

Relatório LIneA / INCT do e-Universo 2018

Com a missão de apoiar cientistas brasileiros participantes de grandes levantamentos astronômicos, o Laboratório Interinstitucional de e-Astronomia em conjunto com o Instituto Nacional de Ciência e Tecnologia do e-Universo tiveram mais um ano de importantes realizações nas suas diversas áreas de atividade. Alguns dos destaques em 2018 são apresentados a seguir.

Gestão

Com os recursos disponíveis do INCT em 2018 foram integrados ao time uma gerente de serviços, uma *Scrum Master* para acompanhamento diário das atividades de *sprints* que são definidos em reuniões quinzenais, uma gerente de produtos responsável pela definição de requisitos e validação dos produtos desenvolvidos e uma engenheira de software que está colaborando na avaliação do desenho e desempenho do atual portal e, baseada na experiência acumulada, está compilando requisitos para o novo portal a ser desenvolvido para o projeto *Large Synoptic Survey Telescope* ([LSST](#)) usando o legado do portal científico do *Dark Energy Survey* ([DES](#)) construído ao longo dos últimos 10 anos. É importante lembrar que o projeto DES é um levantamento fotométrico de uma grande área do céu sendo portanto um precursor do LSST.

Esta equipe tem contribuído enormemente para aumentar a produtividade do time de TI e a qualidade dos produtos, introduzindo metodologias e formalizando processos. Em particular, com a equipe de gestão está sendo possível: 1) criar um cronograma de longo-prazo a partir de fichas de projeto onde os objetivos, escopo, recursos, e um calendário de entregas são descritos a partir de discussões envolvendo todos os *stakeholders* e levando em conta as prioridades de cada time científico e da colaboração internacional; 2) organizar reuniões quinzenais de *sprint*, semanais de acompanhamento e na medida da necessidade de validação; 3) coordenar uma comissão de controle de mudanças (*change control-board*) para acompanhar mudanças na base de código que pode impactar a operação das diversas ferramentas; 4) formalizar a entrega de produtos; e 5) preparar relatórios mensais disponíveis publicamente relatando as atividades de desenvolvimento e serviços, e desta forma mantendo todos os interessados devidamente informados das atividades em andamento.

Por exemplo, em 2018 foram realizados 24 *sprints*, de duas semanas cada, para projetos complexos como o portal envolvendo as equipes técnicas e científicas. Cada *sprint* gera, em média, 25 tarefas distribuídas em projetos como: 1) o *LIneA Science Server* entregue em Janeiro à colaboração DES; 2) um protótipo do *Quick Look Framework* (QLF) desenvolvido para avaliar 15.000 espectros de 5.000 objetos em tempo real para a

colaboração *Dark Energy Spectroscopic Instrument* ([DESI](#)); 3) o portal desenvolvido para a colaboração *Transneptunian Occultation Network* ([TON](#)); 4) um sistema de busca de dados obtidos pelo grupo TON no Observatório do Pico dos Dias no LNA; e 5) um sistema de visualização de imagens rápido para servir ao LSST, permitindo inspecionar os 189 CCDs que formam uma imagem do LSST em menos de 5 segundos.

Uma outra importante tarefa deste grupo tem sido coordenar o esforço de compilar, consolidar e organizar o vasto acervo de documentos acumulados pelo LIneA ao longo dos últimos 14 anos de funcionamento. A documentação inclui diferentes atividades relacionadas a operação, descrição dos vários serviços prestados pelo LIneA, relatórios sobre testes realizados, e documentação dos sistemas de aquisição e acesso a dados em operação no Brasil e no exterior. A documentação também inclui para cada pipeline em funcionamento uma descrição de alto nível, manual de uso e histórico de manutenção, além de sua manutenção num repositório de código (ex., `cgit`, `github`, `gitlab`) o que permite usá-los como código legado para o novo portal sendo desenvolvido para o LSST. Esta documentação, em particular sobre o portal e seus pipelines científicos, e a experiência acumulada na operação de um centro de dados são essenciais para que o aprendizado seja útil no desenvolvimento do chamado portal 2.0 adaptado às demandas do projeto LSST e na expansão do centro de dados para se adequar a escala exigida por este projeto.

Colaborações Científicas

O LIneA/INCT atualmente apoiam cinco projetos internacionais (DES, DESI, LSST, SDSS, TON) para os quais o LIneA desenvolve software para uso dessas colaborações (*Quick Reduce*, *DES Science Portal*, *LIneA Data Server*, *LIneA Science Server*, *Quick Look Framework*, *Solar System Object platform*) ou presta serviços como a distribuição de dados do SDSS (*SkyServer/CasJobs*) e a operação de computadores usados pelo projeto *APOGEE South* no observatório de Las Campanas, Chile. Estes produtos e serviços fazem parte da contribuição do LIneA em troca da participação de pesquisadores brasileiros interessados em fazer parte destas colaborações internacionais. Estas contribuições foram negociadas caso a caso. Para cada projeto foram organizados times científicos formados por pesquisadores de diferentes instituições brasileiras em diferentes estados como Rio de Janeiro (ON, UFRJ, UFF), São Paulo (USP, UNESP, UNICAMP), Paraná (UTFPR), Santa Catarina (UFSC) e Rio Grande do Sul (UFRGS, UFSM).

Durante o ano de 2018 uma das principais atividades do LIneA/INCT foi apoiar a integração de equipes brasileiras nos projetos DESI e LSST. Para o projeto DESI o empenho foi no desenvolvimento do sistema QLF. Este sistema fará a avaliação em tempo real de 15 mil espectros em diferentes faixas de frequência (azul, vermelho e

infravermelho) de 5.000 objetos em intervalos de tempo que variam de 2 a 20 minutos. Uma versão beta já foi disponibilizada ao projeto DESI sendo que a parte de interface com outros subsistemas do equipamento foi avaliada com sucesso em Junho de 2018 numa campanha de observações simuladas de uma semana no Observatório de Kitt Peak no Arizona com a participação de membros do LIneA. Uma nova revisão está sendo feita agora por futuros usuários antes da entrega final do produto para o período de comissionamento do instrumento no segundo semestre de 2019.

Para o LSST, dando sequência ao trabalho de formação do *Brazilian Participation Group* (BPG) iniciado em 2017, o comitê de seleção/acompanhamento foi convocado em 2018 para rever as atividades dos grupos selecionados, emitindo um parecer sobre o desempenho de cada um, incluindo recomendações específicas de como ajudar na preparação do time brasileiro para o início do levantamento. Ao mesmo tempo, uma segunda chamada foi realizada e mais quatro propostas estão sendo consideradas pela comissão de seleção/acompanhamento. Como parte do treinamento dos pesquisadores já selecionados as seguintes atividades foram organizadas: 1) 16 webinars com membros do LSST cobrindo diferentes aspectos do projeto; 2) a semana do LSST no Rio de Janeiro com a presença dos gerentes de dados e de comunicação do projeto; e 3) o encontro de trabalho “*South American Workshop on Cosmology in the LSST era*” realizado na UNESP com o apoio do INCT do e-Universo. O INCT também apoiou: 1) a participação no encontro realizado no [Chile do LSST](#), a participação de cinco pesquisadores na reunião [anual do projeto LSST](#) realizada em Tucson; 2) a participação de membros do BPG-LSST em sete reuniões do Dark Energy Science Collaboration (SLAC, Carnegie Mellon); e 3) reuniões técnicas realizadas sobre [rede](#) (Santiago, Chile), [técnicas de visualização](#) (NCSA) e [engenharia e gerenciamento de dados](#) (Caltech).

O BPG está em pleno funcionamento com os pesquisadores selecionados participando de reuniões, telecons e acompanhando o progresso do projeto pelos diferentes canais de comunicação disponibilizados pelo LSST. Além disso, o LIneA vem realizando estudos e testes de como lidar com a quantidade de dados envolvidos e as características do centro de apoio para o acesso e processamento dos dados do LSST em colaboração com o RNP e LNCC. Neste sentido espera-se submeter um projeto específico para manter a equipe técnica e iniciar a compra de equipamentos para 2022 quando começa o levantamento.

Para as colaborações já em andamento em 2018 foram submetidos para publicação um total de 96 artigos dos quais 68 foram publicados e três aguardam publicação. Entre estes artigos, dois descrevem o complexo sistema implantado para o cálculo de *redshifts* fotométricos e, para a criação de catálogos prontos para diferentes análises científicas desenvolvido pelo LIneA para o portal científico do DES. A figura 1 coloca a produção de

2018 num contexto histórico do LIneA que levou a produção de 337 artigos (291 publicados, 4 aceitos para publicação e 42 submetidos) no últimos 10 anos, uma média de 34 artigos/ano. com a participação de pesquisadores apoiados pelo LIneA e nos últimos dois anos pelo INCT.

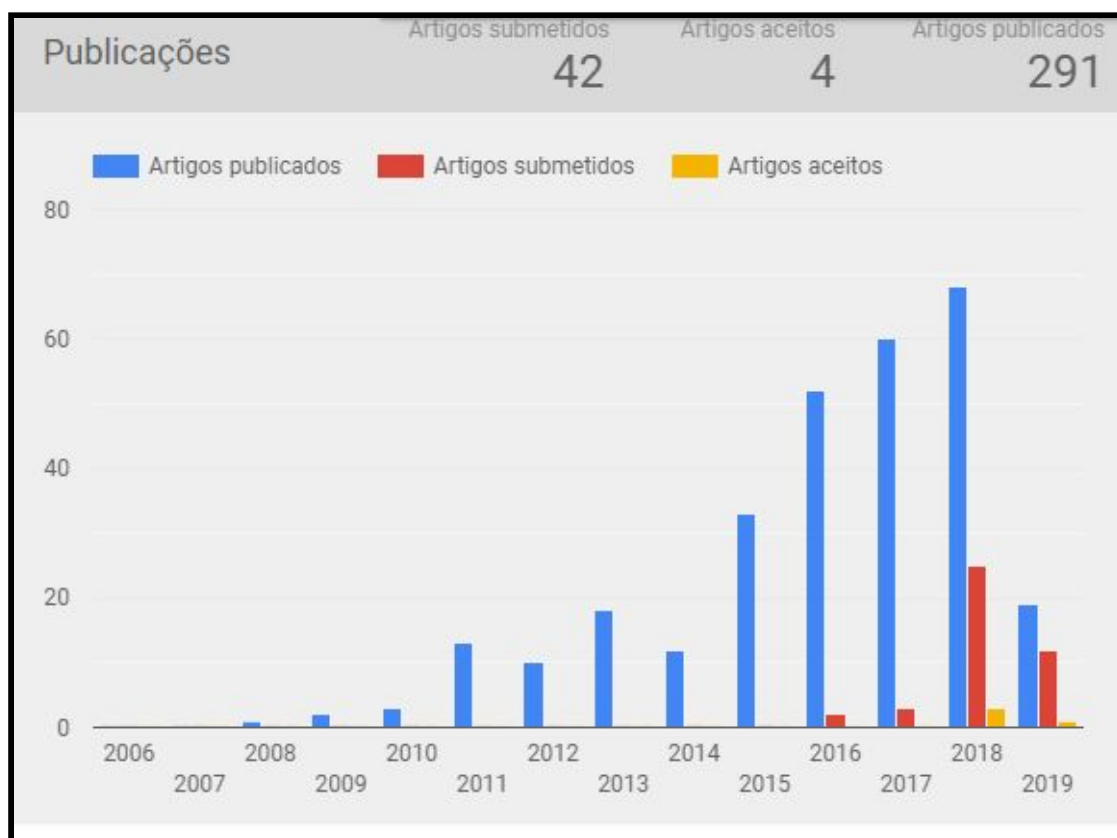


Figura 1- Número de artigos publicados, submetidos e aceitos em revistas especializadas por ano. Na parte de cima do gráfico apresenta o número total nessas categorias. Mais detalhes podem ser vistos na [página do LIneA](#).

Formação de Pessoal

Uma importante missão do LIneA/INCT é a formação de pessoal técnico e científico e como fica evidente na Figura 2 a equipe do LIneA mistura tecnologistas e pesquisadores em diferentes estágios da carreira. Em particular, a grande maioria de membros é de jovens pesquisadores incluindo desde estudantes de graduação à pós-doutorandos. Além disso, a grande maioria dos membros contratados também é formada por jovens que

embora sejam mais experientes ainda tem uma longa carreira pela frente. Isto é importante tendo em vista que os projetos DESI e LSST devem se estender no mínimo até 2027 e 2032, respectivamente.

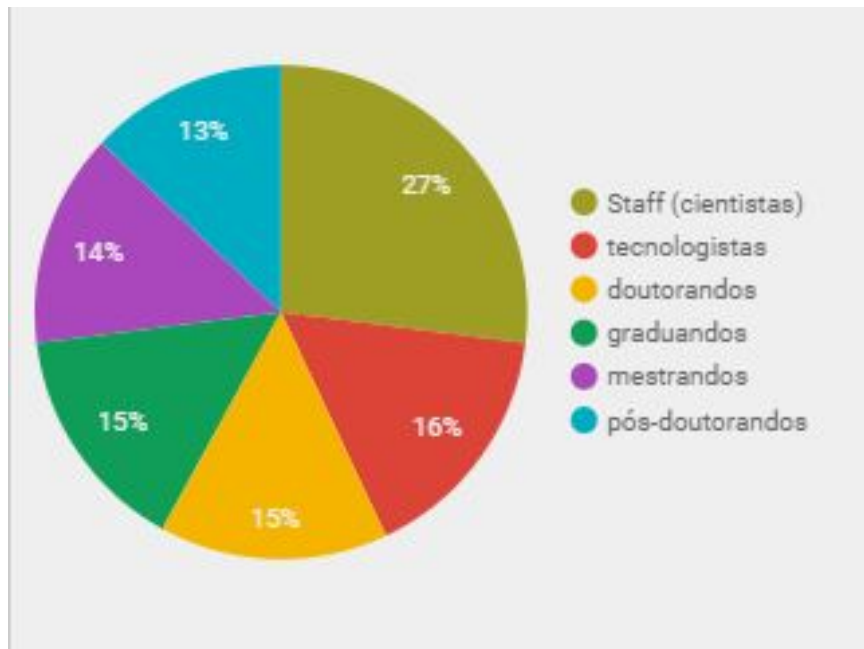


Figura 2 - Distribuição de membros do LIneA por categoria.

É importante enfatizar que os jovens pesquisadores afiliados ao LIneA tem uma formação diferenciada, pois além das suas atividades científicas, eles são expostos regularmente aos problemas da ciência de dados, e a participação em grandes colaborações dá a eles uma importante vivência e inserção internacional. Além disso, participam regularmente de telecons nacionais e internacionais com pesquisadores seniores de outros institutos dando a eles uma visão bem mais ampla da sua área de atuação que não fica restrita ao seu supervisor local. A eles também é solicitado darem pequenas palestras sobre suas atividades científicas em lunch talks e assistirem aos webinars organizados semanalmente pelo LIneA/INCT. Esta formação é fundamental para preparar uma nova geração de pesquisadores para a era do LSST.

Como parte da formação e internacionalização de seus membros o INCT apoiou ao longo de 2018 um total de 22 bolsistas, dos quais quatro de pós-doutorado, e concedeu auxílio integral ou parcial a inúmeros filiados para participarem de 27 eventos a grande maioria no exterior, sendo que pelo menos 11 foram relacionados ao LSST. Desses eventos, três foram realizados no Brasil e foram apoiados pelo INCT: 1) o *DES International*

Collaboration Meeting em colaboração com a UNICAMP; 2) o Workshop em Cosmologia na era do LSST em colaboração com a UNESP; e 3) a Semana LSST. A Figura 3 mostra o histórico de participação de membros do LIneA e do INCT (a partir de 2017) em eventos que totalizam 150 ao longo dos últimos 10 anos. A figura mostra ao mesmo tempo o nível de internacionalização alcançado pelo LIneA e a importância do INCT nos últimos anos em apoiar a participação de tecnólogos e pesquisadores de diferentes níveis nesses eventos. Esta mobilidade é fundamental para assegurar a integração do time nas colaborações internacionais e assim tirar proveito científico das mesmas.

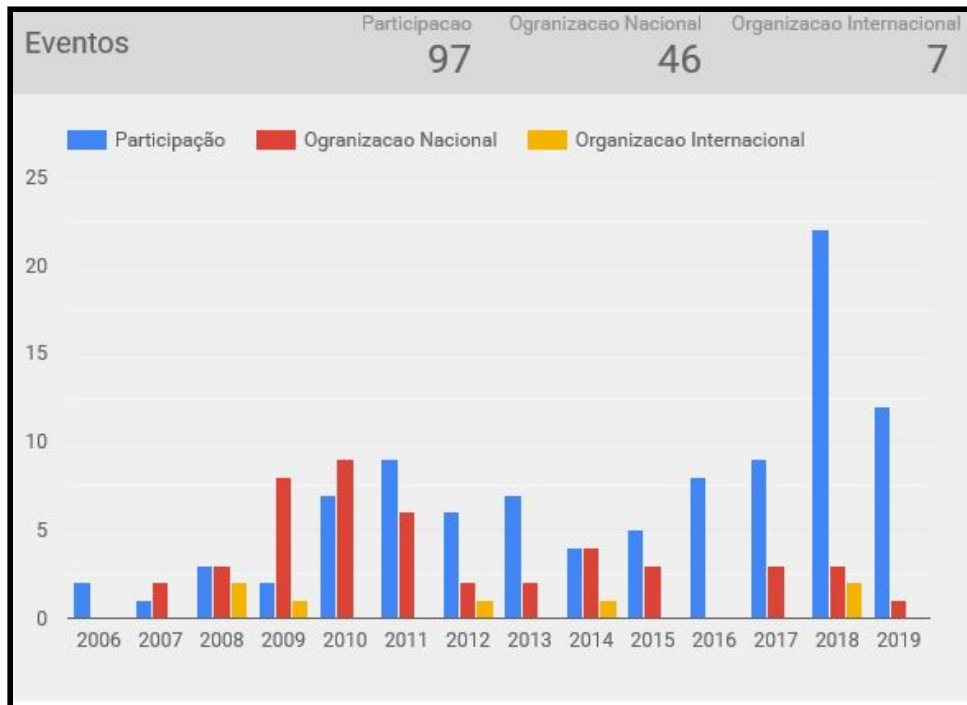


Figura 3 - Número de eventos com a participação de membros do LIneA e reuniões de carácter nacional e internacional, organizadas pelo LIneA.

Entre os eventos destacamos o *LIneA Science Day* dedicado a uma ampla discussão sobre o trabalho sendo realizado nas diferentes colaborações, a reunião *LSST week* dedicada a atividades relacionadas ao LSST com a visita de convidados e o *Design Sprint* organizado pela RNP para discutir a criação de um centro de suporte a e-ciência, com a participação de membros do LIneA compartilhando suas experiências. Complementando essas reuniões mais formais, reuniões informais do tipo *hack sessions* foram organizadas para compartilhar conhecimento entre pesquisadores. Em 2018 foram 17 dessas sessões tendo como foco principal os métodos usados na criação de catálogos para serem usados em diferentes aplicações científicas a partir dos dados do DES. As mesmas técnicas serão usadas no LSST.

O apoio do INCT tem sido importante sendo responsável pelo aumento de participação brasileira nos eventos dos últimos dois anos (Figura 3). Cumprindo a missão do INCT este apoio tem sido distribuído por todos os projetos apoiados pelo INCT como mostra a Figura 4, com o DES no estágio mais avançado e com mais participantes recebendo a maior fração dos recursos.

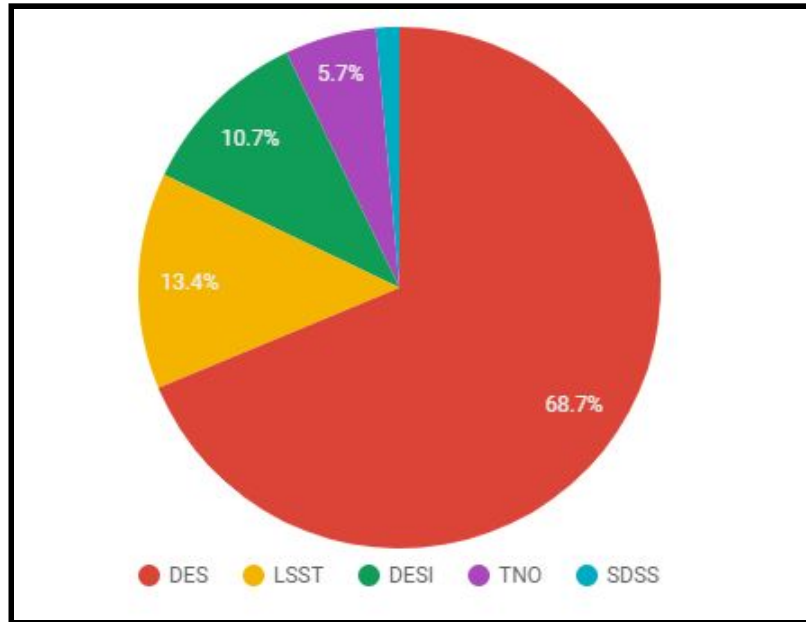


Figura 4 - Fração de recursos do INCT usados para apoiar a participação de pesquisadores nos diferentes projetos apoiados pelo LIneA. A figura inclui bolsas e auxílios.

O LIneA atualmente conta com um grande número (~40) de graduandos, mestrandos e doutorandos cujas atividades científicas dependem da infraestrutura sendo mantida pelo LIneA além de 13 pós-doutorandos sendo financiados por diferente agências incluindo o INCT do e-Universo. A Figura 5 mostra o número de defesas de mestrado e doutorado ao longo dos anos. Até agora foram defendidas 13 teses de doutorado e 18 de mestrado, por estudantes envolvidos nos grandes levantamentos apoiados pelo LIneA. Cinco dessas dissertações de mestrado foram defendidas em 2018.

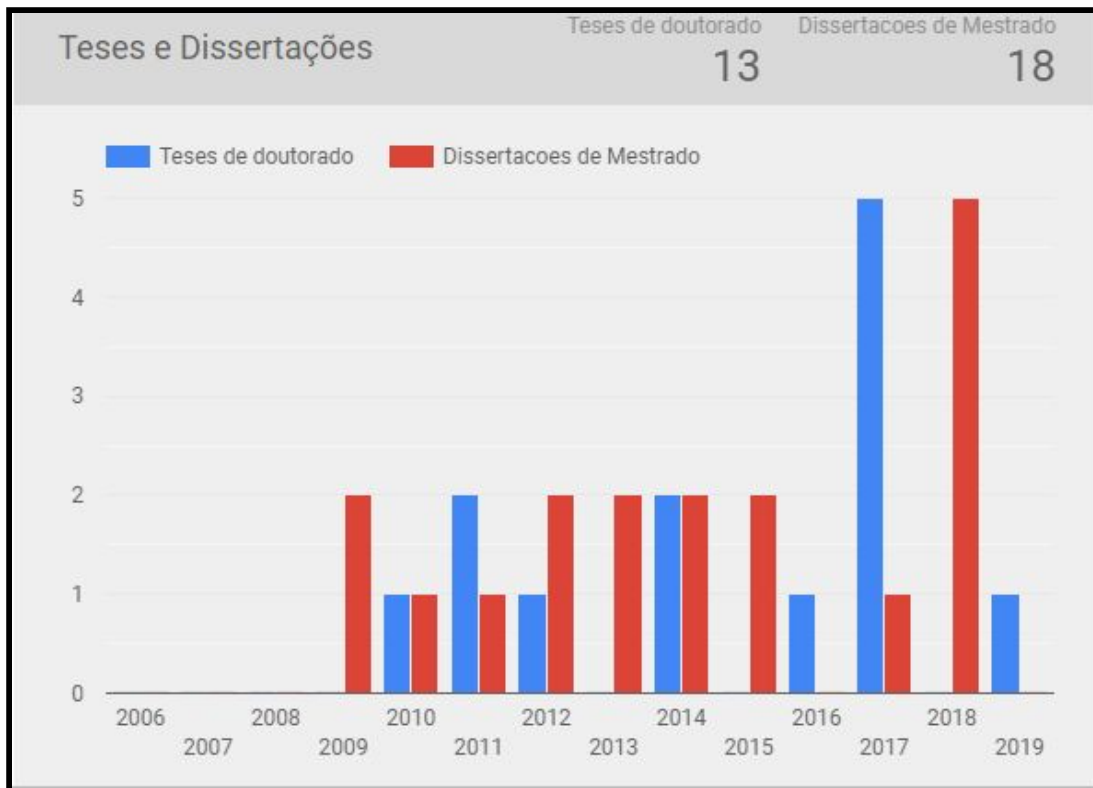


Figura 5 - Número de teses de doutorado (azul) e dissertações de mestrado (vermelho) produzidas por alunos participando das colaborações internacionais promovidas pelo LIneA. Os valores totais são apresentados na parte superior do gráfico disponível no [site do LIneA](#).

Outra atividade promovida desde 2011 são apresentações remotas transmitidas publicamente por palestrantes vinculados a instituições distribuídas por todo o mundo cobrindo os mais variados tópicos em astronomia, cosmologia, instrumentação e computação de alto desempenho, relevantes a missão do LIneA. A série procura mesclar pessoas em vários estágios da carreira com ênfase em nomes internacionais procurando assim manter os pesquisadores brasileiros bem informados sobre o andamento das pesquisas nos grandes centros. Ao longo dos anos foram realizadas um total de 239 webinars (Figura 6), por 227 palestrantes, dando uma média de mais de 30 webinars por ano. Em 2018 foram organizados 38 webinars com destaque para a série sobre o SDSS-V (4) e o LSST (16). Os vídeos dos últimos 84 webinars podem ser acessados no repositório de [vídeos](#) do LIneA (ver Figura 7).

Iniciativa recente foi a criação de uma série de *lunch talks* onde a prioridade é dada para alunos e pesquisadores jovens em geral para apresentar o que estão fazendo em suas

pesquisas e assim estimular a colaboração entre grupos científicos ou de diferentes instituições. Vídeos desses dessas apresentações também estão disponíveis mas apenas internamente para manter a informalidade.

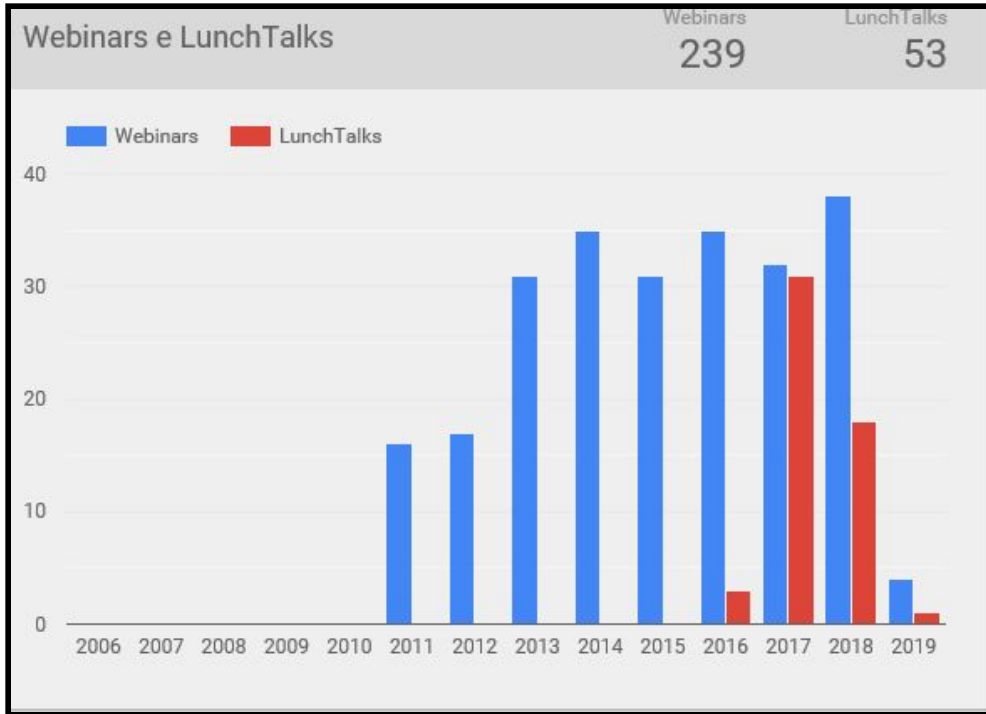


Figura 6 - Número de webinars e lunch talks organizados pelo LIneA desde 2011 e pelo INCT a partir de 2016.

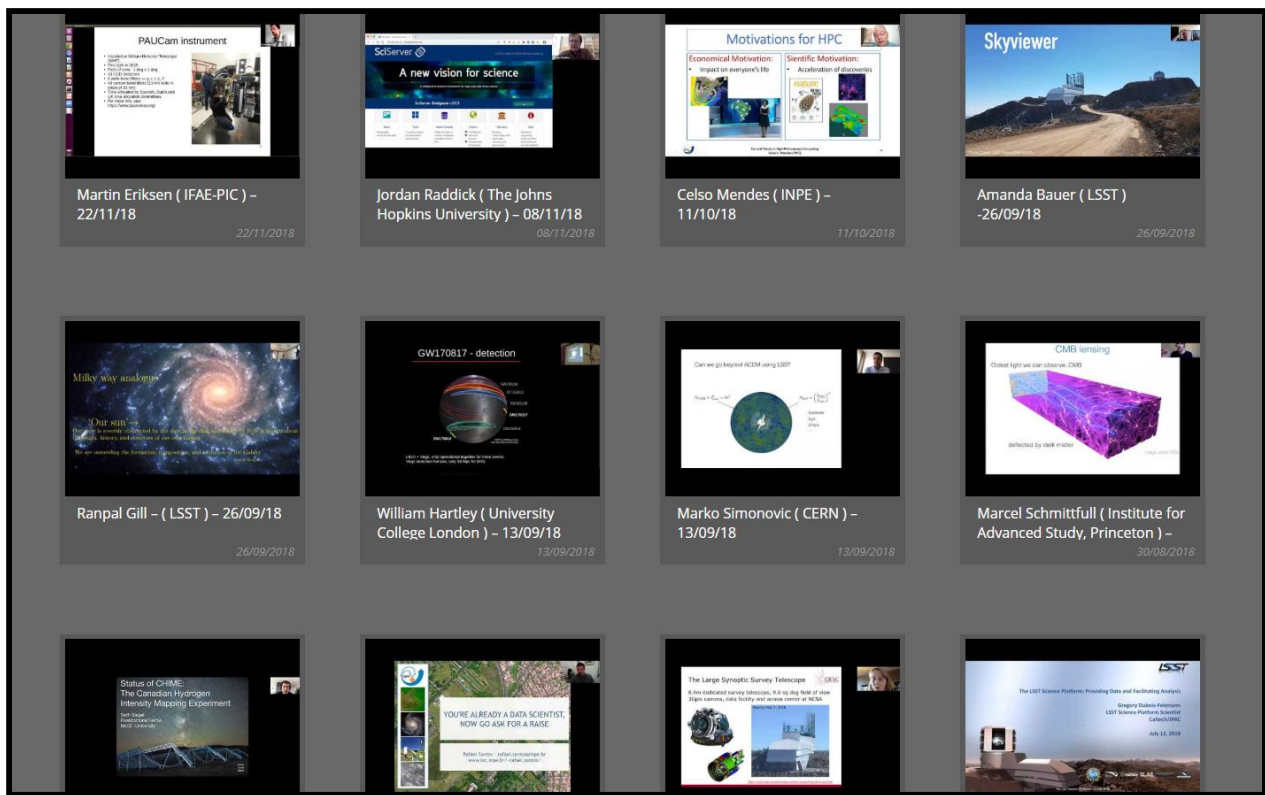


Figura 7 - Captura de tela do [site do LIneA](#) onde os videos dos webinars dos últimos três anos podem ser encontrados.

Divulgação

O principal mecanismo de divulgação do trabalho feito pelo LIneA/INCT é a manutenção de um site atualizado procurando mostrar as diversas atividades científicas e tecnológicas, incluindo uma galeria de vídeos e imagens astronômicas. O site também dá acesso público aos dados dos levantamentos SDSS e DES.

Outro mecanismo utilizado são notícias sobre as atividades de pesquisa e desenvolvimento do laboratório ou de seus membros veiculadas na forma de blogs disponibilizados no site e em redes sociais. Periodicamente estas notícias são agrupadas e em conjunto com informações sobre futuros webinars na forma de um boletim eletrônico chamado *LIneA News* e enviados por mala direta para mais de 1000 pessoas. Como mostra a Figura 8, ao longo dos anos foram publicados mais de 300 blogs, consolidados em 44 *LIneA News*, dos quais 62 blogs circulados em oito boletins foram em 2018,. Os blogs são preparados pelos próprios pesquisadores com a ajuda de uma aluna de jornalismo e validado por pelo menos dois pesquisadores seniors. O conjunto completo das notícias pode ser encontrado neste [link](#).

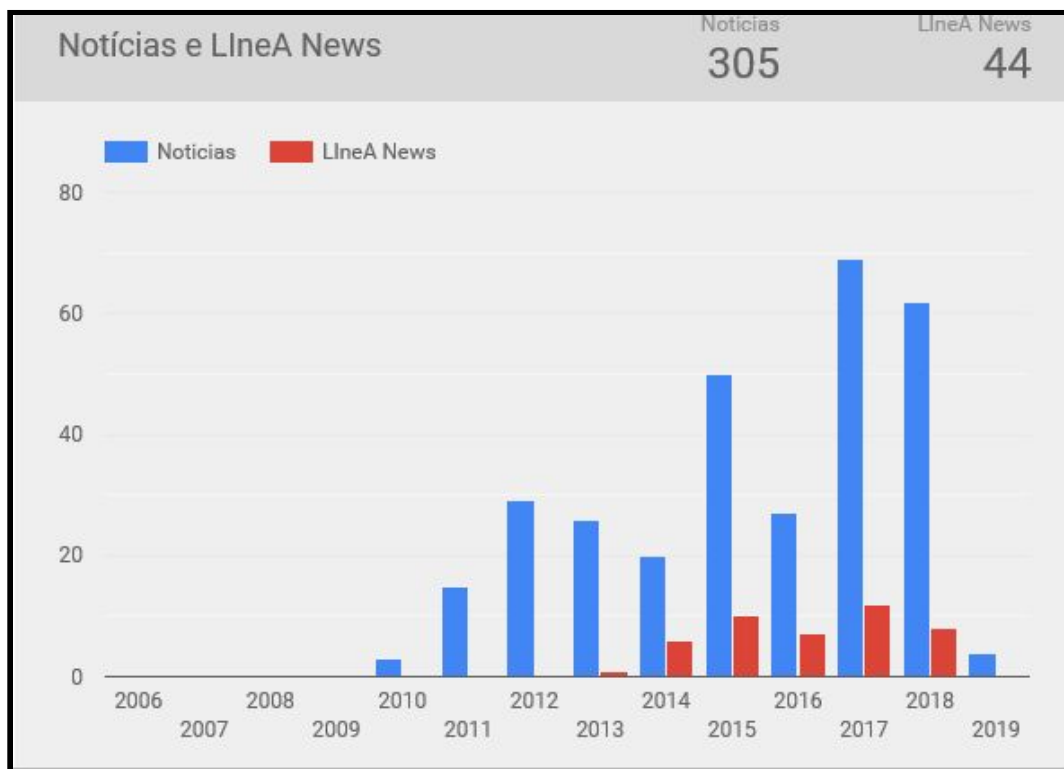


Figura 8 - Número de blogs publicados e *LIneA News* distribuídos.

Outras formas de divulgação para cientistas incluem:

- Webinars internacionais com vídeos disponibilizados ao grande público.
- Apresentação de palestras convidadas. Entre estas destacam-se as apresentações feitas na Academia Brasileira de Ciências em Abril e Setembro sobre grandes projetos e o 2o Encontro Latino-americano de e-ciência em Cartagena, Colômbia.
- Participação em Escolas e Workshops como a Escola de Inverno em Cosmologia patrocinada pela UNICAMP.
- Organização de reuniões internacionais como a semana do LSST, *DES International Collaboration meeting* e o *South American Workshop on Cosmology in the LSST era*.

Para o grande público os mecanismos usados são:

- Entrevistas na rádio MEC e EBC.
- Preparação de pequenos vídeos de entrevistas pelos filiados.
- Participação na exposição comemorativa dos 190 anos do ON realizada no Museu Histórico Nacional.
- Participação na SNCT no ON e na UFRRJ.
- Reuniões informais em bares como "Dose de Ciência" (Santa Maria) e "Chopp comCiência" (Rio de Janeiro)
- Curso para alunos do segundo grau "A Física do Universo: Cosmologia" realizado na UNESP, SP.
- Curso para professores do Ensino Médio no South American Institute for Fundamental Research localizado no IFT-UNESP
- Palestra para o público geral no Planetário de São Paulo -
- Artigos em revistas de divulgação com o *Ciência hoje das Crianças* e *Revista FAPESP*.

Centro de Acesso e Processamento de Dados Astronômicos (CAPDA)

Uma das principais metas tanto do LIneA como do INCT é a de dar aos pesquisadores brasileiros envolvidos em grandes levantamentos acesso a um parque computacional adequado e ferramentas de software para a eficiente e competitiva exploração científica do crescente volume de dados sendo gerados por esses projetos. No momento, estudos estão em andamento procurando desenvolver uma visão do que um centro de acesso e processamento no Brasil deve oferecer em 2022, quando começa o levantamento do

LSST. O desafio é conciliar esta visão de médio-prazo com as necessidades de curto prazo e a disponibilidade financeira. Um primeiro passo nessa direção foi feito em 2018 com a compra do cluster APOLLO 2000 da HPE com recursos do INCT, aumentando significativamente a capacidade de processamento do centro. Este sistema complementa a compra realizada em 2017 do cluster ICE-X da SGI.

Em 2018 as atividades do time responsável pelo centro foram: 1) a participação no projeto Mercúrios e colaboração com pesquisadores do NCE para implantar um ambiente de nuvem integrando os vários clusters locais; 2) a preparação de casos de uso para avaliar a melhor forma para o uso das 500.000 AUs do supercomputador Santos Dumont aprovadas, explorando a possibilidade do LIneA contribuir para sua expansão tanto para processamento como para armazenamento, e sua integração ao Centro de Dados para o LSST; 3) entrevistas com diferentes fornecedores e, com o apoio da RNP, avaliação das soluções de armazenamento apresentadas. Estas soluções têm que satisfazer ao mesmo tempo as necessidades de HPC, armazenamento de grandes volumes de dados e backup, e ter a possibilidade de expansão incremental. Para auxiliar nesse processo de avaliação foi criado um comitê de especialistas para assessorar o time técnico do LIneA e o Comitê Gestor do INCT na compra de novos equipamentos com os recursos disponíveis do INCT e da FINEP.

Em 2018 também foi realizado um grande esforço para melhorar a comunicação com os usuários do centro. Isto foi possível graças a um pequeno aumento do time, o que permitiu entre outras coisas um melhor monitoramento e registro dos problemas de operação, envio de avisos para os usuários de paralisações programadas ou emergenciais, além de melhorias no atendimento de *tickets* para o serviço de *helpdesk*. Isto faz parte da transição do LIneA que está evoluindo e desenvolvendo uma cultura de serviço fundamental para atender um público de usuários maior do que aquele envolvido nas colaborações científicas apoiadas pelo laboratório. Um exemplo deste esforço é apresentado na Figura 9 que apresenta um calendário de 2018 com todos os incidentes que afetaram o uso de suas máquinas e através de *links* indicando as causas dessas paralisações. Em 2018, o número de incidentes foi bastante alto e foram causados por problemas de infraestrutura no CPD do LNCC o que implicou que os sistemas estiveram indisponíveis 13% do tempo. É importante lembrar que medidas corretivas foram tomadas pelo LNCC, o que deve melhorar a qualidade de serviço em 2019.

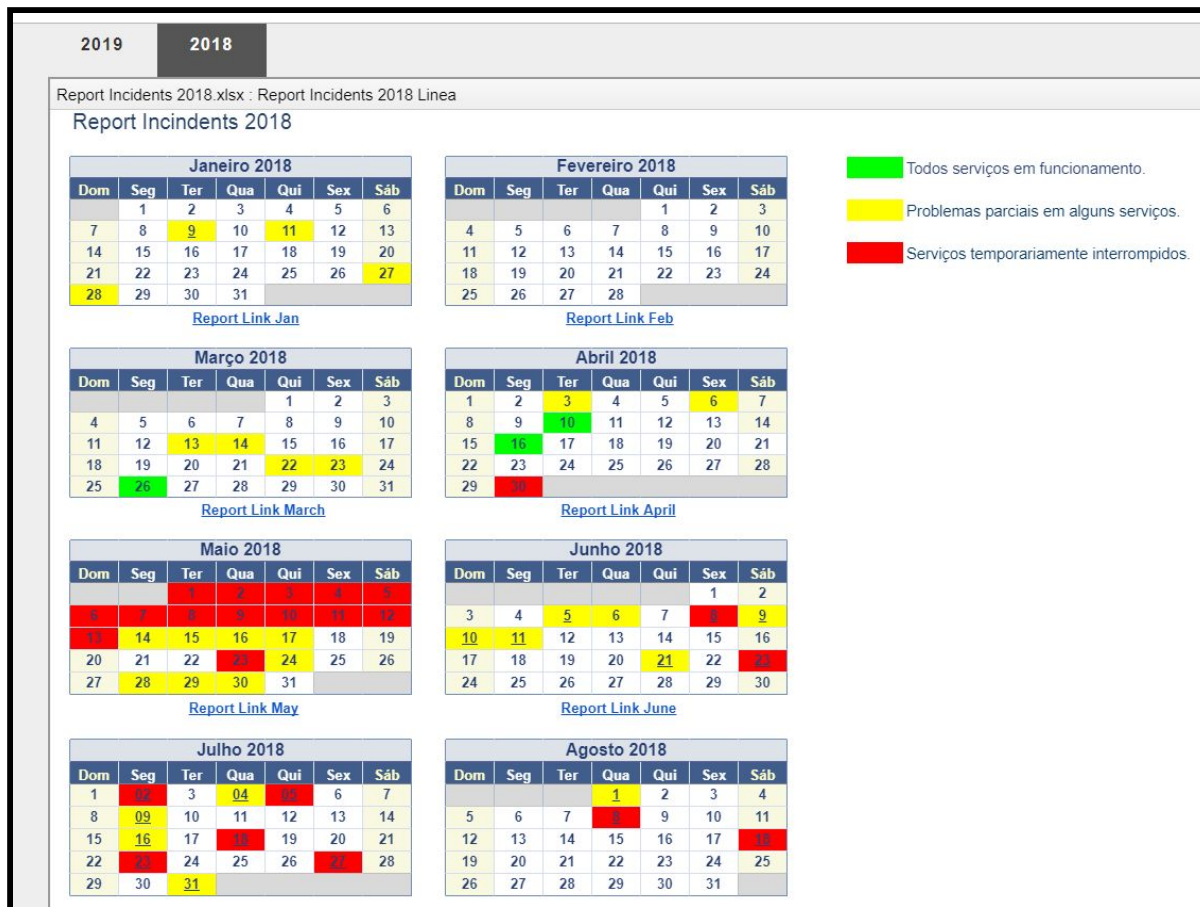


Figura 9 - Calendário de 2018 indicando problemas na operação do CAPDA.

Desenvolvimento de software

Em 2018 foram conduzidos e acompanhados o desenvolvimento de 10 sistemas de software de variado grau de complexidade desde atividades em tempo-real até plataformas hospedando workflows científicos. Estes projetos geraram 2583 tarefas de *sprint* totalizando 13365 horas estimadas de trabalho. Com as 5.9 FTEs disponíveis foi possível completar apenas 53% (1368) destas tarefas que consumiram 9375 horas, correspondendo a 70% do tempo estimado. A diferença de 30% é causada por desvios esperados no atendimento de *tickets*, participação em reuniões de planejamento e acompanhamento, e aprendizado de novas tecnologias. Isto demonstra que o time técnico deveria ter ~10 FTEs para atender a demanda cada vez mais intensa com o começo de novos levantamentos.

O trabalho de desenvolvimento ocorre em duas frentes: 1) sistemas complexos referidos como portais onde vários processos interdependentes que requerem parâmetros de configuração e cujos resultados precisam ser registrados; 2) interfaces interativas envolvendo processos de busca e visualização de dados incluindo sistemas em tempo real. Em 2018 os seguintes projetos foram desenvolvidos:

Portais

Portal Científico: um complexo sistema desenvolvido para a análise de dados de levantamentos fotométricos. Originalmente desenhado para atender as necessidades do levantamento DES, agora foi iniciada sua adaptação para atender ao projeto LSST a partir das lições aprendidas.

O portal hospeda um grande número de pipelines que: 1) criam dados auxiliares como mapas de profundidade e mapas que mostram a variação das condições observacionais, os quais são usados na criação de catálogos; 2) dados espectroscópicos mantidos pelo LIneA que são usados para treinar códigos que usam técnicas de *machine learning* para estimar *redshifts* de galáxias; 3) classificam objetos detectados nas imagens como estrelas ou galáxias; 4) calculam propriedades para as galáxias a partir da medida de fluxo nas bandas observadas; 5) criam catálogos para diferentes aplicações a partir de buscas no banco relacional associado ao portal; e 6) consomem os catálogos produzidos para análise.

Ao longo do último ano vários *workflows* científicos foram desenvolvidos e integrados a instância de produção incluindo: 1) um conjunto de pipelines usados pelo grupo de aglomerados de galáxias para identificar sistemas em dados reais e simulações e comparar com os resultados de outros autores; 2) o do grupo de arqueologia galáctica usado para estudar a estrutura da Via Láctea, gerando modelos que são comparados aos dados; 3) o do grupo que estuda a evolução de galáxias medindo a função de luminosidade de galáxias em diferentes domínios de *redshift* e desenvolvendo jupyter notebooks para facilitar o ajuste destas medidas à modelos, de forma interativa; e 4) um pipeline para medir a função de correlação ou de forma equivalente o espectro de potência da distribuição projetada de galáxias a diferentes *redshifts* e métodos para o cálculo da matriz de covariância usada na estimativa de parâmetros cosmológicos.

A experiência do último ano foi fundamental para demonstrar a necessidade de melhorar a interface entre as equipes científicas e técnicas, em particular de disponibilizar um pesquisador para servir de ligação entre estas equipes e ajudar na preparação de documentação e na definição de requisitos que ajude a equipe técnica na integração de

algoritmos no portal e de um programador científico que reveja os códigos fornecidos de forma a melhorar sua eficiência num ambiente de HPC.

A Figura 10 dá uma dimensão do escopo do projeto mostrando o *dashboard* (painel de controle) do portal científico do DES fornecendo informações sobre a execução de todos os pipelines hospedados. No *dashboard* os pipelines são agrupados em seções que representam diferentes estágios de processamento como: 1) instalação de novos dados; 2) construção de mapas que mostram a variação das condições observacionais em diferentes partes do céu; 3) preparação dos dados que inclui o cálculo de parâmetros adicionais para objeto detectado; 4) criação de catálogos para alimentar diferentes algoritmos científicos e amostras especiais; 5) workflows científicos para diferentes tipos de análise; 6) estimativa de parâmetros cosmológicos envolvendo processos MPI; e 7) outros pipelines utilitários. Existem atualmente mais de 60 pipelines validados na instância de produção e mais de 100 no ambiente de testes. Estes pipelines foram desenvolvidos pela equipe científica e técnica do LIneA e representa uma rica base de conhecimento acumulado.

Data Installation

Pipeline	Start	Duration	Runs	Status
Install Bright Mask	2018-07-20T14:28:08	09:01:14	4	Success
Install Catalogs	2015-11-03T15:01:01	02:26:39	1	Success
Install Depth Maps	2016-09-19T15:16:00	09:45:37	2	Success
Install Mangie Mask	2017-09-19T09:52:05	09:59:01	3	Success
Systematic Maps	2017-05-15T18:18:21	12:41:38	1	Success
Zeropoint Correction	2016-02-25T23:07:01	12:43:52	4	Success

Data Preparation

Pipeline	Start	Duration	Runs	Status
Photo-z Compute	2019-04-30T17:57:52	09:43:40	25	Success
Photo-z Training	2019-04-30T17:50:43	09:03:38	25	Success
Photo-z Validation	2017-11-07T17:36:37	09:49:26	1	Success
SG Separation	2018-09-19T12:53:37	09:30:15	13	Success
Training Set Maker	2019-04-17T12:11:29	09:04:57	33	Success

Science-Ready Catalogs

Pipeline	Start	Duration	Runs	Status
Cluster	2019-04-30T20:08:25	09:14:21	25	Success
GA	2017-08-18T16:33:04	09:17:37	2	Success
GE	2016-07-08T18:00:51	09:28:02	1	Success
Generic	2018-05-23T15:29:49	09:00:20	1	Failure

Simulations

Pipeline	Start	Duration	Runs	Status
No records found				
Utilities				
Catalog Comparison	2016-07-13T12:52:06	09:17:52	2	Success
QA Coadd	2016-07-15T10:40:19	10:33:49	8	Success
Make representative sa...	2018-06-15T21:12:16	09:09:23	1	Success
Template Set Maker	2018-05-23T15:27:33	09:00:25	2	Failure
Upload	2018-12-21T15:36:37	09:00:07	7	Failure
WAZP afterburner	2018-08-19T11:34:14	09:06:42	13	Success
Special Samples				
Cluster and Field Galaxies	2018-09-14T16:43:09	09:01:18	8	Success
RED LSS Sample	2018-03-22T09:39:18	09:00:38	1	Failure
Catalog Association	2017-07-12T18:42:14	09:57:56	1	Success

Science Analysis

Pipeline	Start	Duration	Runs	Status
ACF LSS	2017-11-06T17:52:24	09:22:38	1	Success
MW Fitting	2018-08-03T17:56:48	10:22:06	42	Failure
MW Fitting Decoupled	2018-09-14T15:31:23	09:04:04	1	Failure
StarHorse	2017-06-04T11:20:29	06:48:06	2	Success
WAZP	2019-05-02T18:03:30	09:29:52	14	Success

Figura 10 - Painel de controle (dashboard) do portal científico do DES listando os processos hospedados. Os pipelines são agrupados em diferentes seções (estágios) e para cada pipeline são apresentadas informações sobre a última execução incluindo quando foi iniciado, sua duração, status e o número de vezes que foi executado. A partir desta interface é também possível acessar todos os processos anteriores inclusive a configuração usada, a procedência dos dados de input e os resultados obtidos, graças a um sofisticado banco administrativo associado.


O portal científico envolveu até agora um esforço de mais de 60 FTEs-ano, dos quais 5,9 em 2018, e representa um dos principais legados tecnológicos do LIneA. No momento o enfoque principal é a contínua integração de *workflows* científicos na arquitetura do portal levando em conta os diferentes tipos de processamento e uma análise detalhada do desempenho do sistema para estabelecer requisitos que garantam a escalabilidade do sistema para o volume de dados esperado para o LSST.

Solar System Objects (SSO) portal: plataforma para o uso da colaboração franco-brasileira que estuda pequenos corpos do sistema solar usando a técnica de ocultação. A plataforma recolhe a posição de todas as exposições da câmera DECam feitas desde do início de sua operação em 2013, identifica CCDs que contém objetos do sistema solar examinando um banco de dados (*Skybot*) no LIneA que é mantido atualizado pelo Observatório de Paris, seleciona o tipo de asteroides de interesse, recupera as imagens contendo os estes objetos, procura melhorar o cálculo da posição desses objetos na imagem, combina estas novas determinações com as disponíveis no banco de dados do *Minor Planet Center* para prever a órbita desses objetos que é então comparada com a posição de estrelas na esfera celeste e assim predizer possíveis ocultações.

Este trabalho era feito manualmente pelo grupo, mas tendo em vista o aumento exponencial de ocultações possíveis de serem previstas, graças ao aumento significativo da detecção de objetos do sistema solar pelas novas câmeras grande angular sendo usadas, e a disponibilidade de posições mais precisas das estrelas do satélite GAIA, todo este esforço passa a ter uma escala industrial. Neste sentido desde 2017 o time do LIneA vem desenvolvendo uma plataforma que irá auxiliar neste tipo de estudo. Uma versão beta deve ficar disponível ainda no primeiro semestre de 2019 e será utilizada para analisar dados da DECam acumulados pelos projetos DES e DECALS. A Figura 11 mostra a interface para usuários sendo construída. A parte esquerda da tela lista as diferentes etapas do procedimento mencionadas acima, enquanto à direita o usuário tem acesso a execução dos processos de cada etapa e um resumo dos resultados obtidos. O projeto foi dividido em três partes: a primeira vai até a publicação das predições de ocultações; a segunda diz respeito a coordenação da campanha de observações para

uma ocultação incluindo o gerenciamento dos dados das observações obtidas por astrônomos profissionais e amadores em dezenas de observatórios espalhados pelo mundo; a terceira fase consiste da análise das curvas de luz obtidas pelas diferentes observações e sua interpretação sobre a estrutura do objeto sendo estudado.

Uma outra iniciativa sendo considerada é a de como integrar na plataforma os alertas sendo produzidos pelo projeto *Zwicky Transient Factory* (ZTF) que identifica transientes e serve como modelo para o que será fornecido pelo LSST. Isto permitirá ao time do projeto TON a experiência necessária para capitalizar no tsunami de dados que ficará disponível durante a operação do LSST que entra em testes já em 2020.



- [DASHBOARD](#)
- [REGISTRATION](#)
- [POINTINGS](#)
- [SEARCH SSSO](#)
- [FILTER OBJECTS](#)
- [ASTROMETRY](#)
- [REFINE ORBIT](#)
- [PREDICTION OF OCCULTATIONS](#)
- [OCCULTATIONS](#)
- [LIGHT CURVE ANALYSIS](#)

plot_dynamics

Total number of objects (and their observations) for each



Band	Observations	Objects
g	~1000	~1000
r	~1000	~1000
i	~1000	~1000
z	~1000	~1000
Y	~1000	~1000
J	~1000	~1000
u	~1000	~1000

plot_band

Total number of objects (and observations) for each band (u.g. i, z)



Band	Observations	Objects
Hungaria	~1000	~1000
KBO	~1000	~1000
Mars-Crosser	~1000	~1000
MB	~1000	~1000
NEA	~1000	~1000
Trojan	~1000	~1000

Total of CCDs
amount: 11204

Total number of observations
amount: 11204

object with the largest number of observations
name: KBO

Number of observations of object KBO
amount: 11204

SkyBot Output

complete list with all entries recorded in the database. can search for object name and number

OBJECT NAME	NUMBER	RA (DEG)	DEC (DEG)	EXPOSURE	DYNAMIC CLASS	VISUAL MAGNITUDE	ERROR ON THE POSITION
2007 RS312	378450	35.18575	-4.18804	258829	MB>Outer	20	0.057
2000 AY65	71301	35.78691	-4.26540	359699	MB>Outer	19.4	0.014

Figura 11 - Captura de tela do portal sendo desenvolvido para o projeto TON.

Interfaces para usuários

Além dos portais descritos acima, desde 2010 o LIneA vem desenvolvendo sistemas para a avaliação de dados em diferentes instâncias, tanto em tempo real como depois dos dados reduzidos. O primeiro desses sistemas foi o [Quick Reduce](#) disponibilizado em 2012 para uso no Observatório Interamericano de Cerro Tololo (CTIO) no Chile, para avaliar a qualidade dos dados da câmera DECam usada pelo projeto DES e pela comunidade astronômica internacional. O sistema foi encomendado pelo projeto DES e operou por quase sete anos sendo retirado de operação em Janeiro de 2019, após o processamento de alguns milhões de CCDs.

O sucesso deste sistema levou a novos pedidos como o [LIneA Data Server](#) disponibilizado para a colaboração DES em 2014 no Fermilab para avaliação dos dados reduzidos. Esta ferramenta ainda está em funcionamento após cinco anos.

Mais recentemente as seguintes ferramentas foram entregues para operação ou estão em desenvolvimento:

LIneA Science Server: uma versão moderna da ferramenta [LIneA Data Server](#) mencionado acima. Esta versão incorpora novas tecnologias, resolve problemas identificados na versão anterior e atende novas demandas feitas por usuários. A ferramenta está disponível no *National Center for Supercomputing Applications* (NCSA) desde de janeiro de 2018 e é uma das interfaces oferecidas pela colaboração DES para o acesso aos dados acumulados nos primeiros três anos de observação por toda a comunidade astronômica internacional. Este sistema fará parte do conjunto de serviços que será disponibilizado para toda a comunidade nacional, usando as instalações do LIneA. A Figura 12 mostra a interface apresentada ao usuário após seu registro no sistema. Atualmente, dois sistemas são mantidos no NCSA: um para membros da colaboração DES e outro para o público em geral. No futuro espera-se adaptar o sistema já amadurecido para outros repositórios como o SDSS e futuramente para os dados do LSST.

Quick Look Framework (QLF): ferramenta desenvolvida pelo time de TI para o projeto DESI como contrapartida para a participação de seis pesquisadores brasileiros. Este levantamento pretende observar num período de cinco anos da ordem de 50 milhões de galáxias para determinar a história da evolução do Universo com uma precisão sem precedentes. O QLF é desenhado para atender as necessidades do sistema que irá

avaliar em tempo-real a qualidade de 15.000 espectros de 5.000 objetos. O grande desafio é que o espectrógrafo a ser utilizado pelo projeto DESI será o único instrumento disponível no telescópio o que exige que o software de análise seja robusto e fácil de usar.

O sistema já passou por uma das revisões feitas pela colaboração e está sendo examinado por uma comissão de usuários antes de sua aprovação final e entrega prevista para junho de 2019. O sistema consiste de diferentes serviços que podem ser acessados pela interface mostrada na Figura 13, que servem para monitorar: 1) as observações (*Observing history*); 2) o processamento (*Pipeline Monitor*); 3) acesso aos resultados do processamento (*Processing History*); 4) o funcionamento do instrumento (*Trend Analysis*); 5) as condições atmosféricas e do telescópio (*Observing Conditions*); 6) o progresso do levantamento em diferente escalas de tempo (*Survey Reports*); 7) as configurações usadas (*Configuration*); e 8) a validação do plano de observações para a noite (*Afternoon Planning*).



Figura 12 - Interface do *LIneA Science Server* disponível no NCSA que permite a visualização de imagens (*Sky Viewer*), superposição de lista de objetos (exportada para o NCSA do portal científico ou gerada por buscas no banco de dados) sobre as imagens (*Target Viewer*) e realização de buscas no banco de dados gerando listas de alvos (*User Query*) que podem ser visualizadas com o *Target Viewer*.

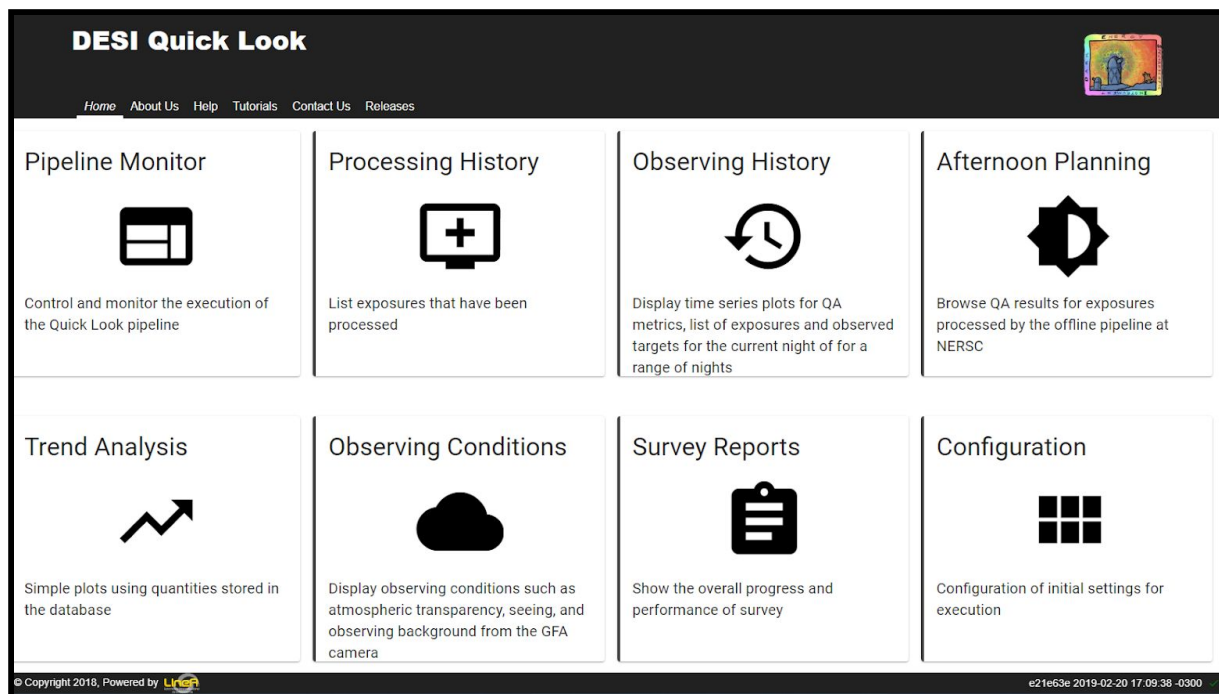


Figura 13 - Interface do sistema de tempo real QLF mostrando os diferentes serviços que podem ser acessados por esta interface o que permite ao usuário construir o seu próprio painel de controle durante a noite de observação.

LInEA Jupyter hub: seguindo sugestões do grupo de software do LSST foi instalado um serviço no LInEA para hospedar *jupyter notebooks*. A primeira implementação foi feita para o treinamento da equipe técnica do LInEA e para os pesquisadores que já participam do [BPG-LSST](#) oferecido por Will O'Mullane em setembro de 2018 durante sua visita ao LInEA patrocinada pelo INCT. Esta plataforma será uma das formas que os dados do LSST serão oferecidos.

O acervo de notebooks inclui o sistema de redução *stack* desenvolvido pelo LSST, entre outros exemplos criados para atender diferentes análises. Desde então o LInEA tem mantido sessões de *hack* onde novos notebooks adaptados às necessidades locais estão sendo criados. Os notebooks serão utilizados: i) no desenvolvimento de pipelines a serem posteriormente integrados ao portal; ii) na redução de imagens de diferentes câmeras; iii) na validação do catálogo bruto produzido pela colaboração; iv) na validação de produtos tornados públicos pelo LInEA como amostras espectroscópicas que consolidam as medidas de *redshift* obtidas por diferentes levantamentos; v) os dados usados no treinamento dos algoritmos que estimam o valor do *redshift* de galáxias a partir dos dados fotométricos (*photo-z*); vi) o catálogo com estas estimativas; e vii) catálogos preparados para a análise científica, chamados de *value-added catalogs (vacs)*.

LNA Archive – motivado pelo grupo de TNO com um grande número de observações realizadas no Observatório do Pico dos Dias e no *Southern Observatory for Astrophysical Research*, criar uma interface de busca de imagens destes observatórios acumuladas ao longo dos anos. Um protótipo já está em funcionamento e é mostrado na Figura 14. Uma das principais dificuldades é traduzir as palavras-chaves que caracterizam imagens de diferentes instrumentos e épocas para um sistema comum o que existe um grande esforço.

LSST exposure viewer – Ferramenta solicitada pelo LSST envolvendo tecnologia que permita a visualização rápida das imagens de 189 CCDs obtidas pelo telescópio a cada 20 segundos. Um protótipo foi desenvolvido pelo time de TI e proposto pelo LIneA ao LSST (ver Figura 15).

LNA Archive

PI Name _____

Observer
Bruno

Target

Target Name _____

Telescope
0.60m(BC)

Instrument
Cam2+Ikon

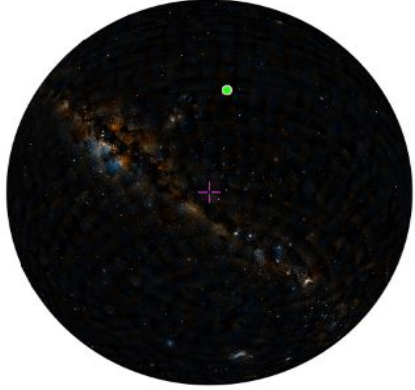
Filter
Any

Exposure Time (s)
time in seconds, greater than or equal to.

Observation Date
 Start Date dd/mm/aaaa
 End Date dd/mm/aaaa

Coordinates
 RA (deg) _____
 Dec (deg) _____
 Radius (deg) _____

J2000 15 43 2.304 -49 14 33.76



FoV: 180°

IMAGES DOWNLOAD HELP

Results: 1741

Download	Filename	Target	RA	Dec	Date Obs	Filter	Exptime	Telescope	Instrument	Observer	Size
	Jupiter_IL_0044.fits	Jupiter	15:21:21	-17:09:57	03/28/2018 3:15:34 AM	Metano	8	0.60m(BC)	Cam2+Ikon	Bruno Morgado	36 KB
	Jupiter_IL_0052.fits	Jupiter	15:21:21	-17:09:57	03/28/2018 3:17:20 AM	Metano	8	0.60m(BC)	Cam2+Ikon	Bruno Morgado	36 KB
	Jupiter_IL_0024.fits	Jupiter	15:21:21	-17:09:57	03/28/2018 3:11:08 AM	Metano	8	0.60m(BC)	Cam2+Ikon	Bruno Morgado	36 KB
	Jupiter_IL_0019.fits	Jupiter	15:21:21	-17:09:57	03/28/2018 5:45:11 AM	Metano	6	0.60m(BC)	Cam2+Ikon	Bruno Morgado	36 KB
	Jupiter_IL_0029.fits	Jupiter	15:21:21	-17:09:57	03/28/2018 3:12:15 AM	Metano	8	0.60m(BC)	Cam2+Ikon	Bruno Morgado	36 KB

Figura 14 - Interface sendo desenvolvida para a busca de imagens obtidas pelo projeto TON nos telescópios do OPD e SOAR.

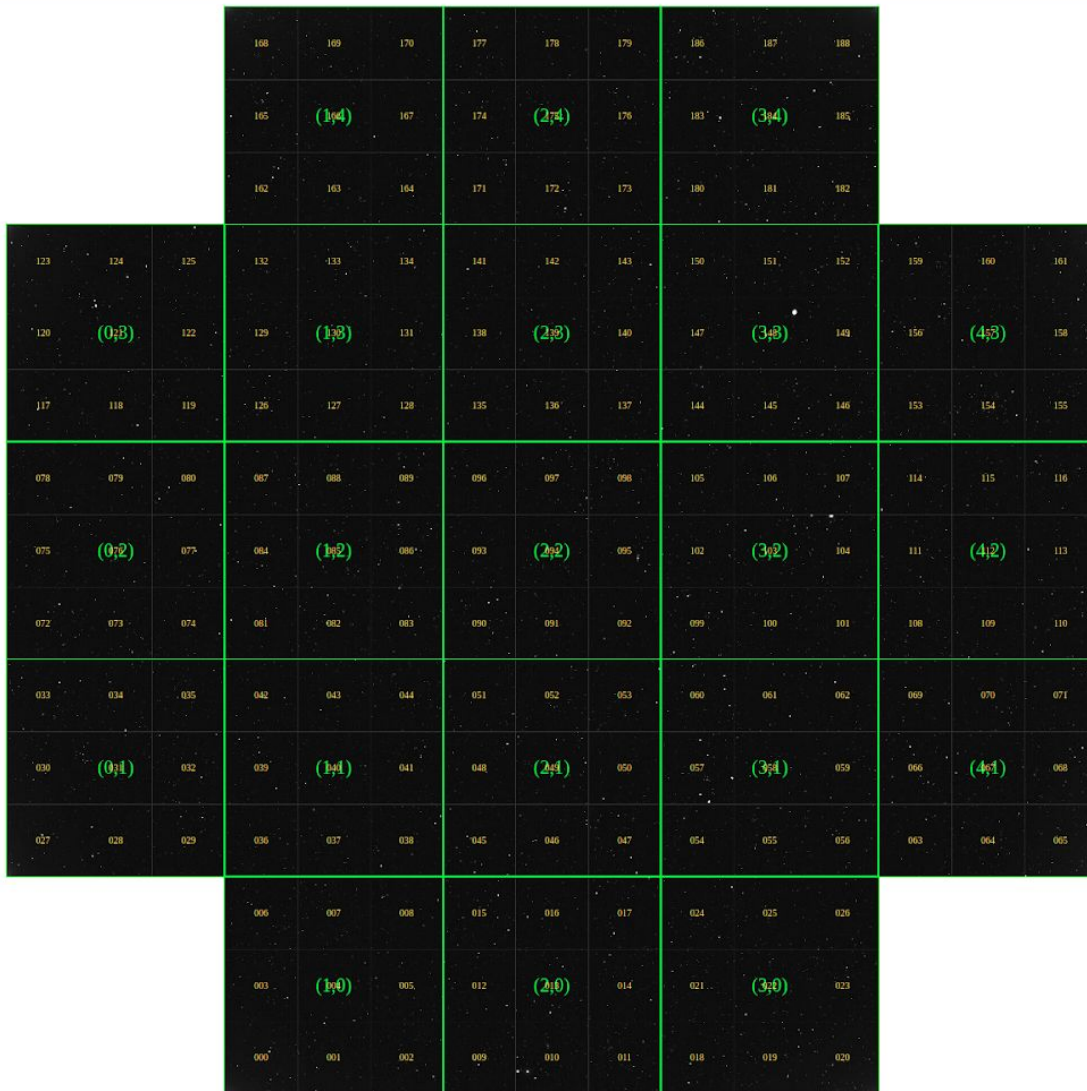





Figura 15 - Imagem simulada da câmera do LSST. A figura mostra em amarelo os 189 CCDs agrupados em 21 conjuntos (*rafts*) de 9 CCDs marcados em verde. A imagem de cada CCD é criada separadamente e o software desenvolvido pelo LIneA compõe a imagem em menos de 5 segundos.




LIneA Services - uma interface que está sendo desenvolvida para dar acesso via o site do LIneA/INCT aos serviços oferecidos, distinguindo o privilégio do usuário em relação ao acesso de dados públicos e privados de acordo com as regras das diferentes colaborações científicas. Os serviços para membros das colaborações incluirá acesso ao portal que hospeda pipelines adequados para a colaboração; um *jupyter hub* contendo *notebooks* com acesso a dados públicos e privados; acesso ao *Science Server* para a busca e visualização de dados públicos e privados da colaboração sendo considerada;

acesso a recursos computacionais em geral; e uma interface de acesso a produtos avançados criados a partir de processos executados pelo portal e *notebooks* associados. O mesmo será oferecido a membros da comunidade científica em geral, limitado apenas aos dados que estarão disponíveis para uso de acordo com as regras de cada colaboração.

LSST Portal – este é um projeto de médio prazo com inúmeras fases. A primeira requer a revisão do sistema de orquestração dos processos que permita um monitoramento mais adequado do processamento e que sirva para melhor diagnosticar ineficiências do sistema. Tudo isso será feito à partir do portal desenvolvido para o levantamento DES, a partir de requisitos para o novo portal. Com o mesmo propósito uma compilação de problemas encontrados com o DES portal está sendo mantido e possíveis soluções estão sendo debatidas. Outros pontos sendo avaliados são o gerenciamento de dados, o tipo de orquestração a ser usado para as aplicações de interesse da equipe brasileira e escalabilidade para o LSST das soluções adotadas para o DES. Neste sentido muitas dessas questões serão exploradas utilizando as simulações numéricas sendo feitas pelo LSST que já permite examinar algumas das questões de integração desses dados ao portal.






[Início](#)
[Ciência](#)
[Data Center](#)
[Software](#)
[Serviços](#)
[Divulgação](#)
[Colaboradores](#)
[Contatos](#)


Jupyter

Provides access to DES / NCSA / LSST notebooks integrated to available data.




Science Products

Provides access to user-generated products created from publicly available data.




Science Server

Provides access to publicly available data from the Dark Energy Survey.



TON publico

Provides access to occultation predictions.



Linea - Laboratório Interinstitucional de e-Astronomia
 Rua General José Cristiano, 77 - Vasco da Gama - Rio de Janeiro - RJ - Brasil - 20921-400
 (21) 3504-9165 - [Mapa de Localização](#)

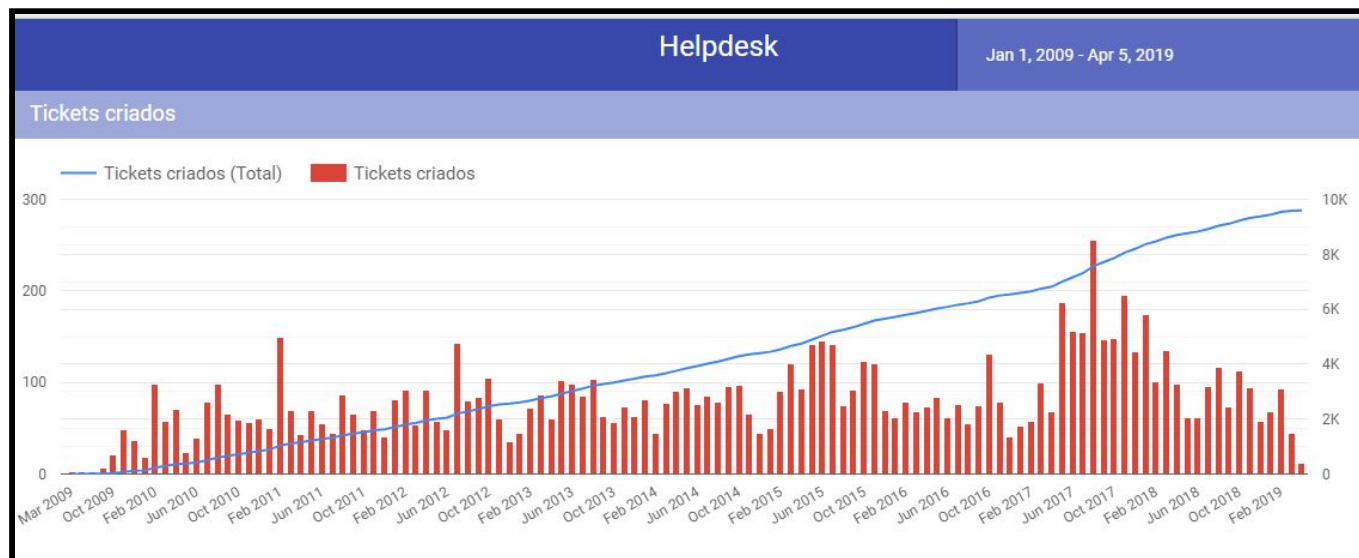
Figura 16 - Projeto em andamento no LIneA para o acesso de pesquisadores a serviços a serem disponibilizados para o público entre os quais: *jupyter hub*, interface para acesso a produtos gerados por usuários para distribuição; interface para busca, visualização de imagens e catálogos.

Durante o último ano um grande esforço de P&D também foi feito para investigar possíveis soluções para viabilizar o processamento distribuído. Através de parcerias com a RNP e a Universidade Federal de Campina Grande membros do LineA participaram de reuniões do Projeto Mercurius apoiado pela FAPESP. O objetivo do projeto é criar um ambiente de nuvem local integrando as diferentes plataformas instaladas como o cluster Altix e ICE-X da SGI e o novo cluster Apollo 2000 da HPE. Neste sentido durante 2018 tentativas foram feitas de instalar no cluster ICE-x o sistema *Openstack* em colaboração com o Núcleo de Computação Eletrônica (NCE) da UFRJ. Infelizmente, isto não foi possível devido a arquitetura da máquina. Uma nova tentativa será feita utilizando o cluster ALTIX, Uma das possibilidades sendo consideradas é a de integrar a nuvem interna a outros sistemas de instituições brasileiras com quem o LIneA tem parceria como a UNESP e UNICAMP. Outra possibilidade é usar containers para exportar os nossos sistemas para diferentes máquinas. Essas soluções estão programadas para 2019.

Serviços

O LIneA oferece inúmeros serviços para seus afiliados bem como para membros das colaborações internacionais. Entre os serviços prestados se destacam: 1) a operação de um centro de acesso e processamento de dados; 2) a operação de um portal científico; 3) a manutenção de repositórios de dados dos projetos DES e SDSS; 4) a manutenção de um serviço de processamento utilizando *jupyter notebooks*; 5) a manutenção de um site; 6) um serviço de *helpdesk* atendendo diferentes demandas (ver Figura 17); 7) a manutenção de uma wiki de apoio aos vários projetos apoiados para o compartilhamento de informações, publicações e documentação em geral; 8) a construção de pacotes de software sob demanda em apoio aos *workflows* científicos; 9) um repositório de códigos com mais de 800 códigos versionados; 10) um repositório com mais de 600 documentos dos mais variados tipos; 11) um repositório com mais de 150 documentos sobre os sistemas desenvolvidos e operação; 12) manutenção de um repositório com mais de 500 vídeos incluindo webinars, lunch talks, hack sessions, reuniões de sprints, tutoriais entre outros, 13) acesso a dezenas de ferramentas colaborativas; e 14) o monitoramento da operação de oito locais onde o LIneA tem máquinas (LNCC, POP-RJ), ou é responsável pelo funcionamento das máquinas (Observatorio de Las Campanas) ou tem sistemas desenvolvidos pelo LIneA em operação (Fermilab, NCSA, CTIO, KPNO, NERSC).

Figura 17 - Número de *tickets* abertos ao longo dos anos totalizando quase 10.000 nos últimos 10 anos.



Os sistemas em operação em 2018 incluíram o *Quick Reduce* no CTIO, o *LIneA Data Server* no Fermilab e o *LIneA Science Server* no NCSA onde foi mantido dois serviços, um para o grande público e outro para a colaboração acessando dados ainda em período de embargo.

Abaixo são apresentados números de uso para os sistemas em que é possível contabilizar visitas individuais. Por exemplo, a Figura 18 mostra o acesso ao site do LIneA que desde 2011 teve mais de 168 mil visitas sendo que durante o ano de 2018 o site foi visitado 37 mil vezes. Como não existe registro os números dados no gráfico são estimados a partir do número de visitas contabilizado usando Analytics.

O mesmo se aplica aos acessos ao [espelho do banco do SDSS](#) mantido e operado pelo LIneA desde 2012. Em 2018 este foi visitado 819 vezes de um total de mais de 19 mil visitas desde julho de 2012 quando foi liberado o serviço (Figura 19). Este serviço está aberto ao público e não requer registro. Existe ainda um serviço mais avançado chamado *CasJobs* com mais de 1300 requisições executadas ao longo dos anos por membros da comunidade brasileira. Uma das vantagens de se ter um espelho local do sistema disponível na Universidade de Johns Hopkins é a de poder adaptar os parâmetros de uso às necessidades locais, o que já foi feito atendendo à diversas solicitações.

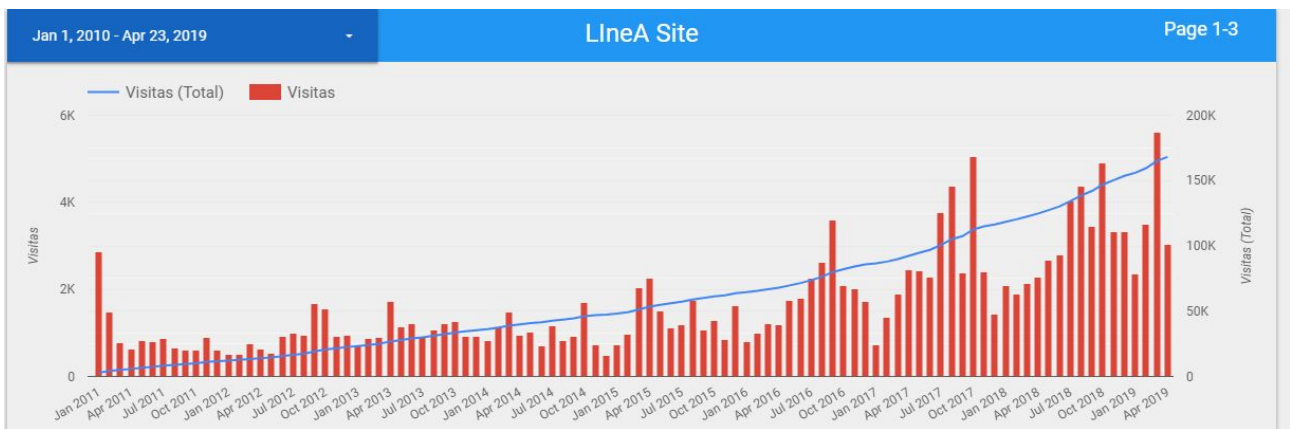


Figura 18 - Visitas feitas ao site do LIneA onde o usuário tem acesso a várias ferramentas e informações. A manutenção do site é de fundamental importância no modelo de operação do LIneA e requer uma atenção especial do time responsável pelos serviços.

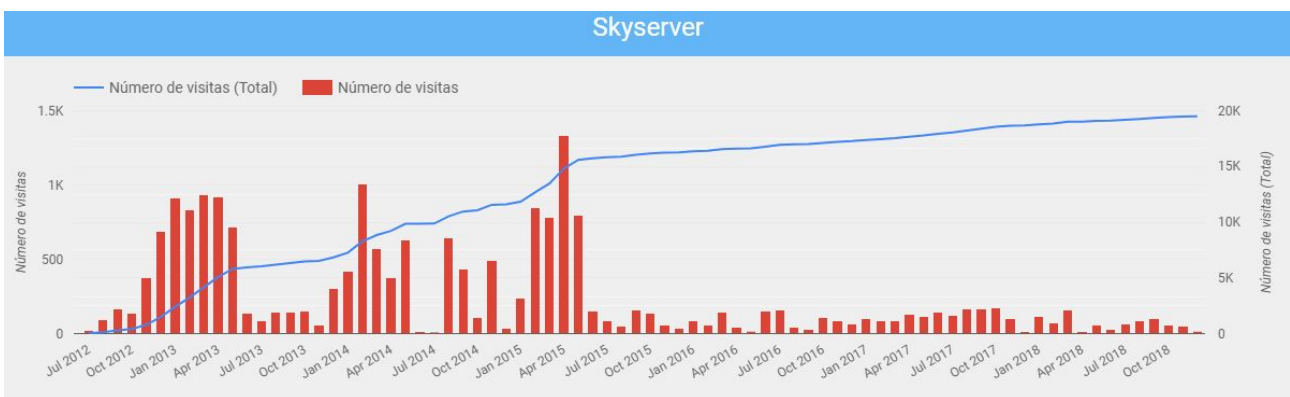


Figura 19 - Número de acessos ao serviço *skyserver* do SDSS hospedado pelo LIneA.

No caso dos serviços feitos a pedido da colaboração DES o número de usuários e visitas pode ser monitorado pois os mesmos exigem login. O *LIneA Data Server* disponível desde 2014 no Fermilab para membros da colaboração (Figura 20) e ainda em funcionamento tem 273 usuários registrados e foi acessado mais de 4700 vezes, uma média de aproximadamente de 900 acessos por ano. A partir do segundo semestre de 2017 teve uma queda pela demora de disponibilização de novos dados e a criação do sistema *LIneA Science Server* que representou uma importante evolução do *Data Server* e já tem o mesmo número total de usuários (278) com 1490 acessos durante 2018 (Figura 21).

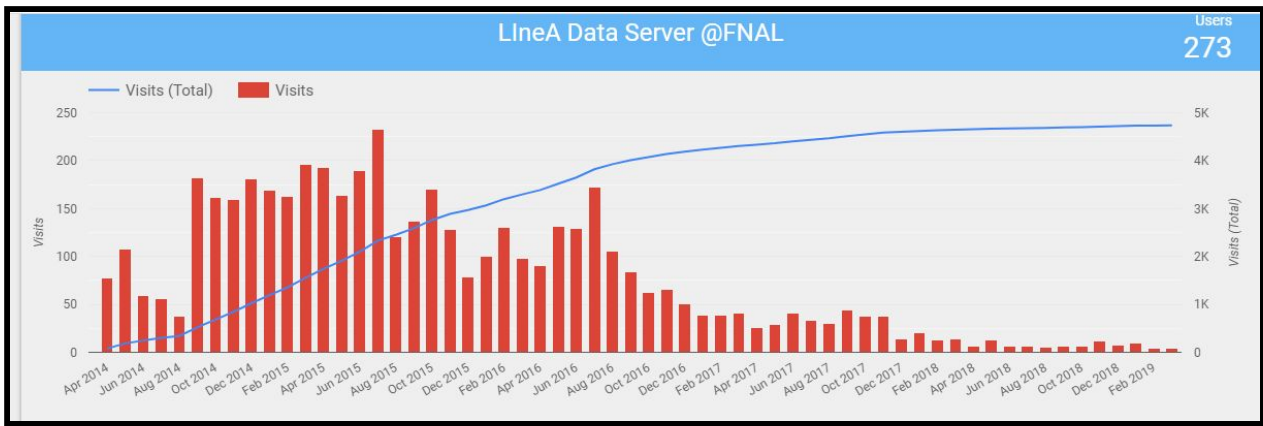


Figura 20 - Número de acessos por mês (histograma em vermelho) e acumulado (curva azul) desde o início da operação do LIneA Data Server no Fermilab.

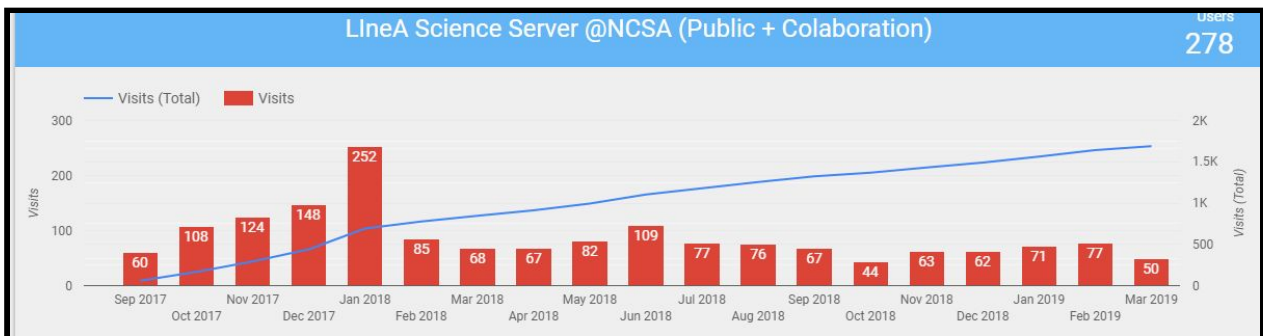


Figura 21 - Número de acessos por mês (histograma em vermelho) e acumulado (curva azul) do serviço LIneA Science Server disponível no NCSA.

Finalmente, a partir de discussões com o grupo técnico do LSST foi implantado em 2018 para a semana do LSST um serviço de *jupyter hub* onde são hospedados *jupyter notebooks*. Atualmente, quase 20 membros do LIneA estão usando este serviço de forma regular e um plano está sendo desenvolvido, baseado na experiência sendo adquirida, para disponibilizar este serviço para a comunidade astronômica.