



Relatório: INCT do e-Universo

(Janeiro 2017 Julho 2019)

Resumo Executivo

A idéia do [INCT do e-Universo](#) surgiu a partir da experiência de trabalho que teve início em 2006 com a entrada de um grupo de pesquisadores brasileiros no Dark Energy Survey (DES). Na ocasião foi negociada a entrada de 10 pesquisadores seniores (contratados) e um número ilimitado de pós-doutorandos e alunos em troca de uma contribuição financeira (US\$ 300 mil) e do compromisso de desenvolver sistemas de software para avaliação de dados em tempo real (Quick Reduce) e de produtos avançados como imagens co-adicionadas e catálogos (Science Server). Isto levou ao projeto estruturante do ON Astrosoft e ao acordo no final de 2010 entre o ON, LNCC, CBPF e mais tarde RNP para a criação do Laboratório Interinstitucional de e-Astronomia (LINeA) com a seguinte missão:

- Dar apoio logístico e financeiro para a participação de cientistas e estudantes em grandes projetos internacionais.
- Implantar e manter um Centro de Dados Astronômico para facilitar a exploração científica dos dados provenientes destes experimentos.
- Desenvolver e manter um portal científico para a validação e mineração de dados e para conduzir de forma eficiente a análise científica destes grandes acervos.
- Ajudar na formação de jovens pesquisadores e na sua preparação para atuar pró-ativamente em grandes colaborações internacionais.
- Ajudar na formação de tecnólogos para lidar com projetos de Big Data.
- Contribuir na divulgação dos resultados de pesquisa para o grande público.

Para seu funcionamento, o LINeA teve o apoio do MCTIC através de projetos submetidos a FINEP e CNPq e de editais da FAPERJ (e.g., Sediadas, Pronex) com a maior parte dos recursos indo para a formação de um time de TI e para a compra de equipamentos de computação para a criação de um Centro de Acesso e Processamento de Dados Astronômicos (CAPDA). No entanto, para cumprir sua missão faltavam recursos para inúmeras outras atividades o que só se tornou possível com a aprovação do INCT do e-Universo. Isto está permitindo o melhor aproveitamento científico dos vários projetos ([DES](#), [SDSS](#), [DESI](#), [LSST](#) e [TON](#)) com o apoio sendo dado a participação de pesquisadores em reuniões internacionais, bolsas técnicas e científicas, viagens de intercâmbio, além de contribuir pontualmente para o bom funcionamento do centro de dados. Após mais de dois anos de funcionamento o INCT teve os seguintes importantes impactos:

Impacto intelectual

- Publicação de mais de 172 publicações nos últimos dois anos nos mais variados tópicos de astronomia e cosmologia. Destas publicações membros do INCT participaram da ordem de dez comunicados de imprensa feitos pelos principais laboratórios mundiais (Fermilab, European Space Agency, National Science Foundation) inclusive de artigos liderados por membros do INCT.
- Seis teses de doutorado e seis dissertações de mestrado.
- Desenvolvimento de sistemas de software feitas pelo time de TI para as colaborações apoiadas sendo usadas por toda a comunidade internacional.
- Criação de um laboratório de ciência de dados voltado para explorar o céu digital.
- Participação de pesquisadores brasileiros em projetos de pesquisa na fronteira científica.

Impacto social

- Internacionalização de pesquisadores e tecnologistas através de colaborações científicas e tecnológicas.
- Formação científica da população através de diferentes iniciativas de divulgação.
- Preparação de projetos de Ciência cidadã utilizando ferramentas desenvolvidas pelo time de TI e usando a infraestrutura do LIneA.
- Projeção internacional de tecnologistas.

Impacto econômico

- Grande economia com a entrada nas colaborações em troca do esforço de desenvolver sistemas de software e serviços de interesse dos projetos.
- Caso de uso para a criação de um centro de suporte de e-ciência em colaboração como o LNCC e RNP que prevê a redução de custos de pessoal e equipamento através do compartilhamento de recursos.
- Transferência de novas tecnologias através do intercâmbio técnico promovido pelas colaborações internacionais (e.g., banco de dados de alto desempenho).
- Formação de jovens pesquisadores na ciência de dados.

Atualmente, as atividades do LIneA e do INCT se complementam, com o LIneA representando o lado mais técnico (desenvolvimento de software, centro de dados, serviços) e o INCT o lado mais científico (bolsas, auxílios, reuniões). A seguir estas atividades (gestão, centro de processamento, serviços, desenvolvimento de software, colaborações científicas, formação de pessoal e divulgação) são descritas e os principais resultados são apresentados.

Informações Gerais

O INCT do e-Universo é formado por 58 pesquisadores e 15 tecnologistas afiliados a 12 instituições espalhadas em cinco estados da federação (RJ, SP, PR, SC, RS). A composição da equipe é bem equilibrada como mostra a figura 1, com aproximadamente um número igual de pessoas em cada estágio da carreira.

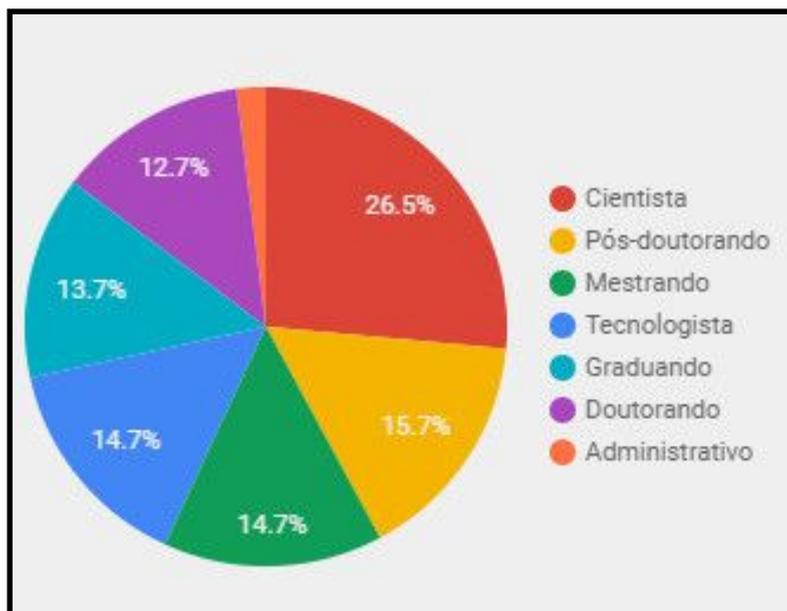


Figura 1 - Composição do time do INCT do e-Universo.

Até agora, dos R\$ 6,8 milhões concedidos, foram gastos após dois anos e meio de funcionamento R\$ 2,2 milhões, correspondendo a aproximadamente a 32% do total. Em condições normais, este perfil de gastos seria apropriado para cobrir o período do INCT. No entanto, é preocupante a situação da FAPERJ onde ainda precisam ser liberados R\$ 3,2 milhões e do CNPq com aproximadamente R\$ 1 milhão a receber. Um visão gráfica da situação global, sem distinguir a fonte e as rubricas, pode ser vista na figura 2.

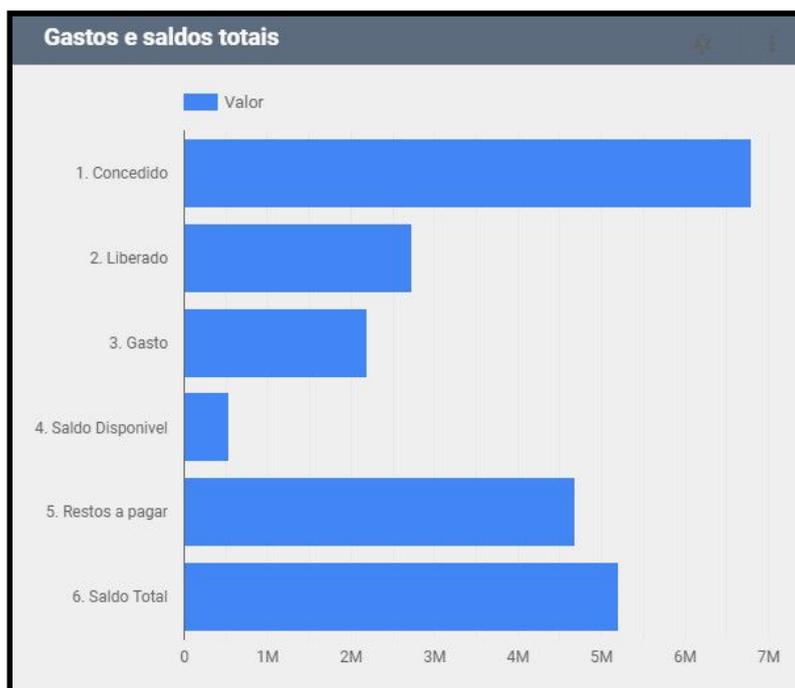


Figura 2 - Situação financeira do INCT do e-Univero sem considerar agências ou rubricas. Em princípio a posição do INCT é confortável caso a situação da FAPERJ seja regularizada.

Os gastos do INCT foram distribuídos entre bolsas, equipamentos e custeio, na sua maior parte gasta em auxílios para viagens. Por exemplo, a figura 3 mostra como os recursos do CNPq foram investidos.

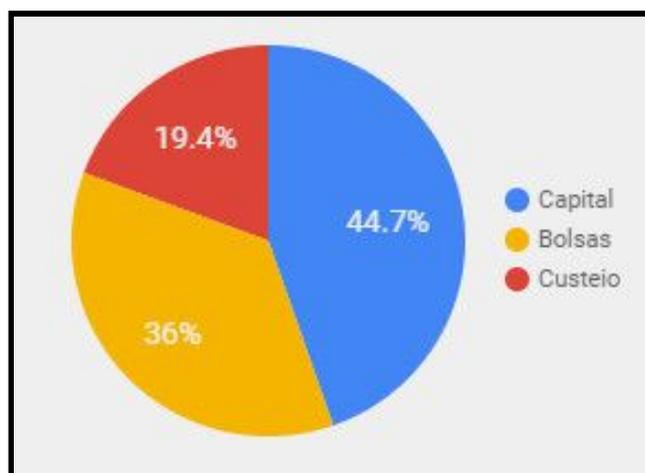


Figura 3 - Tipo de gastos realizados até agora com os recursos do CNPq com a maior parte sendo usada para equipamentos seguido de bolsas. O perfil de gastos deve mudar nesta segunda fase com mais recursos indo para auxílios viagem.

Gestão

A gestão do INCT do e-Universo é feita por seu coordenador assessorado pelo Comitê Gestor que se reúne semanalmente para acompanhar as diversas atividades, aprovar gastos e a concessão de bolsas e auxílios para participar de reuniões nacionais e internacionais seguindo recomendações de um [comitê de especialistas](#), no caso de grandes compras de equipamento de informática, e comitês de bolsas técnicas e científicas.

Os recursos disponibilizados pelo INCT deram uma enorme flexibilidade na participação de pesquisadores em reuniões internacionais, fundamental para este INCT cuja principal missão é apoiar colaborações internacionais. Outra importante contribuição do INCT foi a possibilidade de usar bolsas técnicas para atrair uma equipe com experiência de gestão de projetos que colaborou muito para a organização do INCT e do LIneA. Com os recursos disponíveis do INCT 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 portal científico desenvolvido para o Dark Energy Survey ([DES](#)) e, baseada na experiência acumulada ao longo dos últimos 10 anos, está compilando requisitos para o novo portal sendo desenvolvido para o projeto *Large Synoptic Survey Telescope* ([LSST](#)). 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 sendo criados, introduzindo metodologias (e.g., *Agile*) e formalizando processos. Em particular, com a equipe de gestão está sendo possível: 1) organizar reuniões de [planejamento estratégico](#); 2) 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; 3) organizar reuniões quinzenais de *sprint*, semanais de acompanhamento e na medida da necessidade de validação; 4) monitorar o progresso dos projetos em reuniões (remotas) de *stand-up* diárias; 5) implementar o uso de time tracking software para rastrear o tempo gasto por cada desenvolvedor nos diferentes projetos em andamento. Desta forma é possível estimar o custo em FTEs (*full-time equivalent*) das diferentes ferramentas sendo desenvolvidas, importante para negociar o valor das contribuições *in-kind*; 6) 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; 7) formalizar a entrega de produtos; 8) preparar [relatórios mensais/anuais](#) disponíveis publicamente relatando as atividades de desenvolvimento e serviços, mantendo desta forma devidamente informados das atividades em andamento todos os interessados; e 9) criar e supervisionar “*tiger teams*” para resolver problemas importantes e urgentes. Todas as reuniões são normalmente gravadas e já contabilizam aproximadamente 600 gravações;

Desde 2017 foram realizadas 56 *sprints*, de duas semanas cada, para projetos complexos 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 de 2018 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](#)) entregue para avaliação em 2019; 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; 5) um sistema rápido de visualização de imagens para servir ao LSST, permitindo inspecionar os 189 CCDs que formam uma imagem do LSST em menos de cinco segundos, e que está sendo negociado com o projeto como uma contribuição in-kind; 6) Interfaces para monitoramento de processos do portal e de uso da infraestrutura computacional; 7) Interfaces financeiras e de métricas para o LIneA e para o INCT; 8) implementação de plataforma para *jupyter notebooks*; 9) sistema autenticação e autorização de serviços; 10) atualizações do site.

Uma outra tarefa importante 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 e a possível implantação de um dos Centros de Acesso a Dados independentes sendo previsto pelo projeto LSST. Também foi recentemente iniciado um processo de validação desta documentação e reuniões semanais para discutir documentos específicos de interesse amplo onde pessoas de diferentes background participam.

Para facilitar o processo de gestão foram criadas pela equipe de TI do LIneA interfaces que permitem ao Comitê Gestor do INCT e o público acompanharem os gastos e as métricas de desempenho. A interface financeira permite acompanhar automaticamente o fluxo de caixa do INCT, analisar os gastos de capital e custeio realizados por agência, por colaboração científica e por estado e os gastos com bolsas por tipo, por agência e por estado das instituições participantes. A interface de métricas inclui o número de eventos nacionais e internacionais organizados ou que tiveram a participação de membros do INCT, número de publicações por ano, número de webinars e *lunch talks* organizados, e número de notícias publicadas no site do LIneA/INCT. Combinadas estas interfaces servem para o Comitê Gestor avaliar a aderência do INCT a missão proposta. Estas interfaces estão disponíveis no site do [LIneA/INCT](#). A gestão de recursos humanos da equipe técnica que apoia o INCT foi beneficiada pelo apoio sendo dado desde 2018 pela RNP que administra os recursos concedidos ao LIneA pelo MCTIC.

Internamente as atividades do LIneA/INCT são revistas semanalmente numa reunião para transmitida pela internet usando o sistema de conferência zoom disponível para todos os seus membros. Nesta reunião são revistas as principais atividades da semana pelo coordenador, representantes de diferentes grupos científicos, informes sobre o centro de dados, highlights da área de desenvolvimento de software. Outras formas de comunicação também sendo usadas

rotineiramente entre os membros do INCT são o sistema de mensagens *slack* e listas de correio eletrônico mantidas pelo LIneA.

Além destas reuniões semanais o INCT organiza periodicamente reuniões gerais com diferentes enfoques como:

- [Reunião Geral do INCT \(19/05/2017\)](#);
- [Dia de Ciência \(22/01/2018\)](#).
- [LSST Workshop \(20/03/2018\)](#).
- [Reunião Geral INCT/LIneA \(06/08/2018\)](#).
- [Semana do LSST \(24-28/09/2018\)](#).
- [Planejamento Estratégico \(17-18/01/2019\)](#).
- [Revisão Planejamento Estratégico \(2-3/07/2019\)](#).
- [LIneA Bootcamp \(2-6/09/2019\)](#).

Outra forma de obter um retorno de seus membros uma enquete com 19 perguntas para avaliar os serviços prestados pelo LIneA/INCT foi realizada em Março de 2019 e os resultados foram apresentados num [blog](#) publicado em 30/05/2019 no site. A enquete motivou a organização da reunião LIneA Bootcamp com o objetivo de familiarizar os membros do INCT com o funcionamento e os serviços disponibilizados pelo LIneA. Isto será feito através de apresentações, tutoriais e de aulas práticas que serão preparadas por pesquisadores e membros das equipes técnicas e de gestão do LIneA. Também seguindo sugestões da enquete uma [caixa de sugestões](#) anônimas foi disponibilizada no site.

O trabalho sendo executado pelo LIneA trabalhando na fronteira da pesquisa científica e o tratamento de grandes volumes de dados tem sido usado como caso de uso para a idéia da criação de um Centro de Suporte a e-Ciência numa colaboração LNCC e RNP. Em particular, o envolvimento do LIneA/INCT no projeto LSST e a necessidade de implantar um centro de acesso a dados deste projeto (iDAC) criam novas demandas de infraestrutura que não são facilmente atendidas com as organizações atuais. Esta discussão foi o tópico de um [Design Sprint](#) encomendado pela RNP realizado no período de 2-5 de Outubro de 2018 com a participação de vários membros do LIneA e INCT.

Apesar do enorme esforço realizado pela atual equipe está claro que devido à grande complexidade da operação do LIneA e do INCT do e-Universo o grande número e diversidade de suas atividades requer uma equipe maior e um novo modelo de governança. Estas questões vêm sendo debatidas com o Grupo de Trabalho (GT) nomeado pela Secretaria Executiva do MCTIC e formado pelos diretores do LNA, LNCC, ON e RNP. Como parte deste trabalho, sendo desenvolvido ao longo dos últimos três anos, foram consideradas diferentes possibilidades de institucionalização do LIneA. A proposta mais concreta é a criação de uma associação sem fins lucrativos e um estatuto foi elaborado e deve ser apresentado em breve ao GT e ao MCTIC. Esta decisão é fundamental para que seja possível assegurar a continuidade de seu trabalho essencial para o INCT do e-Universo em particular para atingir sua principal missão de apoiar projetos como DESI e LSST que se estendem por toda a próxima década.

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.

A infraestrutura computacional do LIneA (figura 4), sendo usada pelos membros do INCT, vem sendo construída ao longo da última década através de contribuições do ON e de recursos obtidos por projetos submetidos a FINEP e FAPERJ. Recentemente, o INCT foi fundamental na modernização do centro contribuindo com a compra do cluster ICE-X da SGI realizada em 2017 e em 2018 do cluster APOLLO 2000 da HPE, aumentando desta forma significativamente a capacidade de processamento do centro de ~4,8 TFLOPS para ~24 TFLOPS. O centro tem ainda uma capacidade de armazenamento de 780 TB. O investimento total feito até agora no centro corresponde a ~US\$ 1,5 milhões, mas novos investimentos são necessários tanto para a renovação de equipamentos como também a expansão da infraestrutura para atender às crescentes demandas de processamento e armazenamento devido ao crescimento no número de usuários, no volume de dados e a complexidade das análises.

Um dos problemas mais urgentes é o de armazenamento tendo em vista a idade do sistema atual e de recorrentes falhas nos discos e a necessidade de reconciliar os requisitos para o processamento de alto-desempenho e espaço de armazenamento. Outro desafio é conciliar uma visão de médio prazo, como discutido abaixo, com as necessidades de curto prazo e a disponibilidade financeira. No momento, também estão em andamento estudos procurando definir o que um centro de acesso e processamento no Brasil deve oferecer em 2022, quando começa o levantamento do LSST. A idéia é propor ao projeto LSST a implantação no Brasil de um Centro de Acesso a Dados independente (iDAC em inglês) para apoiar os pesquisadores brasileiros que fazem parte do *Brazilian Participation Group* (BPG) do LSST (ver abaixo) durante o período de dados com embargo e a comunidade brasileira de astronomia e cosmologia para os dados no domínio público, o que aumentaria a competitividade dos pesquisadores brasileiros. O CAPDA seria o precursor desse centro. Este trabalho tem as seguintes vertentes:

- Armazenamento - diferentes fornecedores de sistemas de armazenamento foram entrevistados e as necessidades do centro descritas. No momento, com o apoio de um técnico da RNP as soluções propostas estão sendo revistas e serão apresentadas ao comitê de especialistas que fará uma recomendação Comitê Gestor do INCT. 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. Recursos disponíveis na FINEP serão usados para adquirir o sistema recomendado. Para atingir este objetivo, depois de rever propostas de quatro fornecedores, durante os últimos meses, a equipe de TI do Linea tem trabalhado em conjunto com o time de engenharia da Dell para definir uma solução de storage para suportar todos os requisitos de processamento exigidos pelos pipelines científicos do LIneA. Foram feitas diversas reuniões, trocas de e-mails e testes no laboratório de engenharia da Dell para chegar numa solução escalável, que atenda os requisitos atuais e futuros do LSST e que coubesse no orçamento disponível para compra.

[Link para acesso à figura 4.:](#)

Figura 4 - Diagrama da infraestrutura computacional do LIneA usada pelos membros do INCT.

- Processamento - é consensual que não deverá existir uma solução única para atender toda a demanda de processamento e que provavelmente será uma solução híbrida, dependendo do tipo de processamento e pressão de tempo, envolvendo clusters dedicadas, nuvens privadas ou comerciais e o uso de supercomputadores. Neste sentido a equipe do LIneA e membros do INCT participaram de reuniões com pesquisadores do NCE tentando implantar um ambiente de nuvem integrando os vários clusters locais e participam do **projeto Mercurius** da FAPESP em colaboração com a Universidade de Campina Grande e RNP para estudar o uso de nuvens federadas. Outra solução sendo considerada é o uso de supercomputadores e para isso o desempenho de diferentes pipelines usados pelo LIneA (cálculo de redshifts fotométricos, simulações, medida de correlações de galáxias e aplicações de MPI) estão sendo instalados e estão usando as 500.000 AUs aprovadas para avaliar diferentes casos de uso. O objetivo é explorar a possibilidade do LIneA contribuir para a expansão do supercomputador tanto para processamento como para armazenamento, e fazer parte do o Centro de Dados para o LSST. Para isso foram instalados programas sendo usados na análise do DES como *LePhare*, *MLZ*, *Namaster*, *Treecor*, *Cosmosis* tanto no cluster ICE-X como no Santos Dumont para comparar os tempos de execução alcançados no ICE-X, Santos Dumont e NERSC no LBL também sendo usado por membros do INCT. Mais recentemente, o LIneA participou da proposta *Big Data Research Infrastructure Collaboration Toward the SKA (BRICKS)*, submetida em conjunto com o LNCC ao **BRICS Astronomy Flagship Project**.

A grande maioria das máquinas do LIneA estão instaladas no CPD do LNCC em Petrópolis e são monitoradas e mantidas por uma firma especializada que trabalha com o LIneA há vários anos e que foi a responsável por toda a instalação em Petrópolis. O trabalho de manutenção foi intenso no passado devido a problemas de infraestrutura no CPD do LNCC (falhas no fornecimento de energia) o que causou a indisponibilidade das máquinas por pelo menos 13% do tempo e danos aos equipamentos do LIneA. Na verdade a continuidade da operação só foi possível, em grande parte, aos recursos do INCT que permitiram a rápida substituição do grande número de discos danificados. Mesmo assim, vários nós do cluster Altix tiveram que ser removidos permanentemente e tornaram a compra dos clusters ICE-X e Apollo prioritárias. Felizmente, medidas corretivas foram tomadas pelo LNCC e a qualidade do serviço melhorou significativamente em 2019.

Do ponto de vista do INCT, um grande esforço foi feito para melhorar a comunicação com seus usuários para mantê-los informados do estado do centro e para isso foram implementadas diversas melhorias como: a) o [monitoramento](#) e [registro](#) dos problemas de operação; b) o envio de avisos para os usuários de paralisações programadas ou emergenciais; c) a redução do tempo de resposta de tickets abertos no helpdesk; c) o desenvolvimento de uma interface de acesso ao serviço de monitoramento dos clusters de processamento (figura 5); d) o desenvolvimento de uma interface de acesso rápido ao serviço de monitoramento dos processos do portal científico. De novo essas melhorias só foram possíveis graças à disponibilidade das bolsas técnicas oferecidas pelo INCT que permitiram o crescimento da equipe de TI do LIneA. Isto faz parte do esforço sendo feito pelo LIneA de desenvolver uma infraestrutura e uma cultura de serviço que permita atender

as necessidades de um público maior do que aquele envolvido nas colaborações científicas apoiadas pelo laboratório e atuar como precursor do iDAC. Entre as iniciativas sendo tomadas uma é a reorganização do site do LIneA com o propósito de torná-lo um ponto de acesso a serviços que irão incluir serviços de computação, acesso a dados e *jupyter notebooks*. Para isto o time do LIneA está trabalhando com o grupo de trabalho Gidlab da RNP para implementar um serviço de autenticação e autorização e está em processo de contratação de uma firma de design para a reestruturação do site para facilitar a navegação do público e dos cientistas.

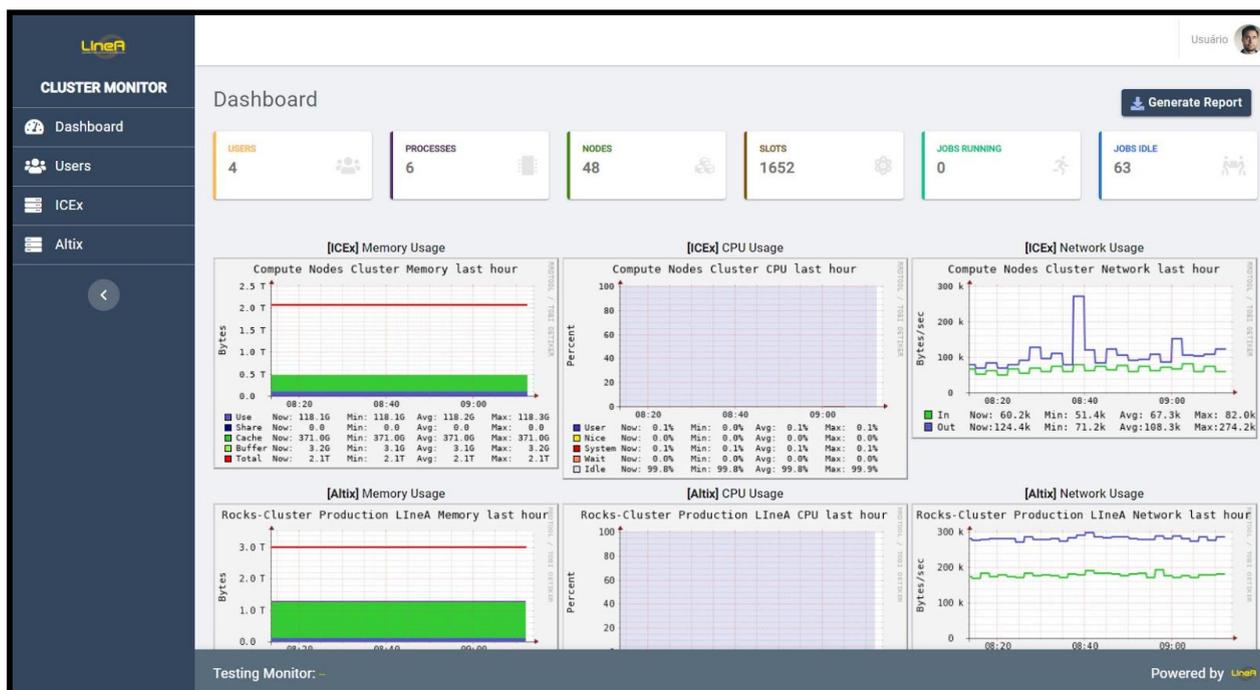


Figura 5 - Sistema de monitoramento dos clusters construída pelo time de TI do LIneA integrando informações de serviços de monitoramento (ganglia), facilitando aos usuários acompanhar seus processos nos diferentes clusters. A interface ainda se encontra em fase de testes. A figura mostra uma de quatro interfaces. **Nota: esta figura como outras apresentadas neste relatório devem ser vistas usando o zoom no arquivo pdf.**

Outro trabalho fundamental que está sendo feito em colaboração com o grupo de trabalho IDS (Infraestrutura definida por software) da RNP é a criação de containers e kubernetes para facilitar a disponibilização e gerenciamento de códigos e *jupyter notebooks* em diferentes ambientes o que tornaria mais eficiente a integração de outros sistemas para analisar dados dos grandes levantamentos apoiados pelo LIneA adotando a solução híbrida mencionada acima. Em particular, a viabilidade do uso de *containers* no Santos Dumont em estudo agora será fundamental para as decisões a serem tomadas a respeito de novos investimentos a serem feitos para expandir a capacidade de processamento do LIneA e que irá impactar o projeto para a implantação do iDAC.

Pela discussão acima fica claro os benefícios da colaboração entre o LIneA, LNCC e RNP graças ao compartilhamento de custos e conhecimento e a economia que isso representa. Esses benefícios poderiam ser expandidos para outras áreas com a criação de um Centro de Suporte de e-Ciência que em conjunto com o LIneA poderia ser responsável pela operação do iDAC do LSST brasileiro como de outros centros similares.

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, SDSS e TON; 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, com mais de 11.000 tickets; 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 (Observatório de Las Campanas) ou tem sistemas desenvolvidos pelo LIneA em operação - Fermilab ([LIneA Data Server](#)), NCSA ([LIneA Science Server](#)) e até janeiro de 2019 no CTIO (*Quick Reduce*) quando as observações do DES foram concluídas.

Alguns desses serviços ainda estão restritos a membros do LIneA/INCT em particular o acesso ao uso de *jupyter notebooks* e processamento em geral. Isto se deve a limitações de infraestrutura que estão sendo corrigidas com a contratação de uma gerente de serviços coordenando o trabalho de 2,5 FTEs, a implantação de um sistema de autenticação e autorização já mencionado acima, a compra de novos equipamentos, o redesenho do site (*landing page*) e a preparação de documentação e tutoriais que facilitem o acesso do público e a implantação da plataforma *jupyter Lab*.

Certamente um dos principais serviços do LIneA e uma das missões do INCT é dar acesso público, após um período de embargo, aos dados disponibilizados pelos grandes levantamentos apoiados pelo LIneA/INCT, multiplexando o potencial científico dos dados.

Os serviços de acesso a dados desenvolvidos pelo LIneA (*Data Server* e *Science Server*) permitem registrar o usuário e desta forma foi possível criar interfaces de gestão para acompanhar o uso dos sistemas ao longo do tempo automaticamente e esta auditoria está disponível no site. Com a implantação do esquema de autenticação federada e autorização sendo implantado será possível manter o mesmo histórico para os outros serviços.

A Tabela 1 apresenta na coluna (1) o serviço; na coluna (2) quando possível o número de pesquisadores registrados na grande maioria estrangeiros; na coluna (3) o número de visitas aos respectivos sites durante o período de funcionamento do INCT; e na coluna (4) o número total de visitas desde a implantação dos serviços com o ano quando iniciado o serviço e/ou o monitoramento em parênteses. Os valores apresentados na tabela foram obtidos das interfaces mencionadas acima

Tabela 1

Serviço	# Usuários	Visitas (2017-2019)	Visitas (Tota)
LIneA/INCT Site	-	95.000	180.000 (2011)
SDSS Sky Server	-	2.371	19.631 (2012)
DES - Data Server	273	548	4.700 (2014)
DES - Science Server	278	1.690	1.690 (2017)

Esta experiência tem sido fundamental em vários aspectos levando a colaborações com as equipes técnicas de *Johns Hopkins University* com larga experiência sendo responsável pela curadoria dos dados do SDSS por quase 20 anos e NCSA que será o hub principal do LSST. Uma das vantagens de se ter um espelho local dos dados do SDSS e do DES é a de poder adaptar os parâmetros de uso às necessidades locais, o que já foi feito atendendo à diversas solicitações. Esta experiência é valiosa para a implantação do iDAC brasileiro.

O LIneA/INCT também mantém um grande número de ferramentas colaborativas (algumas gratuitas outras pagas com recursos do INCT) para uso por seus membros para comunicação (*Zoom, Slack, Mattermost*), redação de documentos (*Overleaf, Grammarly*), planejamento (*Trello, Smartsheet*) e desenho de interfaces (*Lucidchart, Balsamiq*).

Finalmente, a partir de discussões com o grupo técnico do LSST foi implantado no final de 2018, para o evento "[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. O objetivo agora é facilitar o uso desses *notebooks* para acessar e analisar dados públicos e desta forma treinar jovens pesquisadores para o uso destas ferramentas que farão parte da plataforma de ciência (science platform) sendo desenvolvida pela equipe do LSST como interface aos dados. Um infraestrutura similar está sendo projetada para o LIneA integrando serviços como o Portal Científico, *LIneA Science Server* e *Jupyter notebooks*.

A colaboração com a RNP também tem sido muito importante para testar as várias ferramentas de e-ciência sendo desenvolvidas pelos grupos de trabalho apoiados pela RNP.

A equipe técnica do LIneA foi ainda responsável pela instalação, monitoramento e manutenção dos computadores e sistemas usados pelo sistema de aquisição do espectrógrafo [APOGEE-2](#) do SDSS montado no telescópio *du Pont* de 2 metros do [Observatório de Las Campanas](#) no Chile. É importante lembrar que este trabalho em conjunto com o serviços de distribuição de dados do SDSS ([SkyServer/Casjobs](#)) representou uma economia de US\$ 400 mil para a formação do [Brazilian Participation Group \(BPG\)](#) que permitiu a participação de 10 pesquisadores brasileiros mais um número limitado de estudantes patrocinados pelo LIneA.

Desenvolvimento de Software

Uma das principais atividades do LIneA ao longo dos anos tem sido um agressivo programa de desenvolvimento de software que como mencionado acima serviu como moeda de trocas para a entrada de pesquisadores brasileiros no DES e mais recentemente no projeto DESI. Entretanto, por muito tempo o processo de planejamento, acompanhamento, previsão de entrega e a contabilização do esforço exigido por estes projetos não era possível de ser feito de uma maneira profissional. Graças ao INCT foi possível engajar um time de gestão que revolucionou a forma de trabalhar com um enorme impacto. Por exemplo, agora é possível saber que 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. Esses projetos geraram 2.583 tarefas de *sprint* totalizando 13.365 horas estimadas de trabalho. Com as 5,9 FTEs disponíveis foi possível completar apenas 53% (1.368) destas tarefas que consumiram 9.375 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.

Outra importante contribuição da equipe de gestão foi a criação de um processo bem definido de planejar, acompanhar, validar, e liberar para uso e manter sistemas desenvolvidos pela equipe de TI do LIneA que requer a criação de fichas de projetos, descrição detalhada do sistema a ser desenvolvido, detalhamento de tarefas, integração no cronograma geral, programação de sprints, acompanhamento pela *scrum master* e por um comitê de controle de mudanças. Isto representou uma grande mudança que só foi possível com o apoio do INCT.

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. Durante os últimos dois anos os seguintes projetos foram desenvolvidos:

Portal Científico: um complexo sistema desenvolvido para a análise de dados de levantamentos fotométricos. Originalmente desenhado para atender as necessidades do levantamento DES, o esforço agora é usar as lições aprendidas e adaptá-lo ao projeto LSST que representa um grande desafio devido ao volume de dados esperado. É estimado que o tamanho dos catálogos a serem processados serão mais de 40 vezes aqueles do DES como consequência da maior área de cobertura (duas vezes maior), maior profundidade e a inclusão da banda *u*. Três meses de

observação com o LSST corresponderá a 5 anos do DES, e irá funcionar o ano inteiro por 10 anos!

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 a partir de medidas de fluxo em diferentes bandas; 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 (figura 6); e 6) consomem os catálogos produzidos para análise.

Ao longo do último ano vários *workflows* científicos foram desenvolvidos e integrados à 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 conjunto de pipelines 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.

The screenshot shows the 'Cluster' page in the Science Portal. At the top, there are navigation tabs: Dashboard, My Workspace, Pipelines, Tools, Science Products, Data Server, Documentation, and Help. The user's name 'Luiz da Costa' is in the top right. Below the navigation, there are tabs for 'Input Data', 'Configuration', and 'Summary'. The 'Input Data' tab is active, showing 'Release: Y1A1' and 'Dataset: STRIPE82' with an 'Ok' button. Below this, there are sections for 'Pholo-z' and 'Objects Catalog'. The main section is 'S/G Classification', which has a table with 23 entries. The table has columns for Product Name, Product Class, Dataset, Process ID, Configuration, Date, Owner, and Provenance. The table is currently showing 10 entries.

Product Name	Product Class	Dataset	Process ID	Configuration	Date	Owner	Provenance
Y1 Modest v4 2	Y1 Modest v4	STRIPE82	5212		2019-06-04 11:40:30	Michel Agüena	
Y3 Extended Class Spread 2	Y3 Extended Class Spread	STRIPE82	5212		2019-06-04 11:40:30	Michel Agüena	
Y3 Extended Class Wav Spread 2	Y3 Extended Class Wav Spread	STRIPE82	5212		2019-06-04 11:40:30	Michel Agüena	
Y1 Modest v4 1	Y1 Modest v4	STRIPE82	5169		2019-05-28 18:23:19	Michel Agüena	
Y3 Extended Class Spread 1	Y3 Extended Class Spread	STRIPE82	5169		2019-05-28 18:23:19	Michel Agüena	
Y3 Extended Class Wav Spread 1	Y3 Extended Class Wav Spread	STRIPE82	5169		2019-05-28 18:23:19	Michel Agüena	
Y1 Modest v3 9	Y1 Modest v3	STRIPE82	4934		2019-03-15 20:06:31	Ricardo Ogando	
Y1 Modest v3 8	Y1 Modest v3	STRIPE82	4930		2019-03-14 19:23:11	Ricardo Ogando	
Y1 Modest v3 7	Y1 Modest v3	STRIPE82	3881		2018-07-17 15:16:42	Hillysson Sampaio	
Y1 Modest v2 1	Y1 Modest v2	STRIPE82	2812		2017-11-14 16:04:05	Christophe Benoist	

At the bottom of the table, there is a 'Next' button. The footer of the page shows 'Science Portal v0.9-28 (Jun 22 2019)' and 'Powered by LIneA'.

Figura 6 - Exemplo de uma interface do portal científico usada para a criação de um catálogo de galáxias usado para alimentar o algoritmo desenvolvido pela equipe científica do INCT para a identificação de aglomerados de galáxias. A partir desta interface os dados de entrada necessários para executar o processo podem ser escolhidos, parâmetros de configuração selecionados e o processo submetido para execução.

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 ajudem 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.

Os pipelines são agrupados em estágios de processamento que inclui aqueles usados para: 1) instalar dados de diferentes anos; 2) construir para cada ano mapas que mostram a variação das condições observacionais em diferentes partes do céu; 3) calcular parâmetros e atributos adicionais para cada objeto detectado; 4) criar catálogos para alimentar diferentes algoritmos científicos e amostras de dados especiais; 5) executar diferentes tipos de análise; 6) estimar parâmetros cosmológicos envolvendo processos MPI. A arquitetura do portal é bastante sofisticada e permite ao usuário acessar todos os processos executado, os dados de entrada e configurações usadas, e os resultados e produtos criados usando o dashboard apresentado na figura 7.

The screenshot displays the 'DES Science Portal Dashboard' with a navigation bar at the top. Below the navigation bar, there are two dropdown menus for 'Release' and 'Dataset', both set to 'All'. The dashboard is divided into several sections, each containing a table of pipeline runs. The 'Data Installation' section lists pipelines like 'Coarse Depth Map' and 'Data Installation'. The 'Data Preparation' section lists pipelines like 'Data Preparation' and 'Galaxy Properties'. The 'Special Samples' section lists pipelines like 'Catalog Association' and 'Cluster and Field G...'. The 'Simulations' section shows 'No records found'. The 'Science Analysis' section lists pipelines like 'ACF Covariance Ma...' and 'ACF Fullshape'. At the bottom, there is a footer with 'Testing Portal: v1.3.1-7-gaeeec1d0' and 'Powered by LineA'.

Pipeline	Start	Duration	Runs	Status
Coarse Depth Map	2017-06-26	01:01:54	1	Success
Data Installation	2015-08-28	00:20:03	27	Success
Install Bright Mask	2018-07-20	00:01:14	13	Success
Install Catalogs	2016-08-30	00:00:23	16	Success
Install Depth Maps	2017-11-27	00:33:12	11	Success
Install Mangle Mask	2018-01-12	00:13:36	13	Success
Systematic Maps	2019-05-28	00:43:50	23	Success
Zero-point Correction	2016-04-11	00:15:52	10	Success

Pipeline	Start	Duration	Runs	Status
Catalog Association	2017-07-12	00:57:56	1	Success
Cluster and Field G...	2018-09-14	00:01:18	8	Success
Early and Late-type...	2018-03-15	00:00:12	3	Success

Pipeline	Start	Duration	Runs	Status
ACF Covariance Ma...	2017-11-23	00:11:49	1	Success
ACF Fullshape	2018-08-15	00:01:33	9	Success
ACF L55	2017-11-06	00:22:38	8	Success
AddStar	2018-10-30	00:29:43	2	Success
Cluster Cosmology	2018-02-08	00:38:03	2	Success
LF	2019-07-23	01:14:06	34	Success
LF Simulations	2019-07-23	00:03:41	7	Success
ModStar	2019-04-22	03:45:57	8	Success
MW Fitting	2018-04-25	00:11:55	27	Success
StarHorse	2017-06-04	09:48:06	2	Success
WAZP	2019-06-10	00:28:13	18	Success

Figura 7 - Dashboard do portal científico desenvolvido para o projeto DES onde os pipelines são agrupados pelos diferentes estágios de processamento (e.g. instalação, preparação, criação de catálogos *science-ready* genéricos ou especiais, workflow científicos e *utilities*). Nesta interface são apresentados os resultados da última rodada dos pipeline, a data em que foi executado, a duração, o número de execuções prévias deste pipeline e o status da última execução. Pressionando o número de execuções prévias dá acesso, através de um *pop-up*, a outra interface com as mesmas informações (e.g., data, duração) sobre todos os processos bem como informação sobre a proveniência dos dados de entrada, configuração, resultados e produtos. Isto é possível graças ao banco administrativo associado ao portal. O portal é sem dúvida um dos sistemas mais complexos desenvolvidos pelo LineA com uma importante contribuição dos membros do INCT.

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 seu conjunto representa uma rica base de conhecimento acumulado que deve ser preservado e documentado em repositórios de códigos e de documentos. Uma fração significativa de uma FTE do LIneA está encarregada deste esforço e este legado é considerado uma das principais missões do INCT.

O portal científico envolveu até agora um esforço de mais de 60 FTEs-ano, 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 estelar. A plataforma recolhe a posição de todas as exposições da câmera DECam feitas desde do início de sua operação em 2013, identifica CCDs que contêm objetos do sistema solar através do serviço *SkyBoT*, desenvolvido e mantido pelo Observatório de Paris com espelho no LIneA, seleciona a classe dinâmica dos objetos 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 prever 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 durante o segundo semestre de 2019 e será utilizada para analisar dados da DECam acumulados pelos projetos DES e DECALS. 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.

Este modelo de operação deve sofrer profundas mudanças com a entrada em operação do LSST tem em vista o grande número previsto de ocultações. No momento, planos para a criação de uma rede de telescópios robóticos alinhados de norte a sul no Brasil estão sendo discutidos. Em paralelo, em preparação para o LSST os alertas sendo produzidos pelo projeto [Zwicky Transient Facility \(ZTF\)](#) 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 com mais de 1.500 TNOs sendo observados por noite.

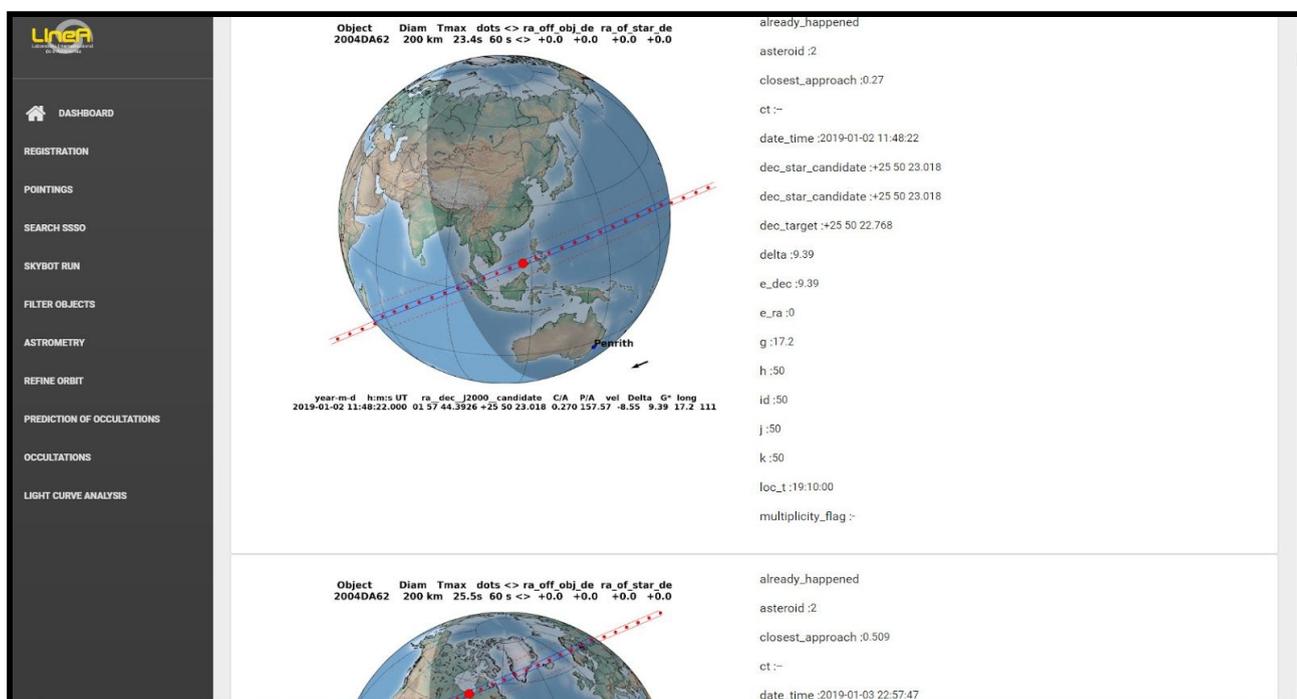


Figura 8 - Uma das interfaces do portal sendo desenvolvido para o projeto TON usando os dados acumulados do DES. Um dos principais objetivos do portal é a predição da ocultação de estrelas por objetos do sistema solar a partir da astrometria realizada com as observações realizadas pelo levantamento DES. Isto envolve os vários processos descritos no texto e listados no lado esquerdo da tela. Essa lista permite o acesso a diferentes interfaces a partir das quais os diferentes processos podem ser executados e monitorados. A figura mostra o resultado final do processamento apresentando, para um objeto, um dos eventos de ocultação previsto (i.e., quando e onde, sobre a Terra, tal evento poderá ser observado).

Portal LSST: com recursos do INCT uma engenharia de software foi contratada para acompanhar a operação do portal científico do DES, entrevistar usuários e desenvolvedores do portal. Integrar o Comitê de Controle de Mudanças (*Change Control Board*) para compilar os problemas reportados ao longo dos dois últimos anos e a partir destas informações e análises dos problemas existentes propor soluções e desenvolver um documento de requisitos para o desenvolvimento do portal para o LSST.

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 ferramentas descritas abaixo foram entregues para operação ou estão em desenvolvimento.

LineA Science Server: o [Science Server](#) (figura 9) é uma versão moderna da ferramenta [LineA Data Server](#) mencionado acima. Esta versão incorpora novas tecnologias, resolve problemas identificados na versão anterior, e atender novas demandas feitas por usuários. A ferramenta está disponível no *National Center for Supercomputing Applications* (NCSA) desde de janeiro de 2018 como 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. 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.



Figura 9 - Interface de acesso aos diferentes serviços disponíveis no *LineA Science Server* e disponível no *National Center for Supercomputer Applications* (NCSA). O serviço permite a visualização de imagens e catálogos gerados pela colaboração ou a partir de buscas no banco de dados Oracle.

Tile Inspection: recentemente foi solicitado pela colaboração DES um serviço que permitisse a rápida avaliação da qualidade das imagens co-adicionadas do sexto ano de observações do DES. Isto foi feito em tempo recorde com a nova interface (figura 10) sendo integrada no *Science Server* e já sendo usada pela equipe responsável pela liberação dos dados. A interface permite a visualização de cada ladrilho, o uso de zoom, a mudança de contraste, a indicação se a imagem é boa ou ruim, a inclusão de comentários, a busca de textos no campo de comentários. O sistema foi também apresentado ao projeto LSST e está sendo considerado para permitir a visualização de imagens tanto na montanha como no NCSA.

Link fig_10: http://www.linea.gov.br/wp-content/uploads/2019/08/fig_10.jpg

Figura 10 - A figura mostra uma captura de tela da interface *tile inspection* que permite a equipe responsável pela redução de imagens e pesquisadores do DES inspecionar as imagens co-adicionadas de 0.25 graus quadrados que compõem o mosaico cobrindo 5.000 graus quadrados. O levantamento completo consiste de 10.000 dessas imagens. Nesta figura foi feito o zoom na posição de um par de galáxias na parte central da figura.

Quick Look Framework (QLF): ferramenta desenvolvida pelo time de TI para o projeto DESI como contrapartida para a participação de seis pesquisadores brasileiros, correspondendo a US\$ 400 mil. 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 (figura 11) 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 consiste de diferentes serviços que podem ser acessados pela interface principal 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*). O sistema passou por três revisões feitas pela colaboração, uma em Columbus e outra no Observatório de Kitt Peak e foi entregue ao projeto no início de 2019. As despesas para a participação da equipe brasileira nesta reuniões foram custeadas pelo INCT.

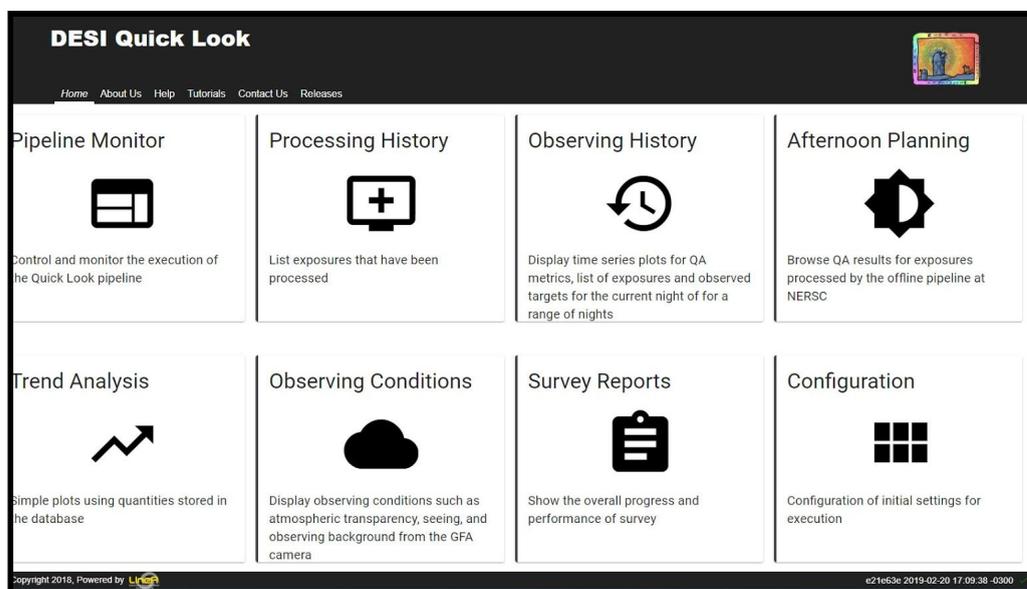
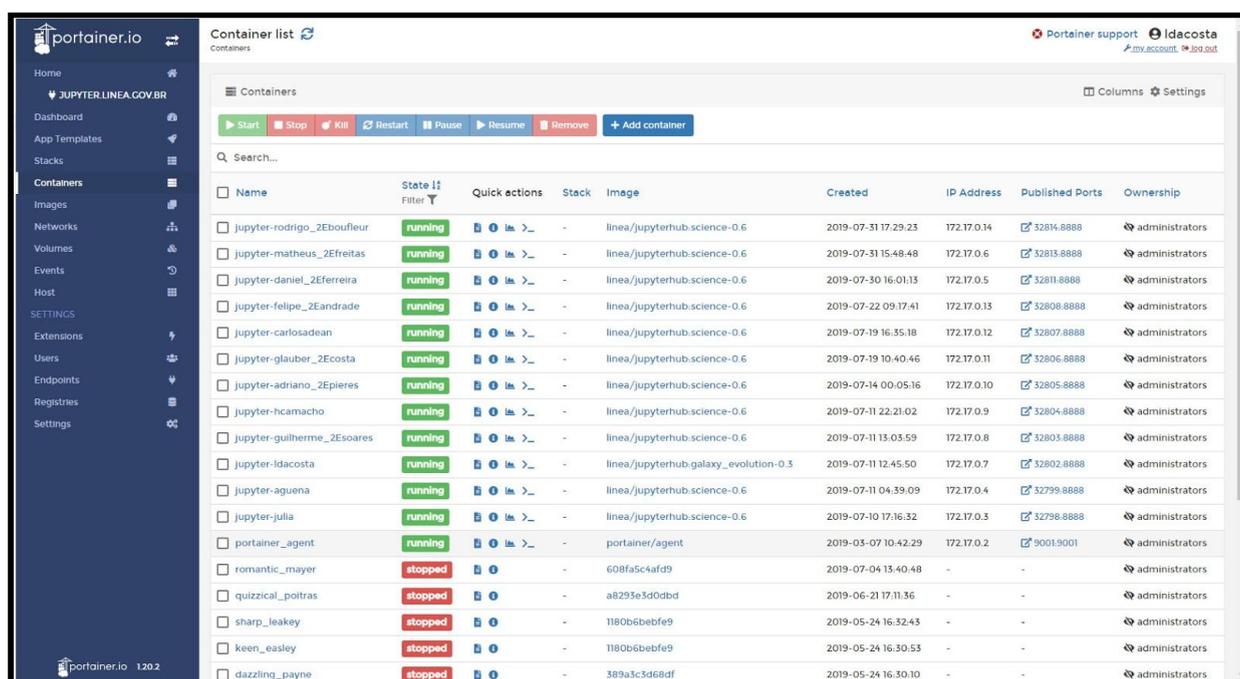


Figura 11 - Interface de acesso aos serviços do sistema de monitoramento dos dados espectroscópicos desenvolvido para o projeto DESI em troca da entrada de seis pesquisadores brasileiros no projeto.

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 serão disponibilizadas para acessar os dados do LSST.

O acervo de notebooks inclui o sistema de redução *stack* desenvolvido pelo LSST, entre outros exemplos criados para atender diferentes análises (figura 12). 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 (figura 13); 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, catálogos preparados para a análise científica, chamados de *value-added catalogs (vacs)*.

Um projeto sendo levado adiante e a integração dos serviços do portal científico. *LIneA Science Server* e os notebooks, de maneira a permitir o compartilhamento dos produtos gerados pelos diferentes serviços.



The screenshot shows the Portainer.io interface for managing containers. The main area displays a list of containers, each representing a Jupyter notebook. The interface includes a sidebar with navigation options like Home, Dashboard, App Templates, Stacks, Containers, Images, Networks, Volumes, Events, Host, SETTINGS, Extensions, Users, Endpoints, Registries, and Settings. The container list table has columns for Name, State, Quick actions, Stack, Image, Created, IP Address, Published Ports, and Ownership. Most containers are in a 'running' state, while some are 'stopped'.

Name	State	Quick actions	Stack	Image	Created	IP Address	Published Ports	Ownership
jupyter-rodrigo_2Eboufleurl	running	[Start] [Stop] [Kill] [Restart] [Pause] [Resume] [Remove]	-	linea/jupyterhub.science-0.6	2019-07-31 17:29:23	172.17.0.14	32814:8888	administrators
jupyter-matheus_2Efreitas	running	[Start] [Stop] [Kill] [Restart] [Pause] [Resume] [Remove]	-	linea/jupyterhub.science-0.6	2019-07-31 15:48:48	172.17.0.6	32813:8888	administrators
jupyter-daniel_2Eferreira	running	[Start] [Stop] [Kill] [Restart] [Pause] [Resume] [Remove]	-	linea/jupyterhub.science-0.6	2019-07-30 16:01:13	172.17.0.5	32811:8888	administrators
jupyter-felipe_2Eandrade	running	[Start] [Stop] [Kill] [Restart] [Pause] [Resume] [Remove]	-	linea/jupyterhub.science-0.6	2019-07-22 09:17:41	172.17.0.13	32808:8888	administrators
jupyter-carlosadean	running	[Start] [Stop] [Kill] [Restart] [Pause] [Resume] [Remove]	-	linea/jupyterhub.science-0.6	2019-07-19 16:55:18	172.17.0.12	32807:8888	administrators
jupyter-glauber_2Ecosta	running	[Start] [Stop] [Kill] [Restart] [Pause] [Resume] [Remove]	-	linea/jupyterhub.science-0.6	2019-07-19 10:40:46	172.17.0.11	32806:8888	administrators
jupyter-adriano_2Epieres	running	[Start] [Stop] [Kill] [Restart] [Pause] [Resume] [Remove]	-	linea/jupyterhub.science-0.6	2019-07-14 00:05:16	172.17.0.10	32805:8888	administrators
jupyter-hcamacho	running	[Start] [Stop] [Kill] [Restart] [Pause] [Resume] [Remove]	-	linea/jupyterhub.science-0.6	2019-07-11 22:21:02	172.17.0.9	32804:8888	administrators
jupyter-guilherme_2Esoares	running	[Start] [Stop] [Kill] [Restart] [Pause] [Resume] [Remove]	-	linea/jupyterhub.science-0.6	2019-07-11 13:05:59	172.17.0.8	32803:8888	administrators
jupyter-ldacosta	running	[Start] [Stop] [Kill] [Restart] [Pause] [Resume] [Remove]	-	linea/jupyterhub.galaxy_evolution-0.3	2019-07-11 12:45:50	172.17.0.7	32802:8888	administrators
jupyter-aguena	running	[Start] [Stop] [Kill] [Restart] [Pause] [Resume] [Remove]	-	linea/jupyterhub.science-0.6	2019-07-11 04:39:09	172.17.0.4	32799:8888	administrators
jupyter-julia	running	[Start] [Stop] [Kill] [Restart] [Pause] [Resume] [Remove]	-	linea/jupyterhub.science-0.6	2019-07-10 17:16:32	172.17.0.3	32798:8888	administrators
portainer_agent	running	[Start] [Stop] [Kill] [Restart] [Pause] [Resume] [Remove]	-	portainer/agent	2019-03-07 10:42:29	172.17.0.2	9001:9001	administrators
romantic_mayer	stopped	[Start] [Stop] [Kill] [Restart] [Pause] [Resume] [Remove]	-	608fa5c4afd9	2019-07-04 13:40:48	-	-	administrators
quizzical_poitras	stopped	[Start] [Stop] [Kill] [Restart] [Pause] [Resume] [Remove]	-	a8293e3d0dbd	2019-06-21 17:11:36	-	-	administrators
sharp_leakey	stopped	[Start] [Stop] [Kill] [Restart] [Pause] [Resume] [Remove]	-	1180b6b6bfe9	2019-05-24 16:52:43	-	-	administrators
keen_easley	stopped	[Start] [Stop] [Kill] [Restart] [Pause] [Resume] [Remove]	-	1180b6b6bfe9	2019-05-24 16:30:53	-	-	administrators
dazzling_payne	stopped	[Start] [Stop] [Kill] [Restart] [Pause] [Resume] [Remove]	-	399a3c5d68df	2019-05-24 16:50:10	-	-	administrators

Figura 12 - Interface listando vários notebooks sendo processados, mostrando o impacto desta plataforma junto aos usuários. Esta plataforma permite o uso da infraestrutura do LIneA de uma forma fácil e organizada e está prevista de ser usada num dos projetos de divulgação junto a escolas de ensino médio atualmente sendo explorados com o INCT de Comunicação Pública da Ciência e Tecnologia.

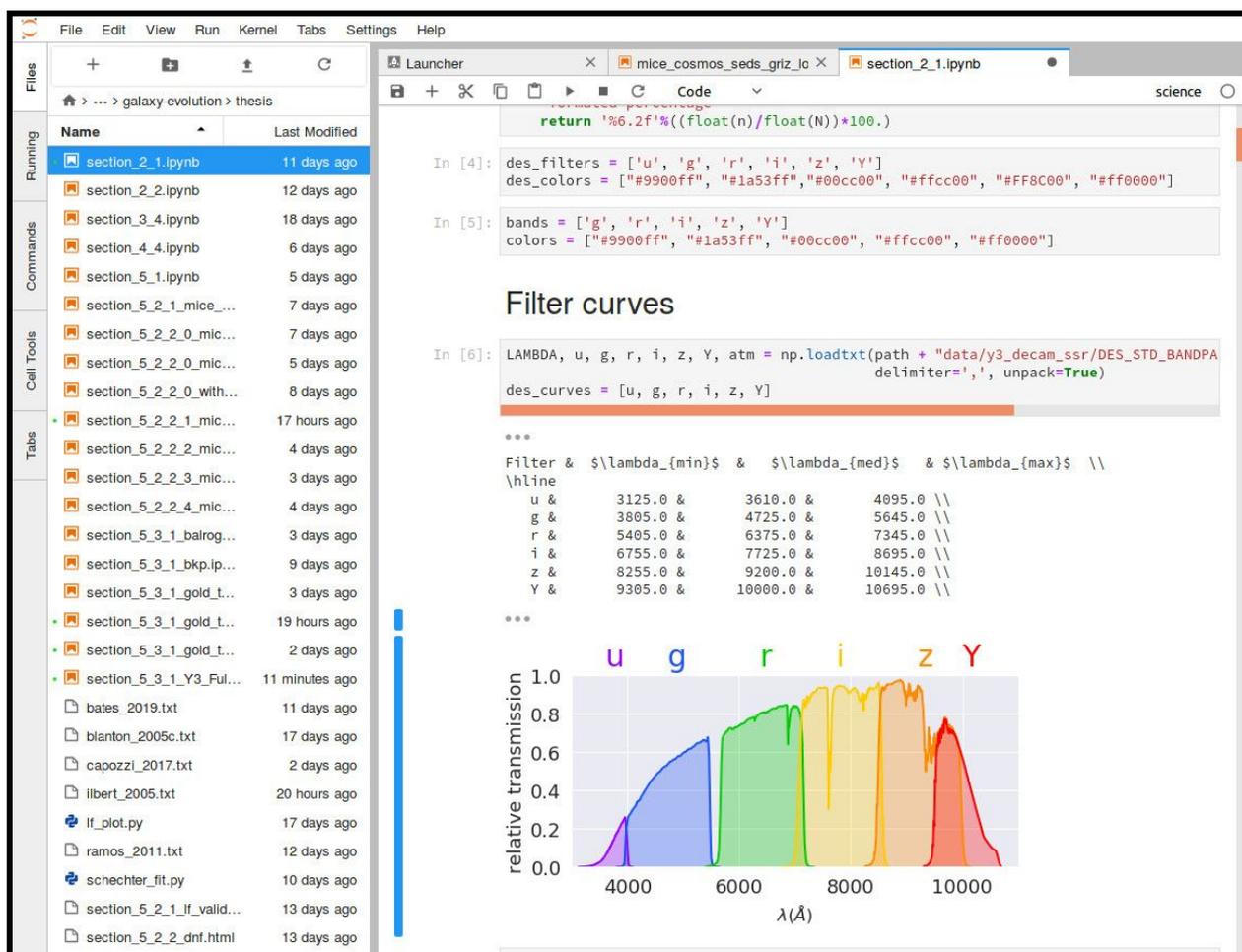


Figura 13 - Exemplo de *jupyter notebook* sendo usado para testar pipelines sendo desenvolvidos pelo grupo de evolução de galáxias.

LNA Archive: interface de busca de imagens de observações realizadas no Observatório do Pico dos Dias e no *Southern Observatory for Astrophysical Research (SOAR)*, disponíveis no repositório mantido pelo LNA. Esta interface está sendo desenvolvida para o projeto TON que tem um grande número de imagens obtidas durante uma década. 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. Uma versão beta do sistema já está disponível (figura 14).

Download	Filename	Target	RA	Dec	Date Obs	Filter	Exptime	Telescope	Instrument	Observer	Size
	Jupiter_I_0044.fits	Jupiter	152121	-17:09:57	03/28/2018 9:15:54 AM	Metano	8	0.60m(BC)	Cam2+ikon	Bruno Morgado	36 KB
	Jupiter_I_0052.fits	Jupiter	152121	-17:09:57	03/28/2018 9:17:30 AM	Metano	8	0.60m(BC)	Cam2+ikon	Bruno Morgado	36 KB
	Jupiter_I_0024.fits	Jupiter	152121	-17:09:57	03/28/2018 9:11:08 AM	Metano	8	0.60m(BC)	Cam2+ikon	Bruno Morgado	36 KB
	Jupiter_IL_0019.fits	Jupiter	152121	-17:09:57	03/28/2018 5:45:11 AM	Metano	6	0.60m(BC)	Cam2+ikon	Bruno Morgado	36 KB

Figura 14 - Interface de busca de observações realizadas pelo projeto TON. Atualmente é possível realizar buscas pelo nome do observador, telescópio, câmera e filtro.

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.

Portal Interfaces: para facilitar usuários do portal inúmeras interfaces estão sendo desenvolvidos para monitorar os processos (portal monitor, figura 15) o histórico de processamento (portal dashboard figura 16), disponibilizar produtos científicos (science products, figura 17), compilação de pipelines, componentes e bibliotecas disponíveis (portal manager, figura 18).

Portal Monitor

Filter: All

Search...

Process ID	Start Date	Start Time	Duration	Pipeline	Instance	Release	Dataset	Owner	Status	Saved	Published
5544	2019-08-04	17:33:47	00:01:15	Spectroscopic Sample	production	-	-	Julia Gschwend	Failure	-	-
5543	2019-08-04	01:45:54	13:10:03	LF	production	Y3A2	Y3 Gold Full	Julia Gschwend	Success	✓	✓
5542	2019-08-02	12:39:06	00:09:46	LF Simulations	production	Micecat v2.0	Micecat v2.0 Small	Julia Gschwend	Success	✓	✓
5541	2019-08-02	11:37:12	01:57:50	Objects Count	production	Y3A2	Y3 Gold Full	Carlos Adean	Success	-	-
5539	2019-08-01	21:49:17	01:58:19	LF Simulations	production	Micecat v2.0	Micecat v2.0 Small	Julia Gschwend	Success	-	-
5538	2019-08-01	16:41:10	01:23:54	Objects Count	production	Y3A2	Y3 Gold Full	Carlos Adean	Failure	-	-
5537	2019-07-31	11:45:40	01:25:04	Objects Count	production	Y3A2	Y3 Gold Full	Carlos Adean	Failure	-	-
5536	2019-07-31	10:05:22	01:21:51	Objects Count	production	Y3A2	Y3 Gold Full	Carlos Adean	Failure	-	-
5535	2019-07-31	09:58:42	00:04:17	Objects Count	production	Y3A2	Y3 Gold Medium	Carlos Adean	Success	-	-
5534	2019-07-31	09:54:04	00:03:28	Objects Count	production	Y3A2	Y3 Gold Medium	Carlos Adean	Success	-	-

Rows per page: 10 | 1-10 of 3884 | 1 2 3 ... 389

Figura 15 - Interface para monitorar processos submetidos pelo portal listando a data e a hora de início, duração, nome do pipeline, o conjunto de dados usados o status e se o processo foi salvo e publicado.

DES Science Portal Dashboard

Release: Y1A1 | Dataset: STRIPE82

Pipeline	Start	Duration	Runs	Status
Data Installation				
Install Bright Mask	2018-07-04	00:01:16	5	Success
Install Catalogs	2016-01-29	03:51:24	1	Success
Install Depth Maps	2017-10-31	00:08:05	7	Success
Install Mangle Mask	2017-09-25	01:14:44	3	Success
QA Coadd	2016-08-01	02:09:47	1	Success
Systematic Maps	2017-05-09	00:50:20	2	Success
Zeropoint Correction	2017-10-10	00:43:30	2	Success
Data Preparation				
Galaxy Properties	2016-11-17	03:20:53	1	Success
Photo-z Compute	2019-04-26	00:15:09	97	Success
Photo-z Training	2019-04-26	00:04:14	82	Success
Photo-z Validation	2018-08-23	00:03:14	4	Failure
SG Separation	2019-06-04	00:29:35	19	Success
Training Set Maker	2019-04-11	00:05:19	52	Success
Science-Ready Catalogs				
Cluster	2019-06-11	03:12:41	121	Success
GA	2017-12-04	00:10:34	4	Success
Utilities				
Catalog Comparison	2016-12-13	00:33:01	9	Success
Concatenate Fields	2017-02-01	00:51:11	5	Success
Depth Map Afterbu...	2017-10-31	00:56:38	5	Success
Export Table	2016-08-01	00:23:40	2	Success
Upload	2016-09-21	00:03:14	6	Success
WAZP afterburner	2018-09-11	00:06:49	7	Success
Special Samples				
Cluster and Field G...	2018-09-14	00:02:46	3	Success
Simulations				
No records found				
Science Analysis				
ACF LSS	2017-11-06	00:20:56	1	Success
MW Fitting	2018-04-20	00:25:55	13	Success
StarHorse	2017-05-12	02:04:36	1	Success
WAZP	2019-06-11	00:37:28	79	Success

Production Portal: v1.3.1 | Powered by LINER

Figura 16 - Dashboard do portal DES onde o usuário tem acesso a todo o histórico de processamento realizado para um conjunto de dados. O dashboard não só mostra o resultado das últimas rodadas de um dado pipeline como também dá acesso a todas rodadas anteriores.

Science Products										
Release	Dataset	Type	Class	Product Name	Product Type	Process ID	Owner	Band	Date	
1	Y3A2	Y3 Gold Full	Measurements	LF Median	LF Median 30	Catalog	5543	Julia Gsch...	---	2019-08-04
2	Y3A2	Y3 Gold Full	Measurements	LF Measurement	LF Measurement 31	Catalog	5543	Julia Gsch...	---	2019-08-04
3	Y3A2	Y3 Gold Full	Region Selection	Footprint map	Footprint map 34	Map	5522	Julia Gsch...	---	2019-07-27
4	Y3A2	Y3 Gold Full	Value-Added Catalogs	GE	GE 22	Catalog	5522	Julia Gsch...	---	2019-07-27
5	Y3A2	Y3 Gold Full	Galaxy Properties	Lephare	Lephare 5	Catalog	5521	Julia Gsch...	---	2019-07-27
6	Y3A2	Y3 Gold Full	Measurements	LF Median	LF Median 29	Catalog	5486	Julia Gsch...	---	2019-07-20
7	Y3A2	Y3 Gold Full	Measurements	LF Measurement	LF Measurement 30	Catalog	5486	Julia Gsch...	---	2019-07-20
8	Y3A2	Y3 Gold Full	Measurements	LF Measurement	LF Measurement 29	Catalog	5488	Julia Gsch...	---	2019-07-20
9	Y3A2	Y3 Gold Full	Measurements	LF Median	LF Median 28	Catalog	5488	Julia Gsch...	---	2019-07-20
10	Y3A2	Y3 Gold Full	Measurements	LF Median	LF Median 27	Catalog	5482	Julia Gsch...	---	2019-07-18
11	Y3A2	Y3 Gold Full	Measurements	LF Measurement	LF Measurement 28	Catalog	5482	Julia Gsch...	---	2019-07-18
12	Y3A2	Y3 Gold Small	Measurements	LF Median	LF Median 3	Catalog	5478	Julia Gsch...	---	2019-07-18
13	Y3A2	Y3 Gold Small	Measurements	LF Measurement	LF Measurement 4	Catalog	5478	Julia Gsch...	---	2019-07-18
14	Y3A2	Y3 Gold Small	Value-Added Catalogs	GE	GE 6	Catalog	5477	Julia Gsch...	---	2019-07-18
15	Y3A2	Y3 Gold Small	Region Selection	Footprint map	Footprint map 20	Map	5477	Julia Gsch...	---	2019-07-18
16	Y3A2	Y3 Gold Full	Measurements	LF Median	LF Median 26	Catalog	5473	Julia Gsch...	---	2019-07-18
17	Y3A2	Y3 Gold Full	Measurements	LF Measurement	LF Measurement 27	Catalog	5473	Julia Gsch...	---	2019-07-18
18	Y3A2	Y3 Gold Full	Value-Added Catalogs	GE	GE 21	Catalog	5464	Julia Gsch...	---	2019-07-15
19	Y3A2	Y3 Gold Full	Region Selection	Footprint map	Footprint map 33	Map	5464	Julia Gsch...	---	2019-07-15
20	Y3A2	Y3 Gold Full	Measurements	LF Median	LF Median 25	Catalog	5470	Julia Gsch...	---	2019-07-15
21	Y3A2	Y3 Gold Full	Measurements	LF Measurement	LF Measurement 26	Catalog	5470	Julia Gsch...	---	2019-07-15
22	Y3A2	Y3 Gold Full	Measurements	LF Median	LF Median 24	Catalog	5469	Julia Gsch...	---	2019-07-15
23	Y3A2	Y3 Gold Full	Measurements	LF Measurement	LF Measurement 25	Catalog	5469	Julia Gsch...	---	2019-07-15
24	Y3A2	Y3 Gold Full	Measurements	LF Median	LF Median 23	Catalog	5465	Julia Gsch...	---	2019-07-15
25	Y3A2	Y3 Gold Full	Measurements	LF Measurement	LF Measurement 24	Catalog	5465	Julia Gsch...	---	2019-07-15

Page 1 of 116 | Displaying 1

Figura 17 - Interface que dá acesso a todos os produtos criados pelo portal que foram salvos e publicados. É desta interface que os produtos podem ser transferidos pelos pesquisadores interessados.

Portal User Interfaces							
PIPELINES	COMPONENTS	PACKAGES	CLASSES	DATASETS	EXTERNAL DATASET	SIMULATIONS	
<input type="text" value="pre"/>							
Pipeline ↓	Name	Version Date	Stage	Owner	Descripti...	User Man...	History
Training Set Maker	training_set_maker	2019-06-21	Data Preparation	Julia Gschwend		-	-
Template Set Maker	template_set_maker	2019-06-21	Data Preparation	Julia Gschwend		-	-
Spectroscopic Sample	centralized_spec_db	2019-06-21	Data Preparation	Aurelio Carnero		-	-
SG Separation	sg_separation	2019-06-21	Data Preparation	Ricardo Ogando		-	-
Photo-z Validation	photoz_validation	2019-06-21	Data Preparation	Julia Gschwend		-	-
Photo-z Training	photoz_training	2019-06-21	Data Preparation	Julia Gschwend		-	-
Photo-z Compute By TI...	photoz_compute_by_titles	2019-06-21	Data Preparation	Julia Gschwend		-	-
Photo-z Compute	photoz_compute	2019-06-21	Data Preparation	Ricardo Ogando		-	-
Make representative sa...	representative_sample_m...	2019-06-21	Utilities	Michel Agüena		-	-
Galaxy Properties MC	galprop_mc	2019-07-29	Data Preparation	Cristiano Singulani		-	-
Rows per page: 10						1-10 of 11	

Production Portal: v1.5.0 | Powered by Unifra

Figura 18 -Interface para usuários do portal onde são listados todos os pipelines, componentes, pacotes e outras informações de interesse para o desenvolvedor.

Interfaces de Gerenciamento: para ajudar na gestão do LIneA e do INCT interfaces foram construídas para monitorar e analisar os gastos, estatísticas de uso dos serviços e métricas de desempenho que são usadas neste relatório e estão disponíveis no site.

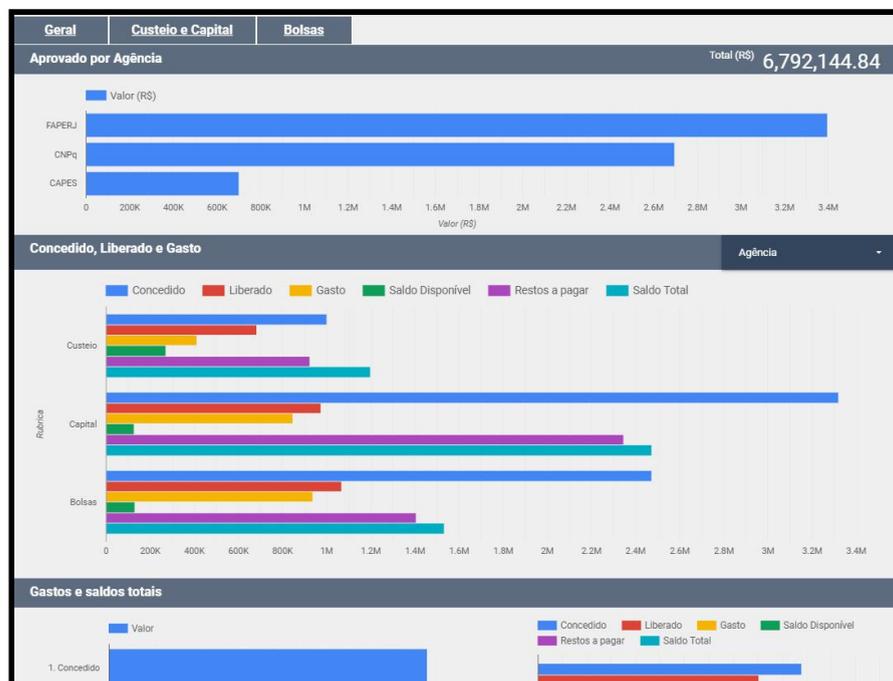


Figura 19 - Uma das interfaces desenvolvidas pelo time de TI do LIneA para facilitar o acompanhamento das finanças do INCT.



Figura 20 - Interface desenvolvida pelo time de TI mostrando algumas métricas de desempenho do INCT que está disponibilizada no site



Figura 21 - O mesmo que figura 20 mas mostrando as métricas do LineA desde 2010.

Landing Page: desenho e implementação de um estrutura de links que permitem acesso aos diferentes serviços do LineA e documentação disponíveis ao público e aos membros das diferentes colaborações científicas que será integrado ao novo site do LineA/INCT para o qual uma firma foi contratada após um processo de seleção realizado pelo comitê gestor do INCT.

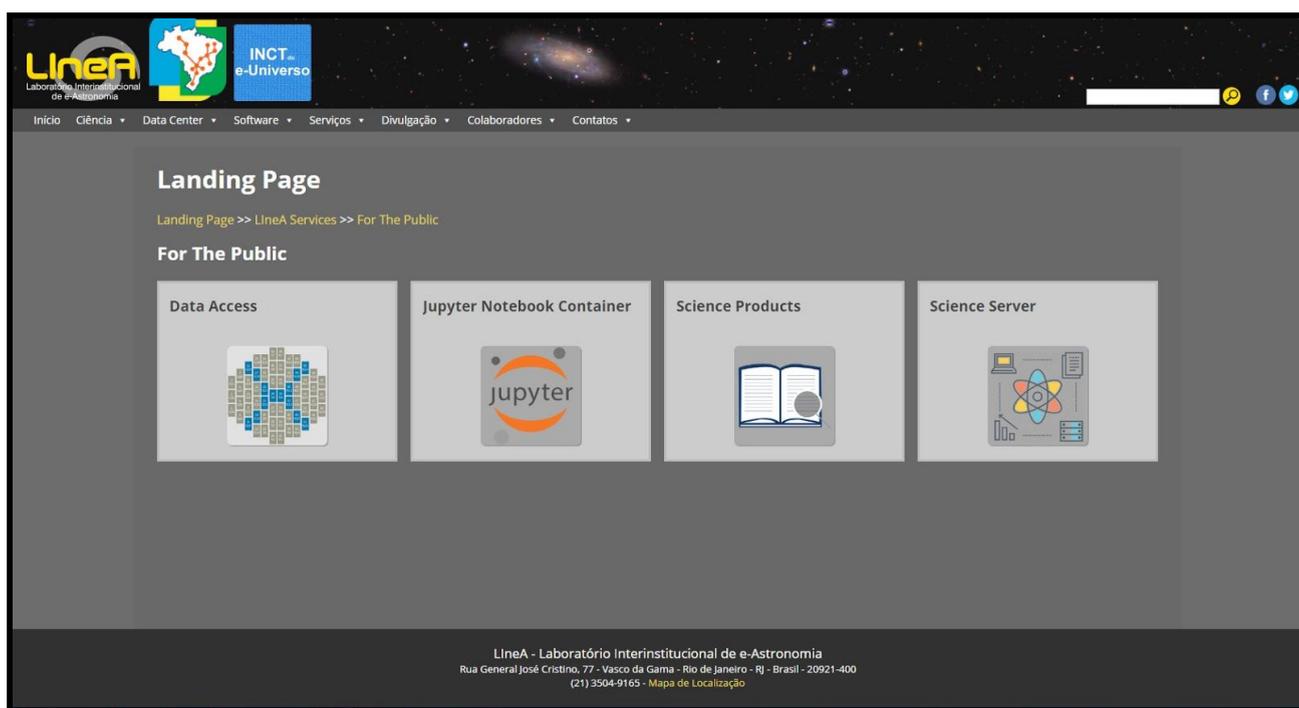


Figura 22 - Página sendo desenvolvida no site para permitir o acesso de pesquisadores não associados a projetos a serviços do LIneA.

Colaborações Científicas

O LIneA e o 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).

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 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. Esta entrega foi feita ao projeto no início deste ano e a equipe de pesquisadores brasileiros já iniciou suas atividades.

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](#); 2) a participação de cinco pesquisadores na reunião [anual do projeto LSST](#) realizada em Tucson em 2018 e de dois pesquisadores e um tecnologista em 2019; 3) a participação de membros do BPG-LSST em oito reuniões do Dark Energy Science Collaboration (SLAC, Carnegie Mellon, Paris); e 3) reuniões técnicas realizadas sobre [rede](#) (Santiago, Chile), [técnicas de visualização](#) (NCSA) e [engenharia e gerenciamento de dados](#) (Caltech).

O [BPG-LSST](#) 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. A tabela 2 resume a produtividade científica do INCT no período 2017-2019

Tabela 2

Artigos publicados e aceitos	172
Teses de doutorado	4
Dissertações de Mestrado	5

Os artigos publicados são os mais variados aproveitando ao máximo os dados dos levantamentos e demonstrando o potencial científico de grandes levantamentos fotométricos e espectroscópicos. Artigos sobre a descoberta de novos objetos do sistema solar, a predição de ocultações, a descoberta de sistemas de estrelas dentro ou na vizinhança da Galáxia, mapeamento da distribuição de estrelas combinando medidas espectroscópicas no infravermelho com medidas de movimento próprio do satélite GAIA, a evolução de galáxias, a identificação de aglomerados de galáxias e o estudo de suas propriedades e comparação com aglomerados identificados por outros métodos, o uso destes sistemas para vincular parâmetros cosmológicos, mapas da distribuição de matéria em grande escala a partir de medidas da distorção de imagens devido a distribuição de matéria escura, a introdução de novas técnicas (3x2pt) combinando medidas de correlação espacial de galáxias e de lenteamento fraco para estimar parâmetros cosmológicos, identificação de supernovas do tipo Ia para estudos cosmológicos e participação em projetos multi-messenger com identificação óptica de eventos associados a detecção de ondas gravitacionais. Alguns destaques que ilustram a diversidade dos artigos produzido podem ser encontrados [aqui](#). Isto mostra o grande efeito multiplicador dos dados provenientes de grandes levantamentos.

Os pesquisadores brasileiros também ocupam posições de destaque nessas colaborações tanto na governança das colaborações (*Management Committee* do DES, *Collaboration Council* do SDSS, *Institutional Board* do DESI, *Membership Committee* do DES, *Builders Committee* DES, *Steering Meeting Committee*) como na áreas científica (*Science Committee*, Coordenador de Grupos/sub-grupos de Trabalho). Outra importante distinção dos pesquisadores brasileiros é que atualmente temos oito pesquisadores considerados *builders* do DES, uma demonstração da contribuição feita (2 FTEs) para a infraestrutura do projeto.

Formação de Pessoal

Uma importante missão do LIneA/INCT é a formação de pessoal técnico e científico e como fica evidente na Figura 1 a equipe do LIneA mistura tecnologistas e pesquisadores em diferentes estágios da carreira. Em particular, 60% dos membros são de jovens pesquisadores indo 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. O mesmo acontece com a equipe técnica que é em sua maioria formada de membros com menos de 40 anos e inclui *trainees*.

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 até agora um total de 26 bolsistas. Entre as bolsas científicas concedidas uma foi de mestrado no Paraná e cinco de pós-doutorado (três em São Paulo, uma no Rio Grande do Sul e uma no Rio de Janeiro). Além disso, o INCT fez um grande investimento para assegurar a participação em reuniões concedendo auxílio integral ou parcial a 28 filiados para participarem de 61 reuniões, das quais 49 no exterior, sendo que pelo menos 20 relacionadas 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](#). O INCT apoiou ainda a ida de tecnólogos em atividades no exterior. Esta mobilidade é fundamental para assegurar a integração do time nas colaborações internacionais e assim tirar proveito técnico/científico das mesmas.

Como parte do processo de formação o INCT se juntou ao LIneA e promoveu de 2017 até agora 90 [webinars](#) (de um total de 260 desde 2011), a grande maioria de palestrantes estrangeiros e 57 [lunch talks](#) com a participação de jovens pesquisadores. Mais recentemente foi organizada uma outra reunião para rever de forma informal o trabalho sendo feito pelos jovens pesquisadores criando um outro fórum para eles apresentarem o seu trabalho e rever e divulgar o material discutido em teleconferências internacionais. O objetivo desses esforços é de promover a interação de jovens pesquisadores lotados em diferentes instituições e trabalhando em diferentes áreas e criar uma atmosfera de colaboração interdisciplinar,.

Continuando o esforço de treinar jovens pesquisadores o INCT está promovendo na semana de 2-6 de Setembro de 2019 a reunião [LIneA Bootcamp](#) com o objetivo de familiarizar novos membros, inclusive do BPG-LSST, com o funcionamento e os serviços disponibilizados pelo laboratório. Isto será feito através de mais de 40 apresentações, tutoriais e de aulas práticas (veja [programação](#)) que serão preparadas por parte de pesquisadores e de membros das equipes técnicas e de gestão do LIneA. Esta reunião é uma ação atendendo a um dos pedidos mais frequentes feitos na consulta realizada junto aos afiliados do LIneA. A reunião terá a participação de 26 pesquisadores e estudantes (21 presenciais e 5 remotos) do Rio, Brasília, Espírito Santo, São Paulo, Santa Catarina, Paraná e Rio Grande do Sul com financiamento parcial ou integral do INCT.

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 e no futuro a outros serviços, como por exemplo jupyter notebooks.

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. Os blogs são preparados pelos próprios pesquisadores com a ajuda de uma aluna de jornalismo com uma bolsa do INCT e validado por pelo menos dois pesquisadores seniors. Infelizmente, para um serviço mais ágil seria importante a participação de uma pessoa com mais experiência principalmente na participação de *press releases* com organizações internacionais. Outras formas de divulgação utilizadas foram:

Para Cientistas

- Ministrando palestras convidadas.
- Participação em Workshops ([LNCC](#), [UNESP](#)).
- Participação em Escolas (e.g. [UNICAMP](#)).
- Organização de encontros/reuniões internacionais (LSST, [DES](#)) (ON, UNICAMP).
- Contribuição para o livro *The Dark Energy Survey: The Story of a Cosmological Experiment* (no prelo).
- [Palestras na Academia Brasileira de Ciências](#) sobre grandes projetos internacionais.
- Cursos em e-Ciência (PPG ON).
- Participação em [Design Sprint](#) para criação de centro de e-ciência.
- [Webinars](#) (> 30/ano, séries LSST, SDSS-V, outros grandes projetos).
- Produção e [acesso ao acervo de dados](#) dos projetos apoiados (SDSS, DES, TON).

Para o Público

- [Acordo com Planetário RJ](#).
- Entrevistas em Rádio (MEC, EBC).
- [Participação da Semana Nacional de Ciência e Tecnologia](#) (ON, UFF, UFRGS, UFRRJ, CEFET).
- Participação em exposições (Museu Histórico Nacional, [Universidade Federal Fluminense](#)).
- Artigos em Revistas ([Ciência Hoje das Crianças](#), FAPESP).
- Cursos para professores do Ensino Médio (IFT-UNESP).
- Curso para o público (Planetário Ibirapuera).
- Encontros informais (e.g. [Dose de Ciência](#), Santa Maria; Chopp com Ciência, RJ).
- [Apresentação do programa de EPO do projeto LSST](#).
- [Blogs/eNews](#) (60/ano, 10/ano) (Site, facebook, twitter).
- [Galeria de imagens astronômicas](#).
- [Organização da LSST Week](#).

- Participação em *press releases* em conjunto com organizações internacionais (NSF, ESA entre outros).

Várias outras atividades estão sendo consideradas entre as quais:

- Preparação de encarte e melhorias no site do INCT (em andamento).
- Preparação de vídeos sobre o LIneA/INCT.
- Criação de conteúdo para Planetário RJ (full dome).
- Montagem de exposição do INCT no Planetário do RJ.
- Acordo de colaboração com o projeto LSST e NOAO na área de divulgação.
- Treinamento de professores de ensino médio no uso das plataformas que dão acesso aos dados e seu uso de forma didática (LSST).

Também em estudos é um projeto para a implantação de uma Rede de Ocultação com pequenos telescópios robóticos instalados em escolas secundárias e universidades espalhadas de norte a sul pelo Brasil. Estes grupos seriam responsáveis não só pela supervisão do equipamento mas poderiam participar da redução e análise de dados utilizando *jupyter notebooks* disponibilizados pelo LIneA. Desta forma os professores locais poderiam ter material para treinamento em programação, redução de imagens e análise de dados com a apoio de pesquisadores e tecnólogos do LIneA/INCT. Isto seria um excelente exemplo de Ciência Cidadã. Com esse objetivo foi submetido agora em Agosto, com o apoio do [LSST Corporation](#), um projeto de US\$ 500 k a [Fundação Templeton](#) para a implementação de uma primeira fase consisting de 10 do s50 telescópios projetados.

Também foram iniciadas conversas com o INCT de Comunicação Pública da Ciência e Tecnologia para a realização de exposição e com a Fundação Serrapilheira sobre um possível apoio aos projetos de divulgação sendo considerados. Nomes de dois membros do INCT foram dados ao CNPq como pontos de contato para estas atividades.

Conclusão

O INCT tem contribuído de forma decisiva para atingir sua missão de apoiar a participação de pesquisadores do INCT em grandes colaborações científicas nas áreas de Astronomia e Cosmologia. Em particular foi fundamental: 1) em manter a operação do centro de dados, permitindo a compra de novos clusters numa hora crítica; 2) na concessão de bolsas técnicas usadas para organizar toda a operação do LIneA seja na área de serviços como na área de desenvolvimento de software permitindo cumprir acordos estabelecidos; 3) bolsas científicas e auxílios que permitiram o engajamento científico do time do INCT. O impacto deste apoio foi extremamente positivo e já foi mencionado no resumo executivo.

Infelizmente, existem alguns pontos negativos que devem ser destacados:

1. A falta de previsibilidade quanto a liberação dos recursos e a rubrica associada, dificultando o planejamento.
2. A situação incerta da FAPERJ que pode comprometer todo o planejamento.
3. A perda de bolsas causada na transferencia de recursos entre o CNPq e a CAPES.
4. As dificuldades encontradas para a liberação de bolsas da CAPES.
5. Incerteza quanto a liberação do saldo do CNPq.
6. Falta de apoio institucional para a administração de recursos que até agora envolveram da ordem de 150 operações financeiras.
7. As dificuldades encontradas com o uso do cartão de pesquisa.

Apesar dessas dificuldades o INCT foi, como dito antes, fundamental para atingir os objetivos propostos no projeto original e que permitiu uma poupança da ordem de US\$ 4 milhões pela estratégia de usar serviços e software como moeda de troca nos grandes projetos internacionais na fronteira científica.

Finalmente, vale ressaltar mais uma vez que a estratégia adotada de se juntar a estes projetos de ponta tem sido enormemente bem sucedida seja: 1) cientificamente; 2) na formação de pessoal; 3) para alavancar outras colaborações; e 4) para o intercâmbio tecnológico. Garantir a liberação dos recursos de forma previsível, recuperar o seu valor original e estender seu prazo seriam medidas importantes para garantir o sucesso do programa como originalmente previsto.

No momento, outra forma de assegurar a continuidade do trabalho do LIneA/INCT, que será apresentado ao MCTIC, prevê a criação de uma organização - LIneA-ICT - com a missão bem definida de preencher a lacuna que hoje existe na infraestrutura de pesquisa básica oferecendo serviços de e-ciência para tratar projetos de *big data* nas áreas de Astronomia e Cosmologia, sendo o LSST o principal caso de uso.