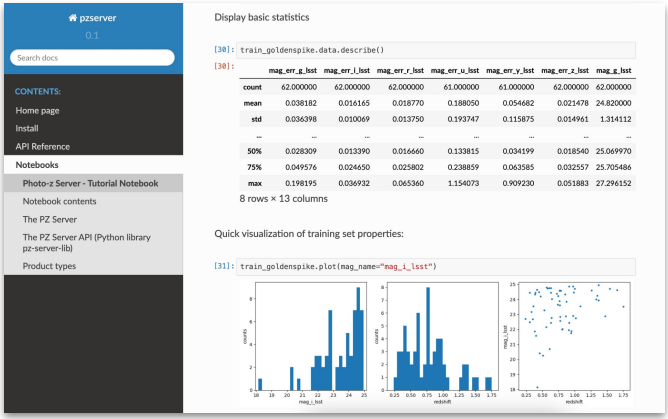
(test environment at https://pzserver-dev.linea.org.br)



Repo: linea-it/pzserver_app

https://linea-it.github.io/pzserver/

Python Library (API) `$ pip install pzserver`



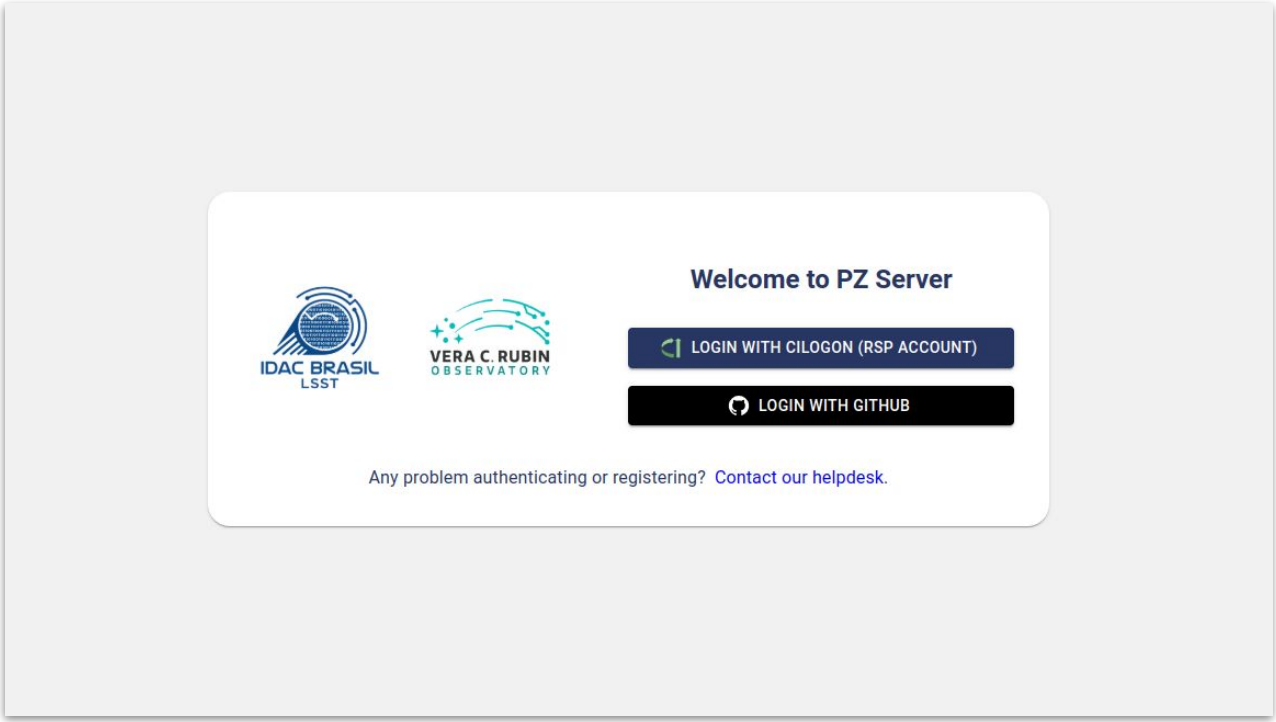
Repo: linea-it/pzserver

PZ Server - A&A with CILogon and Rubin Auth. API

Glauber Vila-Verde
Front-end developer

Cristiano Singulani
Back-end developer

+ IDAC IT team



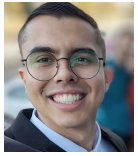
Training Set Maker

PZ Server pipelines

The *Training Set Maker* service

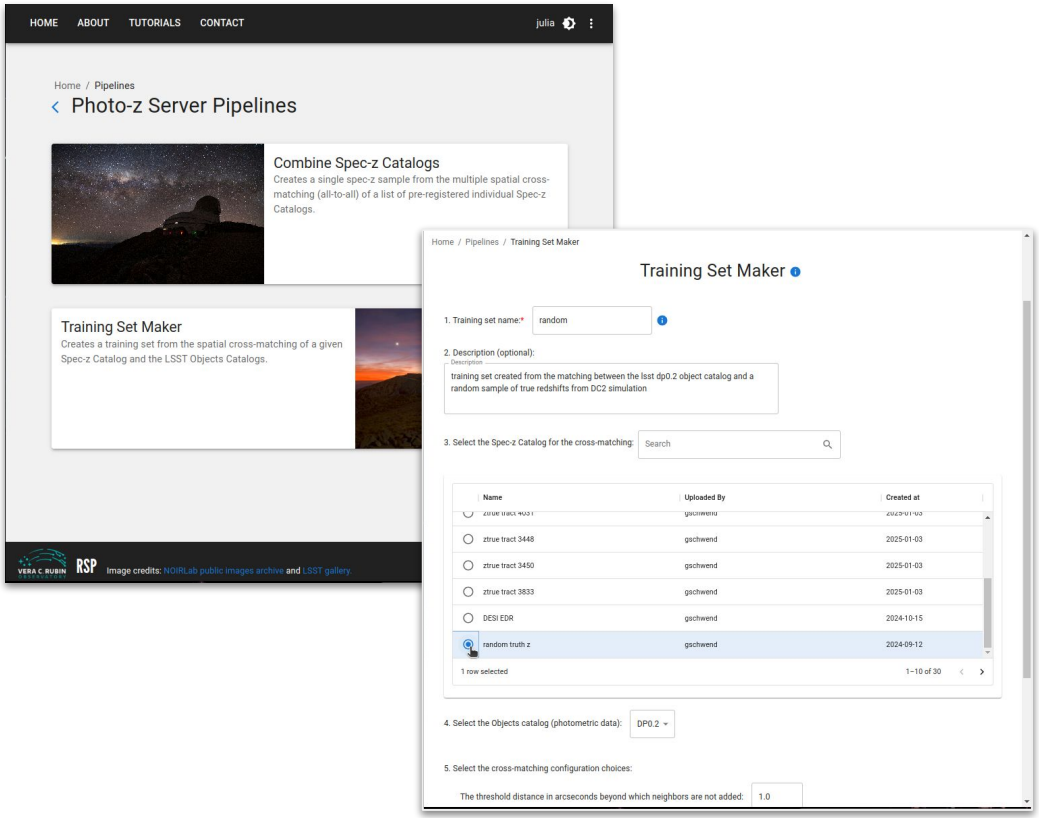
Jandson Vitorino
Front-end developer

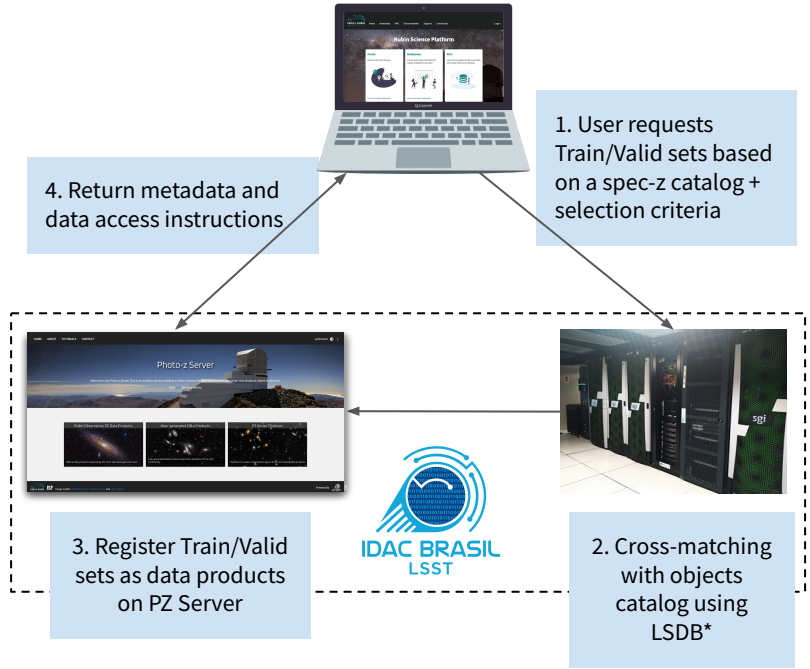
Cristiano Singulani
Back-end developer



Luigi Silva
Scientist

Julia Gschwend
Scientist





PZ Server pipelines

The *Training Set Maker* service

Cross-matching service to create Training Sets on demand

Back-end:

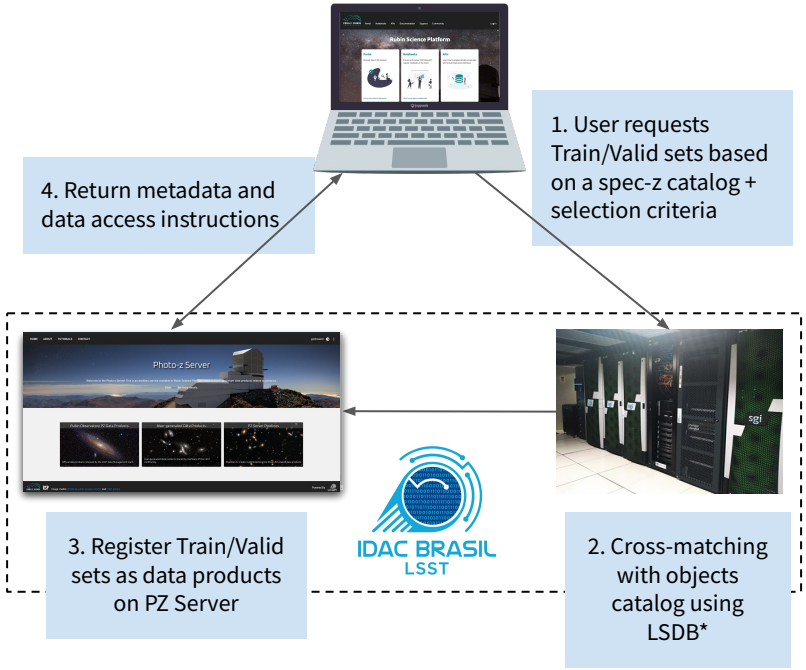
- Orchestration system to receive users requests and run processes at the IDAC
- Data partitioning based on HATS
- LSDB as cross-matching tool

} developed by LINCC team

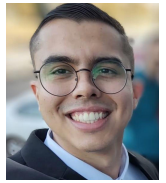
- Training Sets created become available as a PZ Server data product

Front-end :

- PZ Server API (methods in PZ Server class)
- Pipelines page on the PZ Server



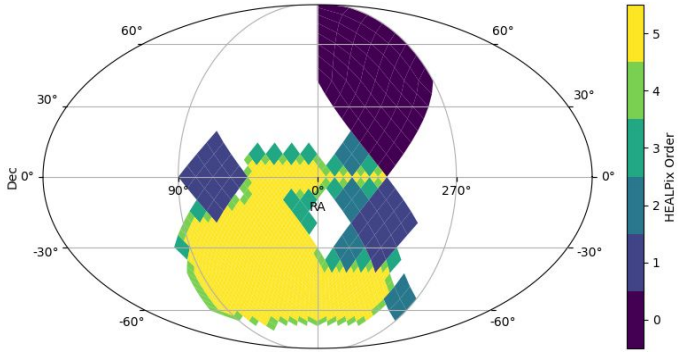
HATS and LSDB @LIneA [report]



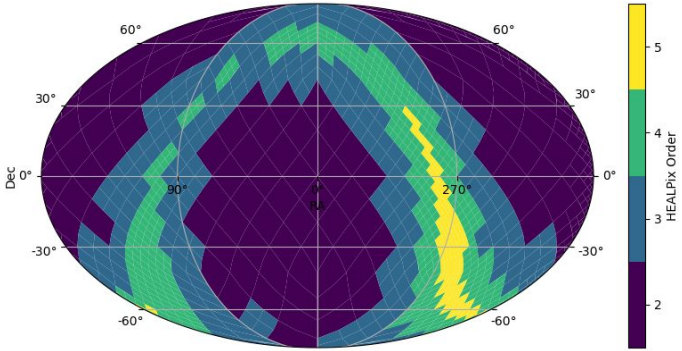
Luigi Silva
Scientist

LSDB

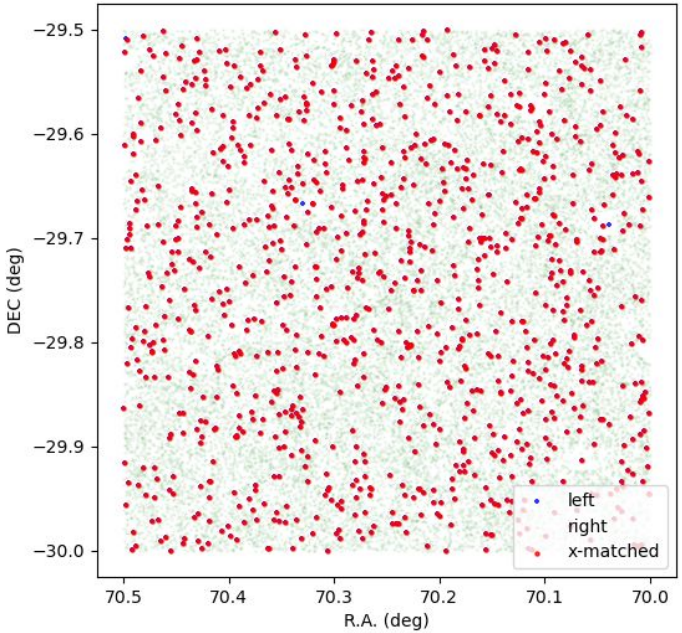
DES DR2



Gaia DR3



DES DR2 x Gaia DR3



https://docs.lsdb.io/en/stable/

LIneA

PZ Compute

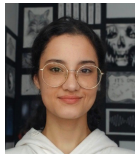
PZ tables as federated datasets

PZ Compute

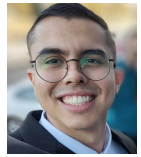
PZ Tables as federated datasets

Henrique Dante
Software engineer

Cristiano Singulani
Back-end developer



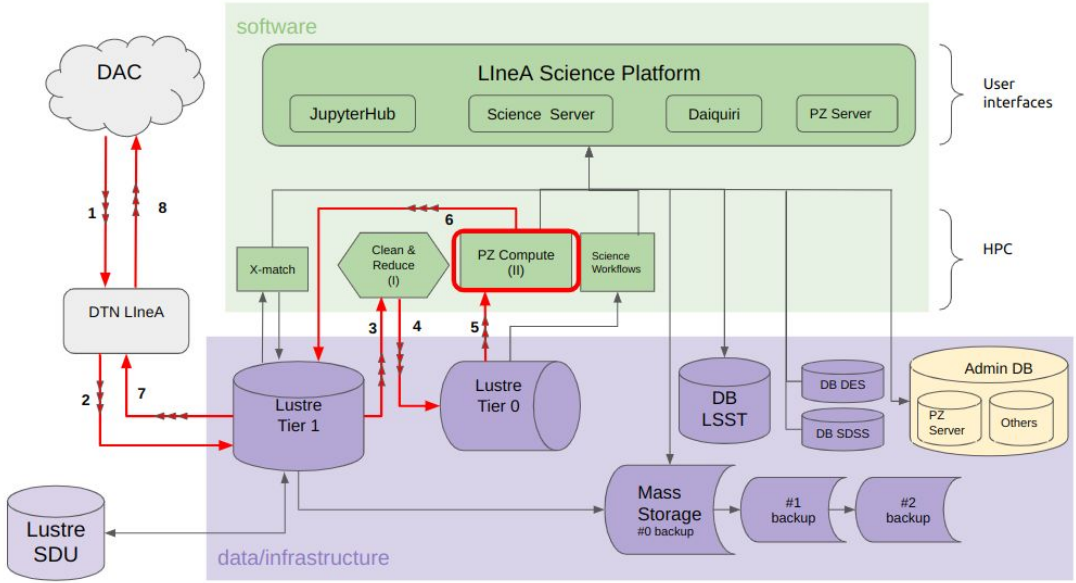
Heloisa Mengisztki
Back-end Developer



Luigi Silva
Scientist

Julia Gschwend
Scientist

PZ Compute pipeline optimizes the usage of LSSTDESC/RAIL's PZ estimators on the IDAC's cluster HPE Apollo 2000 using Slurm



linea-it/pz-compute

PZ Compute

PZ Tables as federated datasets

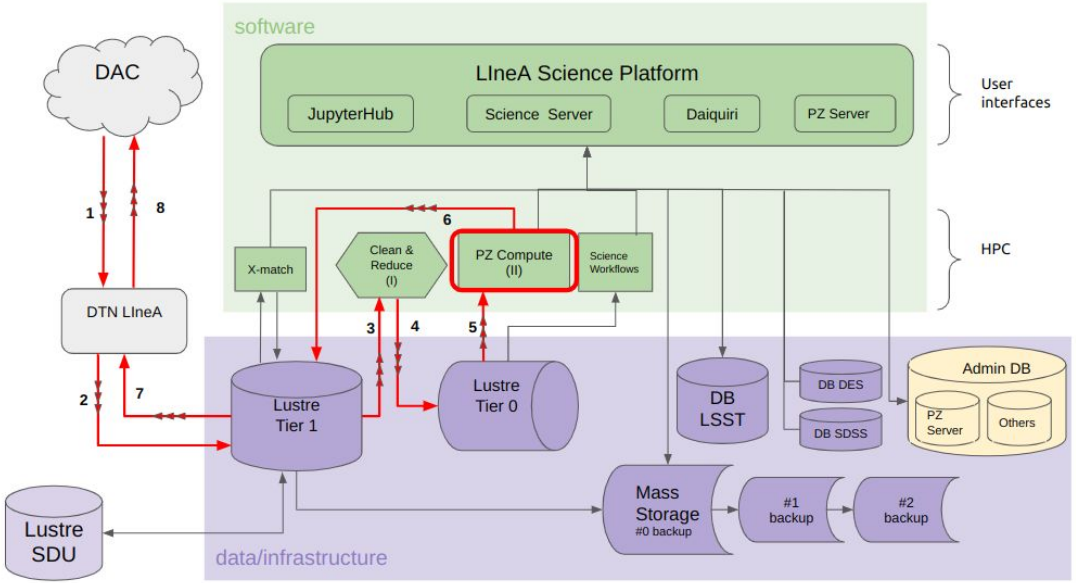
Summary of PZ Compute stages

Stage I - Data acquisition
- Data transfer from USDF
- Organize/split into data partitions, data cleaning
- Download and store ancillary files

Stage II - Photo-z pre-processing
- Training PZ machine learning methods
- PZ scientific validation

Stage III - Photo-z computing
- Execute PZ codes (use RAIL as code wrapper)
- Validate results with Rubin PZ Coord.

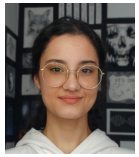
Stage IV - Photo-z post-processing
- Data transfer from BR IDAC to USDF
- Register on RSP as a federated dataset.



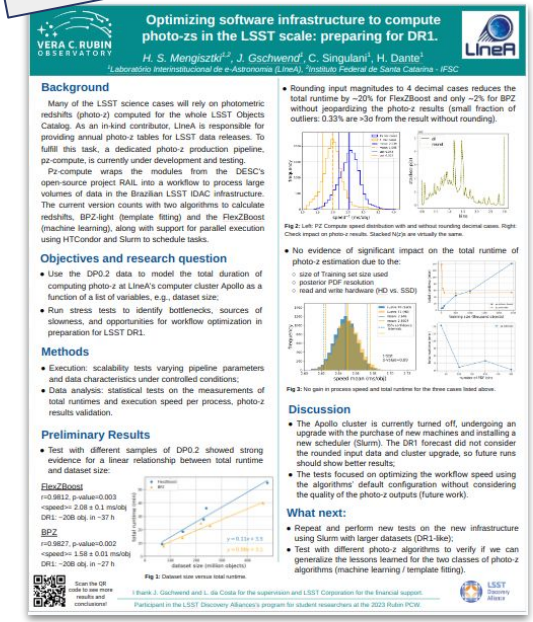
linea-it/pz-compute

Undergrad science projects

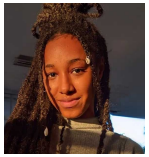
PCW 2023



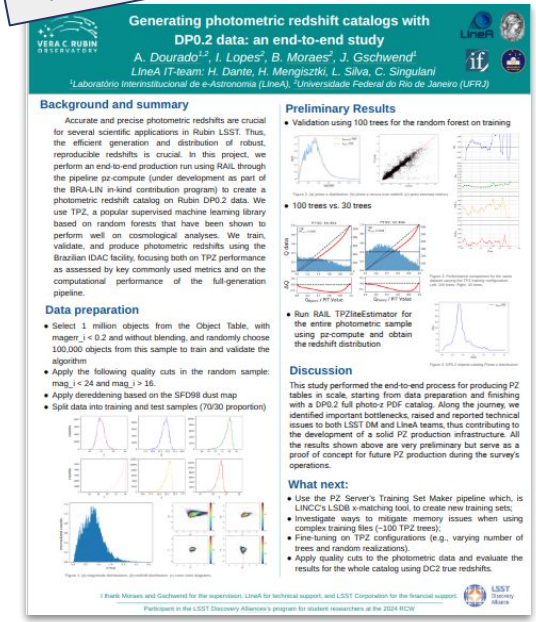
Heloisa Mengisztki
Back-end Developer
(supervisor: Dr. Julia Gschwend)



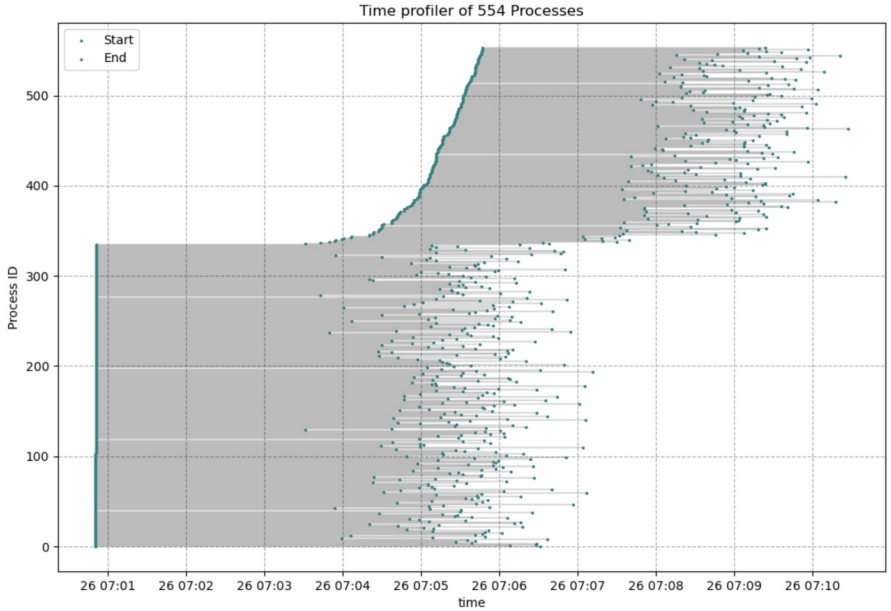
RCW 2024



Andreia Dourado
Undergrad Student
(supervisor: Prof. Bruno Moraes)



PZ Compute performance tests



Latest results of performance tests (DR11* forecast):

BPZ**	FlexZBoost	TPZ Lite	LePhare
~7 hours	~9 hours	~9 hours	~24 days

* multiple copies of DP0.2 to mimic a **40 billion objects** sample
** unrealistic configuration (default = 8 templates)

Optimal † input data partition (skinny tables):

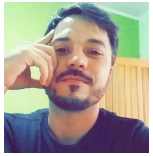
- ❖ 130k objects/parquet file or ~15 MB

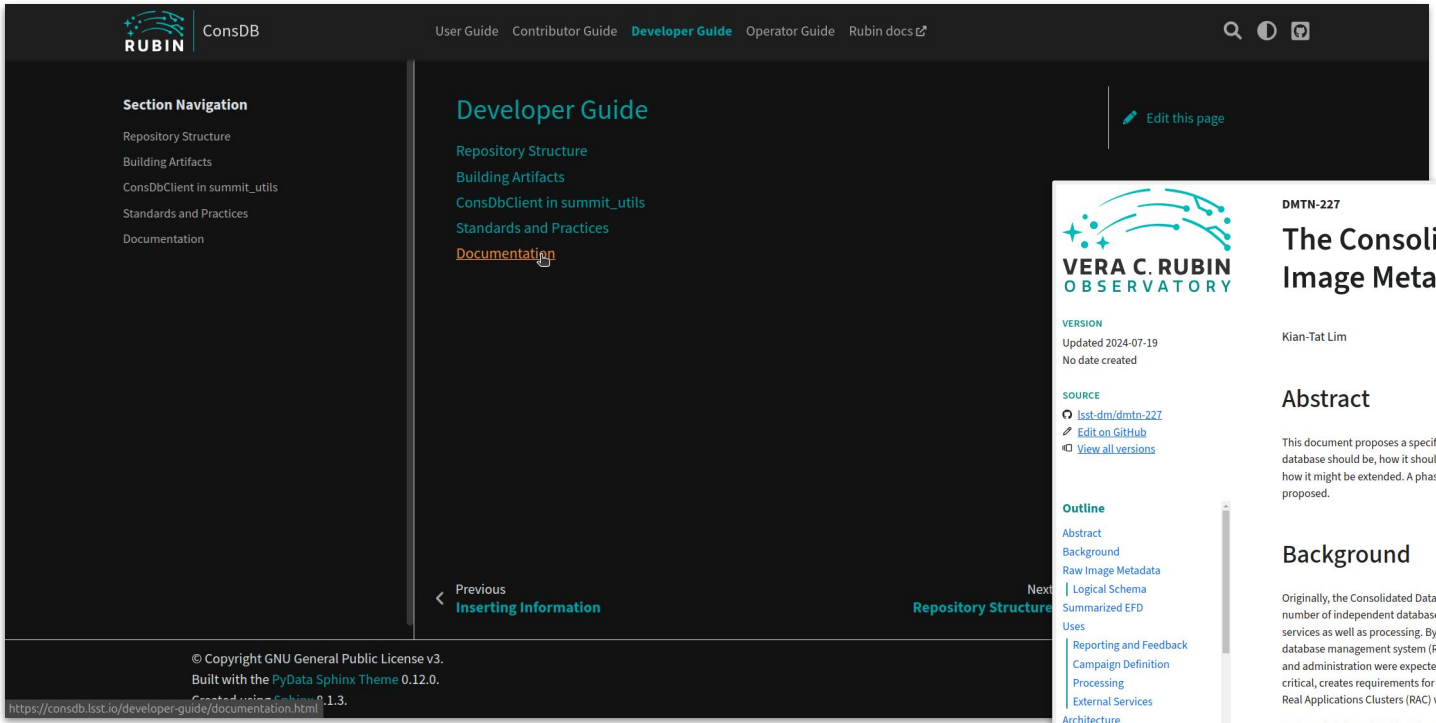
† considering IDAC's resources, RAIL's requirements and memory usage requirements including outputs and temporary files.

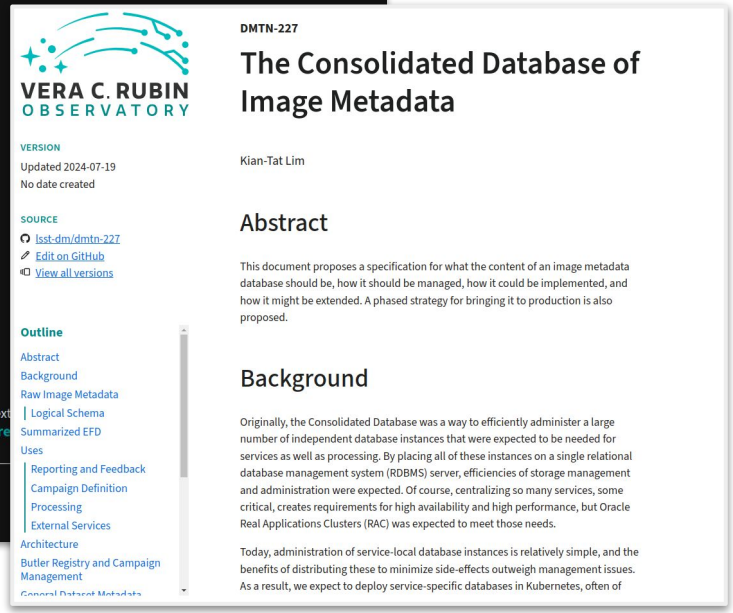
Next algorithm to test: **DNF**

Contribuição não coberta pelo acordo

Além da obrigação:
Rubin Data Management's ConsDB





https://dmtn-227.lsst.io/

Considerações Finais

Resumo

O programa de contribuições in-kind BRA-LIN inclui:

- Um centro regional de dados com infraestrutura própria, cópias locais dos dados e plataformas científicas dedicadas para a comunidade LSST.
- Diversos projetos de desenvolvimento de software para apoiar a ciência feita com dados do LSST.
- Produção de produtos científicos sob demanda e com entregas periódicas previstas para todo o período de comissionamento e operação do LSST.
- Recursos humanos - desenvolvedores experientes em software para astronomia e cientistas de dados com mestrado e doutorado em astronomia.

Além disso, vale ressaltar que:

- Todo software produzido como contribuição in-kind é código aberto (disponível na organização linea-it no GitHub)
- Suas sugestões são bem vindas. Conte-nos sobre o seu caso de uso: julia@linea.org.br

LIneA é apoiado por: