

Laboratório Interinstitucional de e-Astronomia  
(LIneA)

**Relatório de Atividades e Produtos  
2011-2015**

20 de Fevereiro de 2016

## Índice

Resumo Executivo.....	5
1 Histórico .....	8
2 Missão .....	10
3 Organizações Envolvidas .....	12
4 Projetos Apoiados.....	14
4.1 Dark Energy Survey (DES) .....	14
4.2 Sloan Digital Sky Survey III (SDSS-III) .....	15
4.3 O Sloan Digital Sky Survey – IV (SDSS-IV).....	17
4.4 Transneptunian Occultation Network (TON).....	18
4.5 Large Synoptic Survey Telescope (LSST).....	20
4.6 Dark Energy Spectroscopic Instrument (DESI) .....	25
4.7 Comentários gerais.....	28
5 Organização Interna.....	30
6 Afiliação .....	33
7 Ciência .....	37
7.1 Principais destaques.....	38
7.2 O time científico.....	55
7.3 Publicações .....	57
7.4 Webinars.....	58
7.5 Resumo da atividade científica .....	59
8 Centro de Processamento e Distribuição de Dados.....	61
8.1 Infraestrutura física .....	61
8.2 Transferência de dados .....	68
8.3 Armazenamento de dados.....	71

8.4	Processamento e análise de dados .....	74
8.5	Banco de dados .....	75
8.6	Indicadores do centro de dados.....	81
9	Desenvolvimento de Software.....	82
9.1	Portal Científico.....	82
9.2	Quick Reduce .....	84
9.3	Data Server .....	87
9.4	Infraestrutura para criação de catálogos científicos.....	88
9.5	Workflows científicos .....	91
9.6	Instâncias do portal em operação.....	92
9.7	Avaliações internacionais .....	93
9.8	Novos projetos.....	94
9.9	Metodologia de desenvolvimento .....	99
9.10	Documentação .....	102
9.11	A equipe técnica.....	104
10	Suporte aos usuários e outros serviços.....	107
11	Eventos.....	112
11.2	Participação .....	112
11.2	Organização .....	112
12	Divulgação .....	115
13	Acordos e compromissos .....	120
13.1	Acordos Projetos Internacionais .....	120
13.2	Acordos Técnicos.....	121
13.3	Solicitações de Serviços .....	121
14	Manifestação de Apoio .....	123
14.1	Cartas Institucionais de apoio ao LIneA.....	123

14.2 Cartas Individuais de apoio ao LIneA.....	123
14.3 Outras cartas de apoio .....	123
15 Finanças.....	124
15.1 Orçamentos Consolidado .....	124
15.2 Projetos Submetidos .....	127
16 Dificuldades e riscos.....	129
16.1 Científicos .....	129
16.2 Técnicos .....	129
16.3 Gestão e financiamento.....	130
17 Perspectivas.....	131
18 Necessidades .....	133
18.1 Pessoal .....	133
18.2 Equipamento.....	134
18.3 Projetos Científicos.....	135
19 Previsão Orçamentária.....	136
20 Resumo.....	141
Apêndices .....	143

## Resumo Executivo

O desafio de poder atuar de forma decisiva no *Dark Energy Survey* (DES) motivou, há quase 10 anos, a formação de uma rede de pesquisa composta por pesquisadores de diferentes instituições e a formação de um time de tecnologistas capaz de dar suporte técnico para o desenvolvimento de ferramentas e procedimentos que facilitassem o manuseio de volumes de dados muito maiores do que aqueles normalmente obtidos em pequenos projetos individuais. Estas necessidades levaram à proposta da criação de um projeto estruturante no ON conhecido como Astrosoft, encomendas à FINEP, um projeto transversal do CNPq, vários projetos FAPERJ, e finalmente, à criação do Laboratório Interinstitucional de e-Astronomia (LIneA), um centro de e-Ciência voltado para a Astronomia. Este centro é uma ideia inovadora e transformadora, marcando o início de uma nova era para a astronomia brasileira permitindo aos pesquisadores brasileiros se engajarem na ciência de grandes levantamentos astronômicos e na ciência de dados.

A missão do LIneA é criar soluções de software e montar uma infraestrutura física para lidar com o problema de *Big Data* gerado por projetos de *Big Science*, onde por *Big Data* entende-se projetos que envolvam a aquisição de dados que satisfaçam um dos seguintes quesitos: velocidade, volume ou variedade.

A evolução do LIneA nos últimos 5 anos foi surpreendente. Apesar de inúmeras dificuldades, após cinco anos tornou-se um centro de dados de médio porte atendendo às demandas de seus mais de 50 pesquisadores afiliados. Desenvolveu um sofisticado portal científico incorporando o que há de mais moderno em TI, está em constante evolução, e possui instâncias de operação em vários centros internacionais como CTIO, Fermilab e NCSA. Estabeleceu colaborações técnicas com importantes centros como Berkeley, Fermilab, NCSA, LSST e SLAC. Expandiu enormemente suas metas promovendo a entrada de pesquisadores brasileiros em novos projetos internacionais como SDSS-IV, DESI e LSST em colaboração com os principais centros de pesquisa no mundo, apoiando pesquisa de ponta de alta qualidade e impacto. Também promove a formação de um novo tipo de profissional capaz de atuar como protagonista em grandes colaborações internacionais e capacitado para a nova ciência de dados.

Neste documento vários aspectos da operação do LIneA são revistos bem como os principais resultados obtidos ao longo dos últimos cinco anos, entre os quais se destacam:

- 1) A formação de jovens pesquisadores produzindo até agora 12 dissertações de mestrado e seis teses de doutorado. Além disso, cinco pesquisadores jovens foram reconhecidos como “construtores” no projeto DES pela contribuição equivalente a 24 meses de trabalho dada para a infraestrutura do projeto, no caso o desenvolvimento de ferramentas de análise integradas ao portal científico desenvolvido pelo LIneA.

- 2) A produção de 95 publicações em revistas arbitradas, e a organização de duas reuniões internacionais e 127 *webinars*. Com quase 20 artigos por ano e da ordem de 5.000 citações, a produtividade e o impacto científico dos projetos são inquestionáveis. A razão custo/benefício é também positiva quando comparada com outras estratégias adotadas pela a Astronomia brasileira. Esses números deverão crescer ainda mais tendo em vista que as análises realizadas até agora com os dados do DES só incluíram as observações feitas durante o período de verificação científica do DES.
- 3) A criação e a operação de um laboratório multiusuário de médio porte para o processamento e distribuição de dados, dedicados aos grandes levantamentos astronômicos e integrados a vários centros internacionais. O laboratório tem mais de 100 usuários cadastrados, atendendo pesquisadores de universidades no Rio de Janeiro, São Paulo, Paraná e Rio Grande do Sul e instituições no exterior.
- 4) O desenvolvimento de um complexo e abrangente sistema de software para atender de forma eficiente as necessidades de validação, análise e mineração do grande volume de dados envolvido. Estes sistemas estão em operação no observatório do CTIO no Chile, atendendo a todos os usuários da câmera DECam construída pelo projeto DES, no Fermilab com mais de 180 usuários registrados, no LIneA atendendo aos pesquisadores brasileiros e, em breve, no NCSA e em Berkeley.

Como consequência dos termos acordados pelo LIneA com os vários projetos, o laboratório assumiu compromissos que se estendem pelos próximos cinco anos. Além disso, a experiência adquirida mostra o papel vital que o laboratório pode e deve ter na preparação da comunidade brasileira para a era do LSST. Esta experiência lhe confere a credencial de liderar a criação de um Centro de Dados Regional para o LSST e de continuar o trabalho de desenvolver as ferramentas necessárias para lidar com os grandes volumes de dados. Além disso, podemos servir de modelo e trabalhar para **a criação no MCTI de um Centro de e-Ciência com o intuito de apoiar grandes projetos em outras áreas de interesse estratégico**, do qual o LIneA seria apenas um dos participantes.

Para atingir estes objetivos, o LIneA requer uma análise crítica dos termos do acordo e de seu funcionamento, encaminhando ao MCTI recomendações que reconheçam sua singularidade, complexidade, e a necessidade de um orçamento próprio adequado que garanta a continuidade de seu trabalho para atingir seus objetivos de longo prazo. Os recursos estimados para a operação do laboratório são de R\$ 7 milhões por ano. Nela estão incluídos: 1) o custo de pessoal (time de TI, serviço de manutenção, apoio administrativo e de gerência de projetos); 2) o pagamento de taxas de participação no SDSS-IV e DESI; e 3) a compra de mais três PB de armazenamento e de um cluster com 42 nós.

Além de um orçamento adequado, também é necessário ter a facilidade de contratação de especialistas nas mais variadas áreas de TI, o uso de consultoria pontualmente, e a

agilidade na compra/manutenção de equipamentos de informática, condições que infelizmente não existem nas Unidades de Pesquisa do MCTI.

## 1 Histórico

O Laboratório Interinstitucional de e-Astronomia (LIneA) foi criado por uma portaria do Ministério de Ciência, Tecnologia e Inovação (MCTI) publicada no Diário Oficial da União em 8 de Novembro de 2010, seção 213, página 13 (apêndice 1.1) celebrando um acordo entre as Unidades de Pesquisa (apêndice 1.2) que apoiaram a formação do consórcio de pesquisadores denominado DES-Brazil que viabilizou a entrada de pesquisadores brasileiros no projeto *Dark Energy Survey* (DES) em 2007. As Unidades de Pesquisa envolvidas no acordo foram o Centro Brasileiro de Pesquisas Físicas (CBPF), Laboratório Nacional de Computação Científica (LNCC), Observatório Nacional (ON) e, em 2012, a Rede de Pesquisa e Ensino (RNP), já refletindo a natureza interdisciplinar que caracteriza a pesquisa científica moderna (apêndice 1.3). O acordo teve uma duração de cinco anos e sua renovação deve ser avaliada à luz dos resultados alcançados pelo LIneA e de sua relevância em futuros empreendimentos. Em caso de renovação, os termos do acordo e a situação institucional do laboratório também devem ser revistos levando em conta a experiência acumulada ao longo dos cinco anos de sua operação.

Apesar de sua operação oficial ter se iniciado efetivamente a partir de 2011 o embrião do LIneA começou em 2006 com diferentes ações. A primeira foi quando os diretores, na época, do CBPF (R. Galvão), LNCC (M. Raupp) e ON (S. Fontes) apoiaram junto à FINEP e ao Secretário-Executivo do (MCT) o pedido de uma encomenda que permitiu a negociação da entrada de ~10 pesquisadores incluindo físicos e astrônomos de diferentes instituições (CBPF, IF-UFRJ, ON, UFRGS, OCA, INAF) no projeto DES. Reconhecendo: 1) a necessidade desde logo de desenvolver uma infraestrutura de TI que permitisse ao time brasileiro participar proativamente do projeto tendo os meios de lidar com o volume e a complexidade dos dados envolvidos; 2) A experiência acumulada ao longo de 25 anos por membros do time em levantamentos astronômicos no desenvolvimento de sistemas para redução e análise de dados; e 3) a oportunidade identificada durante as discussões, com o então diretor do DES J. Peoples, da redução do custo de entrada por pesquisador, estimado na época em US\$ 100 mil, através da contribuição em forma de trabalho. Para isto, foi proposto o desenvolvimento de um sistema de redução (*Quick Reduce*, QR) a ser utilizado durante as observações para avaliar a qualidade das imagens da câmera DECam formada por um mosaico de CCDs com um total de 512 megapixels. Esta contribuição somada a outras atividades a serem descritas abaixo representou uma poupança de US\$ 700 mil, sendo pago pelo MCTI apenas a quantia de US\$ 300 mil. A segunda importante ação foi a submissão do projeto Astrosoft como um projeto estruturante do ON. A terceira ação foi uma encomenda submetida à FAPERJ com o apoio do MCTI (apêndice 1.4) que possibilitou a compra de alguns computadores, inicialmente instalados no ON, e mais tarde transferidos para as dependências do POP-RJ com o apoio do pessoal do LNCC e do CBPF. A quarta ação foi o acordo feito com o projeto *Sloan Digital Sky Survey* (SDSS-III) em 2008 para a participação de pesquisadores brasileiros neste projeto que envolvia levantamentos com diferentes objetivos científicos como a descoberta de exo-planetas (MARVELS), a estrutura da Via

Láctea (SEGUE, APOGEE), a evolução de galáxias e quasares, e cosmologia (BOSS). Neste acordo, novamente uma redução do valor a ser pago foi obtida em troca da instalação e manutenção de um sistema de distribuição de dados do projeto no Brasil para uso dos pesquisadores brasileiros e como um sítio alternativo àquele montado na Universidade de Johns Hopkins. De maneira similar ao modelo adotado no caso do projeto DES, foi criado um consórcio de pesquisadores denominado Grupo de Participação Brasileiro, conhecido internacionalmente com *Brazilian Participation Group* (BPG) formado por pesquisadores do Observatório do Valongo (OV) da UFRJ, IF-UFRGS, IF-USP, e ON que responderam a uma chamada pública feita na época em nome do ON, que se responsabilizou pelo pagamento de US\$ 950 mil em 6 parcelas, atuando em sua capacidade de laboratório nacional.

A criação do LIneA resultou da necessidade de formalizar a relação entre as UPs envolvidas nestes projetos, e do entendimento de que era necessário criar uma estrutura capaz de atender aos diferentes compromissos entre os quais: i) criar um ambiente colaborativo de forma a otimizar o retorno científico; ii) montar uma infraestrutura central para o armazenamento, processamento e distribuição de dados; iii) promover o intenso esforço exigido para o desenvolvimento de software relacionado a proposta feita para o DES e para atender as necessidades dos pesquisadores do projeto tendo em vista o grande volume de dados previsto. A seguir, vários aspectos da operação do LIneA são revistos bem como os principais resultados obtidos ao longo dos últimos cinco anos.

## 2 Missão

A missão original do LIneA como enunciada na portaria de sua criação:

- Gerenciar a participação brasileira nos experimentos científicos DES e SDSS-III.
- Prover a infraestrutura computacional necessária para a implantação de um site terciário para distribuição de dados de ambos os projetos.
- Promover reuniões de trabalho, workshops e encontros nacionais e internacionais.
- Desenvolver um portal científico para uso da comunidade astronômica.
- Interligar os clusters disponíveis no ON, LNCC e CBPF relacionados aos projetos mencionados na cláusula 1, numa grade para uso pelos experimentos do LIneA criando a base necessária para apoiar outros futuros projetos de Astronomia.

Como serão discutidas ao longo do presente relatório, estas metas não só foram cumpridas, como também foi necessário adaptá-las à realidade operacional do laboratório, que incluem:

- 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 bem como de outras fontes que possam complementar estes dados.
- 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 proativamente em grandes colaborações internacionais.
- Ajudar na formação de tecnologistas e de pesquisadores com um perfil mais moderno (*data scientists*) preparando-os para atuar em projetos de *big data*.
- Contribuir na divulgação dos resultados de pesquisas para o grande público.

Em resumo, como ilustrado na Figura 2.1 o foco do LIneA é preencher uma importante lacuna que é prover a infraestrutura necessária para lidar com os desafios impostos por projetos de *Big Science* envolvendo *big data*, e no processo formar uma nova geração de pesquisadores e tecnologistas, além de desenvolver uma cultura de serviço na área de distribuição e mineração de dados ainda não existente no Brasil.

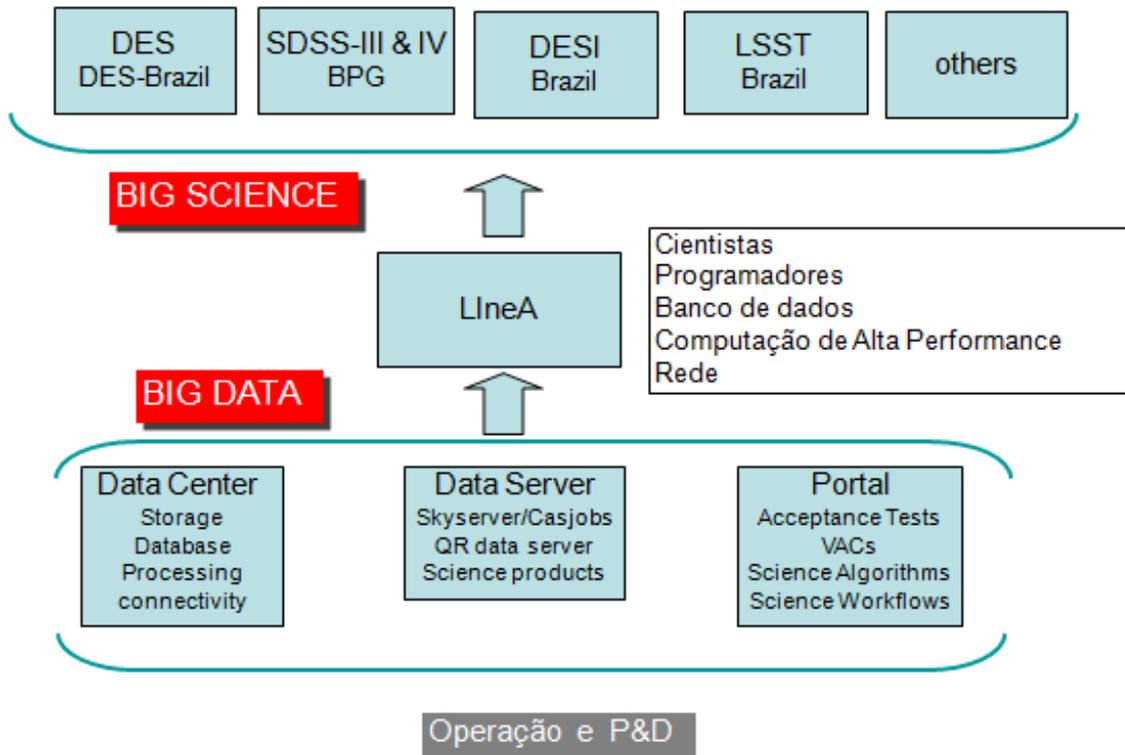


Figura 2.1- O LIneA e o contexto de Big Science e big data.

Para cumprir sua missão, o LIneA ao longo dos últimos tornou-se um laboratório nacional **multiusuário**, com um centro nacional de distribuição de dados astronômicos originais gerados pelos projetos nos quais seu afiliados participam, um centro de pesquisa e de desenvolvimento de software aplicado à astronomia na busca de soluções para problemas de *big data*, e uma rede de pesquisa nacional em astronomia. A atuação do laboratório se sobrepõe às atividades sendo realizadas em diferentes institutos do MCTI, envolve diferentes áreas de conhecimento e transcende barreiras geográficas envolvendo pesquisadores e tecnologistas espalhados pelo Brasil e por diversos países do mundo. Por definição, sua natureza é virtual e interinstitucional, e a experiência acumulada e suas parcerias internacionais na área de e-ciência são relevantes para outras áreas de atividade e devem ser compartilhadas.

### 3 Organizações Envolvidas

Como mencionado acima as organizações que inicialmente fizeram parte do acordo foram o CBPF, LNCC e o ON. Entretanto, em 2013 com a saída do único pesquisador do CBPF que fazia parte tanto do DES-Brazil como do SDSS-III o CBPF decidiu não mais participar do acordo. Em paralelo, a entrada da RNP estava sendo negociada e para isso um aditivo foi elaborado onde a responsabilidade de cada instituto ficou assim designada:

#### Caberá ao ON:

1. Contribuir para a terceirização do time atual de software.
2. Alocar o orçamento operacional mencionado na Cláusula 3, como disposto na proposta orçamentária aprovada.
3. Hospedar o LIneA em seu campus, atendendo as necessidades logísticas e operacionais do mesmo.
4. Contribuir para a formação de pessoal na área de TI.
5. Estabelecer e operar o site terciário para os experimentos do DES e SDSS-III, através do LIneA.
6. Ampliar o portal científico para atender às necessidades de ambos os projetos assim como de outros projetos de interesse, através do LIneA.

#### Caberá ao LNCC:

1. Colaborar na implantação física do sistema de armazenamento e de soluções para o resgate eficiente dos dados.
2. Hospedar em suas dependências parte estável dos dados dos experimentos.
3. Colaborar para a criação de uma grade interligando máquinas do ON, LNCC e RNP.
4. Contribuir para a absorção e terceirização de novos membros da equipe de TI.
5. Contribuir na implantação dos sites terciários no LIneA gerenciando os sistemas de processamento e armazenamento através do SINAPAD – Sistema Nacional de Processamento de Alto Desempenho.
6. Realizar pesquisa e desenvolvimento para apoio ao gerenciamento de dados e imagens, bem como processamento de experimentos científicos requeridos ou produzidos pelos projetos DES e SDSS-III.

#### Caberá a RNP:

1. Disponibilizar o acesso ao serviço de provisionamento de conexões para realização de transferência de grandes volumes de dados entre LIneA e as instituições com as quais tem contato constante (Fermilab, NCSA, SLAC, JHU).
2. Realizar diagnósticos no LIneA com o objetivo de identificar eventuais problemas de conectividade tanto entre o LIneA (PoP-RJ) e o ON quanto entre o LIneA e seus parceiros internacionais propondo mudanças nas configurações de rede para melhorar as taxas de transferências.

3. Instrumentar a infraestrutura do LIneA para a realização de medições regulares, nos principais enlaces utilizados pelo sistema, permitindo a realização de avaliações para crescimento de suas atividades e com isso garantir um acesso eficiente aos usuários do LIneA para acessar o portal e o sistema de desenvolvimento.
4. Elaborar e coordenar programas de P&D para o desenvolvimento de ferramentas de apoio a transferência de dados.
5. Contribuir com apoio nas áreas de sua expertise.
6. Apoiar o LIneA no uso dos serviços disponíveis pela RNP em produção.

## 4 Projetos Apoiados

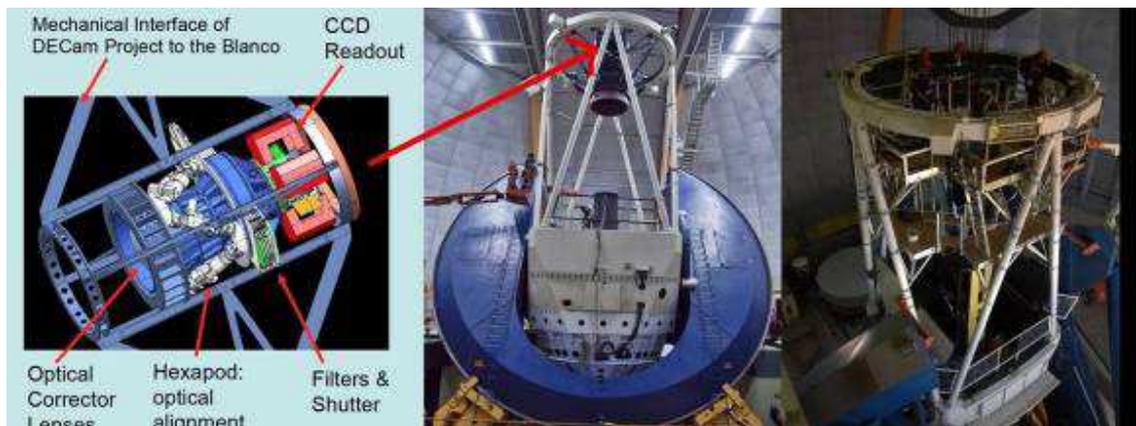
Como descrito acima, a criação do LIneA foi motivada pela entrada de pesquisadores brasileiros nos seguintes projetos descritos a seguir.

### 4.1 Dark Energy Survey (DES)

O levantamento DES tem por objetivo estudar a natureza da energia escura, uma componente descoberta recentemente, que representa da ordem de 70% do conteúdo do Universo, sendo esta energia a suposta responsável pela aceleração de sua expansão. O projeto procura determinar a abundância da energia escura e a sua variação ao longo da história do Cosmos, através das seguintes observações:

1. do número de aglomerados de galáxias em função do desvio para o vermelho (*redshift*);
2. da caracterização da distribuição de galáxias em grande escala;
3. das distorções nas imagens das galáxias causadas pelo efeito de lentes gravitacionais fracas;
4. da variação da distância de supernovas do tipo Ia com *redshift*.

Para cumprir com esta tarefa foi construída uma supercâmera de 520 Megapixels (Figura 4.1), o equivalente a ~50 câmeras digitais de 10 Megapixels. Ela está instalada no telescópio Blanco de 4m, localizado no *Cerro Tololo Inter-american Observatory* (CTIO), Chile (Figura 4.2). As observações começaram no segundo semestre de 2012 e utilizarão um total de 512 noites de telescópio durante os cinco anos de coleta de dados.



**Figura 4.1** – Diagrama mostrando as principais partes da câmera DECam construída pelo projeto DES e instalada no foco primário do telescópio Blanco no CTIO, Chile.

A entrada brasileira deu-se em 2007 e a conclusão do projeto está prevista para 2020. Entretanto, a responsabilidade de manter os dados deverá se estender além desta data. O custo previsto foi de US\$ 1,0 milhão do qual US\$ 300 mil foram pagos em espécie. O

restante foi aceito na forma de trabalho envolvendo o desenvolvimento de sistemas de software para a colaboração, descritos mais abaixo.



**Figura 4.2** – Vista externa do telescópio Blanco de 4 metros usado pelo projeto DES

Como mencionado anteriormente, a entrada do Brasil neste projeto se deu através da formação de um consórcio de pesquisadores criado através de um protocolo de colaboração (apêndice 4.1) que já antecipava a criação do LIneA. A escolha dos pesquisadores envolvidos no projeto foi feita após a realização de uma reunião organizada no CBPF e aberta a toda a comunidade.

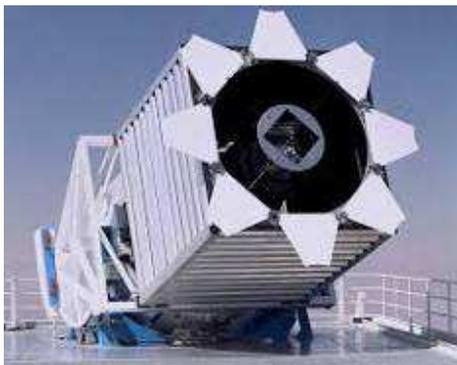
## **4.2 Sloan Digital Sky Survey III (SDSS-III)**

O SDSS-III é uma colaboração internacional, que produziu vários levantamentos espectroscópicos de extensas regiões do céu, criando amostras estatísticas sem precedentes para estudos em diferentes áreas. O projeto se beneficiou da infraestrutura montada pelos projetos SDSS-I e SDSS-II usando o telescópio de 2.5 metros do *Apache Point Observatory* (APO) e o espectrógrafo alimentado por mais de 500 fibras ópticas, um dos mais competitivos instrumentos da época (Figura 4.3). Para o projeto APOGEE foi construído um espectrógrafo trabalhando no infravermelho (Figura 4.4) o que permite se observar em regiões com grandes quantidades de poeira e impossíveis de se observar no óptico.

Seguindo o exemplo inovador do DES, um anúncio de oportunidade foi circulado pela comunidade astronômica brasileira (apêndice 4.2) e um consórcio de pesquisadores denominado *Brazilian Participation Group* (BPG) foi formado com pesquisadores do CBPF, ON, UFF, UFRJ, UFRGS, USP e de pesquisadores brasileiros no exterior, como convidados.

As observações se encerraram em 2014, mas análises científicas ainda continuam sendo feitas. Quatro grandes temas foram estudados a partir de dados obtidos por diferentes sub-levantamentos (Figura 4.5) com os seguintes objetivos: 1) MARVELS – monitorar 11 mil estrelas para detecção de planetas gigantes; 2) SEGUE-2 – mapear a estrutura, cinemática e composição química das partes mais externas da Via Láctea; 3) APOGEE – mapear as partes centrais da Via Láctea feitas no infravermelho, de forma a se evitar o obscurecimento pelas nuvens de poeira nestas regiões; 4) BOSS – mapear a distribuição de galáxias em grande escala para estudar a história da expansão do Universo.

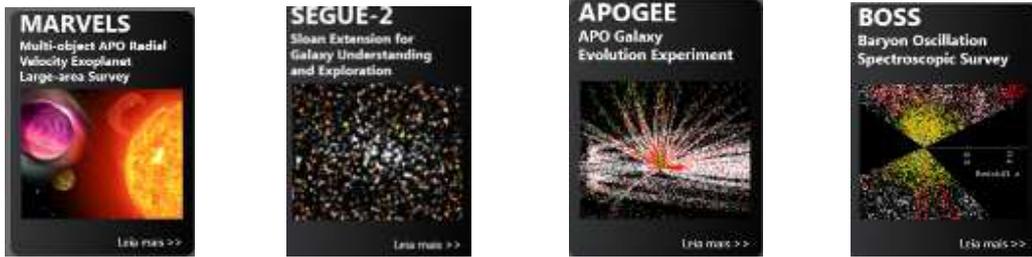
O projeto terminou oficialmente em Dezembro de 2015 um ano após a conclusão das observações. Entretanto o acervo continuará a ser mantido. A entrada do Brasil se deu em 2008, e foi feita com um anúncio de oportunidade e recrutamento procurando formar um time que melhor pudesse explorar os quatro levantamentos. Como no caso do DES, um acordo foi feito para que a entrada do BPG fosse em parte uma contribuição financeira feita em 5 parcelas e uma parte em trabalho. Esta parte foi feita pelo LIneA e envolveu a implantação e a manutenção de um centro para a distribuição de dados do projeto. Isto envolveu a compra de máquinas e software (MS SQL) como exigido para ser uma réplica do centro disponível na *Johns Hopkins University* (JHU).



**Figura 4.3** - Telescópio de 2.5 metros dedicado aos levantamentos do projeto SDSS (painel esquerdo) e cartucho com as fibras ópticas (painel direito), já colocadas nas posições dos objetos a serem observados, que é montado no foco do telescópio.



**Figura 4.4** – Espectrógrafo infravermelho construído para o projeto APOGEE.



**Figura 4.5-** Levantamentos executados pelo projeto SDSS-III.

Desde então, o LIneA vem negociando e apoiando a entrada de pesquisadores brasileiros em outros projetos internacionais ou projetos liderados por pesquisadores brasileiros mas que necessitam de uma infraestrutura operacional existente no laboratório. Abaixo, são apresentados os novos projetos.

### 4.3 O Sloan Digital Sky Survey – IV (SDSS-IV)

O projeto SDSS-IV é em grande parte continuação dos trabalhos feitos no SDSS-III. Em particular, dois dos levantamentos (eBOSS e APOGEE-2) são extensões naturais de levantamentos realizados no SDSS-III. O projeto inclui ainda um novo levantamento (MaNGA) de 10.000 galáxias próximas utilizando pacotes de fibra óptica que geram cubos de dados, onde a terceira dimensão é o espectro obtido por cada fibra na posição sobreposta na galáxia, permitindo estudar de uma só vez sua distribuição de luminosidade e química, bem como suas propriedades cinemáticas. (Figura 4.6)



**Figura 4.6 –** Levantamentos sendo executados pelo projeto SDSS-IV.

Alguns comentários devem ser feitos. O primeiro é de que o projeto APOGEE-2 além de aumentar significativamente a amostra sendo observada no hemisfério norte, também está construindo uma réplica do espectrógrafo montado no APO para ser colocado no telescópio *Dupont* do *Las Campanas Observatory* (LCO) operado por *Carnegie Observatories*. Isto permitirá a observação na direção do bojo de nossa galáxia, só visível do hemisfério sul o que proporcionará pela primeira vez uma visão panorâmica de nossa galáxia penetrando regiões obscurecidas por poeira. O segundo ponto é que o levantamento eBOSS não é apenas a continuação do projeto BOSS, usando diferentes tipos de galáxias para mapear diferentes intervalos de *redshift*, mas também é o precursor do projeto DESI descrito abaixo, um dos mais promissores para determinar a natureza da

energia escura. Projetos combinados entre eBOSS e DES também foram realizados para calibrar métodos para a escolha de populações de galáxias a serem observadas pelo DESI. Finalmente, o terceiro ponto é que com a filosofia adotada pelo LIneA as ferramentas e os algoritmos de análise utilizados para um projeto servem de base para a próxima geração de projetos como é o caso dos levantamentos BOSS, eBOSS e DESI. Portanto, manter um repositório de códigos, construir *pipelines* que assegurem a retenção de conhecimento e prepararem jovens pesquisadores que possam dar continuidade ao trabalho ao longo de duas décadas é vital para tornar os pesquisadores brasileiros protagonistas internacionais em projetos na vanguarda do conhecimento e de grande impacto como é o caso do estudo da energia escura, entre outros tópicos de interesse.

Seguindo o exemplo de seu antecessor, um anúncio de oportunidade foi circulado tendo uma resposta expressiva da comunidade (apêndices 4.3-4.4), que também manifestaram o seu interesse junto à SCUP (apêndices 4.5-4.6), e um grupo de participação (BPG) foi formado. Neste caso, levando em conta a experiência como o SDSS-III, foi solicitado aos novos membros concordarem com um conjunto de regras estabelecidas no Termo de Compromisso apresentado no apêndice 4.7. Uma das principais exigências foi o comprometimento do pesquisador em dedicar uma fração de seu tempo ao projeto, que justificasse os recursos sendo investidos em seu nome, e o reconhecimento do apoio sendo dado pelo LIneA. Embora inédito na Astronomia brasileira, é importante enfatizar que o conjunto de regras estabelecidas é similar ao adotado nas colaborações internacionais e cujo objetivo é proteger o investimento feito, assegurar o bom funcionamento do grupo, estimular a cooperação entre seus membros e garantir o retorno científico.

#### **4.4 Transneptunian Occultation Network (TON)**

Objetos transnetunianos (TNOs) constituem uma população de pequenos corpos planetários situados além da órbita de Netuno, ou seja, estão afastados do Sol mais de 30 vezes a distância que separa a Terra do Sol (cerca de 150 milhões de quilômetros). Isto faz com que tais objetos tenham sofrido poucas alterações ao longo do tempo e, por isso, são preciosas fontes de informação sobre a história e a evolução do sistema solar exterior.

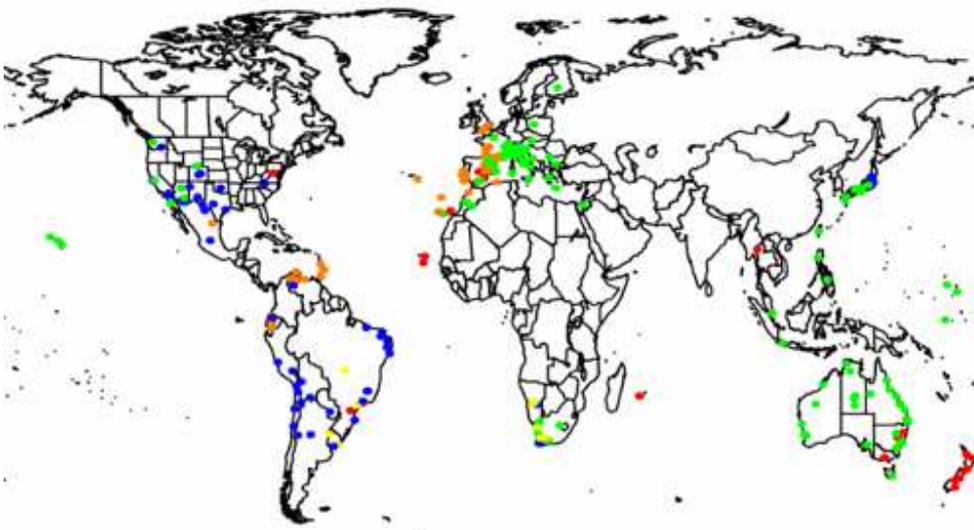
Uma forma de se estudar esta população é através das ocultações estelares, ou seja, quando eles cruzam a linha de visada formada por um observador – normalmente na Terra – e uma estrela. Através da variação da luz (curva de luz) que resulta do evento de ocultação, pode-se determinar a forma e a dimensão do TNO com precisão de poucos quilômetros, saber mais sobre estruturas em suas vizinhanças imediatas, bem como detectar e estudar atmosferas cuja pressão é da ordem de 1 bilhão de vezes menor que a da Terra. Tal estudo requer um grande esforço para, entre outros, prever quando e onde uma dessas ocultações ocorrerá, para observá-la.

Este projeto, já em andamento há alguns anos e com bastante sucesso, solicitou o apoio do LIneA reconhecendo a necessidade de ter um suporte mais sólido de TI em preparação ao esperado crescimento exponencial do número de ocultações que acontecerá em breve com a combinação dos resultados da missão espacial GAIA e dos grandes levantamentos fotométricos como o DES, DECam e LSST. O projeto GAIA fornecerá posições e movimentos próprios estelares com precisão sem precedentes para cerca de 1 bilhão de estrelas espalhadas por todo o céu, o que implicará num grande salto na qualidade da predição dos eventos de ocultação.

As necessidades do projeto TON são variadas e incluem:

- O armazenamento e análise de dados obtidos ao longo de grandes intervalos de tempo. Estas observações são realizadas periodicamente em vários telescópios (OPD, Pic-du-Midi, ESO, entre outros) para determinar com mais precisão órbitas de TNOs. Tendo em vista o grande número de dados, da ordem de 10.000 exposições já disponíveis, será necessário montar um *pipeline* para redução de imagens, calibração astrométrica, e o armazenamento dos resultados em um banco de dados.
- Desenvolver *pipeline* para identificar a posição de TNOs conhecidos nas imagens obtidas com a DECam. Para permitir que este trabalho seja feito com a maior rapidez possível, membros do time foram propostos e aceitos para serem colaboradores externos do levantamento DES.
- Implantar infraestrutura que permita administrar campanhas de ocultação com maior controle e eficiência. Receber, armazenar e analisar dados em vários formatos (vídeo, imagens fits, planilhas) provenientes de uma grande rede de observadores profissionais e amadores.
- Desenvolver no portal os *pipelines* de análise dos dados obtidos durante um evento de ocultação.

O mapa da Figura 4.7 mostra, através dos pontos coloridos, observatórios profissionais e telescópios amadores (fundamentais!) que colaboraram ao menos uma vez com as campanhas observacionais organizadas pela equipe. Além dos resultados científicos, o projeto é ainda de grande interesse social engajando uma enorme comunidade incluindo astrônomos amadores que têm a oportunidade única de contribuir para uma pesquisa científica de ponta.



**Figura 4.7** Os pontos indicam locais onde observações foram feitas para diferentes campanhas de ocultação, indicadas pelas diferentes cores.

#### 4.5 Large Synoptic Survey Telescope (LSST)

O LSST é um telescópio sendo construído em *Cerro Pachón*, no Chile. Previsto para entrar em operação no início da próxima década, o LSST irá mapear, em seis bandas, quase a metade do céu por um período de 10 anos (Figura 4.8). O telescópio com um diâmetro de 8,4 metros (Figura 4.9) tem um campo de visada de quase 10 graus quadrados podendo mapear toda região do céu ao qual ele tem acesso, em apenas algumas noites. Sua câmera consiste de um mosaico de CCDs com 3,2 bilhões de pixels e, cada exposição cobre uma área que corresponde a 40 vezes o tamanho da Lua cheia. A cada noite serão acumulados ~15 TB de dados que devem ser transmitidos para diferentes centros para redução e análise. O sistema será o mais poderoso coletor de luz no óptico e a velocidade das observações fornecerá aos astrônomos pela primeira vez uma visão dinâmica do Universo, onde variações de posição ou fluxo serão registradas numa cadência de algumas noites. Estima-se que o LSST gerará da ordem de 10 milhões de alertas a cada noite que devem ser classificados e os casos mais interessantes serão observados em outros telescópios. Ao término de 10 anos o levantamento obterá informações para 37 bilhões de estrelas e galáxias explorando um volume de espaço sem precedentes e gerando da ordem de 100 PB de produtos. Este projeto representa um grande desafio na área de TI para gerenciar a transferência, processamento, armazenamento, análise e exploração científica da grande quantidade de dados que será gerada de forma ininterrupta. Este é um problema de *big data* que começa a ser enfrentado de uma forma sistemática pelo projeto procurando novas soluções para as áreas de rede, processamento de alto-desempenho e desenho de banco de dados.



**Figura 4.8-** Projeto do prédio que irá abrigar o telescópio do LSST, já em construção.

Com os dados acumulados cientistas irão explorar o Sistema Solar, descobrir novos fenômenos, e estudar a estrutura de nossa galáxia, a formação e evolução de estruturas no Universo em função do tempo cósmico e determinar as propriedades da matéria e energia escura que permeiam o Universo, sendo esta última responsável por sua expansão acelerada. Este fenômeno foi confirmado independentemente em 1998 por duas equipes, as quais receberam o prêmio Nobel de Física de 2011.

Tendo em vista a grande variedade de ciência que pode ser extraída a partir dos dados do LSST, e o possível grande impacto no estudo da matéria e energia escura, o projeto LSST é considerado prioritário e já se encontra em construção (ver Figuras 4.10 e 4.11) com o apoio do Departamento de Energia do governo americano (DOE) e do *National Science Foundation* (NSF), seguindo o exemplo adotado para o DES. O apoio para o LSST vem das comunidades de Astronomia e da Física de Altas Energias americana como manifestado nos respectivos planejamentos estratégicos *New Worlds*, *New Horizons* e *Building for Discovery* (Figura 4.12). O início do levantamento está previsto para 2022.



**Figura 4.9** – Espelho primário do LSST na fábrica em Tucson, Arizona.

Repetindo a história de outros importantes momentos singulares (*Palomar Sky Survey*, *Sloan Digital Sky Survey*) o LSST será um marco sem precedentes para a Astronomia. O impacto que este projeto terá é incomensurável não só por aquilo que pode ser feito diretamente com os dados do levantamento, mas em termos de sua capacidade de gerar amostras inéditas para serem estudadas com outros instrumentos, em diferentes comprimentos de onda. Também é importante ter-se uma visão abrangente de forma a definir políticas de uso para os outros telescópios para que estes possam fazer parte de um sistema integrado possibilitando o estudo detalhado e complementar de possíveis descobertas feitas com os dados do LSST.



**Figura 4.10** – Lançamento da pedra fundamental do observatório em Cerro Pachón com a presença de várias autoridades, incluindo a Presidente chilena. Crédito: LSST



**Figura 4.11** – Obras para a construção do prédio que irá abrigar o LSST.

É importante enfatizar que o maior desafio a ser enfrentado pela comunidade brasileira para usufruir deste projeto será ter a infraestrutura de hardware e software capaz de lidar com a grande quantidade de dados que ficará disponível diariamente.



**Figura 4.12** – Capa dos relatórios preparados pela comunidade astronômica (painel esquerdo) e de física (painel direito) que apoiaram o LSST (e o DESI) como sendo projetos prioritários para a próxima década. Ambos já se encontram em construção com o apoio do NSF e DOE.

Tendo em vista a importância do LSST para o futuro da Astronomia, o LIneA e o LNA vem explorando formas de integrar a comunidade brasileira neste projeto. Com este objetivo foram feitas as seguintes iniciativas: i) a organização em 2008 das reuniões **A Glimpse into the Future of Astronomy** e **New Astronomy: the Data Challenge** organizadas pelo grupo atualmente associado ao LIneA; ii) a reunião **Science with the LSST: A Brazilian/US joint Workshop** organizada em 2012 por várias instituições entre as quais o LNA; iii) a participação do Diretor do LNA e o Coordenador do LIneA na reunião do LSST em 2013; iv) a **Workshop XLDB South America** organizada pelo LIneA em 2014; v) a participação do coordenador do LIneA na reunião do LSST em 2014; vi) a organização patrocinada pelo LIneA e o LNA com o apoio da SAB sobre um levantamento do interesse da comunidade brasileira neste projeto respondido por 170 pesquisadores.

Além destas iniciativas vale a pena mencionar ainda a participação da equipe do laboratório na reunião **Joint DES-LSST Workshop** em 2014 (apêndice 4.8), as inúmeras reuniões técnicas realizadas no Rio, Palo Alto, Tucson e Urbana, e os acordos de colaboração técnica entre o LIneA e o LSST, incluindo a ida de um pesquisador do LIneA para trabalhar na equipe técnica do LSST, e o acordo entre o LIneA e o NCSA, lembrando que o NCSA será o Centro Regional de Dados do LSST nos Estados Unidos (ver seção 13 para maiores detalhes sobre estes diversos acordos). É importante salientar que o DES é visto pelo NSF e o DOE como precursor do LSST, e a experiência científica dos pesquisadores atuando no DES e o trabalho técnico sendo realizado pelo LIneA para o DES será extremamente valioso para a participação brasileira no LSST.

Todo este esforço foi recentemente recompensado com a assinatura de um acordo firmado entre o LIneA, LNA, RNP e a ANSP de São Paulo que vai permitir a participação

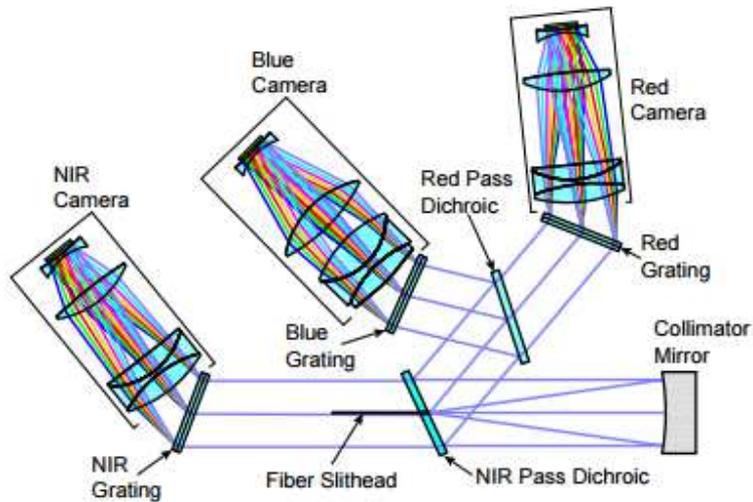
de pesquisadores brasileiros no projeto LSST. O Memorando de Entendimento (ver seção 13) assinado pelas instituições prevê investimentos em conexões de fibra óptica, ampliando a conexão entre a América do Sul e a América do Norte. Além disso, um grupo de 50 pesquisadores brasileiros participará do projeto, considerado revolucionário para a Astronomia, cabendo ao LIneA e ao LNA criar um processo para a seleção dos pesquisadores que deverão fazer parte deste primeiro grupo denominado BPG-LSST. No momento, o comitê executivo do LIneA (ver seção 5) e a direção do LNA discutem uma proposta de regimento interno para este grupo (apêndice 4.9). Em paralelo, explora-se a possibilidade da inclusão de pelo menos dois cientistas do LSST na comissão que escolherá os pesquisadores seniores do grupo.

#### 4.6 Dark Energy Spectroscopic Instrument (DESI)

O DESI é um levantamento dedicado a observações espectroscópicas de dezenas de milhões de objetos para o estudo da energia escura. As observações serão feitas durante cinco anos com o telescópio *Mayall* de 4m do *Kitt Peak National Observatory* (Figura 4.13) e devem começar em 2019. Para atingir os objetivos científicos será construída uma lente corretora que permitirá um campo de visada de 8 graus quadrados. Além disso, será construído um espectrógrafo de média resolução com três divisores de luz que criam feixes encaminhados para detectores no infravermelho, no vermelho óptico e azul óptico. O instrumento terá um posicionador robótico de cinco mil fibras configuradas em menos de um minuto, permitindo uma alta cadência e baixo *overhead*. Esta combinação torna o DESI 15 vezes mais poderoso que o instrumento usado para o levantamento BOSS. Recentemente uma cooperação técnica foi estabelecida entre o LIneA e o projeto DESI para o desenvolvimento da ferramenta *Quick Look Framework* (QLF) para ser usada na avaliação dos 15.000 espectros obtidos em cada exposição da câmera, e desta forma otimizar as observações. As 5000 fibras irão alimentar 10 espectrógrafos onde o espectro de cada objeto será dividido em três partes como mostra a Figura 4.14.

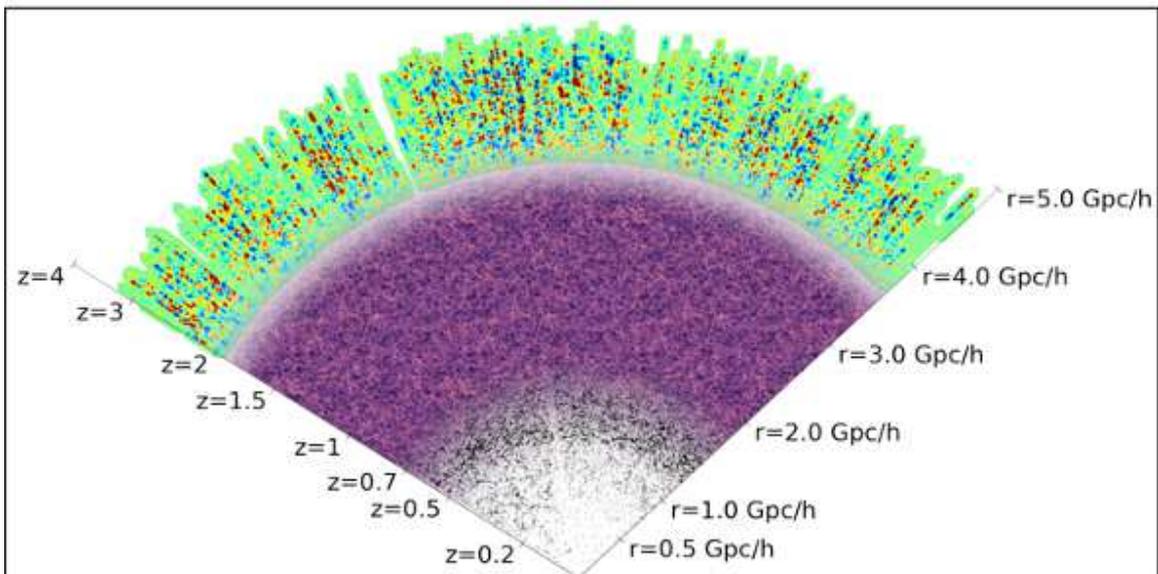


**Figura 4.13-** Observatório de Kitt Peak. O maior prédio é do telescópio Mayall de 4 metros aonde o DESI será montado. O telescópio é gêmeo daquele utilizado pelo DES no CTIO e o projeto contará com o apoio da equipe do Fermilab que trabalhou na montagem da DECam.



**Figura 4.14-** Desenho esquemático do espectrógrafo do DESI. Como descrito no texto o instrumento contará com dez destes espectrógrafos.

Para mapear o Universo espectroscopicamente até *redshifts*  $z \sim 3,5$  serão utilizadas distintas classes de objetos. Para o intervalo de  $z=0,4-1,0$ , serão observadas 4 milhões de LRGs; no intervalo  $z=0,7-1,7$ , 23 milhões de ELGs; no intervalo  $z=1,0-2,2$ , 1,4 milhões de QSOs; e no intervalo  $z=2,2-3,5$ , 0,6 milhões de sistemas *Lyman-alpha*. A área total de céu coberta pelas observações será de 14 mil graus quadrados. Isto permitirá mapear a distribuição espacial de galáxias cobrindo um volume de espaço sem precedentes como mostra a Figura 4.15.

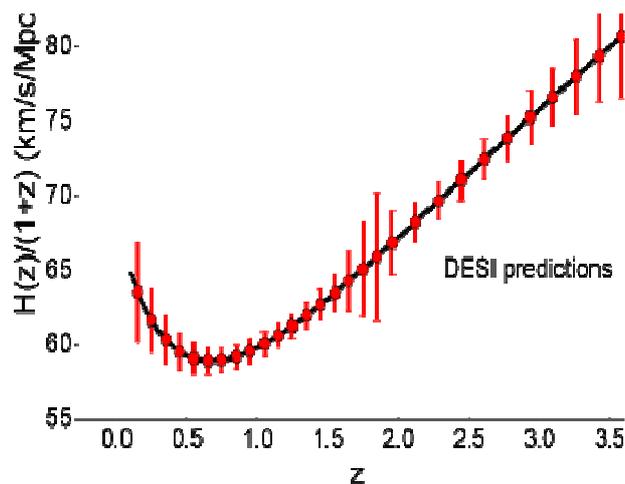


**Figura 4.15 -** Distribuição espacial das galáxias a serem observadas pelo projeto DESI. No gráfico cada cor representa diferentes tipos de galáxias que serão observadas a diferentes intervalos de *redshifts* incluindo LRGs, ELGs, QSOs e *Lyman-alpha*.

Este experimento, da Fase IV como definido pelo *Dark Energy Task Force* (DETF), é sem precedentes e se propõe a impor vínculos fortes sobre modelos de energia escura e de gravidade modificada que tentam explicar os efeitos observados da expansão acelerada do Universo. Com os dados do DESI, por exemplo, será possível determinar a escala do Universo com uma precisão de 1% em 35 intervalos de *redshift* como ilustra a Figura 4.16. Além disso, analisando as distorções no espaço de *redshift* será possível determinar a taxa de crescimento de estruturas e, em princípio, distinguir entre as possíveis explicações sobre a origem da aceleração.

Convém ressaltar, que apesar do foco principal do projeto ser a ciência da energia escura, em noites enluaradas o telescópio será utilizado para estudos da Via Láctea, complementando os dados do APOGEE, e do Universo local.

Reconhecendo o enorme impacto que este tipo de projeto terá em várias áreas e em particular na ciência da energia escura o coordenador do LIneA em sua função de explorar oportunidades científicas participou desde 2012 de inúmeras reuniões sobre a possibilidade da construção de um espectrógrafo com as características propostas pelo DESI para a realização de um grande levantamento espectroscópico. Na época existiam dois projetos sendo discutidos um para o sul (*Dark Energy Spectrograph*, DESpec) e um para o norte (*Dark Energy Spectroscopic Instrument*, DESI) sendo que a decisão final em Dezembro de 2012 recaiu neste último. Desde então o coordenador do LIneA faz parte do *Institutional Board* do projeto como ouvinte, participando de reuniões anuais do projeto e procurando formas de cooperação e estimulando a participação de grupos brasileiros, incluindo uma proposta feita a FAPESP por pesquisadores paulistas. No momento, esta sendo negociada a entrada de um pequeno grupo em troca do desenvolvimento de um sistema para validar os 15.000 espectros que serão produzidos a cada 20 minutos (ver seção 13 em Acordos Técnicos).



**Figura 4.16** – Diagrama de Hubble que irá ser construído a partir das observações de DESI.

## 4.7 Comentários gerais

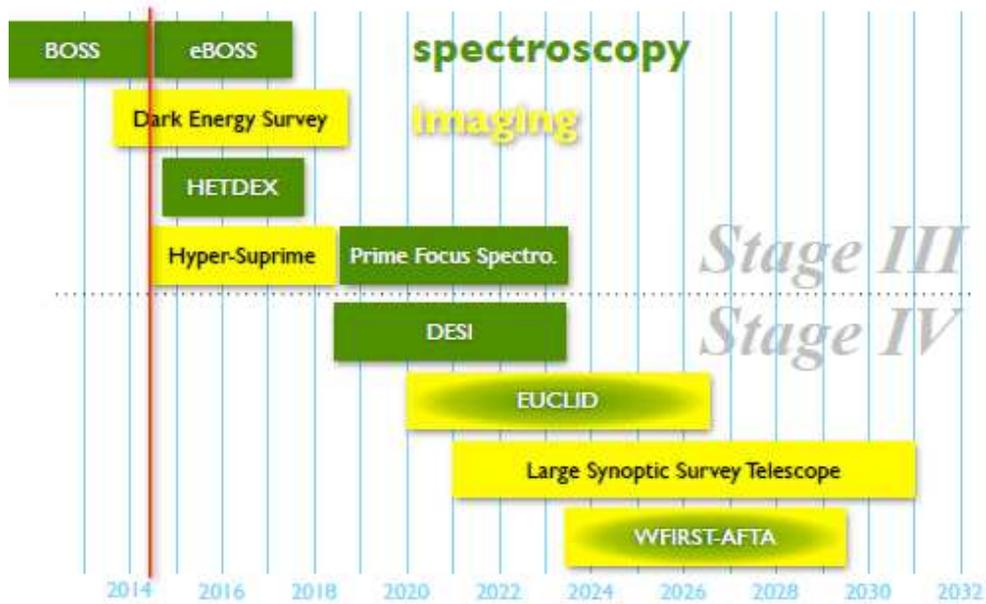
Todos os projetos internacionais acima mencionados têm o apoio financeiro do *National Science Foundation* (NSF) e do Departamento de Energia (DoE) do governo americano e deles participam os principais laboratórios nacionais americanos como o *Argonne National Laboratory* (ANL), *Fermi National Accelerator Laboratory* (Fermilab), *Lawrence Berkeley National Laboratory* (LBL), *National Center for Supercomputing Applications* (NCSA), *National Optical Astronomy Observatory* (NOAO), *SLAC National Accelerator Laboratory* (SLAC) e os principais centros de pesquisa no mundo. A combinação de uma sólida base financeira, a experiência na gestão de grandes projetos nas áreas de física, astronomia e TI, de uma cultura de acompanhamento sistemático e da qualidade científica de seus membros tornam estes projetos de baixo risco e, em geral, de baixo custo/benefício, sendo uma excelente oportunidade de aprendizado técnico/científico. Outro ponto em comum é que estes são todos projetos de *big data* seja pela velocidade, variedade e/ou volume dos dados gerados e que precisam ser gerenciados para garantir o retorno científico.

Os projetos acima também envolvem diferentes técnicas que permitem cobrir um amplo espectro de tópicos como o estudo de objetos do sistema solar, a detecção de exoplanetas, a estrutura de nossa galáxia, a evolução de galáxias, a evolução de estruturas do universo, matéria escura e energia escura. Isto tem o benefício de atender pesquisadores com diferentes expertises e, em princípio, servir como um polo para atrair investimentos, evitando assim a pulverização dos recursos públicos numa variedade de pequenos projetos todos com dificuldades de manutenção.

Como descrito acima todos os projetos também envolvem uma colaboração técnica, seja no desenvolvimento de software, seja na operação de um centro de distribuição de dados, seja no processamento o que traz grandes benefícios para os participantes brasileiros que tem acesso privilegiado aos dados e uma equipe técnica trabalhando com os parceiros internacionais de forma a manter os serviços, as técnicas e as ferramentas disponíveis sempre atualizadas no que há de mais moderno e com experiência no gerenciamento dos dados sendo distribuídos.

Mais importante ainda é que todos se encaixam numa sequência que define um roteiro para os próximos 20 anos como mostra a Figura 4.17. Em particular, no caso dos projetos apoiados pelo LIneA pode ser visto que o DES é o precursor do LSST e o SDSS (BOSS, eBOSS) o precursor do DESI. Portanto, todo o aprendizado e algoritmos desenvolvidos nos projetos atualmente em andamento servirão de base para os projetos futuros e o aluno de hoje será o investigador principal de amanhã. Além disso, as colaborações estabelecidas ao longo dos anos podem em alguns casos possibilitar o envolvimento de pesquisadores brasileiros em projetos espaciais como *Euclid* e *WFIRST*.

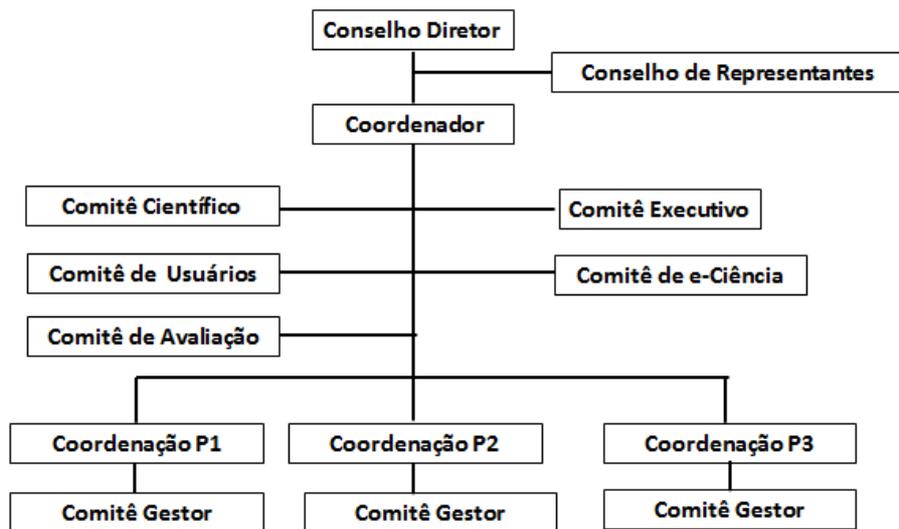
## The dark energy facilities roadmap



**Figura 4.17** – Cronograma dos principais projetos com foco no estudo sobre a natureza da energia escura cobrindo o período de 2013-2032.

## 5 Organização Interna

Durante os últimos cinco anos várias tentativas de organização interna foram feitas para estruturar o LIneA, incentivar o trabalho colaborativo, e dividir o ônus do trabalho administrativo associado. Um rascunho de uma proposta de estatuto para o laboratório chegou a ser discutida internamente (apêndice 5.1), mas não chegou a ser terminada aguardando uma resposta do Ministério sobre o processo de renovação do acordo. A Figura 5.1 apresenta um modelo de organização discutido neste documento, sendo que alguns dos conselhos considerados fazem parte do acordo de criação do LIneA ou já foram criados como fruto das necessidades operacionais do laboratório.



*Figura 5.1- Modelo preliminar de organização interna do LIneA.*

Entre os conselhos previstos no acordo estão:

- Coordenação Geral - cabe ao coordenador geral a responsabilidade de manter o funcionamento do laboratório, preparar o orçamento anual, prestar contas das atividades ao Comitê Gestor sempre que solicitado.
- Conselho Diretor - formado pelos diretores das UPs signatárias do acordo, para a supervisão das atividades do laboratório, aprovar Memorandos de Entendimento entre o LIneA e outros órgãos e instituições nacionais e internacionais, e aprovar seu orçamento.
- Conselho de Representantes - formado por membros de cada instituição e indicados pelos respectivos Diretores e presidido pelo Coordenador. Sua função é acompanhar o dia-a-dia do laboratório e preparar o orçamento anual.

Durante o período 2011-2014 o Conselho Diretor se reuniu várias vezes para discutir formas de financiamento do laboratório e sua situação institucional. Sua última ação foi encaminhar ao MCTI uma carta solicitando uma avaliação do trabalho realizado pelo laboratório e discutir o seu futuro (apêndice 5.1). A atuação do Conselho de representantes foi menos sistemática embora alguns representantes tenham tido uma participação ativa e construtiva.

Outros conselhos em funcionamento ou previstos são:

- Comitê Executivo (EC) - atualmente composto de 6 membros representando diferentes instituições e os diversos levantamentos apoiados mais o coordenador. O EC se reúne semanalmente para rever junto ao coordenador geral temas que afetam o funcionamento do laboratório.
- Comitê Científico - responsável pela vida científica do laboratório estimulando a participação dos pesquisadores em reuniões dos grupos de trabalhos nacionais e internacionais, estimulando a redação de artigos, compilando nomes para a série de *webinars*, e organizando workshops virtuais.
- Comitê de e-Ciência - composto por representantes das áreas de ciência e TI para avaliar periodicamente o trabalho de desenvolvimento de software, rever prioridades, aconselhar no uso de novas tecnologias e aprovar plano de compras para o centro de dados. Convocado quando necessário, este comitê poderia ainda servir como uma referencia para serviços de consultoria junto à *experts* da comunidade para resolver problemas específicos enfrentados pelo laboratório, devidamente remunerados.
- Comitê de Usuários - formado por jovens pesquisadores indicados por cada projeto e os coordenadores de TI para encaminhar suas necessidades à coordenação do laboratório. Periodicidade trimestral ou quando convocado extraordinariamente.
- Comitê de Avaliação - formado por membros do comitê científico e comitê de e-ciência convocados pelo coordenador para avaliar propostas para a entrada de novos projetos e fazer recomendações ao Conselho Diretor.

Outras funções importantes para o funcionamento do centro são:

- Coordenação de Projeto - representante brasileiro junto aos projetos internacionais sendo apoiados oficialmente pelo LIneA ou investigador principal de projeto liderado pelo Brasil. Atualmente estes projetos incluem: DES, SDSS-IV, DESI, e LSST, e TON como descrito acima.
- Líder de Grupos de Trabalho - como os dados dos projetos acima mencionados podem ser usados para uma grande gama de temas científicos, a equipe científica é organizada em grupos de trabalho que se reúnem periodicamente. Estes grupos de trabalho em geral equivalem aos grupos de trabalho internacionais.
- Coordenação de *Webinar* - um serviço importante prestado pelo LIneA é o de manter uma série internacional de seminários virtuais internacionais. É de responsabilidade

do coordenador, elaborar uma programação semestral, fazer os contatos e testes necessários.

- Coordenação de Divulgação - outro aspecto importante do LIneA é o de assumir a responsabilidade e o compromisso de divulgar o trabalho sendo feito para o público em geral. No esforço de divulgação existe um responsável para rever os blogs, a integridade do site, manter a galeria de imagens renovada para atrair a atenção do grande público e a preparação do LIneA News enviado para mais de 1000 endereços eletrônicos. O coordenador de divulgação também é o porta-voz do LIneA sendo o ponto de contato com a imprensa na medida do possível.
- Coordenação de TI - é formada atualmente por dois *data scientists* (astrônomos com grande experiência em TI), dois tecnologistas e presidido pelo coordenador. Sua função é definir estratégias e prioridades do desenvolvimento de software, definir cronograma de atividades e acompanhar sua execução, validar entregáveis, planejar a expansão da infraestrutura computacional.
- Secretaria - dar apoio às atividades do laboratório sendo responsável pela manutenção do calendário de atividades, contato com fornecedores, acompanhamento das despesas, atualização do repositório de documentos.

É importante frisar que a implantação da estrutura proposta também é função do número de projetos e pessoas envolvidas. No entanto, isto não impede que seja feita uma proposta mais adequada para era do LSST. Outro ponto importante da proposta é que para cada projeto também está previsto um comitê gestor para tratar de assuntos internos ao projeto. O coordenador de cada projeto será escolhido por seus pares e fará parte do Comitê Científico do LIneA e membro do Comitê Executivo.

A organização interna e o estatuto do LIneA deve ser um dos assuntos de debate uma vez que a situação institucional do LIneA seja definida. Solicitações neste sentido vêm sendo feitas desde 2013, envolvendo discussões com a SCUP e uma carta do Conselho do LIneA à secretaria executiva de 15/02/2015 (apêndice 5.2), finalmente atendida com a portaria de 03/12/2015.

## 6 Afiliação

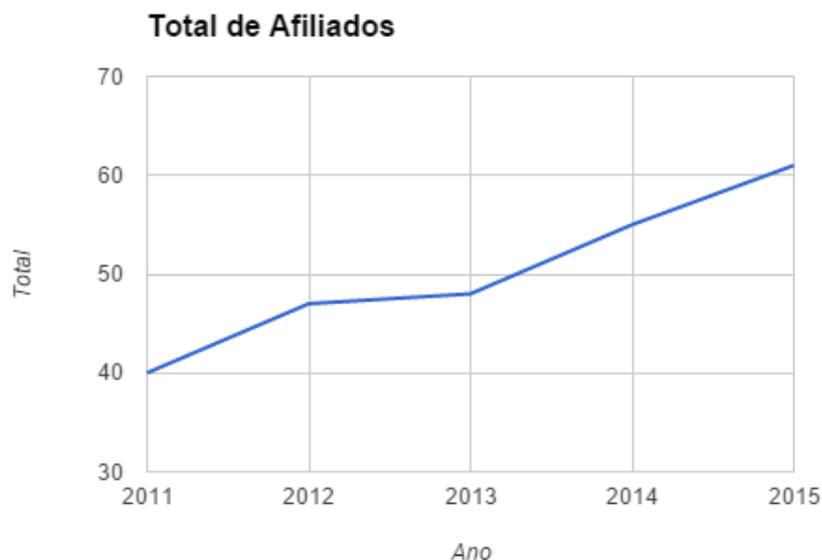
Na área de pesquisa existem as seguintes categorias de usuários do laboratório:

- Membros dos consórcios de pesquisadores brasileiros (e.g., DES-Brazil, e os diferentes BPGs) que participam dos projetos apoiados pelo LIneA seja financeiramente ou seja porque utilizam as instalações do LIneA na execução de seus trabalhos para esses projetos. Em contrapartida, os participantes assumem os seguintes compromissos: i) colocar a afiliação do LIneA nos trabalhos associados a estes projetos. ii) dedicar não menos que 25% do seu tempo trabalhando nos projetos; iii) dedicar uma pequena fração de seu tempo a atividades relacionadas a operação do laboratório.
- Pesquisadores que mesmo sem participar dos projetos apoiados usufruem da infraestrutura do laboratório. Neste caso considerados como membros associados aos quais se solicita que se afilem ao LIneA. Para se tornar um membro associado o pesquisador deve submeter uma solicitação ao laboratório descrevendo suas necessidades. Estes casos serão examinados pelo EC e serão atendidos dependendo da natureza da solicitação, justificativa e disponibilidade de recursos.
- A comunidade em geral que tem acesso aos dados públicos das colaborações apoiadas pelo LIneA e tem apoio da equipe do LIneA aos quais se solicita a inclusão de um agradecimento ao laboratório em artigos que envolvam dados obtidos do centro de dados.

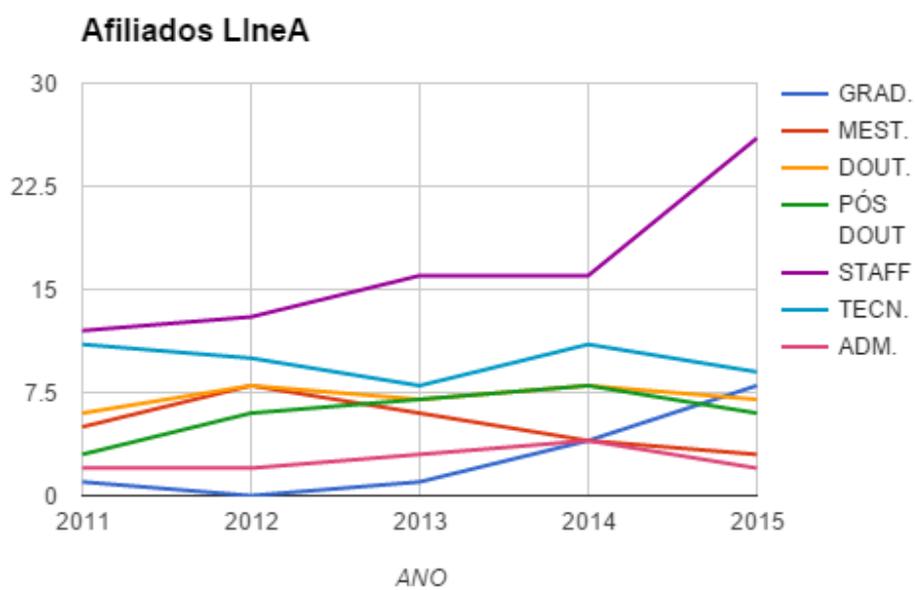
Além das categorias acima mencionadas fazem parte do LIneA o time de TI contratado diretamente pelo laboratório e, como membros associados, tecnologistas de outras instituições como do LNCC, RNP e CEFET que colaboram com o time e participam de projetos conjuntos.

A Figura 6.1 apresenta a evolução do número total de afiliados do LIneA no período de 2011-2015, e na Figura 6.2 discriminado por categorias.

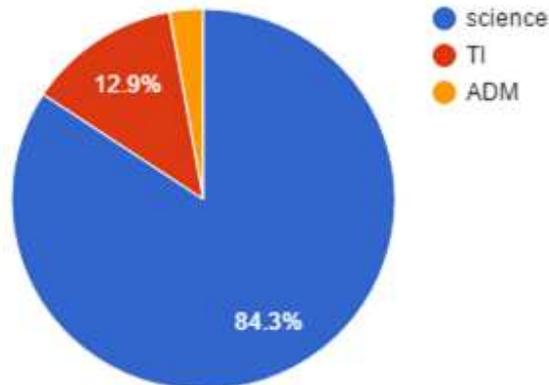
Como pode ser visto na Figura 6.3 o time é composto principalmente de pesquisadores sendo que a equipe de pesquisa tem uma saudável mistura de pesquisadores experientes e jovens pesquisadores como mostra a Figura 6.4. Enfatizamos que mesmo os pesquisadores contratados são em sua grande maioria jovens com menos de 50 anos, e, portanto, com longo futuro de atividades podendo dar contribuições importantes para o LSST.



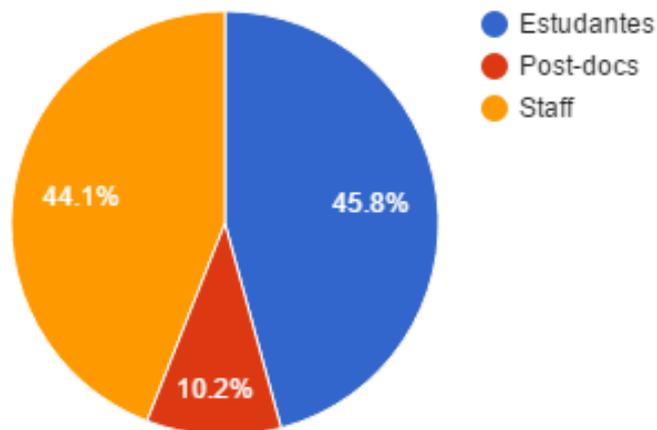
**Figura 6.1** - Número total de afiliados em diferentes categorias que participaram em um dado ano das atividades do LIneA.



**Figura 6.2** - Número de participantes nas atividades do LIneA nas diferentes categorias: Graduandos, Mestrandos, Doutorandos, Pós-doutorandos, Staff (professores e pesquisadores), Técnicos e Apoio Administrativo.



**Figura 6.3** – Composição da equipe do LIneA.



**Figura 6.4** – Composição do time científico do LIneA, considerando todo o período de 5 anos

Incluindo o período que antecede sua criação oficial, ~110 pessoas já participaram das atividades do LIneA.

Os pesquisadores e tecnologistas que já participaram ou participam atualmente das atividades do LIneA vem de instituições nacionais (Figura 6.5) e internacionais (Figura 6.6), enfatizando novamente o caráter **multiusuário** e internacional do laboratório.

Um total de 18 instituições das quais 12 nacionais dos estados de Rio de Janeiro, São Paulo, Paraná e Rio Grande do Sul, e seis internacionais localizadas na Alemanha, Bélgica, França, Inglaterra e Itália estão envolvidas com as atividades do LIneA. Além destas, diretamente ligadas ao LIneA é bom lembrar o enorme número de instituições internacionais e consórcios participando em cada projeto acima apresentado. No total

são 15 países: Alemanha, Austrália, Brasil, Canadá, Chile, China, Coreia do Sul, EUA, Espanha, França, Inglaterra, Itália, Japão, México e Suíça, distribuídos da seguinte forma: SDSS-III (26), SDSS-IV (33), DES (24) e DESI (71). A lista completa de participantes nestes projetos pode ser vista nos apêndices 6.1 a 6.4.



**Figura 6.5** – Logos das instituições brasileiras que tem/tiveram pesquisadores e tecnologistas participando do LIneA.



**Figura 6.6** – Logos das instituições internacionais (Alemanha, Bélgica, França, Inglaterra, Itália) que tem pesquisadores participando de projetos apoiados pelo LIneA.

## 7 Ciência

Uma das principais características dos projetos científicos apoiados pelo LIneA é a grande variedade de tópicos que podem ser cobertos com os mesmos dados, pois serve como um fator multiplicativo e, desta forma reduzindo a razão custo/benefício dos projetos. Por outro lado, isto representa um grande desafio, pois exige a formação de uma equipe científica multidisciplinar, como será mostrado nesta seção com projetos envolvendo as mais diversas áreas como o sistema solar, exo-planetários, a formação e evolução da Via Láctea, a evolução de estruturas (galáxia e aglomerados de galáxia) no Universo, quasares e cosmologia. O modelo adotado foi o de buscar junto à comunidade pesquisadores de qualidade independente de suas instituições de origem e criar consórcios de pesquisadores como DES-Brazil e os BPGs do SDSS e futuramente do LSST. Estes consórcios foram criados através de anúncios públicos e de recrutamento induzido procurando sempre formar um time qualificado, competitivo e comprometido, ao mesmo tempo balanceando experiência e juventude, tendo em vista a ênfase em preparar uma nova geração de pesquisadores nas áreas de astronomia, física e ciência da computação.

A equipe científica é dividida em grupos de trabalho temáticos coordenados preferencialmente por pós-doutores que se reúnem pelo menos a cada 15 dias para rever as atividades realizadas pelo grupo. Estas atividades incluem: **i)** a participação nas teleconferências internacionais; **ii)** o desenvolvimento (aprimoramento) e testes de algoritmos de análise; **iii)** a leitura de material pertinente; **iv)** o acompanhamento da construção, integração e testes de *pipelines* de interesse do grupo no Portal Científico; **v)** a participação na validação dos dados para as atividades do grupo; **vi)** a indicação de nomes para os *webinars*; **vii)** a indicação de nomes para a participação em reuniões nacionais e internacionais; **viii)** a validação de apresentações feitas em nome do consórcio, seguindo as normas de cada colaboração internacional; **ix)** redação de artigos.

Esta filosofia de trabalho representa uma importante mudança em relação à forma tradicional do pesquisador brasileiro trabalhar incentivando a colaboração. Isto contribui para aumentar a eficiência da pesquisa e treinar os jovens pesquisadores a atuarem em projetos internacionais de forma construtiva mas altamente competitiva.

Fazendo parte do LIneA, os alunos e pós-doutorandos tem ainda a oportunidade de participar de todas as fases de um projeto científico como, por exemplo, as considerações envolvidas na construção de um instrumento, das observações, do desenvolvimento de ferramentas de análise, da extração da ciência, da preparação de apresentações internacionais, e da redação de artigos sob o escrutínio não só de seu supervisor mas de toda uma comunidade de *experts* nas áreas consideradas. Tudo isto ajuda na formação de profissionais mais completos do que aqueles formados em isolamento.

Em resumo, o LIneA oferece um ambiente fértil para a formação de pesquisadores voltados a ciência de grandes levantamentos astronômicos, de grande impacto científico, e a ciência de dados que irá dominar o futuro de diversas áreas de pesquisa. Tudo isto inserido desde cedo no cenário internacional. Este modelo, dirigido pela ciência, é bem diferente do modelo tradicionalmente adotado pela astronomia brasileira voltada para o

aluguel de tempo de telescópio para uso individual ou na operação de telescópios e no desenvolvimento instrumental. Ao contrário, o modelo adotado pelo LIneA dá ênfase ao desenvolvimento de software onde países como o Brasil podem competir em condições de igualdade como já demonstrado pelo trabalho realizado pelo LIneA, particularmente para o DES.

No processo, esses jovens pesquisadores são também expostos à prática de gestão, participação em teleconferências e apresentações públicas para diferentes audiências. Isto produz um amadurecimento sem paralelo aos estudantes e aos jovens pesquisadores brasileiros. Também importante é a experiência de trabalhar em colaboração, tanto ao se beneficiar do trabalho da equipe como contribuindo para objetivos comuns. Isto é possível graças à infraestrutura do LIneA, que permite o uso de um repositório de códigos e a integração de *workflows* científicos validados no portal, preservando desta forma o conhecimento adquirido. Este legado pode então ser capitalizado por novos membros que podem dar continuidade ao trabalho sobre uma base já existente. Tudo isto é uma mudança de paradigma na formação de profissionais, onde a comunidade brasileira não está plenamente preparada para atender este novo tipo de demanda, e o LIneA se propõe a preencher esta lacuna.

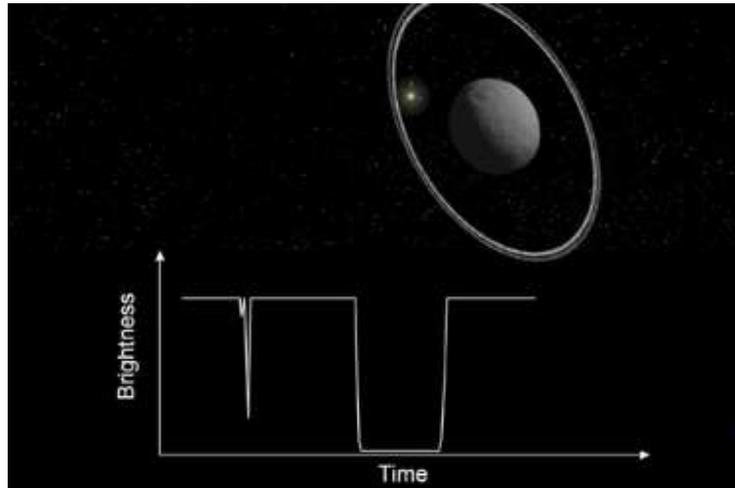
Como mencionado anteriormente os pesquisadores do LIneA se dividem em grupos de trabalho que se reúnem periodicamente e coordenam o trabalho sendo feito pela equipe brasileira reforçando sua inserção nos grupos de trabalho internacionais. Esta seção apresenta alguns destaques resultantes do trabalho alcançados por estes grupos como reportado pelos seus líderes.

## 7.1 Principais destaques

### 7.1.1 Estudos do Sistema Solar (J. Camargo, ON)

Um primeiro destaque é a recente entrada do grupo *Transneptunian Occultation Network* (TON) no LIneA como discutido na seção 4. Recentemente, os resultados deste grupo tiveram uma enorme repercussão com a descoberta de um anel entorno de um asteroide (Figura 7.1) demonstrando o grande potencial desta técnica para o estudo da estrutura e atmosfera destes objetos. O sucesso foi acompanhado do reconhecimento da necessidade de implantar uma infraestrutura mais sólida e eficiente para estes estudos, motivando a entrada do grupo no LIneA.

Recentemente, com o apoio do LIneA e do grupo DES-Brazil, membros deste grupo foram aceitos como colaboradores externos do DES (apêndice 7.1). Isto permitirá ao grupo dar um passo ao mesmo tempo ambicioso e necessário: conseguir estudar TNOs de brilho muito tênue, aumentando consideravelmente o número de eventos de ocultação, bem como obter o *know-how* de trabalho num ambiente de grandes volumes de dados. Este último, em particular, é requisito para uma participação bem sucedida no LSST.



**Figura 7.1** - Descoberta feita em 2013 de um sistema de anéis ao redor de Chariklo. Trata-se do primeiro corpo do sistema solar, além dos planetas gigantes (Júpiter, Saturno, Urano, Netuno) no qual se detecta claramente tal estrutura. Chariklo pertence a uma classe de objetos denominada Centauro (e não TNO). Centauros, no entanto, são “primos” dos TNOs.

### 7.1.2 Infraestrutura Científica (com contribuições de J. Gschwend, aluna de doutorado ON; M. Maia, ON)

Uma das principais tarefas dos times científicos e de TI, discutida abaixo tem sido o desenvolvimento de um portal científico para ser usado na análise de dados simplificando o acesso aos dados e aos *pipelines* científicos por parte do usuário. Isto é necessário tendo em vista a complexidade dos processos envolvidos na preparação dos dados para análise e a crescente quantidade de dados envolvida. A preparação de catálogos prontos para serem usados em diferentes análises científicas requer um time científico e técnico trabalhando em conjunto, e este tem sido um dos focos do time de TI do LIneA nos últimos anos.

Para a preparação destes catálogos é necessário remover regiões afetadas pela presença de objetos brilhantes, levar em consideração as variações espaciais na profundidade dos dados devido a variações nas condições observacionais, classificar as fontes detectadas como estrelas ou galáxias e no caso de um levantamento fotométrico como DES e como será o LSST, estimar o *redshift* das galáxias baseado apenas no fluxo medido em cinco ou seis filtros de banda larga.

Particularmente importante é anexar estimativas de *photo-z's* aos catálogos produzidos, o que é essencial para a maioria das aplicações. Isto é feito no portal por uma série de *pipelines* encadeados de forma a assegurar consistência dos resultados e manter a proveniência dos dados e das escolhas feitas durante o processo. Como pré-requisito deste processo, o LIneA mantém um repositório atualizado de *redshifts* de amostras provenientes de diferentes levantamentos espectroscópicos concluídos ou em andamento

(e.g., OzDES). Em particular, nos últimos dois anos, membros do LIneA vem contribuindo ativamente no trabalho de validar *redshift* obtidos automaticamente pelo levantamento OzDES, utilizando o software MARZ (figura 7.2). Em breve, além dos *redshifts*, os espectros obtidos neste levantamento também estarão acessíveis pelo portal e a ferramenta acima mencionada incorporada a interface *Target Viewer* discutida abaixo, o que permitirá examinar o espectro e a imagem de um objeto simultaneamente, um aspecto importante no processo de validação de *redshifts*.



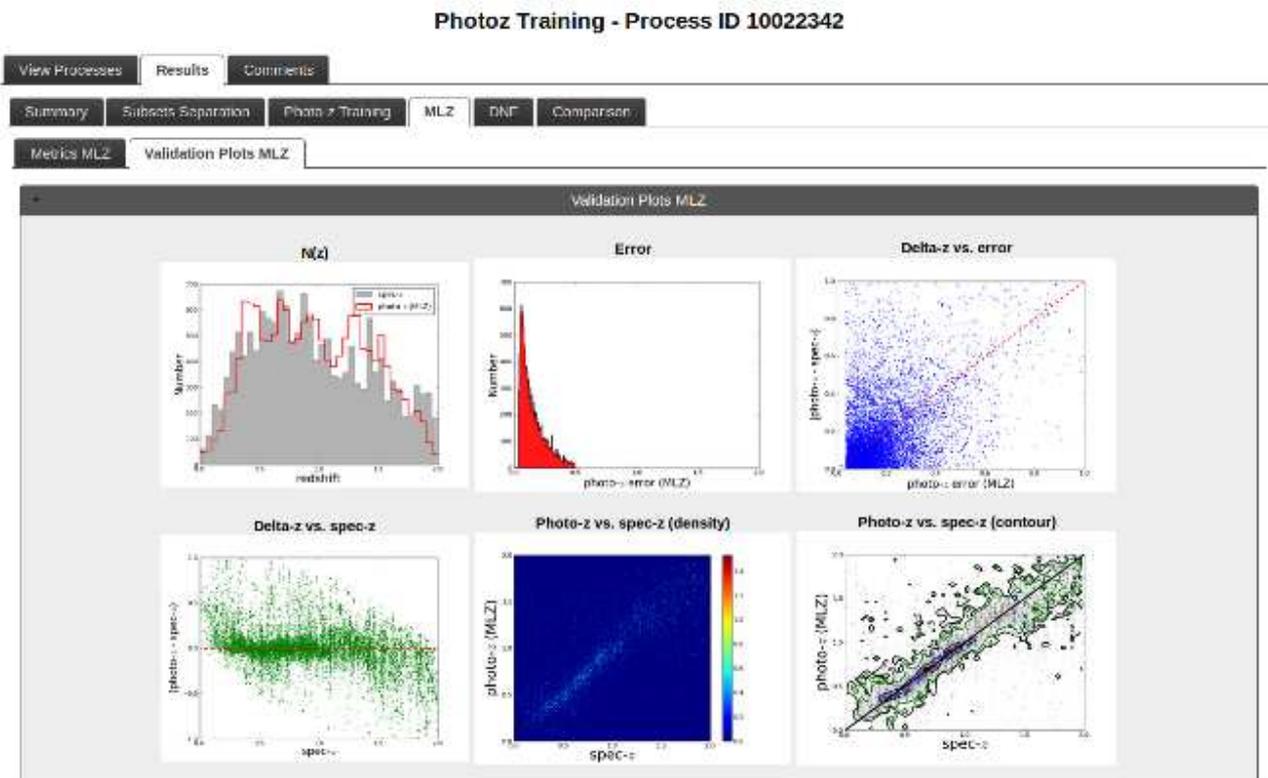
**Figura 7.2** - Captura de tela do software MARZ mostrando o espectro observado (preto) e o melhor template usado para uma estimativa do redshift do objeto considerado. A ferramenta permite ao usuário verificar o redshift vindo de um pipeline automático e atribuir o grau de confiança na medida. Esta interface está prevista de ser integrada ao portal a pedido da colaboração OzDES.

O primeiro *pipeline* a ser executado tem como objetivo consolidar as medidas de *redshift* espectroscópicos vindos de diferentes levantamentos numa única amostra levando em consideração a qualidade das medidas e atribuindo um único valor para cada objeto em caso de múltiplas observações. Atualmente, o repositório tem mais de 1,3 milhões de medidas de *redshift* provenientes de 32 levantamentos. Cada levantamento corresponde a uma tabela no banco de dados do DES mantido pelo LIneA, criada pelo processo de *upload* do portal que permite homogeneizar a informação e facilitar a construção da amostra de *redshift* unificada usada pelos *pipelines* subsequentes.

O segundo estágio da produção de *photo-z's* é a construção dos conjuntos de treinamento, feita pelo *pipeline Training Set Maker*. Estes conjuntos serão responsáveis por preparar os algoritmos empíricos que utilizam técnicas conhecidas como *neural network* e/ou *machine learning* para o posterior cálculo de *photo-z's* nas amostras fotométricas do DES. Este *pipeline* consiste em fazer a correspondência entre as

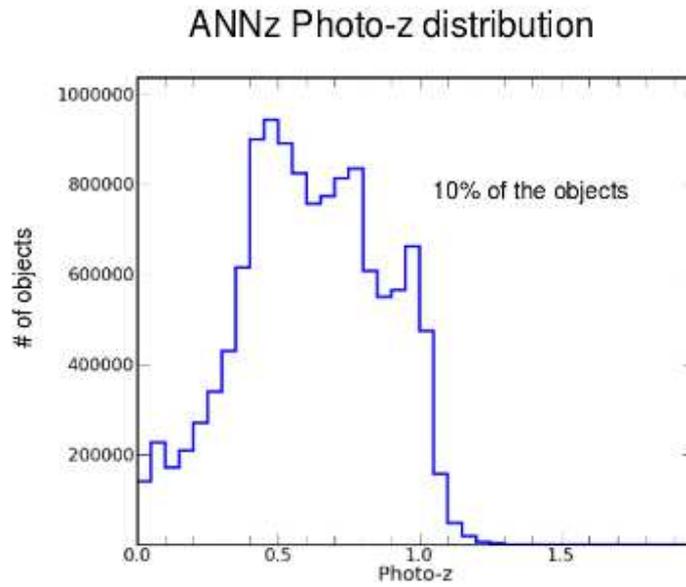
coordenadas espaciais dos objetos nos dois conjuntos, espectroscópico e fotométrico, criando assim um conjunto que contém informação combinada.

O terceiro estágio consiste da execução do *pipeline Photo-z Training*, que realiza o treinamento dos códigos de *photo-z*, gerando, os chamados *training files* - arquivos gerados pelos diversos algoritmos, contendo toda a informação necessária para calcular os *photo-z*'s em uma amostra fotométrica, cada um com suas características e peculiaridades. Até o momento, foram desenvolvidos oito códigos de *photo-z* neste *pipeline*. Como parte deste *pipeline* a amostra criada pelo *training set maker* combinando os dados espectroscópicos com os dados fotométricos do DES é dividida em dois com metade usada para o treinamento dos algoritmos e a outra para uma validação. Também neste estágio, é feita a validação em uma fração do conjunto de treinamento, a fim de acessar a qualidade dos *photo-z*'s que serão produzidos e garantir o cumprimento dos requisitos científicos impostos pela colaboração DES. Nesta etapa também é possível fazer a comparação entre os desempenhos dos diversos métodos. A Figura 7.3 apresenta os resultados de uma rodada deste *pipeline* mostrando os resultados da validação de um treinamento do código MLZ sendo utilizado pela colaboração DES.



**Figura 7.3** - Captura de tela do pipeline *Photo-z Training* mostrando os resultados da comparação dos *photo-z*s calculados partir dos dados fotométricos com os redshifts espectroscópicos disponíveis.

No quarto e último estágio, composto pelo *pipeline Photo-z Compute*, os *photo-z*'s dos objetos observados pelo DES são finalmente calculados, a partir da aplicação dos *training files* gerados na etapa anterior. A Figura 7.4 mostra, como exemplo, a distribuição de *photo-z*s calculados pelo algoritmo ANNz na região conhecida como STRIPE 82, utilizando os dados do *release* Y1A1.

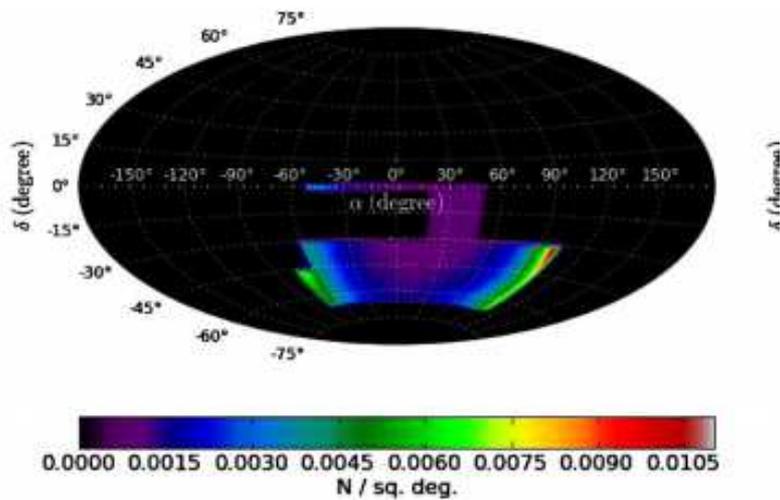


**Figura 7.4** - Distribuição de *photo-z*s obtidos pelo *pipeline Photo-z Compute* para a região Stripe 82 utilizando o algoritmo ANNz. Os dados fotométricos são dados produzidos pelo NCSA para o primeiro ano de observações do DES.

O grupo trabalha agora num artigo cujo objetivo é demonstrar a utilidade do portal para este tipo de trabalho e gerar *photo-z*s produzidos por diferentes algoritmos para a colaboração DES.

Este trabalho está sendo complementado por uma infraestrutura bastante complexa visando à produção de catálogos *science-ready* para diferentes aplicações científicas, como será discutido na seção 9. Como resultado deste desenvolvimento, membros do DES-Brazil contribuíram para elaboração de diversos artigos analisando dados do DES, entre eles: Sánchez, C. et al. **2014**, [MNRAS](#), 445, 482; Banerji, M. et al. **2015**, [MNRAS](#), 446, 2523; Yuan, F. et al. **2015**, [MNRAS](#), 452, 3047.

Outra importante contribuição de infraestrutura feita pelos participantes do DES-Brazil com o apoio do time de TI do LInEA foi o desenvolvimento de simulações para a distribuição de estrelas a serem observadas pelo DES. A Figura 7.5, por exemplo, ilustra o que se deve esperar para densidade superficial de estrelas do tipo O e A levando em consideração a cobertura espacial do DES e os limites de magnitude do levantamento. Estas simulações vão ser usadas para determinar os parâmetros característicos que descrevem a estrutura da Via Láctea combinando os dados do DES e do SDSS. Também importante é que tudo que está sendo feito para o DES poderá ser reutilizado para o LSST, incorporando as lições aprendidas nesta fase.



**Figura 7.5** - Distribuição espacial de estrelas simuladas na área sendo observada pelo DES.

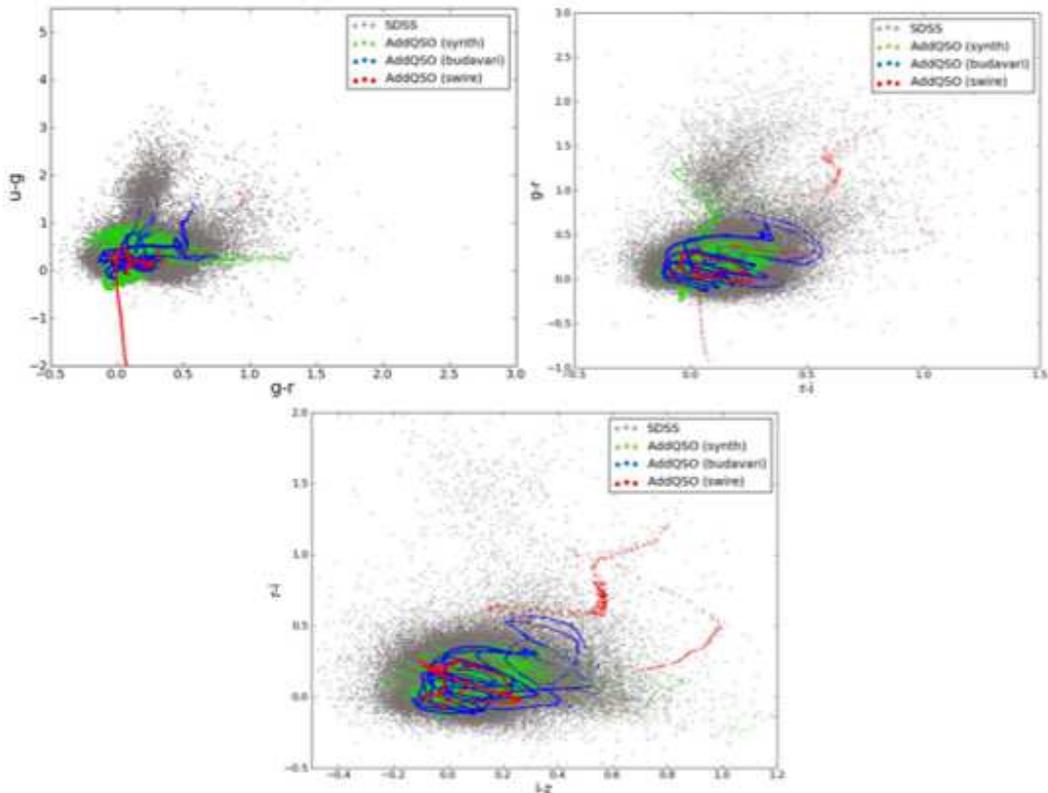
Um grande esforço foi feito para simular amostras de quasares que devem ser observados pelo DES, fazendo previsões sobre suas distribuições numéricas, de cor e de *redshift*. Com a geração de catálogos simulados, ferramentas de análise para a seleção de QSOs e cálculo de *redshifts fotométricos* (photo-z) foram testadas antes de suas aplicações em dados reais. Para realizar estas simulações foi desenvolvida uma suíte de algoritmos chamada AddQSO integrada ao Portal Científico do DES.

O AddQSO integra as funções de luminosidade de quasares de Richards et al. (2006) para simular objetos de  $z < 5$ , 25 e de Willott et al. (2010) para o restante da faixa de  $z$ , o que fornece, além da distribuição em  $z$  dos quasares, a distribuição em magnitude absoluta na banda  $i$  e, conseqüentemente, a distribuição de magnitude aparente nessa mesma banda. Através da convolução entre distribuições espectrais de energia e as curvas de transmissão dos filtros, as magnitudes aparentes nas outras bandas são simuladas. Erros fotométricos são acrescentados seguindo as distribuições desse parâmetro em uma simulação de estrelas realizada pela colaboração DES. A suíte foi validada através de uma comparação com dados do *Sloan Digital Sky Survey*, que mostra conformidade na distribuição de cores, evolução de cores com *redshift*, e número de quasares gerados.

Destas comparações foi concluído que o AddQSO é uma ferramenta robusta para simular catálogos de quasares. Para o levantamento DES, está previsto a detecção de  $\sim 10^7$  quasares em pelo menos uma de suas cinco bandas, sendo quase 1000 objetos com  $z > 6,0$  e 40 com  $z > 7,0$ .

O código ANNz apresenta resultados satisfatórios para o cálculo de *photo-z* e não parece ser afetado drasticamente pela ausência da banda  $u$  como acontece com o algoritmo LePhare. Finalmente, é importante enfatizar que apenas uma pequena fração dos quasares pode ser separada através da aplicação de limites de cor e que a seleção de

objetos apenas pelo método de ajuste de modelos não é suficiente para criar uma amostra de quasares devido à grande taxa de contaminação por outros tipos de objetos. Este estudo de QSOs resultou na dissertação de mestrado de Fernanda Ostrovski (ver exemplo de simulação na Figura 7.6), e na contribuição para o artigo de Davies, L.J.M. et al. **2013**, [MNRAS](#), 434, 296. Novamente, este ferramental continua disponível no ambiente do portal e pode ser utilizado no futuro por interessados na ciência de quasares.



**Figura 7.6** - Comparação entre os diagramas cor-cor dos quasares do catálogo do QDR7 (cinza) e dos quasares simulados pelo AddQSO. As diferentes bibliotecas de SEDs estão codificadas pelas cores indicadas nos painéis.

### 7.1.3 Populações Estelares e Estrutura Galáctica (B. Santiago, UFRGS)

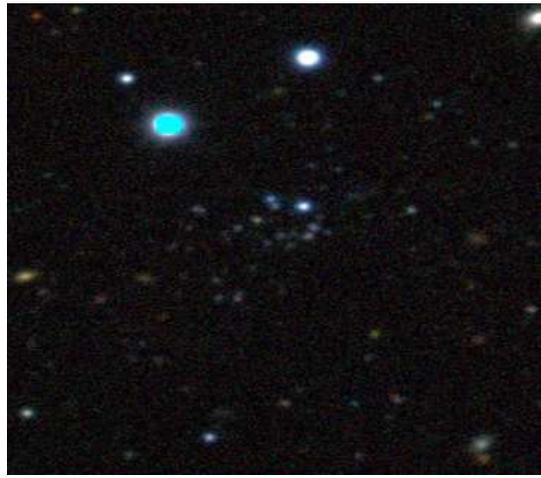
Os principais destaques nos últimos cinco anos na área são:

- Primeiro artigo do DES: foi liderado por pesquisadores afiliados ao LInEA do grupo DES-Brazil, o primeiro artigo oficial da colaboração DES (Rossetto, B. et al. **2011**, [AJ](#), 141, 185). O trabalho faz prognósticos sobre a amostra estelar do DES e a ciência em potencial que poder ser feita com esta amostra. Essas previsões fizeram uso de

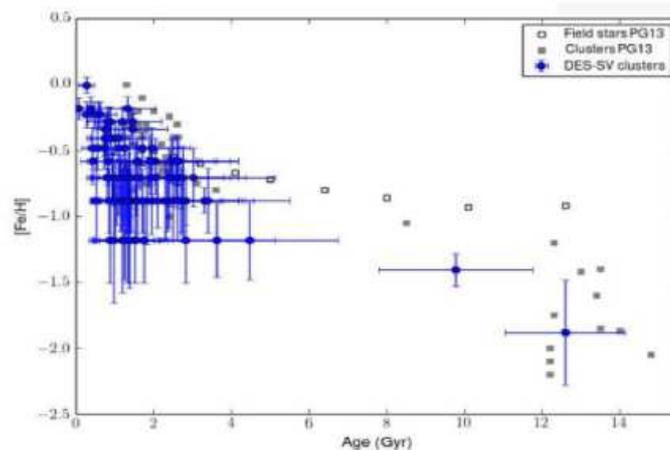
extensas simulações de amostras estelares Galácticas realizadas com a infraestrutura computacional do LIneA.

- Descoberta de satélites da Via Láctea (MW): o grupo de Via Láctea internacional do DES, liderado desde 2008 por um pesquisador do DES-Brazil, identificou 17 novos sistemas estelares do halo da Galáxia, entre aglomerados estelares e galáxias anãs. Um desses sistemas é apresentado na Figura 7.7. O esforço contou com forte participação de pesquisadores e estudantes do LIneA, que fazem parte do consórcio DES-Brazil (Bechtol, K. et al. **2015**, [ApJ](#), 807, 50; Drlica-Wagner, A. et al. **2015**, [ApJ](#), 813, 109; Luque, E. et al. 2016, submetido ao MNRAS). Os resultados com o DES foram precedidos por um desenvolvimento de técnicas variadas de detecção e caracterização desses sistemas (Balbinot, E. et al. **2011**, [MNRAS](#), 416, 393; Balbinot, E. et al. **2013**, [ApJ](#), 767, 101), que serão incorporadas ao portal.
- Descoberta de correntes estelares da MW: além de satélites da Galáxia, os dados do DES revelaram a existência de nuvens e correntes estelares no halo, as quais permitem impor melhores vínculos ao potencial gravitacional e à massa da MW. Esses trabalhos tiveram também protagonismo de membros do DES-Brazil (Li et al. 2015, [ApJ](#), aprovado para publicação; Balbinot et al. 2016, submetido ao [ApJ](#)).
- Análise quimio-dinâmica do disco da MW: um dos projetos do SDSS-III, o APOGEE, permitiu analisar a distribuição de velocidades e abundâncias químicas, bem como suas variações espaciais, de estrelas do disco da Galáxia. Um dos primeiros artigos sobre o tema (Anders, F. et al. **2014**, [A&A](#), 564, 115) foi liderado por pesquisadores do LIneA, usando distâncias espectrofotométricas determinadas pelo grupo de participação brasileira no SDSS-III (BPG; Santiago B.X. et al. **2016**, [A&A](#), 585, 42).
- Descoberta de sistemas binários com secundárias de baixa massa: o projeto SDSSIII/MARVELS foi responsável pela identificação de diversos sistemas binários vizinhos ao Sol, cujas estrelas secundárias tinham massa provável inferior ao limite de queima de hidrogênio. Os pesquisadores, pós-doutorandos e estudantes associados ao BPG do SDSS-III contribuíram para esses trabalhos através da determinação de parâmetros estelares e obtenção de espectros complementares aos do Marvels. Algumas publicações sobre o tema foram: Lee, Brian L. et al. **2011**, [ApJ](#), 728, 32; Schuler, S. et al. **2011**, [ApJ](#), 737, L32; Fleming, S.W. et al. **2012**, [AJ](#), 144, 72; Ma, B. et al. **2013**, [AJ](#), 145, 20; Jiang, P. et al. **2013**, [AJ](#), 146, 65; De Lee, N. et al. **2013**, [AJ](#), 145, 155; Ghezzi, L. et al. **2014**, [AJ](#), 148, 105.
- Análise das regiões externas da Grande Nuvem de Magalhães: Em trabalhos protagonizados por alunos de doutorado do DES-Brazil, as regiões externas da Grande Nuvem de Magalhães (LMC) foram estudadas, levando a vínculos sobre a estrutura da LMC e sobre seu sistema de aglomerados estelares como mostra a Figura 7.8 (Balbinot, E. et al. **2015**, [MNRAS](#), 449, 1129; Pieres, A. et al. 2016, submetido ao MNRAS).

Em resumo, através do DES, e com considerável contribuição de afiliados ao LIneA, é possível afirmar que nossa visão do halo Galáctico no hemisfério sul mudou radicalmente, tendo sido identificadas subestruturas que irão servir para aperfeiçoar os modelos dinâmicos da Galáxia. Além disso, o projeto APOGEE, bem como os projetos SEGUE e MARVELS, permitiram melhor caracterizar as propriedades de estrelas nos ~5 kpc em torno da posição solar, sendo que o APOGEE está sendo estendido para o Hemisfério Sul, também com participação de um BPG associado ao LIneA.



**Figura 7.7** - Imagem de DES 1, aglomerado estelar descoberto por Luque et al. (2016) utilizando dados do DES. A imagem foi obtida com uso da ferramenta Tile Viewer produzida pelo grupo DES-Brazil para o portal científico da colaboração.

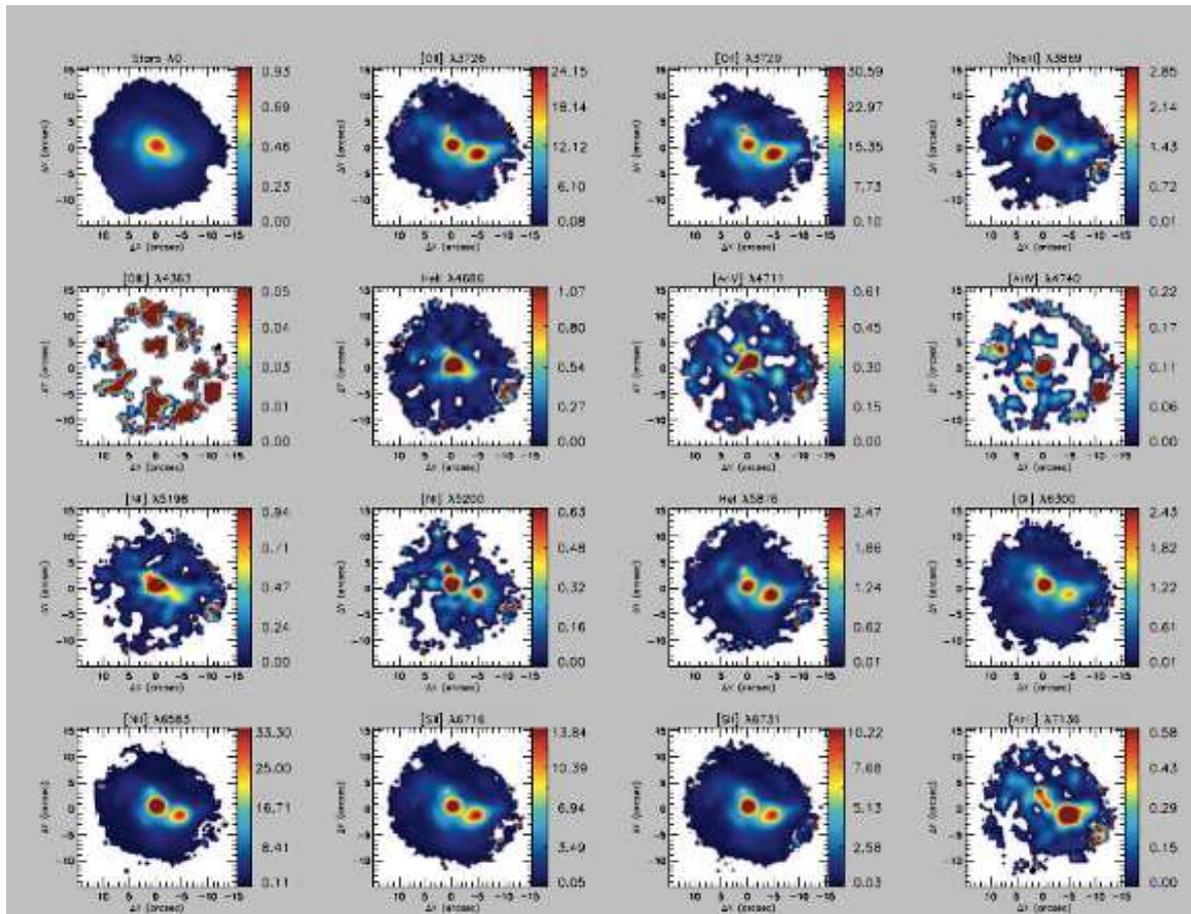


**Figura 7.8** - Relação idade-metalicidade inferida para uma amostra de 116 aglomerados estelares das regiões externas da LMC (pontos azuis). São também mostrados dados de outras amostras de campo e de aglomerados. A amostra de aglomerados foi identificada visualmente usando o portal científico do DES, criado pelo LIneA. Figura tirada de Pieres et al. (2016).

### 7.1.4 Evolução de Galáxias (T. Storchi-Bergman, UFRGS)

Durante o ano de 2015, os pesquisadores trabalhando no levantamento MaNGA dedicaram parte do tempo no aprendizado do manuseio dos cubos de dados, disponibilizados através do portal do LInEA. Vencida esta etapa, foram obtidos os seguintes resultados:

- Sob a liderança de Rogemar Riffel, foram feitas as primeiras medidas de fluxos nas linhas de emissão, com a confecção de mapas de fluxo e mapas da cinemática do gás emissor em uma pequena amostra de galáxias ativas, como ilustrado na Figura 7.9.



**Figura 7.9** - Mapas de fluxo em escala logarítmica na banda do contínuo (1o. painel superior esquerdo, mapeando contribuição das estrelas) e nas diferentes linhas de emissão identificadas acima de cada painel de uma galáxia ativa da amostra do MaNGA.

- Rogério Riffel e o estudante de mestrado Nicolas Mallman desenvolveram uma ferramenta (MEGACUBE) que permite a aplicação do código STARLIGHT (Cid Fernandes, 2005) aos cubos de dados do MaNGA e posterior anexação dos resultados dos ajustes ao Cubo original. O código já foi utilizado com sucesso em uma pequena amostra (10 fontes) e será aplicado a toda amostra do MaNGA utilizando o Cluster de computadores do LInEA.

- O pós-doutorando Jáderson Schimoia desenvolveu uma rotina para o cálculo da razão sinal/ruído S/N spaxel a spaxel e a construção de uma máscara que delimita no campo de observação a região espacial em que a razão S/N é alta o suficiente ( $\sim 10$ ) para que o processo de síntese espectral dê resultados confiáveis e robustos. Tal função já foi integrada ao MEGACUBE descrito acima.
- Sob a liderança de Sandro Rembold, foram selecionados os cubos de todas as galáxias ativas já observadas até a presente data, e, para cada uma foram selecionadas duas galáxias não ativas de controle, como ilustrado na Figura 7.10. Os principais critérios utilizados para selecionar as galáxias não ativas da amostra de controle foram o mesmo *redshift* (distância) e a mesma massa estelar da galáxia ativa, seguidos da morfologia da galáxia e de sua inclinação. Estamos no momento trabalhando na redação do primeiro artigo científico que trata da caracterização da amostra de galáxias ativas e da amostra de controle.

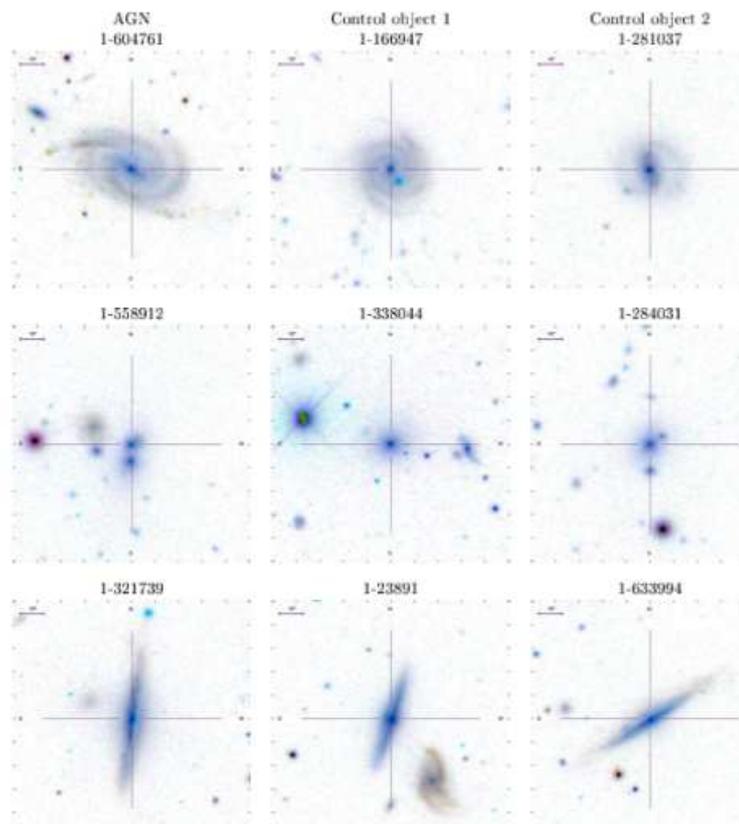
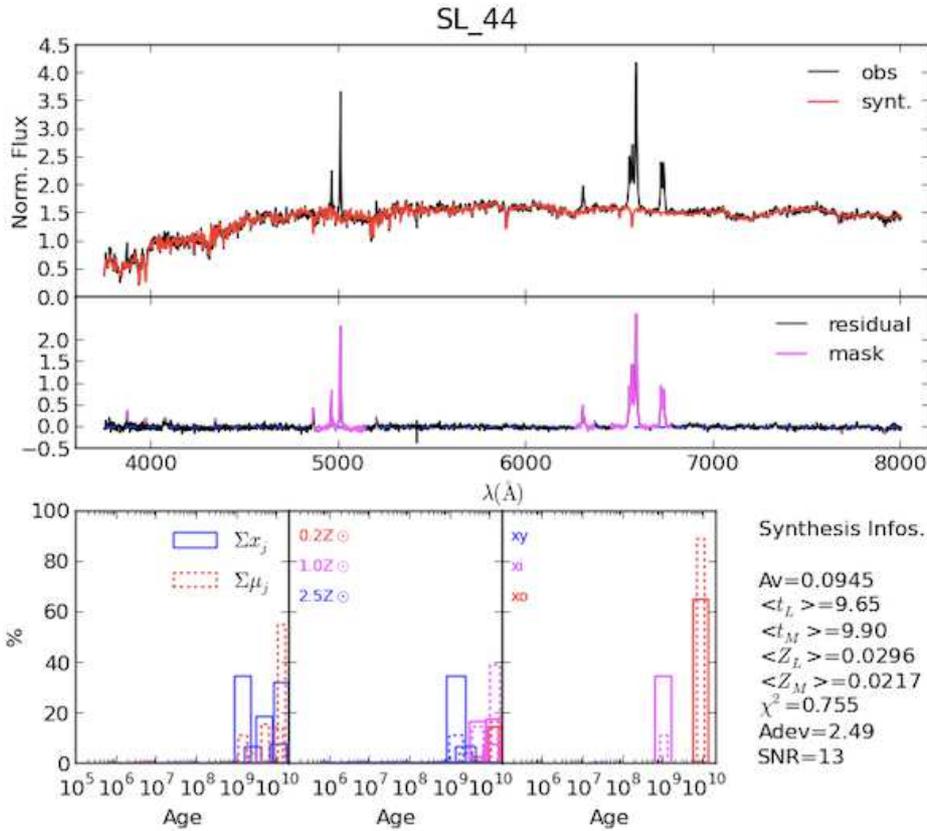


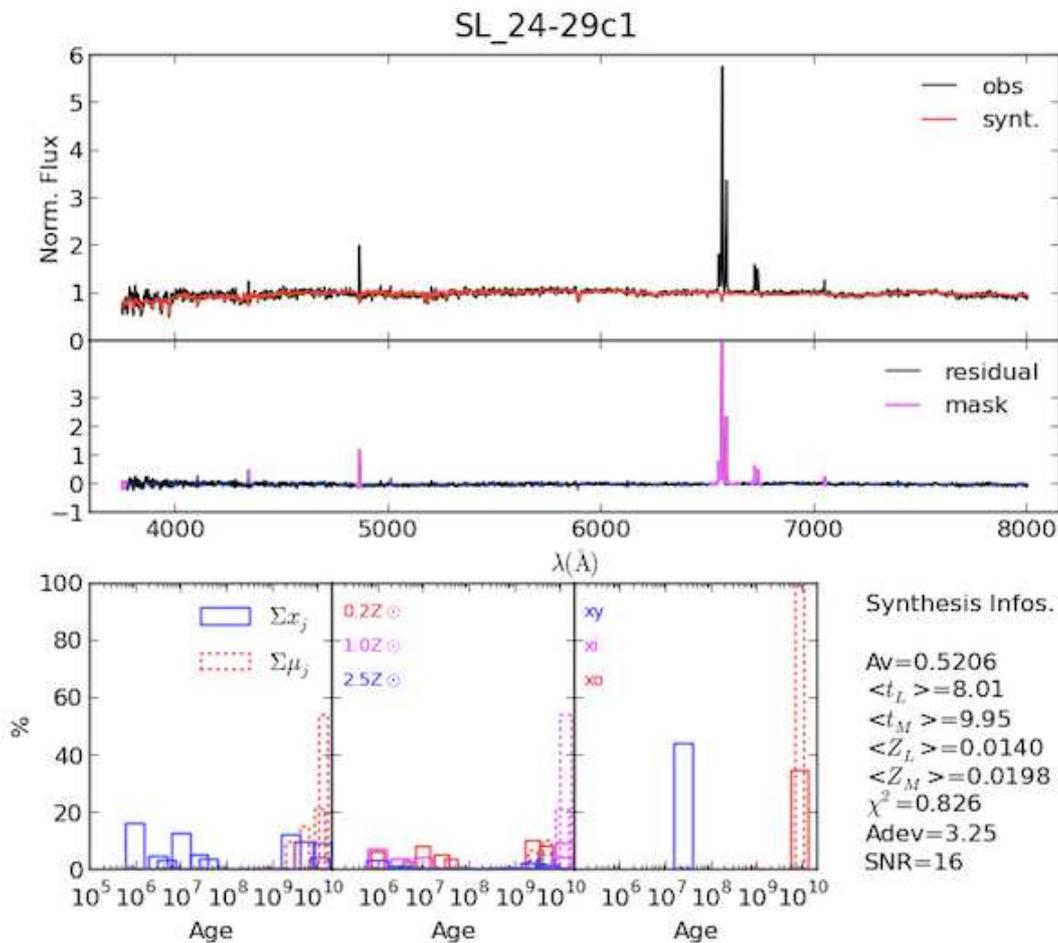
Figura 7.10 - Imagens (do SDSS-III) de 3 galáxias ativas (AGN) e duas galáxias não ativas da amostra de controle segundo o critério de terem a mesma massa estelar, distância, morfologia e inclinação que a galáxia ativa.

- Sob a liderança do pós-doutorando Jáderson Shimoia e supervisão de Thaisa Storchi Bergmann e Rogério Riffel, a aluna de doutorado Janaína Nascimento realizou síntese espectral utilizando o programa *Starlight* dos espectros de regiões nucleares

integradas, para todas as galáxias da amostra (ativas + controle) disponíveis no SDSS-III. Estes resultados serão utilizados para complementar a caracterização da amostra descrita acima e servirão de referência para o estudo da distribuição espacial das características da população estelar a partir dos cubos de dados. Os dados resultantes da síntese foram analisados com o código *Pylight*. Exemplos de resultados da síntese para duas galáxias da amostra observada no projeto MaNGA, uma *Seyfert 2* e outra uma *Starburst* são apresentados nas Figuras 7.11 e 7.12.



**Figura 7.11:** Resultado da síntese espectral (através do programa *Starlight*) do espectro SDSS-III de uma galáxia Seyfert 2 da amostra. Painéis de cima: espectro observado em preto, espectro sintético em vermelho; logo abaixo os resíduos da síntese aparecem em preto e as linhas de emissão mascaradas no processo de síntese espectral aparecem em cor magenta. Painéis de baixo mostram a contribuição percentual de populações de diferentes idades: (painel da esquerda) ao fluxo total em azul e à massa total em vermelho; (painel do centro) distribuição de metalicidades; (painel da direita) agrupamento da contribuição em 3 populações, cada uma cobrindo um intervalo de idades: jovem ( $\sim 10^6 - 10^8$  anos), de idade intermediária ( $10^8 - 10^9$  anos) e velha ( $\sim 10^9$  anos).



**Figura 7.12-** Semelhante à figura anterior mas para o resultado da síntese espectral do espectro SDSS-III de uma galáxia Starburst da amostra.

O grupo realiza reuniões semanais, para acompanhamento das diferentes tarefas em desenvolvimento acima citadas.

O LIneA é essencial para o projeto pelos seguintes fatores:

- Possui repositório dos dados obtidos pelo projeto, facilitando e tornando acessível e rápido o *download* por parte de todos os membros do grupo. Além disso, o banco de dados do LIneA será também o repositório dos resultados de nossas análises (em particular, de nossos Megacubos, como discutido abaixo).
- O Cluster de Computadores do LIneA permitirá um processamento muito mais ágil dos milhares de espectros com que necessitamos trabalhar, em particular, para a realização da síntese espectral spaxel a spaxel. Cada espectro deve ser processado milhares de vezes para a realização da síntese através da variação de um grande número de parâmetros e condições. E como muitas vezes a redução de dados é atualizada ao longo do projeto, é comum ser necessário repetir todo o processamento

várias vezes. Só com um Cluster como o do LIneA será possível fazer um trabalho competitivo nestas condições.

### **7.1.5 Estrutura em Grande Escala e Cosmologia (com contribuições de M. Lima, IF-USP; A. Carnero, pós-doutorando ON; R. Rosenfeld, IFT-UNESP)**

O grupo do LIneA participou do BOSS via o BPG desde 2008 até 2015. Os membros do BOSS no BPG foram Luiz da Costa, Aurelio Carnero, Flavia Sobreira, Fernando de Simoni, Marcio Maia, e Nicolás Busca. A principal contribuição do BPG foram:

- Dentro do grupo de trabalho sobre aglomeração de galáxias, a participação foi principalmente nas medidas de não-gaussianidades primordiais e de dependências da taxa de crescimento de perturbações usando amostras selecionadas através de cores das galáxias. Em ambos os casos, não foi encontrado sinal significativo com os dados do levantamento BOSS.
- Dentro do levantamento da floresta de *Lyman Alpha*, a contribuição principal foi na medida das oscilações acústicas de bárions (BAO) na amostra de quasares do DR11. Mais de uma dezena de artigos foram produzidos pelo grupo, entre eles: White, M. et al. **2011**, [ApJ](#), 728, 126; Ross, A. et al. **2011**, [MNRAS](#), 417, 1350; Sobreira, F. et al. **2011**, [PRD](#), 84, 10300; Ahn, C.P. et al. **2012**, [ApJS](#), 203, 21; Ho, S. et al. **2012**, [ApJ](#), 761, 14; Anderson, L. et al. **2012**, [MNRAS](#), 427, 3435; Dawson, K.S. et al. **2013**, [AJ](#), 145, 10; Ross, A. et al. **2013**, [MNRAS](#), 428, 1116; Mack, Claude E., III et al. **2013**, [AJ](#), 145, 139; Guo, H. et al. **2013**, [ApJ](#), 767, 122; Zasowski, G. et al. **2013**, [AJ](#), 146, 81; Simoni, F. et al. **2013**, [MNRAS](#), 435, 3017; Gontcho A Gontcho, Satya. et al. **2014**, [MNRAS](#), 442, 187; Delubac, T. et al. **2015**, [A&A](#), 574; Bautista, J. et al. **2015**, [JCAP](#), 5, 60; Blomquist, M. et al. **2015**, [JCAP](#), 11, 034; Aubourg, É. et al. **2015**, [PRD](#), 9213516.
- Participação nos artigos descrevendo os *releases* oficiais dos dados do BOSS.

O grupo do LIneA participa também do eBOSS via o BPG, desde o início de 2015 até o presente. Os membros do eBOSS no BPG são Marcos Lima, Luiz da Costa, Aurélio Carnero, Flavia Sobreira, Hugo Camacho, Marcio Maia. Neste curto período, as atividades incluíram:

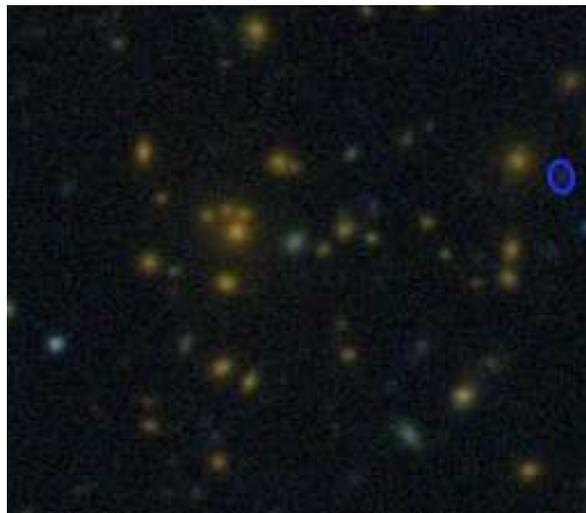
- Contribuição importante dos membros do BPG em 2 trabalhos (Jouvel, S. et al. **2015**, [MNRAS](#), submetido; e Comparat, J. et al. **2015**, [A&A](#), submetido) de interface do DES e eBOSS. Este trabalho foi resultado de um acordo entre o DES e o eBOSS, que compartilharam dados, o que permitiu a realização de dois artigos que investigaram as propriedades de galáxias com linhas em emissão (ELGs), que são traçadores do campo de matéria do universo e permitem importantes estudos da estruturas em grande escala do Universo. O grupo do LIneA trabalhou exatamente na caracterização do nível de aglomeração destas galáxias usando dados do DES e do SDSS. O fato de o LIneA fazer parte tanto do DES como do eBOSS foi fundamental para que os pesquisadores brasileiros pudessem ter uma participação importante nestes trabalhos, que também abriram possibilidades de interação com outros grupos internacionais. Além disso, o este trabalho é visto

como uma preparação para o levantamento DESI que deve começar ao final do eBOSS.

- O BPG está no momento colaborando com Rita Tojeiro em um trabalho usando dados do BOSS/eBOSS para as chamadas galáxias vermelhas luminosas (*Luminous Red Galaxies*, LRGs). O objetivo é caracterizar as propriedades de aglomeração destas galáxias, e vincular os modelos de evolução de galáxias que sejam compatíveis com os resultados obtidos. Esta análise, bem como a do item acima, só está sendo possível (dentro do ambiente internacional competitivo do eBOSS) devido a todas as ferramentas que foram desenvolvidas em trabalhos de Estruturas em Grande Escala feitos anteriormente no BOSS e no DES.
- O BPG foi responsável pela criação de uma máscara de estrelas brilhantes para a análise de LRGs do eBOSS. A primeira versão desta máscara já foi aplicada aos dados do eBOSS, aumentando a eficiência da seleção de LRGs em 2%. No momento, existe um esforço de melhorar o algoritmo para mascarar áreas em torno de estrelas brilhantes, e melhor remover estas áreas que podem afetar a fotometria dos objetos, bem como medindo o efeito das estrelas brilhantes na aglomeração de galáxias, e como corrigir este efeito de forma eficiente.

O grupo de aglomerados de galáxias é formado por Marcos Lima (IF-USP), Christophe Benoist (Nice, França), Ricardo Ogando (ON), Luiz da Costa (ON), Marcio Maia (ON), Vinicius Busti (pós-doutorando, USP), Michel Aguena (doutorando, USP). O grupo atua no DES, e ao longo dos últimos 5 anos, uma série de ferramentas de análise foram desenvolvidas. Estas incluem:

- Desenvolvimento, em fase final, de um identificador de aglomerados de galáxias, chamado WAZP, baseado na identificação de picos na distribuição espacial e de *photo-zs*. Testes preliminares indicam que este método tem o potencial de superar os demais em certos regimes de *redshift* para prover catálogos a serem usados em análises astrofísicas e cosmológicas. A Figura 7.13 mostra um exemplo de aglomerado identificado pelo WAZP.



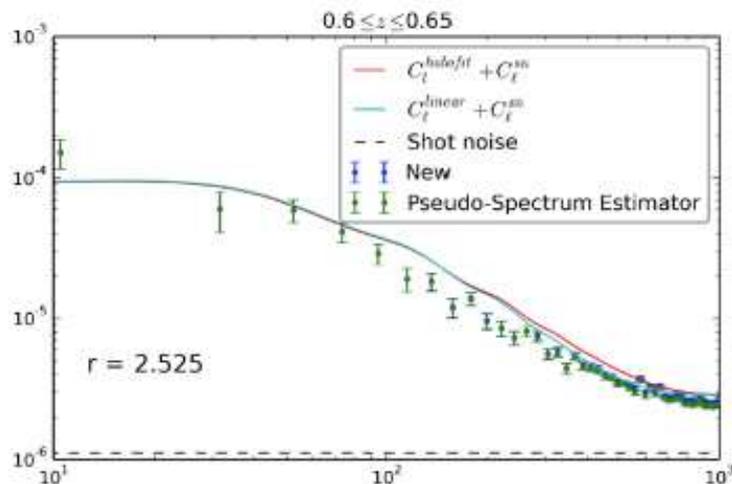
**Figura 7.13** – Exemplo de um aglomerado de galáxias relativamente rico identificado pelo algoritmo WAZP na região conhecida como Stripe 82 com um redshift  $z=0.55$ . A visualização dos sistemas identificados é feita com a ferramenta target viewer parte do sistema de Data Release sendo construído para a colaboração pelo LIneA.

- Desenvolvimento de um *pipeline* para análise dos dados produzidos pelo WAZP, permitindo a caracterização imediata dos catálogos gerados.
- Desenvolvimento de um *pipeline* para avaliação do potencial de previsões (*forecast*) de diferentes experimentos com suas características específicas, e incluindo efeitos de função de seleção de clusters, os quais não têm sido modelados de maneira apropriada na literatura recente.
- Desenvolvimento de um *pipeline* para análise cosmológica dos catálogos produzidos, e a subsequente vinculação de modelos de energia escura, que constituem um dos objetivos finais mais importantes desta análise.
- Aplicação das ferramentas a simulações do DES aos dados de verificação científica (SVA1), bem como aos dados do primeiro ano oficial de observações (Y1A1). Esta análise produzirá ao menos cinco artigos liderados pelo grupo de clusters do DES-Brazil neste contexto. Ela também está conectada ao trabalho de outros grupos de trabalho do LIneA que produzem, a partir dos dados brutos, catálogos científicos (*Value-added Catalogs*, VACS) com *redshifts* fotométricos calculados por diferentes métodos. Esta integração permite que os dados brutos passem por todo o caminho necessário até gerarem produtos científicos e vínculos em modelos específicos, com todas as etapas sendo controladas pelo Portal Científico, de forma independente do restante da colaboração DES, e desta forma podendo servir estes catálogos para outras análises que não as feitas pelos membros do LIneA.

O grupo de trabalho em Estrutura em Grande Escala do Universo participando do DES-Brazil tem desenvolvido diversos estudos que descreveremos a seguir.

- Determinação da distribuição de galáxias em halos de matéria escura - Este é um projeto oficial do DES liderado pela pós-doutoranda Flávia Sobreira (IFT-UNESP). Usando dados de *Science Verification (SV) Survey* um catálogo limitado em volume foi construído e regiões com grandes erros sistemáticos foram excluídas para poder realizar a determinação de parâmetros que caracterizam os chamados modelos HOD (*Halo Occupation Distribution*) que descrevem como galáxias ocupam halos de matéria escura. A participação de Flávia Sobreira em um projeto anterior *Galaxy clustering, photometric redshifts and diagnosis of systematics in the DES Science Verification data*, (Crocce, M. et al. 2016, [MNRAS](#), 455, 4301) permitiu sua liderança nesse novo projeto.

- Limites em Matéria Escura Morna - A presença de matéria escura morna, em contraste com modelos de matéria escura fria, altera a estrutura de halos em pequenas escalas. É possível que a análise do SV possa trazer informações importantes sobre essa característica fundamental da matéria escura. Esta análise está sendo desenvolvida pela doutoranda Jéssica Martins.
- Comparação entre a Função de Correlação Angular (FCA) e o Espectro de Potência Angular (EPA) na determinação de parâmetros cosmológicos - Este projeto, liderado por Flávia Sobreira, trata de um estudo em simulações desenvolvidas para o DES (as simulações MICE), onde há um perfeito controle dos dados, para verificar qual observável entre FCA e EPA tem maior precisão na determinação de parâmetros cosmológicos. A FCA e EPA foram medidas nas simulações e *pipelines* foram criados para a extração de parâmetros cosmológicos e incorporados na plataforma Cosmosis. No projeto o doutorando do IF-USP Hugo Camacho participa ativamente.
- Análise do impacto da inclusão da covariância cruzada entre EPA e o número de aglomerados de galáxias - O DES irá determinar o melhor modelo que descreve o Universo a partir de várias observáveis cosmológicas. A análise estatística necessariamente deve envolver covariâncias cruzadas entre os diferentes observáveis. O pós-doutorando Fabien Lacasa tem liderado um estudo teórico do impacto da inclusão da covariância cruzada entre EPA e o número de aglomerados de galáxias desenvolvendo uma modelagem de HOD e realizando uma análise usando matriz de Fisher. A validação dessa implementação está sendo testada nas simulações MICE em colaboração com Kai Hoffmann de Barcelona, que visitou o DES-Brazil por três meses em 2015. O estudante de iniciação científica Nickolas Kokron e o doutorando Michel Aguená também estão envolvidos no projeto.
- Medidas do EPA no Y1 - O DES recentemente concluiu um catálogo a partir dos dados do primeiro ano de observação (Y1) e um dos chamados *key papers* do DES está relacionado com a medida da função de correlação de 2 pontos. Fabien Lacasa e Nickolas Kokron estão liderando dentro do grupo de trabalho da colaboração internacional a medida do EPA no Y1 (Figura 7.14). Usando a máscara fornecida pela colaboração, um *pipeline* para a medida do EPA está sendo desenvolvido incluindo o efeito da máscara e os erros sistemáticos. O *pipeline* está sendo testado com simulações gaussianas.

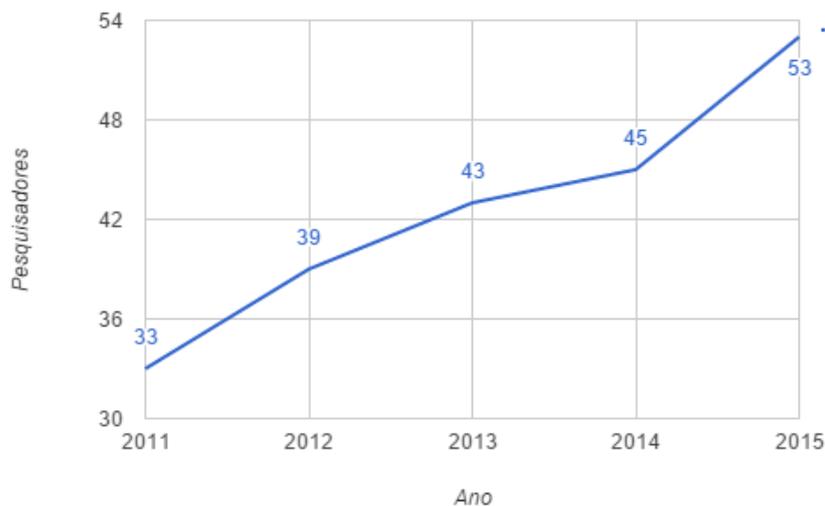


**Figura 7.14** - Medida preliminar do EPA das galáxias vermelhas do catálogo do Y1 no intervalo de redshift  $0.6 < z < 0.65$ , comparado com previsões teóricas.

Todos estes algoritmos de análise sendo desenvolvidos deverão ser integrados ao Portal Científico e fazer parte do legado do projeto DES que poderá servir de base para o projeto do LSST.

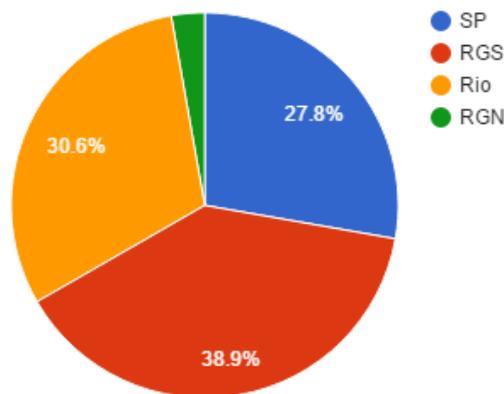
## 7.2 O time científico

A principal preocupação na formação dos times de pesquisadores foi o de encontrar pessoas qualificadas que pudessem aproveitar das oportunidades científicas oferecidas pelos projetos apoiados e o comprometimento de dedicarem uma fração de seu tempo a eles. Os projetos foram anunciados publicamente e a formação dos times foi um misto daqueles que demonstraram interesse espontâneo ou, que foram induzidos tendo em vista uma avaliação da possível contribuição que poderiam dar. Já participaram do LIneA pesquisadores do Rio de Janeiro (CBPF, ON, UFRJ, UFF), de São Paulo (USP e UNESP) e do Rio Grande do Sul (UFRGS e UFSM). Além desses, foram apoiados pesquisadores brasileiros ligados a instituições internacionais na Alemanha, França e Itália, tendo em vista a contribuição singular que poderiam dar a equipe brasileira. Com o apoio do LIneA, o DES-Brazil promoveu ainda a entrada de oito pesquisadores como colaboradores externos de maneira a complementar a expertise do grupo no estudo de objetos do sistema solar, anãs vermelhas, e modelos para a estrutura da Via Láctea. O número de pesquisadores afiliados ao laboratório ao longo dos últimos cinco anos é apresentado na Figura 7.15.



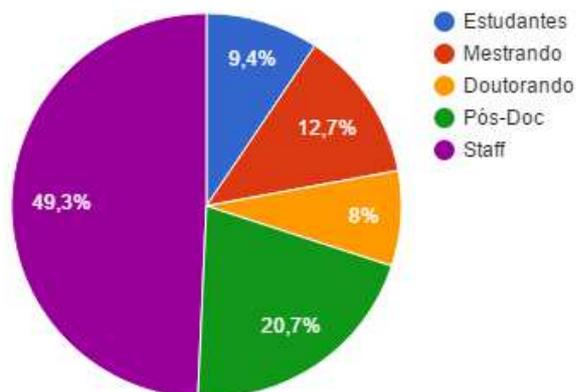
**Figura 7.15** – Evolução do número de pesquisadores associados ao LIneA ao longo dos últimos cinco anos. A figura mostra o gradual aumento de pesquisadores envolvidos, com um crescimento de 60% nos últimos cinco anos.

A Figura 7.16 mostra a distribuição por estados da equipe científica do LIneA. Um ponto preocupante é a pequena participação de centros emergentes, que provavelmente mais se beneficiariam com o acesso a infraestrutura disponível no LIneA, tanto em termos de dados como de recursos de computação. Esta é uma questão que deve ser analisada no futuro.



**Figura 7.16** – Distribuição regional dos pesquisadores associados ao LIneA. A figura mostra uma grande uniformidade entre os principais centros.

Também interessante é ver a distribuição de acordo com a posição na carreira científica. Isto é apresentado na Figura 7.17, onde pode ser visto que existe um saudável balanço de pesquisadores contratados e jovens pesquisadores. Esta distribuição mostra que o LIneA está cumprindo a sua função de ajudar na formação de jovens pesquisadores e de promover oportunidades para quem está no início de carreira, já que muitos dos pesquisadores contratados tem menos de 50 anos.



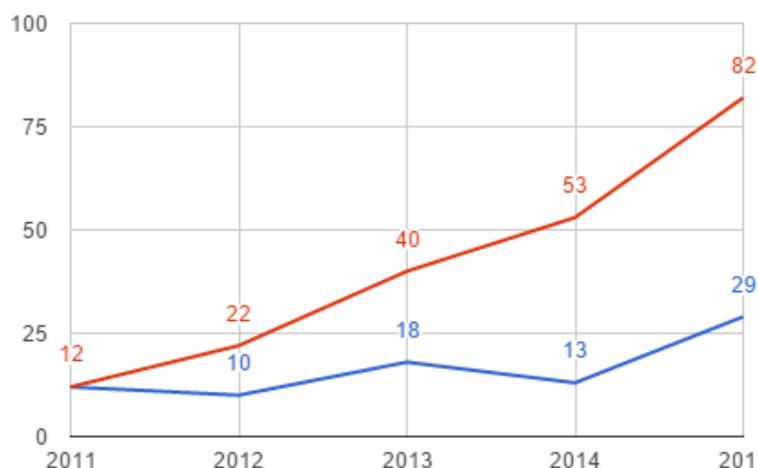
Revisão 09/11/2016

**Figura 7.17-** Distribuição por nível de carreira dos pesquisadores associados ao LIneA em 2015.  
Esta distribuição tem se mantido relativamente constante ao longo dos anos.

### 7.3 Publicações

A produção da equipe científica do LIneA é apresentada na Figura 7.18 onde são mostrados o número de artigos publicados por ano e o número acumulado a cada ano. A estatística inclui apenas os artigos dos membros afiliados ao LIneA e que usam os dados dos projetos apoiados pelo LIneA. É importante frisar que pela natureza dos projetos um artigo pode conter vários membros como coautores.

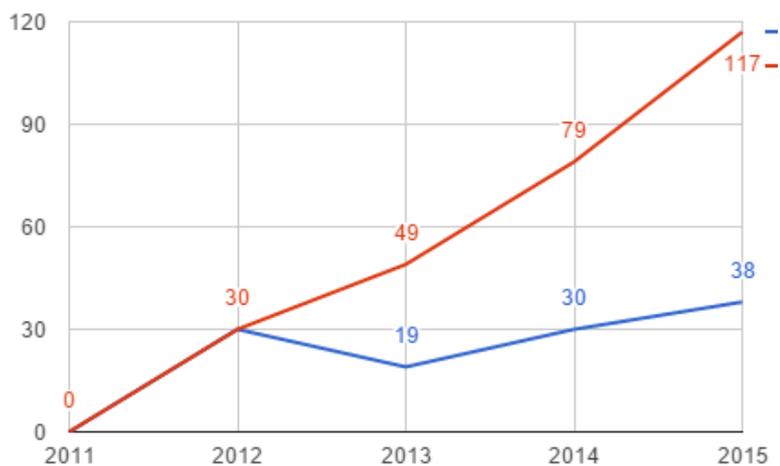
Além dos 82 artigos já publicados (2011-2015) existem ainda mais 28 artigos submetidos, 7 publicados em 2016, totalizando 127 artigos desde 2008 (124 desde 2010) com mais de 5000 citações. Usando este tipo de métrica para avaliar projetos que o LIneA tem promovido, mostra que temos uma excelente produção, da ordem de 20 artigos/ano, e de grande impacto considerando o número de citações. Atualmente, o grande número de artigos se deve principalmente ao DES e, ao contrário do que acontece em outras áreas, só participam dos artigos aqueles diretamente envolvidos nos cálculos e os chamados “construtores” (*builders*). *Builders* são nomeados pela colaboração após análise de uma comissão indicada pela direção do projeto que verifica se a contribuição do candidato em trabalho de infraestrutura representa mais de 24 meses (2 FTEs-ano) e que esta contribuição seja importante para o projeto como um todo. O consórcio DES-Brazil já tem oito *builders*, sendo quatro jovens pesquisadores (três pós-doutorandos), a maioria devido a importantes contribuições feitas para o Portal Científico do DES sendo desenvolvido pelo LIneA. Um aspecto importante desse processo é a enorme visibilidade que é dada ao trabalho sendo realizado por estes jovens pesquisadores e as oportunidades proporcionadas para novas colaborações internacionais, como já acontece.



**Figura 7.18** – Número de artigos publicados por ano (linha azul) e o número acumulado (linha vermelha)

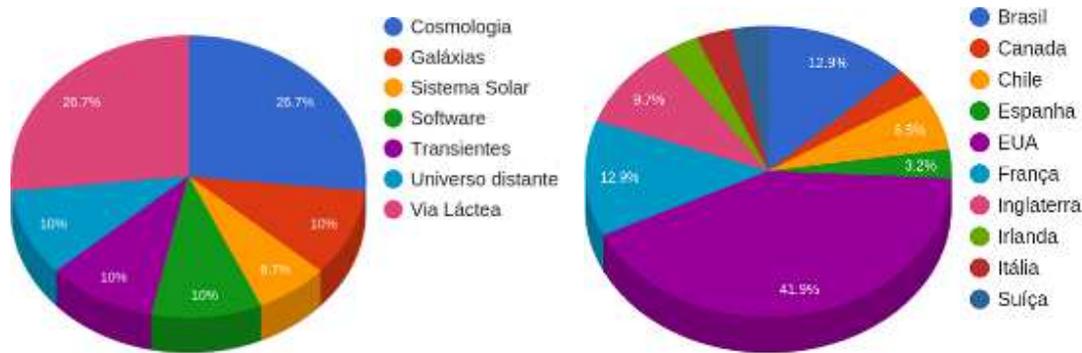
## 7.4 Webinars

Há quatro anos o LIneA deu início a uma série inovadora de seminários utilizando um sistema de conferência via web (*GoToMeeting*) chamado *webinars*. Os *webinars* são transmitidos pela internet em tempo real, e atualmente, aberto a participação de toda a comunidade. Os seminários são anunciados pelo site do LIneA e através de correio eletrônico para uma lista com mais de 700 nomes. O número de *webinars* realizados por ano pode ser visto na Figura 7.19.



**Figura 7.19** – Evolução no tempo do número de webinars por ano (linha azul) e o acumulado (linha vermelha).

Os seminários são ministrados, em sua maioria por palestrantes internacionais, PIs de grandes projetos ou experts de áreas científicas. A Figura 7.20 mostra como esses *webinars* foram distribuídos em termos de tópicos e o país de origem dos palestrantes dos quase 40 *webinars* realizados em 2015.



**Figura 7.20-** Distribuição de tópicos (painel esquerdo) e país de origem (painel direito) dos palestrantes na série de 2015.

Apesar do principal foco dos seminários estar relacionado a ciência dos projetos apoiados, os *webinars* procuram dar uma visão ampla da área de astronomia inclusive convidando os investigadores principais dos grandes projetos de astronomia em andamento ou previstos. Os *webinars* incluem também palestrantes trazendo novidades sobre a área de TI. Esta é uma oportunidade ímpar para estudantes e pesquisadores tomarem conhecimento da pesquisa sendo feita nas mais diversas áreas da astronomia, bem como as novidades na área de TI. Isto é importante para manter o grupo atualizado superando desta forma as antigas barreiras geográficas.

## 7.5 Resumo da atividade científica

A Tabela 7.1 apresenta números que resumem as atividades científicas dos projetos apoiados pelo LIneA durante o período 2011-2015. Estes números demonstram o alcance do projeto que envolveu mais de 62 pesquisadores e alunos, levou a 127 artigos com mais de 5000 citações em apenas cinco anos, cobrindo uma enorme gama de tópicos. Sem dúvida os números apresentados na tabela mostram o potencial aglutinador e multiplicador da estratégia adotada pelo LIneA em apoiar consórcios de pesquisadores participando de grandes levantamentos astronômicos, com times comprometidos, e com a expertise necessária para o melhor aproveitamento dos dados. Ao mesmo tempo oferece uma plataforma para aqueles interessados em se envolver em tópicos pouco desenvolvidos no Brasil, através da interação com cientistas de outros centros. Como será apresentado mais adiante, este tipo de ação é a que traz o melhor retorno científico pelo investimento de recursos feito.

Tabela 7.1 – Ciência em números

Artigos (publicados e aceitos)	127
Número de citações	5000+
Média de artigos por ano	20+

Teses e dissertações	17
Afiliados: Pesquisadores contratados	45
Afiliados: Estudantes	17
Número de webinars	128
Organização de reuniões internacionais	5
DES Builders	8

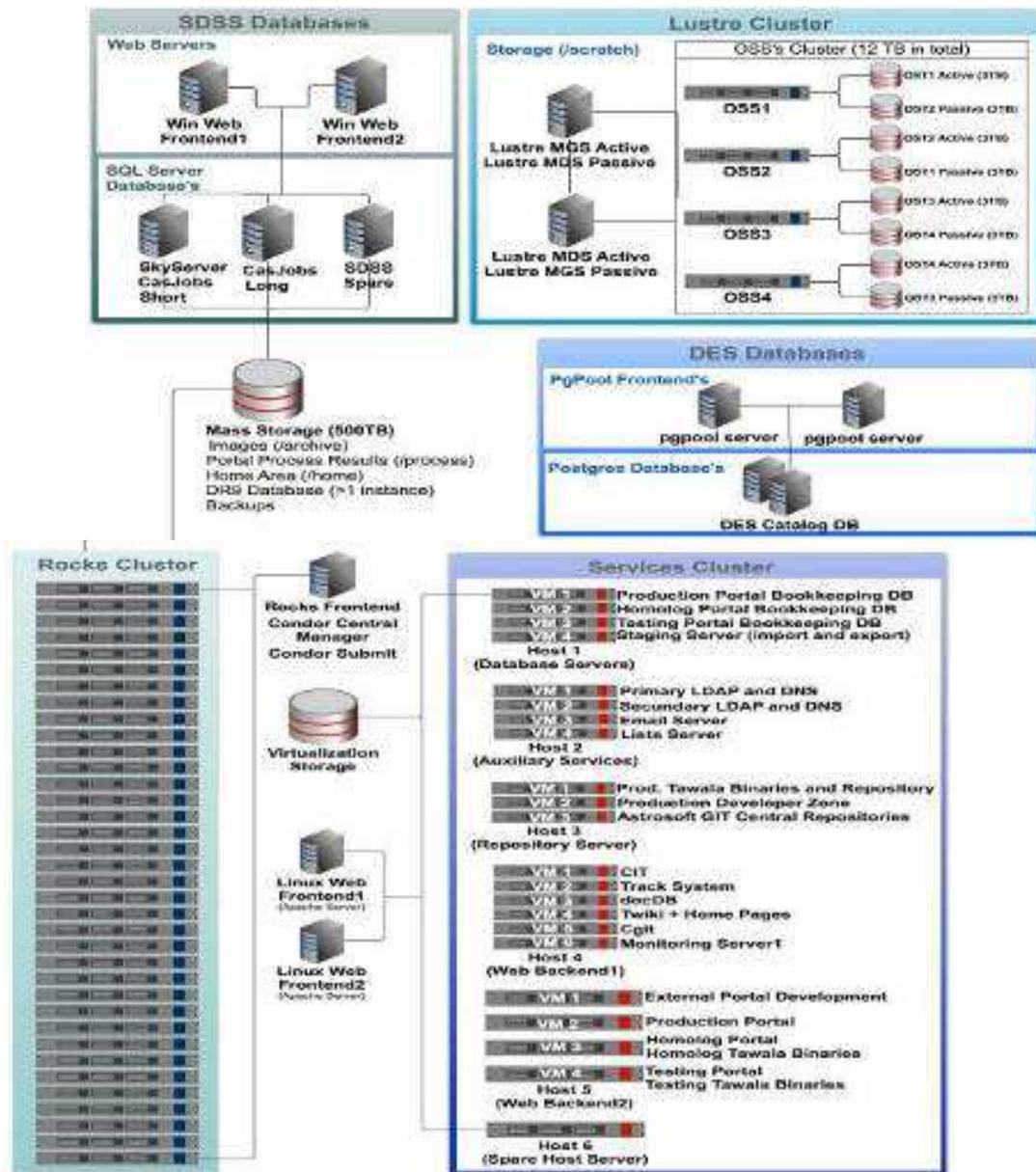
## 8 Centro de Processamento e Distribuição de Dados

### 8.1 Infraestrutura física

Para atender as mais variadas necessidades computacionais dos projetos apoiados pelo LIneA, um centro de armazenamento, processamento, análise e distribuição de dados vem sendo construído ao longo dos anos. Atualmente o centro consiste de mais de 100 equipamentos entre máquinas e switches. Uma visão esquemática do atual centro é apresentada na Figura 8.1. A arquitetura foi desenhada para atender as principais demandas das aplicações científicas previstas e incluem: 1) as altas taxas de transferência de dados; 2) o armazenamento de grandes quantidades de dados; 3) o resgate eficiente de dados armazenados em banco de dados; 4) o processamento paralelo para dados de diferentes partes do céu; 5) a hospedagem de serviços críticos para a operação e para o trabalho colaborativo. Levando estas necessidades em consideração, a arquitetura do centro consiste dos seguintes elementos chaves:

1. Cluster de processamento de produção com 38 nós (912 núcleos)
2. Cluster de processamento para testes de integração com 4 nós (72 núcleos)
3. Cluster Lustre com 6 nós (sistema de arquivos compartilhado de alto desempenho).
4. *Mass storage* (armazenamento de dados).
5. Banco de dados do SDSS (MS SQL).
6. Banco de dados do DES (PostgreSQL).
7. Cluster de serviços (VMs) onde se encontram as ferramentas colaborativas git (repositório de códigos), twiki, repositório de documentos, correio eletrônico.
8. Máquina dedicada ao desenvolvimento.
9. Sistema de Transferência de Dados (DTS).
10. Firewall.

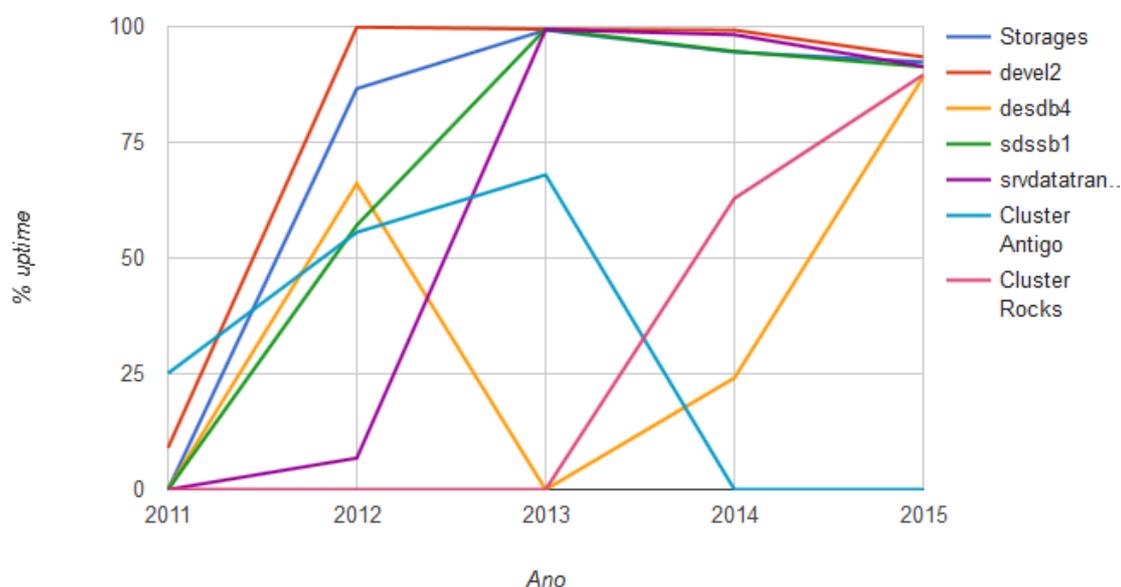
Após um começo em 2006 com máquinas montadas por pequenas firmas, estas foram gradualmente sendo substituídas por máquinas de maior qualidade. Atualmente, a maioria das máquinas é da marca DELL ou SGI. Inicialmente, as máquinas foram instaladas no ON e apesar de investimentos em aparelhos de ar condicionado, no-breaks e de um gerador, os riscos de uma operação 24/7 ficaram claros e a transferência das máquinas para o POP-RJ foi negociada com o LNCC e CBPF.



**Figura 8.1** Visão esquemática do centro de dados do LIneA onde se destacam: um cluster de processamento de 1000 cores, um sistema de armazenamento de 500 TB, um cluster de virtualização, um banco de dados e máquina dedicada a transferência de dados.

Os equipamentos foram então transferidos para o POP-RJ, mas mesmo assim o LIneA continuou sofrendo com os sérios problemas de infraestrutura locais (falta de espaço físico, de energia, de refrigeração, de no-breaks e de geradores), demonstrando a necessidade urgente de prover os institutos do MCTI com uma infraestrutura física adequada para atender as crescentes necessidades computacionais. Estes problemas afetaram sobremaneira tanto o desenvolvimento de software como os trabalhos científicos, o que acarretou uma enorme frustração. O problema foi agravado com a compra de novos equipamentos que por muitas vezes tiveram que ficar desligados por longos períodos de tempo (Figura 8.2).

No final de 2013 o CBPF solicitou a retirada das máquinas do LIneA do POP-RJ, alegando a necessidade do espaço para seus equipamentos. Isto levou o LIneA a fazer, no início de 2014, um longo estudo sobre possíveis alternativas para a realocação das máquinas. As seguintes opções foram consideradas: 1) o aluguel de um container; 2) o aluguel de um espaço de *colocation*; 3) a possibilidade de usar o IDC da RNP em Brasília; e 4) a migração para o CPD do LNCC na época sendo renovado. Finalmente, com as melhorias efetuadas no centro de processamento do LNCC e com a promessa de melhorias na conexão de internet disponível, o que é crítico para o LIneA, as máquinas foram transferidas durante o mês de Julho de 2015. Um dos motivos que pesou nesta decisão foi de que o custo de operação no IDC seria muito alto tendo em vista que a garantia dos equipamentos não incluía visitas a Brasília.



**Figura 8.2-** Evolução da fração de tempo em que as máquinas de diferentes serviços estiveram disponíveis no POP-RJ e mais recentemente LNCC.

A Figura 8.2 resume graficamente a complicada história do Centro de Dados do LIneA. Ela mostra para cada serviço a fração de tempo em funcionamento. Talvez o mais impressionante desta figura seja a fração de tempo parado dos clusters de processamento tanto antes como depois da instalação do cluster sendo usado atualmente, adquirido em 2013, mas que só recentemente pôde começar a ser usado regularmente. O impacto disto ainda repercute, pois atrasou toda a instalação de um ambiente estável de produção forçando os times de TI e a equipe científica a compartilharem o ambiente de teste com consequências adversas.

Desde 2011 os equipamentos têm sido instalados e toda a infraestrutura do LIneA é mantida e monitorada por uma empresa (SLACAM) com a qual o LIneA realiza reuniões semanalmente e que também participa do planejamento de longo prazo e mantém atualizada uma lista de itens críticos a serem adquiridos a curto prazo. A SLACAM também é responsável pelo monitoramento da rede em conjunto com a equipe técnica da

RNP. As responsabilidades da firma estão expressas no termo de referência apresentado a Finep (apêndice 8.1). Vários sistemas de monitoramento estão disponíveis como mostra a Figura 8.3, com as seguintes funções:

- **OCS** - Inventário de hardware dos servidores.
- **CACTI** - Controle do tráfego de rede.
- **IPPLAN** - Controle de distribuição das sub-redes IP.
- **BACULA** - Relatório de backups realizados.
- **Ganglia** - Monitoramento de uso de CPU, Memória e tráfego de rede do cluster, databases e servidores de apoio.
- **HTCondor** - Uso dos *slots* do condor no cluster de produção.
- **ZoneMinder** - Câmeras de Vigilância do POPRJ.
- **Awstats** - Estatísticas de acessos aos sites da web entre os quais do LIneA e do Skyserver.
- **Nagios Developer** - Monitoramento do nagios do site P&D.
- **HA Proxy** - Controle de autenticações do LDAP.
- **SIE** - Monitoramento do Condor do cluster de produção.
- **Temper Prod** - Controle de temperatura site produção.
- **Temper Dev** - Controle de temperatura do site P&D.
- **Helpdesk** - Atalho para *helpdesk* da SLACAM.



**Figura 8.3** - Sistemas disponíveis ao time de TI do LIneA para monitoramento das máquinas, da execução dos processos submetidos pelo portal científico, acesso aos sites e o status da rede.

A Figura 8.4 ilustra o resultado do monitoramento dos nós do cluster de processamento. Neste exemplo, o cluster está com todos os nós sendo utilizados, a menos de dois nós que se encontram fora de serviço. Neste caso específico uma chamada foi aberta para SLACAM que verificou um problema com a fonte, provavelmente causada por repetidas quedas no gerador e no-break do LNCC, agora já resolvidos. Este é um exemplo dos problemas que ocorrem ao longo da operação do centro para os quais deve estar previsto uma reserva técnica. Só em 2015 tivemos problemas com uma controladora do *mass storage* (US\$ 27k), discos, fonte e a máquina usada para a transferência de dados, representando um total de mais de US\$ 50 mil só para reposição. Como os recursos não estão disponíveis, isto exigiu um remanejamento de máquinas removendo uma máquina do sistema Lustre. Infelizmente, agora o Lustre apresenta problemas. Esta é a realidade de um centro de processamento que precisa ser considerada já que dele depende toda uma comunidade para realizar seu trabalho, com poucas ou talvez sem nenhuma alternativa, tendo em vista que o Portal Científico não é trivialmente transportado para outras arquiteturas.

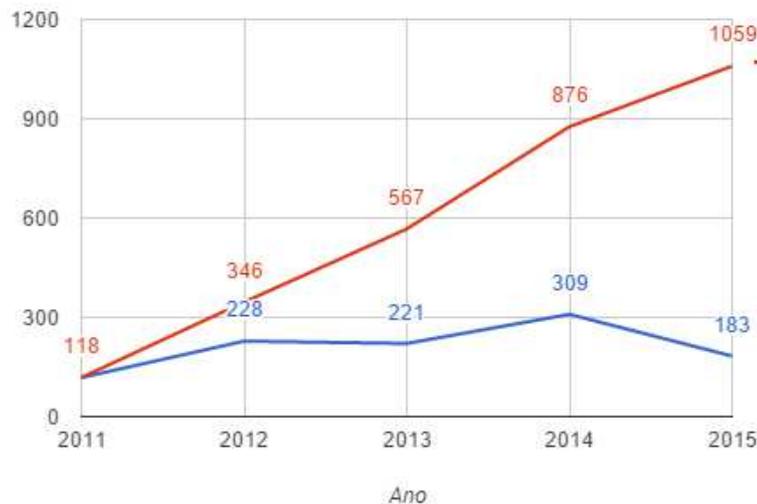


**Figura 8.4** – Uma captura de tela apresentada pelo GANGLIA mostrando a carga de diversos nós do cluster de processamento e reportando o problema em dois deles.

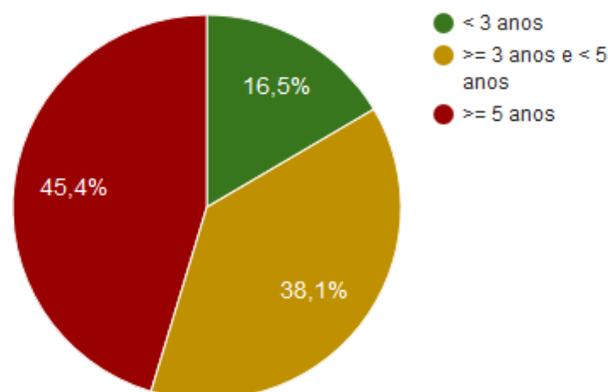
A importância do serviço prestado pela SLACAM pode ser avaliada pelo número de chamados abertos pela equipe de TI. Isto é apresentado na Figura 8.5 que mostra o número de chamados por ano (em média 200), quase uma a cada dois dias. É importante lembrar que a cobertura temporal é de 8/5, muito longe do ideal, pois várias paralisações

acontecem durante o final de semana. O ideal seria poder compartilhar estes serviços com outros grupos, reduzindo o custo e estendendo para uma cobertura de 24/7.

Outro motivo de grande preocupação é a idade dos equipamentos e o fato de que o prazo de garantia dos equipamentos é de apenas três anos e o custo de um contrato de manutenção é na maioria dos casos proibitivo. Na Figura 8.6 a distribuição por idade dos equipamentos é apresentada. Ela mostra que mais de 80% dos equipamentos do LIneA se encontram fora da garantia. Portanto, além da expansão do centro é urgente considerar uma estratégia de redundância, um plano de reposição, e criar uma reserva técnica que leve em consideração a criticidade dos diferentes serviços. Um plano como este deve envolver todas as partes interessadas e com a experiência necessária na operação de CPDs. A criação de um Centro de e-Ciência que provesse uma infraestrutura compartilhada poderia representar um grande avanço neste sentido.

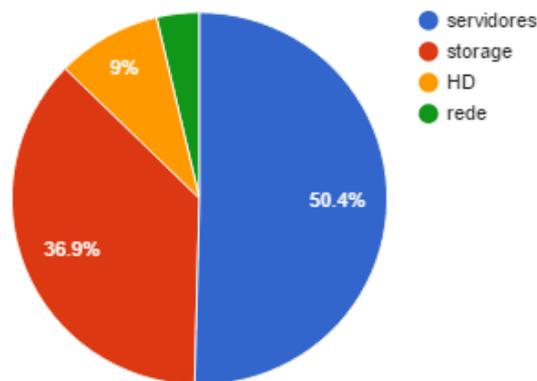


**Figura 8.5** – Número de tickets chamados abertos com a SLACAM por ano (azul) e o acumulado (vermelho).



**Figura 8.6** – Fração de equipamentos em uso em diferentes faixas de idade.

A Figura 8.7 apresenta os tipos de problemas encontrados, sendo que a principal causa de defeitos nos últimos quatro anos foram nos servidores mais antigos. Outro problema recorrente são falhas em discos. Talvez o caso mais sério até agora em termos financeiros tenha sido o defeito apresentado numa das controladoras do *mass storage*. A Figura 8.8 mostra o crescimento dos gastos que seria exigido para reposição de equipamentos. A tendência de crescimento de defeitos mais recentemente é preocupante e provavelmente está correlacionada com o envelhecimento dos equipamentos. Nos últimos quatro anos isto representou um custo de US\$ 80 mil. Portanto, a incorporação de uma reserva técnica na faixa de US\$ 20-50 mil por ano parece ser recomendada para evitar paralisações que podem afetar o bom funcionamento do centro. Igualmente importante é ter a flexibilidade de comprar quando necessário e não depender de uma compra coletiva como atualmente é o caso.



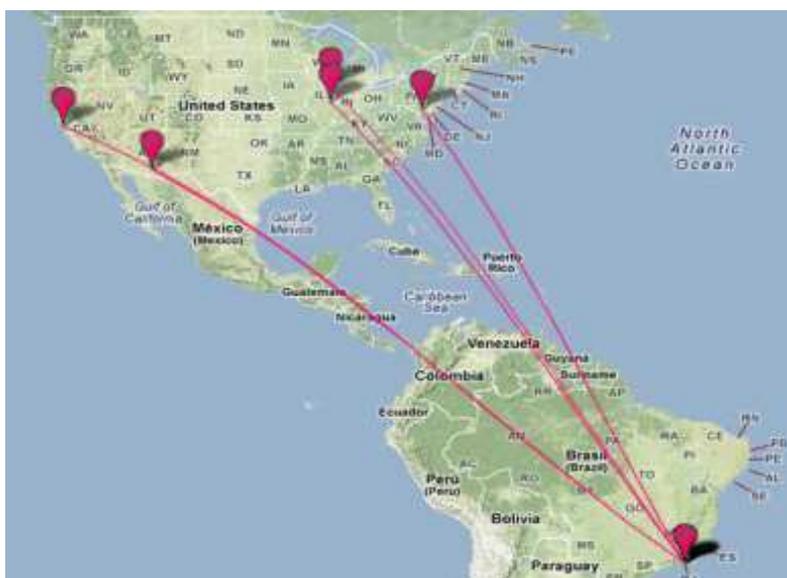
**Figura 8.7** - Tipo de equipamentos que apresentaram defeitos durante os últimos quatro anos.



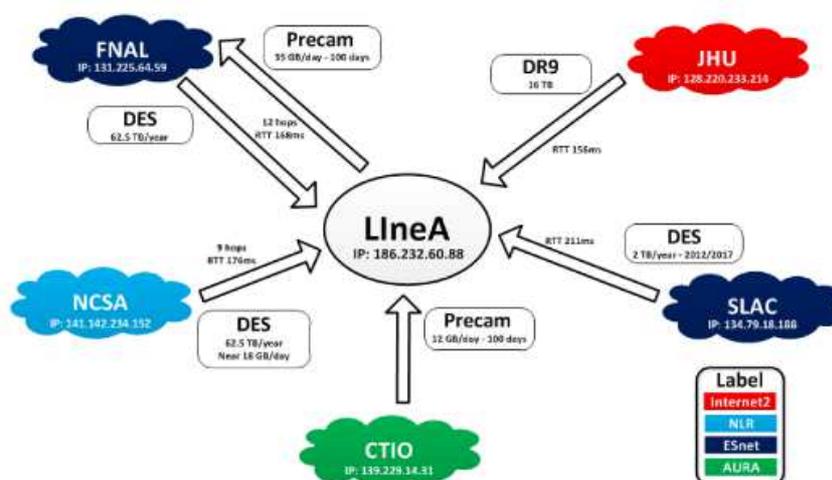
**Figura 8.8** – Custo anual de reposição de equipamentos e peças defeituosas.

## 8.2 Transferência de dados

Outra área de vital importância para o LIneA é a de transferência de dados com os vários parceiros internacionais. A Figura 8.9 mostra a localização geográfica de alguns destes parceiros entre os quais o SLAC (simulações numéricas), Fermilab (metadados), NCSA (bidirecional, dados do DES) e Johns Hopkins University (JHU, banco de dados do SDSS). A figura não inclui novos parceiros como LBL (DESI) e *University of Utah* (SDSS-IV, imagens, espectros e cubos de dados do SDSS). A Figura 8.10 mostra estas instituições e as organizações responsáveis pelas linhas de transmissão entre as quais ESNET e INTERNET2.



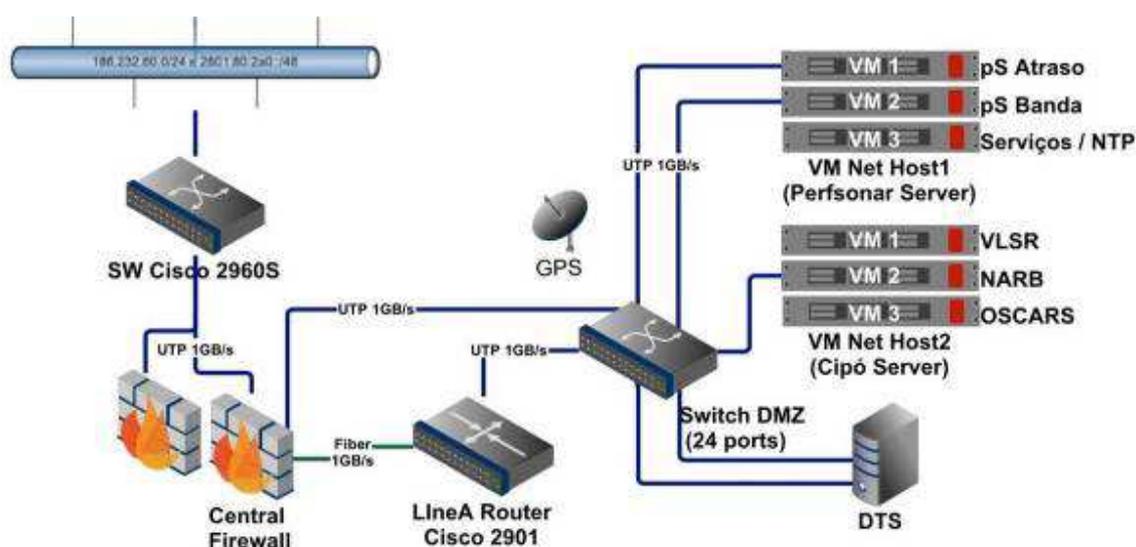
**Figura 8.9-** Localização geográfica dos principais parceiros do LIneA com os quais existem transferência de dados regulares.



**Figura 8.10-** Instituições parceiras do LIneA identificando as organizações responsáveis pela rede.

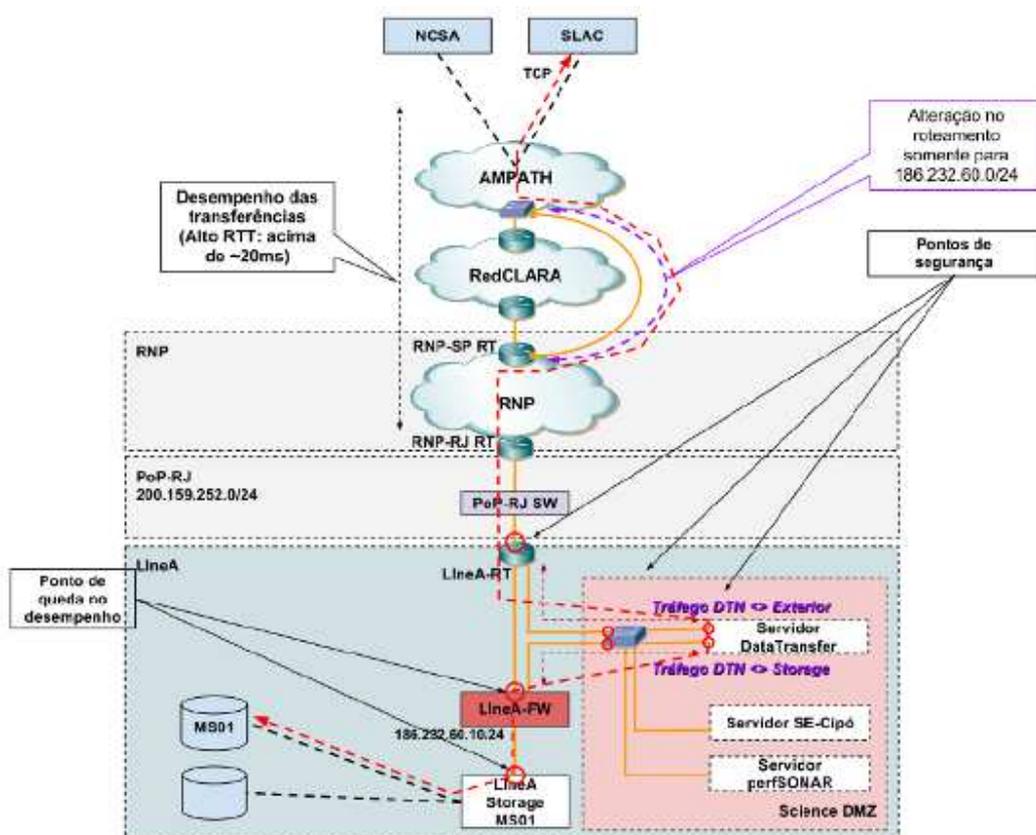
A forma mais frequente destas transferências se dá ao término de uma temporada de observações quando os dados reduzidos são liberados para uso da colaboração. No caso do SDSS isto acontece pelo menos duas vezes ao ano e exige a transferência da ordem de 15 TB. Já no caso do DES isto acontece uma vez por ano e envolve da ordem de 50 TB. Como isto é um pré-requisito para as análises, a transferência de pelo menos 3 TB na forma de catálogos tem que ser feita o mais rápido possível. Além destas grandes transferências, outras estão previstas de diferentes localidades em particular da Austrália de onde se pretende transferir os espectros obtidos pelo projeto OzDES.

Um esforço para melhorar a transferência foi feito com o apoio da RNP quando os equipamentos ainda estavam no POP-RJ. Na ocasião, foi apresentada a solução de criação de uma DMZ (Figura 8.11) parcialmente implantada pelo LIneA dentro de suas limitações orçamentárias. Mais importante ainda foi a identificação de problemas em uma das conexões e a implantação de rotas alternativas que triplicaram a taxa de transferência indo de ~200 Mbps para mais de 600 Mbps (ver Figura 8.12).



**Figura 8.11-** Solução de uma DMZ proposta pela RNP, parcialmente implantada pelo LIneA.

Com a mudança para o CPD do LNCC e com a estabilização do nosso sistema, um novo ciclo de testes foi iniciado pela RNP. Para servir como referência para as transferências futuras, testes foram realizados entre os centros de maior interesse no momento. Os resultados destes testes estão resumidos na Tabela 8.1 que lista a organização de origem, o nome do conjunto de dados, tamanho total, número de arquivos envolvidos na transferência, o número de *streams*, a taxa de transferência obtida e a duração da transferência.



**Figura 8.12** - Rota adotada para a transferência de dados do SLAC e NCSA e identificação de problema com a rede Clara.

Nestes testes foram usadas as ferramentas *aria2* e *wget* para transferência dos dados do projeto MaNGA. Estas transferências foram feitas diretamente para o LNCC que tem uma conexão de 1 Gbps. Como pode ser visto principalmente para os casos dos dados do SDSS vindos de JHU e Utah, o desempenho foi muito baixo. Mesmo no caso do NCSA para trazermos três TB leva-se atualmente um dia, e os 40 TB envolvidos num release do DES implicaria em três semanas. Certamente um novo esforço precisa ser feito para melhorar a taxa de transferência.

**Tabela 8.1**- Resultado de testes de transferência de dados entre os principais centros de interesse do LineA

Origem	Catálogo	Tamanho	# arquivos	streams	Taxa	Duração
SLAC	<a href="#">BCC v1.0</a>	~185 GB	204	5x1	200 Mb/s	2:20 hs
JHU	<a href="#">DR12</a>	96 GB	1	10x5	~30 Mb/s	7 hs
NCSA	<a href="#">Y1A1_COADD_SPT</a>	1.9 TB	16.862	10x5	277 Mb/s	18 hs
Utah	<a href="#">MaNGA</a>	4.7 TB	>316.000	1	~110 Mb/s	123 hs

Para melhorar o desempenho da transferência internacional foi submetido uma proposta (apêndice 8.2) em conjunto com a RNP para o programa ***Enlighten Your Research Global*** e reuniões com representantes da *ESnet* e *Internet2* já estão acontecendo.

Numa segunda fase, deverá ser analisado o desempenho de transferências diretas para o LNCC, ou usar como alternativa transferir os dados primeiro para a RNP conectada a 10 Gbps para posterior transferência do Rio para Petrópolis.

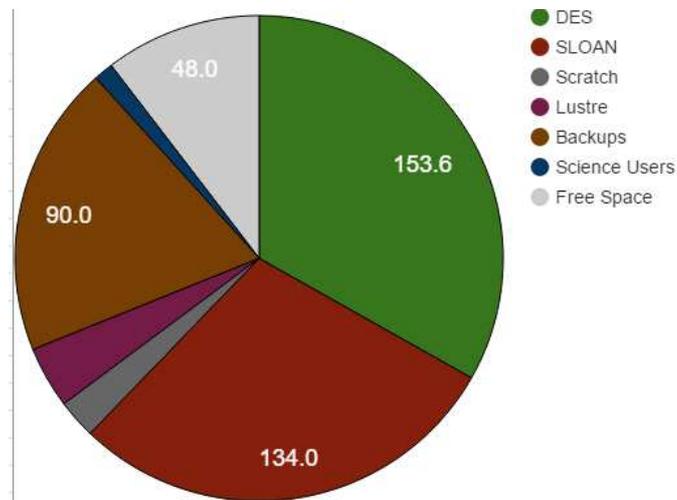
### 8.3 Armazenamento de dados

Um dos principais serviços oferecidos pelo centro de dados do LIneA é o armazenamento dos dados gerados pelos levantamentos apoiados para, em primeiro lugar, facilitar e agilizar a análise científica dos pesquisadores brasileiros envolvidos e, em segundo lugar distribuir estes dados para a comunidade brasileira, quando permitido pelas respectivas colaborações. Este modelo já é adotado para os levantamentos SDSS e DES, e será ainda mais importante na era do LSST. Note que o objetivo não é só o de armazenar os dados brutos, mas também as imagens e catálogos já processados. Prevendo esta demanda foi adquirido um sistema de armazenamento (ms01) contendo 180 discos de 3 TB (540 TB). Foram alocados 12 discos em *hot spare*. O sistema foi montado com 6 volumes com 28 discos em modo RAID6 (perda de 2 discos). Os seis volumes foram agrupados resultando em uma capacidade líquida de 425 TB. Essa configuração foi estudada com a SGI levando em consideração as três prateleiras do sistema e as posições das gavetas dos discos. Foi considerada também a limitação do sistema operacional de agrupar somente 28 discos por volume.

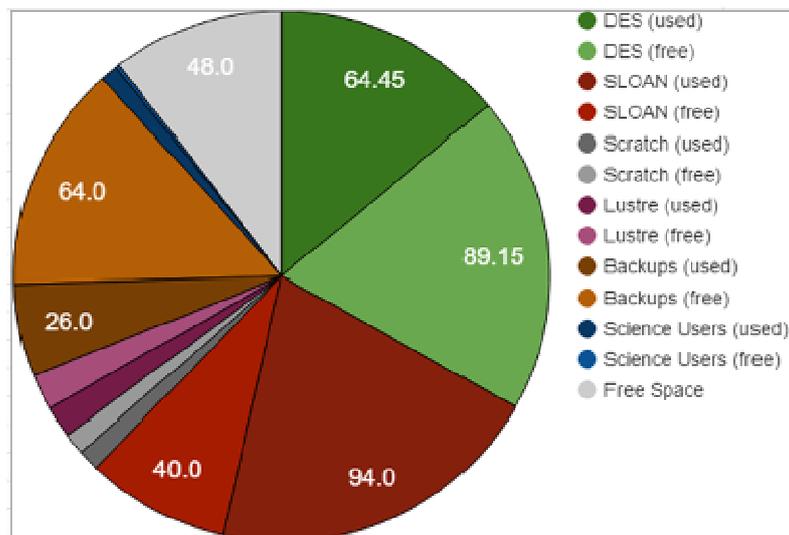
Figura 8.13 mostra como o espaço do *storage* foi inicialmente alocado para cada atividade e na Figura 8.14 a ocupação em cada partição. Como pode ser visto a partição do projeto SDSS está 70% ocupada com os *releases* DR8, DR9 e DR10 do SDSS-III, cada qual com 30 TB. Examinado a Tabela 8.2 para os *releases* previstos do SDSS-IV é necessário mais 150 TB.

Para o DES até agora, só estão armazenados os dados do período da verificação científica e primeiro ano de observação. Considerando que ainda será necessário armazenar dados de mais quatro anos de observação, isto implica a necessidade de ~260 TB adicionais. Na Tabela 8.3 os produtos básicos a serem produzidos anualmente estão discriminados. É importante notar que mais de 70% do volume terá que ser transferido do NCSA para o LIneA.

Para o LSST estimativas preliminares sugerem que será necessário armazenar pelo menos ~400 TB por ano, o que em 10 anos vai exigir da ordem de 5 PB, para manter pelo menos dois releases disponíveis para o grupo brasileiro.



**Figura 8.13** – Diagrama mostrando a atual alocação do espaço em disco do mass storage.



**Figura 8.14** – Diagrama mostrando a porcentagem de espaço livre e ocupado para cada partição do disco (projeto).

Em resumo, para acomodar as necessidades dos projetos SDSS e DES são necessários ~250 TB e ~320 TB para o SDSS e DES, respectivamente. Isto sem contar com outros produtos que serão produzidos durante a fase de análises. Além disso, deve-se prever

algum investimento para 2020 quando começará o período de verificação científica do LSST.

**Tabela 8.2-** Cronograma oficial de liberação de dados do SDSS-IV.

Release Name	Release Date	Data Through	eBOSS content	MaNGA content	APOGEE-2 (N) Content	APOGEE-2 (S) Content
DR13	July 2016	July 2015	New photometry, targeting data, SEQUELS spectroscopy, uniform spectro redux	DRP products for year-one data, trial public-release Marvin interface	Re-redux of SDSS-III DR12 APOGEE	(none)
DR14	July 2017	July 2016	New spectro data	New spectro data	New spectro data	(none)
DR15	July 2018	July 2017	(none)	New spectro data, new DAP products	(none)	(none)
DR16	July 2019	July 2018	New spectro data	New spectro data	New spectro data	New spectro data (maybe different cutoff date)
[DR17, internal]	[December 2020]	July 2019	TBD	TBD	TBD	TBD
DR18	December 2020	July 2020	New spectro data	New spectro data, all final DRP and DAP products	New spectro data	New spectro data (maybe different cutoff date)

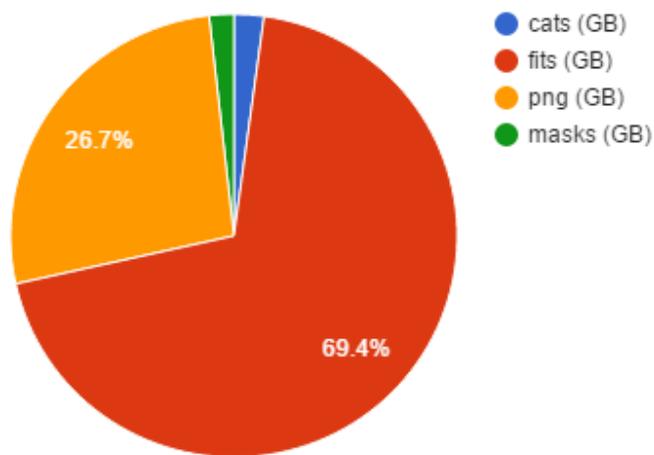
Levando em consideração as projeções do crescimento de dados para os próximos anos fica claro que o sistema tem que ser expandido rapidamente. A urgência é maior tendo em vista que este equipamento já está fora da garantia, e a recente perda de uma das controladoras do *mass storage*. É importante enfatizar que uma falha com a segunda controladora significa a perda completa do equipamento. Além disso, há o fato de que o equipamento já se encontra fora de linha e sem contrato de manutenção. A substituição desse equipamento e a constante expansão da capacidade de armazenamento é uma prioridade para o bom funcionamento do centro de dados.

**Tabela 8.3 -** Volume de dados do DES para cada produto usado em suporte à análise científica indicando sua origem de produção.

Produto	Volume (TB)	Origem
Catálogos	1,4	NCSA
Imagens (fits)	44,0	NCSA
Imagens (png)	17,0	LIneA
Máscaras	1,1	NCSA
<b>Total</b>	<b>63,5</b>	-

Os problemas foram acentuados com a disponibilização dos dados do primeiro ano de observações do DES (Y1) que gerou um total de 22 TB de dados correspondendo a 1800 graus quadrados do céu dos 5000 graus quadrados a serem observados pelo levantamento. Estes dados são todos gerados no NCSA e precisam ser transferidos para o LIneA o mais rápido possível para viabilizar a análise pelos pesquisadores brasileiros que, ao contrário dos outros parceiros internacionais, não tem acesso a grandes centros de processamento.

Os diferentes produtos transferidos do NCSA e os produzidos pelo Portal Científico antes mesmo de qualquer análise e o volume que eles representam são apresentados na Figura 8.15. É importante frisar que os arquivos em formato *fits* correspondem a imagens que são o resultado da co-adição de todas as exposições obtidas naquela parte do céu com um dado filtro. Isto diminui consideravelmente o volume dos dados do levantamento. Estas imagens são transferidas para alimentar as ferramentas desenvolvidas pelo LIneA, para a visualização, validação e exploração dos dados.



**Figura 8.15** – Fração do volume total ocupado por cada tipo de produto.

## 8.4 Processamento e análise de dados

O cluster de processamento de 42 nós adquirido em 2012, foi instalado somente no início de 2014 devido a problemas de infraestrutura no POP-RJ e por problemas burocráticos. Estes problemas persistiram até a mudança do equipamento para o LNCC e, como discutido acima, só recentemente foi possível usar este equipamento de forma regular. O impacto disto no trabalho foi enorme tanto diretamente como indiretamente tendo em vista que o ambiente de desenvolvimento, normalmente instável, era também usado pelos pesquisadores dificultando enormemente o trabalho.

A análise conduzida com os dados do Y1 demonstrou que sem um cluster com um grande número de núcleos será quase impossível dar prosseguimento ao trabalho de pesquisa. Por exemplo, para produzir um catálogo adequado para análise é necessário rodar 16 *pipelines* produzindo 64 produtos, um dos quais necessitou 42 horas de processamento. Isto para um catálogo com 1/3 do tamanho daquele esperado para o segundo ano em diante. Recentemente, uma nova versão do código reduziu este tempo para 10 horas o que significa que para apenas um código seriam necessários 30 horas de processamento.

Além disso, como mencionado abaixo, este é um de oito algoritmos disponíveis no Portal Científico para o mesmo fim, e cujos resultados devem ser comparados.

Além da fase de preparação dos dados, em breve, diferentes grupos irão executar *workflows* científicos para: i) ajustar modelos da Via Láctea; ii) buscar sistemas estelares na vizinhança de nossa galáxia; iii) estudar a evolução das propriedades das galáxias usando para isso a função de luminosidade e massa; iv) calcular correlação angular das galáxias em grande escala; v) identificar aglomeração de galáxias; e vi) estimar os parâmetros cosmológicos utilizando a técnica de *Monte Carlo Markov Chain* (MCMC). Isto sem contar com outras atividades em andamento como: i) processamento de modelos de evolução do sistema solar, um tópico extremamente atual com a sugestão da existência de um novo planeta; ii) a previsão de ocultações baseados em dados do DES e de imagens obtidas em observatórios como o ESO, Pic-du-Midi, SOAR, e OPD nas campanhas de acompanhamento; iii) a redução de dados provenientes de diversos observatórios durante uma campanha de ocultação para o estudo da estrutura e possível existência de atmosfera de objetos do sistema solar.

O compartilhamento racional da infraestrutura permitiu, por exemplo, ao pesquisador Rodney Gomes responsável pela modelagem do Sistema Solar (e um dos proponentes do nono planeta) aumentar a velocidade da sua análise por um fator 50 comparado ao que era possível com as máquinas a sua disposição.

Fica claro que um investimento na infraestrutura do LIneA representa uma economia de escala tanto do ponto de vista da infraestrutura física quanto do uso otimizado do equipamento além do compartilhamento do pessoal de apoio especializado num esforço coordenado.

A infraestrutura do laboratório oferece ainda aos seus usuários acesso a uma máquina de desenvolvimento e a um repositório de códigos onde estes podem ser versionados e compartilhados com seus colaboradores. Novamente este é o fator multiplicador de um centro de e-Ciência do qual o LIneA é um exemplo. Este compartilhamento pode ser feito em nível de uma aplicação ou de infraestrutura.

## 8.5 Banco de dados

A SLACAM, é responsável pela instalação e manutenção do banco de dados do SDSS (MS SQL), incluindo a transferência de dados que são disponibilizados periodicamente ao público, envolvendo ~15 TB em cada *data release*. Esta firma também efetua as mudanças exigidas na interface de usuários a cada liberação. É importante salientar que a escolha do produto MS SQL foi feita pelo consórcio SDSS e não pode ser alterado, pois todo o sistema de busca foi construído no seu entorno. Até agora o LIneA já disponibilizou quatro destes conjuntos de dados (DR8 ao DR12, sendo o DR11 apenas internamente a colaboração) referentes ao projeto SDSS-III. Agora o laboratório se prepara para a

primeira liberação de dados do projeto SDSS-IV, incluindo novos tipos de dados como os sendo produzidos pelo projeto MaNGA.

O acesso à interface específica de um *release* é feito a partir do site do LIneA como mostra a Figura 8.16. A partir desta chega-se a interface pela qual os dados podem ser acessados (Figura 8.17).



**Figura 8.16** - Interface que dá acesso aos dados dos diferentes releases. Apesar de cada release conter os resultados de todos os levantamentos anteriores, mudanças de software exigem que os produtos sejam versionados para garantir a informação de proveniência, (<http://www.linea.gov.br/020-data-center/acesso-a-dados-2/>).

Para montar o sistema no LIneA usando os recursos disponíveis, uma visita foi realizada a JHU para entender o arranjo usado e adaptá-lo as condições locais. Para cada *release*, JHU usa quatro servidores com discos internos para atender as demandas das diferentes ferramentas como *skyserver*, *casjobs*, serviço de *cutout* e mais uma máquina sobressalente (*hot spare*) para evitar paralisações. No caso do LIneA o sistema montado consiste de dois *front-ends* para processar buscas com os dados localizados no sistema de armazenamento. Embora a demanda do LIneA para servir aos pesquisadores brasileiros e mesmo como alternativa ao site americano, seja bem menor, a solução implementada, e em uso até recentemente, teve um desempenho abaixo do desejado. Isto foi porque o *storage* é usado para muitas outras atividades e a solução dependia da largura de banda disponível no momento da consulta.

Recentemente, para o release DR12, esta solução foi abandonada com a instalação de um servidor dedicado com discos internos como é feito no JHU. Infelizmente, no caso do LIneA apenas uma máquina foi adquirida e é compartilhada pelos diferentes serviços. A aquisição de novas máquinas será considerada dependendo da demanda que será

avaliada periodicamente. Convém mencionar que o antigo sistema continua disponível para os *releases* anteriores.

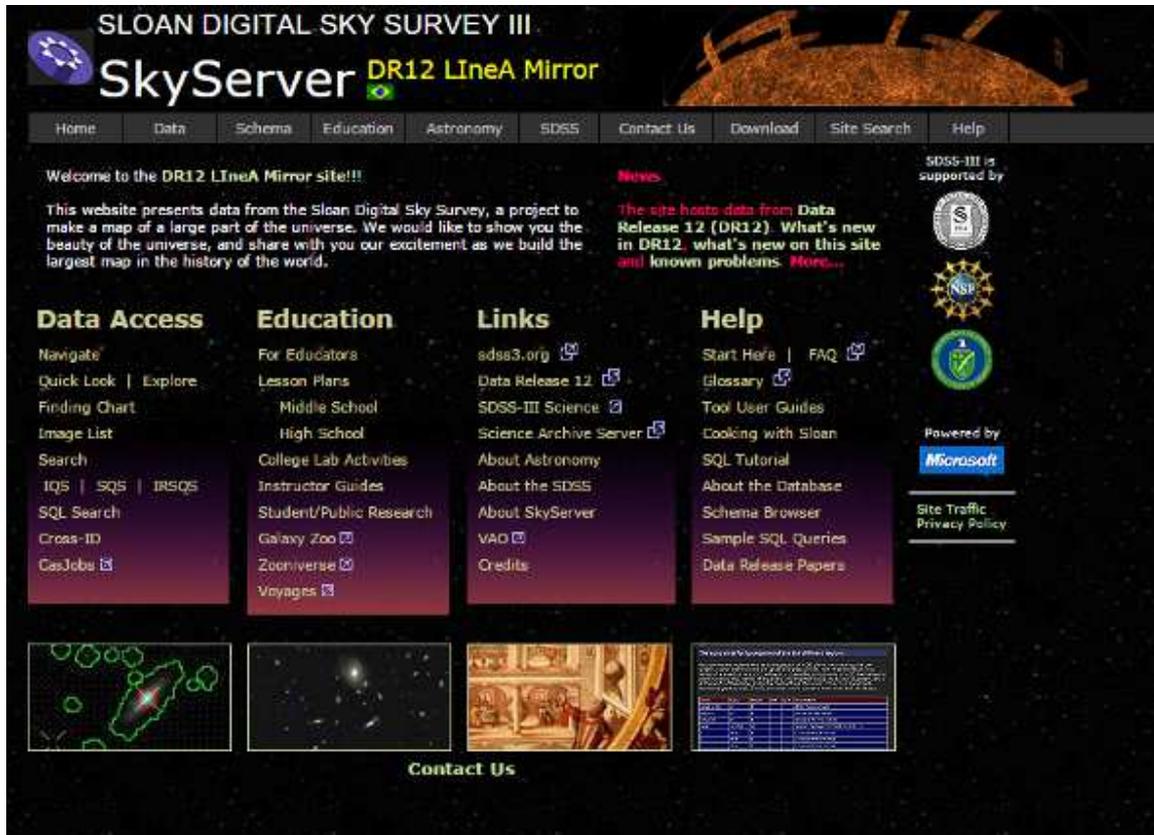
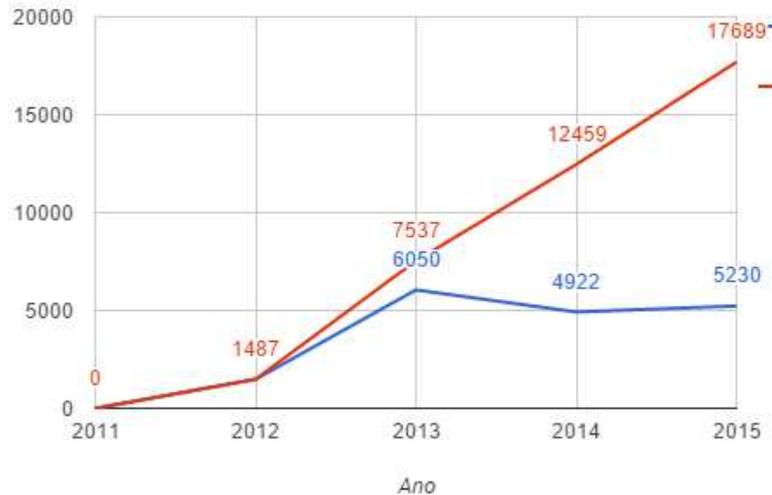


Figura 8.17 - Interface de acesso aos dados do DR12 do SDSS.

A Figura 8.18 apresenta o número de acessos ao serviço do *Skyserver* aberto para consultas ao banco pelo grande público. Ele teve mais de 17.000 visitas de todo o mundo desde 2011 quando foi instalado o monitoramento destes acessos. Nos últimos 3 anos a média de acessos por ano é ~5.000 visitas (22.000 hits) ou ~400 visitas por mês. Já o acesso ao *CasJobs*, aberto para profissionais é bem menor devido em parte, talvez ao baixo desempenho comentado anteriormente. Outras causas podem ser o pouco hábito da pesquisa astronômica brasileira de usar este tipo recursos para sua ciência ou de estar acostumado a usar o sistema do JHU. Com novos levantamentos como, por exemplo, o MaNGA isto deve mudar e a demanda será monitorada.

Para o projeto DES o LIneA está utilizando *PostgreSQL*, que é *open source*, tanto para o banco administrativo do Portal Científico discutido abaixo, como para o armazenamento dos dados do DES. Como no caso do SDSS originalmente os dados eram localizados no *mass storage* e as buscas eram processadas em duas máquinas utilizando o *middleware pgpool-II*. Novamente a experiência de trabalho demonstrou que esta solução não era adequada para os problemas enfrentados tanto para busca como para alimentação dos

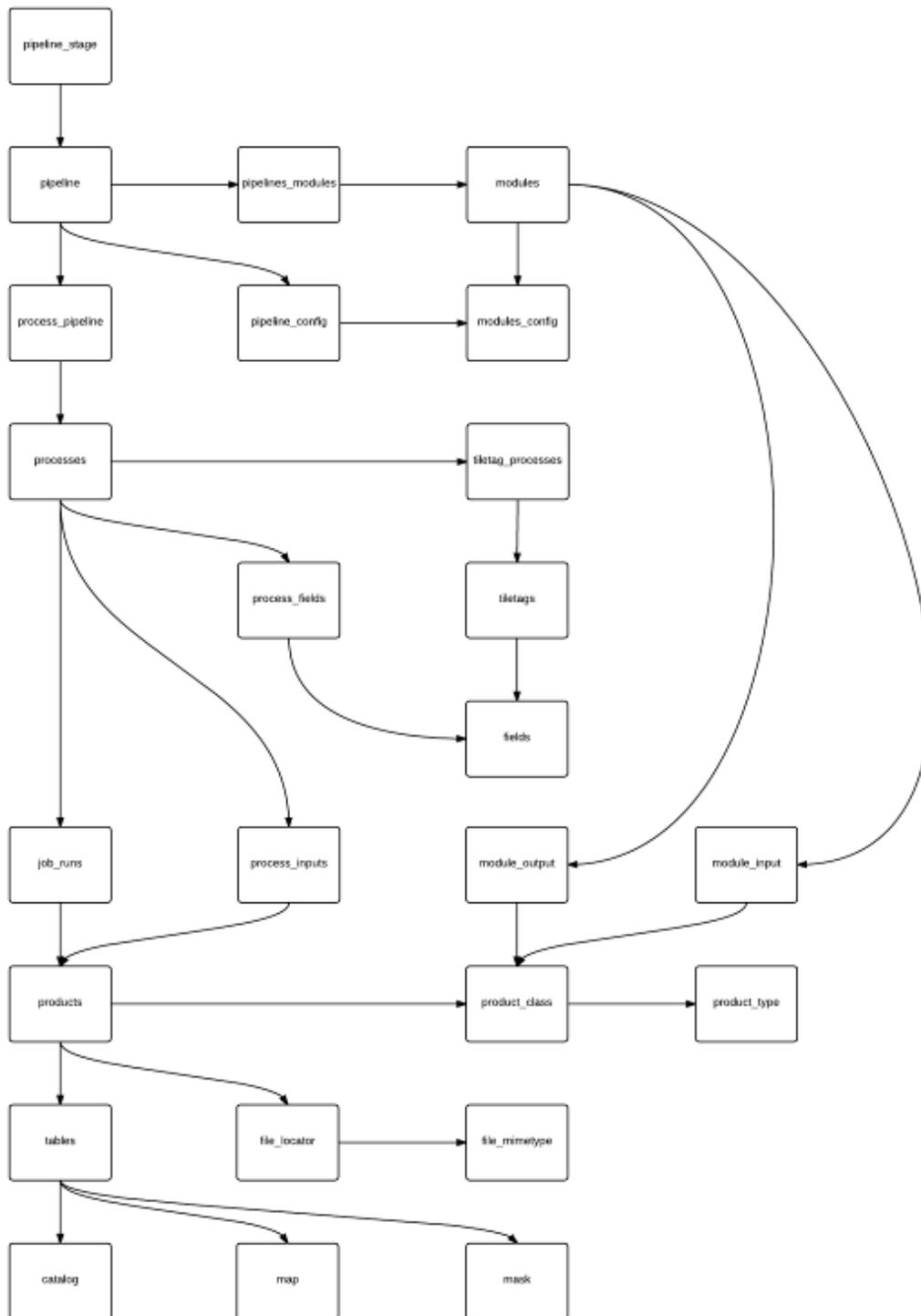
*pipelines* do portal. É importante frisar que este estudo foi feito pelo time de TI do LIneA em consulta com pessoal especializado do LNCC e com o apoio do administrador de banco de dados da SLACAM.



**Figura 8.18** - Número de acessos à página do Skyserver por ano (azul) e acumulado (vermelho).

O banco de dados desenhado para o DES pelo LIneA, e que sustenta a operação do portal consiste de um banco administrativo com 312 tabelas ocupando 21 GB e um banco de catálogos atualmente com 8.487 tabelas ocupando 5.2 TB. É bom salientar que isto representa o resultado de um longo período de desenvolvimento e testes e pode não refletir o que é estritamente necessário para a operação do sistema. De qualquer maneira o banco administrativo representa uma fração pequena do volume de dados total. A figura 8.19 mostra um diagrama usado para a instalação da parte do banco administrativo responsável pelo gerenciamento dos processos executados pelo portal, os dados de *input* e *output*, e configuração associada de maneira a garantir a propagação da proveniência no encadeamento de *pipelines*.

O banco do DES é mantido nos ambientes de teste e produção no LIneA e no ambiente de produção no Fermilab. No Fermilab está sendo desenvolvida uma interface hospedando vários serviços que serão utilizados pela colaboração DES para o seu primeiro *release* público previsto para 2017. Esta será uma versão moderna do CASJobs introduzido pelo SDSS, mas com uma nova filosofia e usando *open source*. Isto faz parte de um compromisso assumido pelo LIneA com o NCSA numa proposta submetida e aprovada pelo NSF. No momento, uma versão beta está disponível localmente para os pesquisadores brasileiros e no Fermilab para toda a colaboração. A Figura 8.20 mostra a interface atualmente disponível no Fermilab.



**Figura 8.19** – Esquema de parte do banco de dados administrativo desenhado pelo time de TI do LInEA para apoio ao Portal Científico.

Observations Data Releases Footprint Tile Viewer Catalog Server Science Products User Query Upload Help Luiz da Costa

[Release Notes](#)

## DES Science Portal: Data Server

The DES Science Portal hosts tools for Quality Assessment (QA), Value-Added Catalogs (VACs) preparation and Science Analysis.

From the **Data Server** instance @ FNAL you have access to following tools:

- **Observations:** information about DES observations from the Night Summary and Quick Reduce
- **Data Releases:** list of the releases currently installed and associated data
- **Footprint:** spatial coverage and overlapping with external catalogs
- **Tile Viewer:** visual inspection of co-add images and catalogs
- **Catalog Server:** access to VACs produced by the portal, uploaded catalogs, reference catalogs and simulations
- **Science Products:** access to science products produced by the portal or uploaded by other authors

The system is designed to be self-evident, use the help icon "?" available on each page.

The Science Portal is a facility developed by [LIneA](#). If you have any question please contact us through the [helpdesk@linea.gov.br](mailto:helpdesk@linea.gov.br)

Tweets [Follow](#)

DES Science Portal @des\_portal  
QR results for night 2016-01-21 are available

DES Science Portal @des\_portal  
QR results for night 2016-01-20 are available

DES Science Portal @des\_portal  
QR results for night 2016-01-18 are available

Tweet to @des\_portal

**Figura 8.20** - Interface de usuários construída pelo time de TI do LIneA disponível no Fermilab para acessar dados do DES armazenados num banco de dados, também desenhado pelo LIneA.

A partir desta a interface os usuários podem monitorar o progresso das observações, obter informações sobre o *release* de dados, analisar as regiões cobertas pelo levantamento, examinar a qualidade das imagens, acessar catálogos produzidos pelo portal e pela colaboração, fazer buscas no banco para a escolha de diferentes tipos de objetos e, visualizar a imagem destes objetos. Este sistema está em processo de migração para o NCSA, onde estará conectado ao banco de dados principal do DES que utiliza *Oracle*. Este sistema tem sido um grande sucesso com mais de 180 pesquisadores registrados (mais da metade dos membros do DES). Desde o início do serviço em Março de 2014 até a data deste documento houveram mais de 3.000 visitas dando uma média de 125 visitas por mês. A Figura 8.21 mostra o número de visitas por mês ao longo de quase 24 meses.



**Figura 8.21**- Visitas de pesquisadores do DES ao banco de dados e interface implementados pelo LIneA e disponível no Fermilab.

Para melhorar o desempenho das consultas ao banco de dados de catálogos está sendo considerada uma solução que vem sendo desenvolvida pelo SLAC para o LSST. Esta

solução denominada *qserv* se baseia na partição dos dados em vários nós, o que exigiria mais um cluster cujo dimensionamento ainda está sendo avaliado. Ainda com relação a banco de dados existe uma nova demanda para hospedar Skybot um serviço para buscar e identificar objetos do sistema solar em imagens astronômicas desenvolvido e mantido pelo *Observatoire de Paris*. O objetivo é dar sustentação ao projeto *Transneptunian Occultation Network* (TON), apoiado pelo LIneA, integrando esse serviço ao pipeline em desenvolvimento para identificar TNOs nas imagens de levantamentos como o DES e DECals, Este trabalho é de fundamental importância como preparação para o LSST. Uma máquina acaba de ser disponibilizada ao grupo e aos seus colaboradores para instalação deste banco.

## 8.6 Indicadores do centro de dados

A Tabela 8.4 resume algumas das informações sobre o Centro de Dados do LIneA, e demonstra o seu caráter multiusuário, sua importância para apoiar a participação efetiva de pesquisadores brasileiros nos grandes projetos apoiados. O laboratório oferece uma infraestrutura dedicada e dirigida a esta finalidade, além de serviços prestados a comunidade brasileira como um todo atuando na curadoria dos dados provenientes dos grandes levantamentos dos quais o LIneA participa.

Tabela 8.4- Centro de Dados em números

Número de equipamentos	110
Número de usuários LIneA (ldap)	108
Acesso ao <i>skyserver</i>	14.927
Acessos ao site do LIneA	112.591
Volume de dados transferidos 2014-2015 (TB)	100
Volume de dados armazenados (TB)	250
Número de processos do portal	23.000
Número de logins na devel (2012-2015)	11.129

## 9 Desenvolvimento de Software

### 9.1 Portal Científico

O volume de dados sem precedentes acumulado pelos levantamentos SDSS, DES e futuramente DESI e LSST exigem uma infraestrutura computacional capaz de analisar grandes volumes de dados e distribuir os resultados dessas análises de forma eficiente. Para isto, ao longo dos últimos anos, o LIneA vem atuando em duas frentes. A primeira montando um centro de dados cuja arquitetura visa atender as necessidades específicas destes projetos, como apresentado na seção anterior. A segunda desenvolvendo um portal científico que oferece os seguintes serviços:

- um repositório de códigos centralizado (git);
- disponibilidade de bibliotecas científicas e pacotes de software atualizadas;
- instalação e verificação automatizada da qualidade dos dados;
- ferramentas para inspeção visual de imagens e catálogos associados;
- banco de dados de catálogos astronômicos incluindo dados de outros levantamentos;
- automação da criação de catálogos para análise científica;
- integração de algoritmos científicos na forma de *workflows*;
- interfaces para acesso uniforme aos dados, metadados e resultados.

Os algoritmos desenvolvidos pelos diferentes grupos de trabalho científicos são mantidos em repositórios de código onde as mudanças são versionadas, permitindo a contribuição de diferentes desenvolvedores de forma organizada e garantindo que algoritmos desenvolvidos por alunos e pós-doutorandos sejam preservados. Esse aspecto é de fundamental importância para programas de longo prazo como é o caso dos grandes levantamentos sendo considerados..

A Figura 9.1 ilustra os principais conceitos do Portal Científico, que são:

- Interoperabilidade de diferentes conjuntos de dados, como por exemplo, os oriundos de observações e simulações;
- Proveniência dos dados de entrada, configuração e versão dos algoritmos utilizados em cada etapa da análise;
- Validação dos resultados em cada etapa da análise;
- Reprodutibilidade dos resultados;
- Comunicação e compartilhamento dos resultados.

## Science Portal Grand View

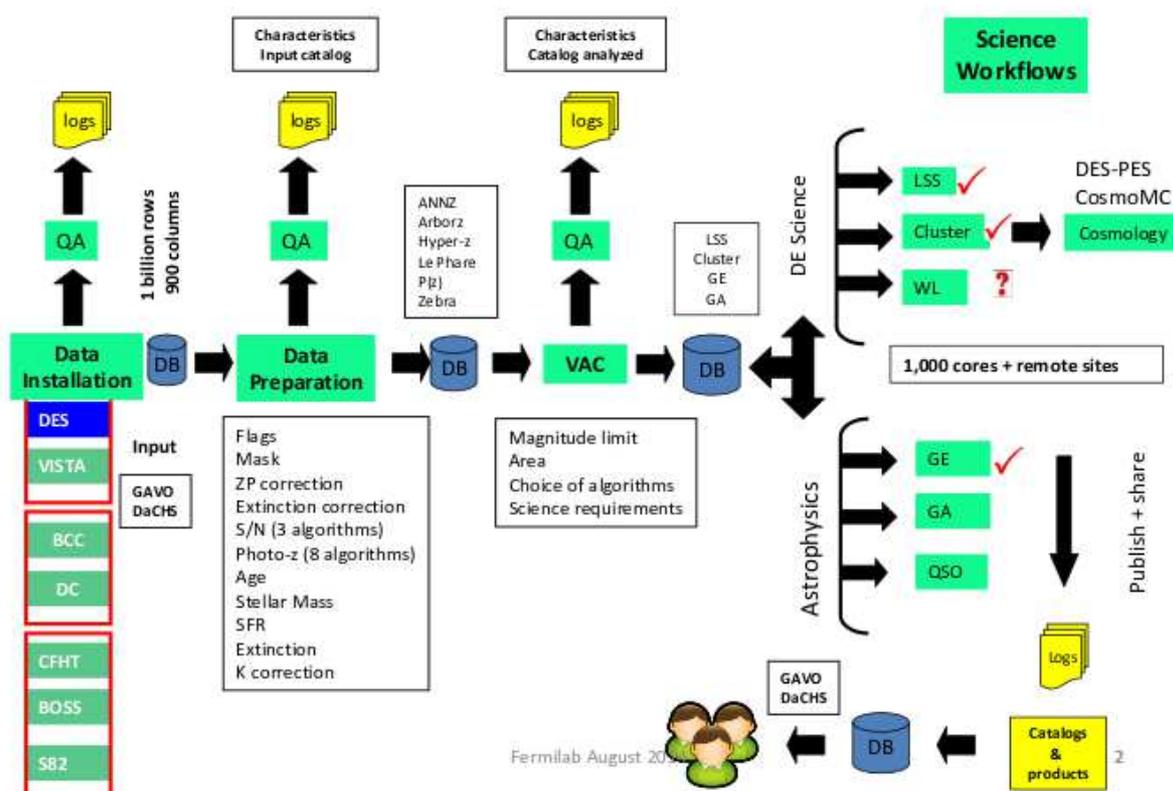
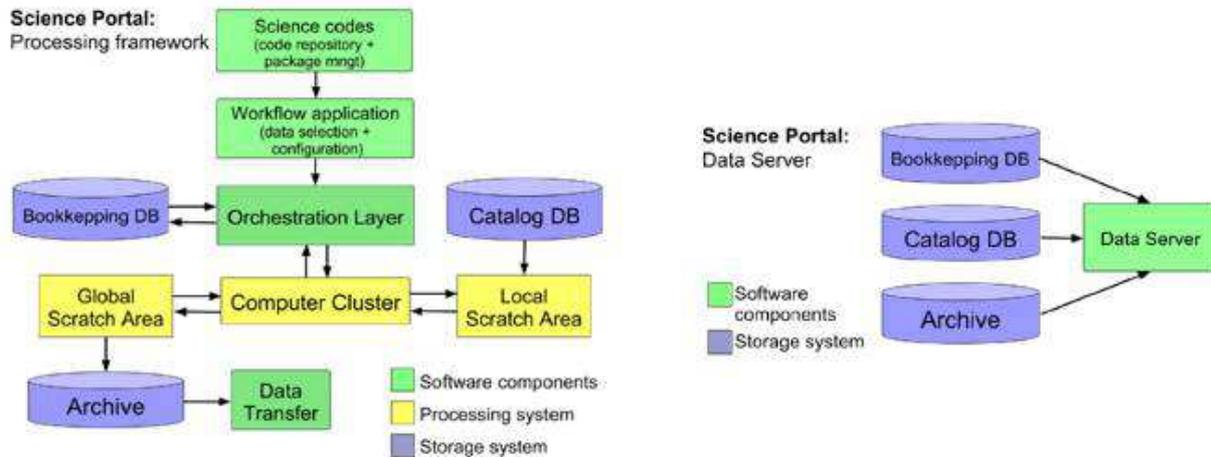


Figura 9.1 – Diagrama esquemático do Portal Científico.

Nesse modelo, as etapas de instalação e de preparação dos dados, que representam uma fração substancial do tempo gasto na análise de dados convencional, são realizadas de forma estruturada e eficiente. A base de dados centralizada e a integração dos algoritmos científicos ao portal são cruciais para minimizar a movimentação de grandes volumes de dados. Os algoritmos científicos integrados ao portal se beneficiam da infraestrutura de processamento e acesso aos dados disponíveis. Esta visão de um sistema *end-to-end* indo dos dados básicos a resultados científicos, rodando pipelines o máximo possível de forma não supervisionada, é fundamental para tornar possível a análise eficiente de grande volume de dados. Isto se aplica tanto ao DES e certamente ao LSST. Esta é uma das razões para a criação de um BPG para o LSST e um dos critérios para a escolha dos pesquisadores deve ser o desejo de compartilhar seus códigos e integrá-los a este sistema.

Além dos algoritmos científicos, o sistema de gerenciamento de *workflows* e a camada de orquestração são os principais componentes de software do portal. Associados a esses componentes estão o banco de dados administrativo, o banco de dados de catálogos, o cluster de processamento e o sistema de armazenamento, como mostrado na Figura 9.2.



**Figura 9.2** – Principais componentes do Portal Científico.

O desenvolvimento do portal teve início em 2007 e seu desenho básico tem evoluído muito ao longo dos anos. Até agora se estima que já foram investidas o equivalente de 60 FTEs (*full-time equivalent*), como normalmente é contabilizado a contribuição *in-kind* nas colaborações internacionais. Já são nove anos de desenvolvimento, sendo que neste período cerca de 30 profissionais de TI contribuíram para o seu desenvolvimento, sem contar com a contribuição de alunos, pós-doutorandos e pesquisadores. Até agora vários produtos já foram entregues e estão disponíveis para diferentes comunidades. Mencionamos alguns deles a seguir.

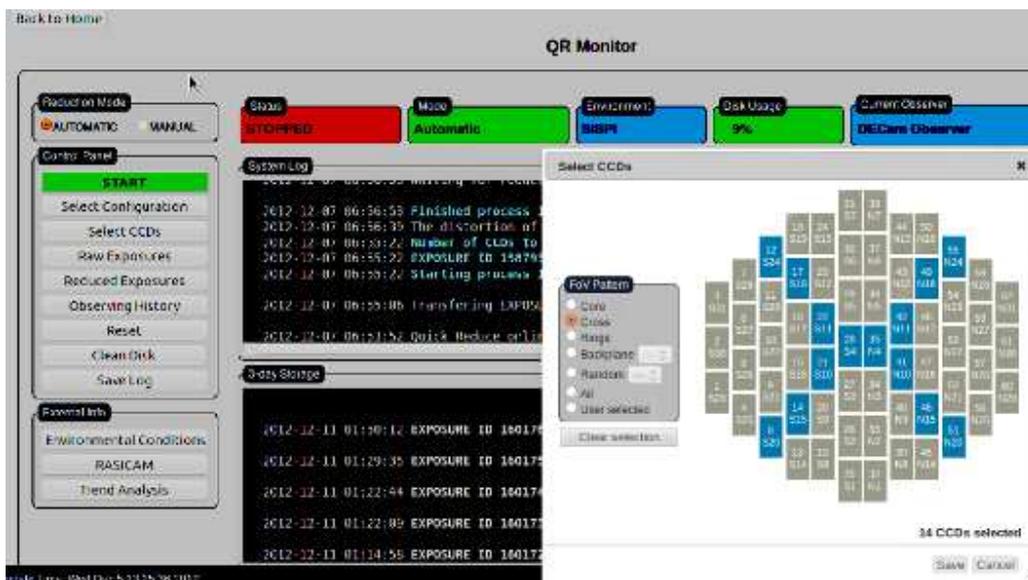
## 9.2 Quick Reduce

O *Quick Reduce* (QR) é um sistema para diagnóstico em tempo real das exposições da câmera do DES (DECam) instalada no telescópio Blanco de 4 metros no *Cerro Tololo Inter-American Observatory* (CTIO), Chile. O QR está em operação desde 2012 e é utilizado não só por membros da colaboração DES, mas também por usuários da DECam da comunidade internacional, inclusive da brasileira que se beneficia com a troca de tempo entre telescópios Blanco e SOAR. Até o momento, só para o levantamento DES, foram reduzidos e examinados 1,2 milhões de CCDs de 117 mil exposições obtidas em 345 noites de observação. O sistema embora sofisticado é robusto com o número de intervenções exigidas da equipe do LIneA sendo extremamente baixo (<20). Ao final de cada noite os resultados obtidos são transferidos para o Fermilab e estes podem ser examinados por todos da colaboração evitando o firewall do CTIO. A Figura 9.3 mostra a sala de observação do telescópio Blanco. À direita estão os monitores que são de responsabilidade do assistente noturno, no centro os monitores registrando informações sobre o sistema de aquisição usados pelo observador e à esquerda os monitores usados para acessar o QR usados pelo observador 2.



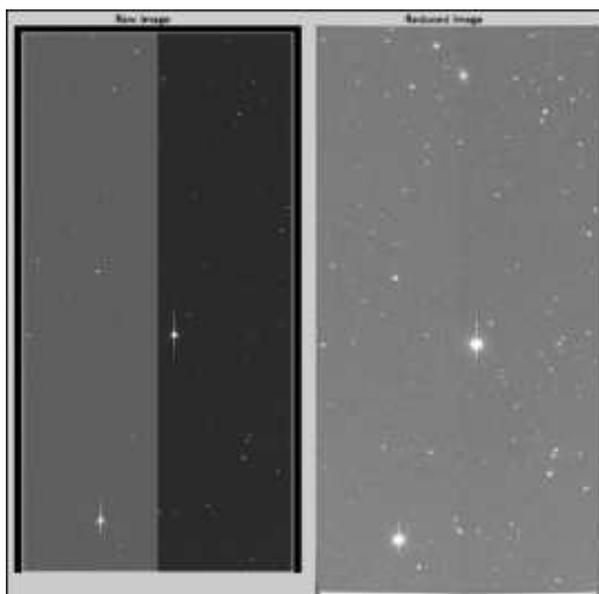
**Figura 9.3** – Sala de observação do telescópio Blanco do CTIO, Chile.

A Figura 9.4 mostra a tela principal do QR onde o observador pode acompanhar o processamento e acessar as várias funcionalidades disponíveis como, por exemplo, a configuração que permite aos observadores de escolher a posição e o número de CCDs a serem examinados em cada exposição.



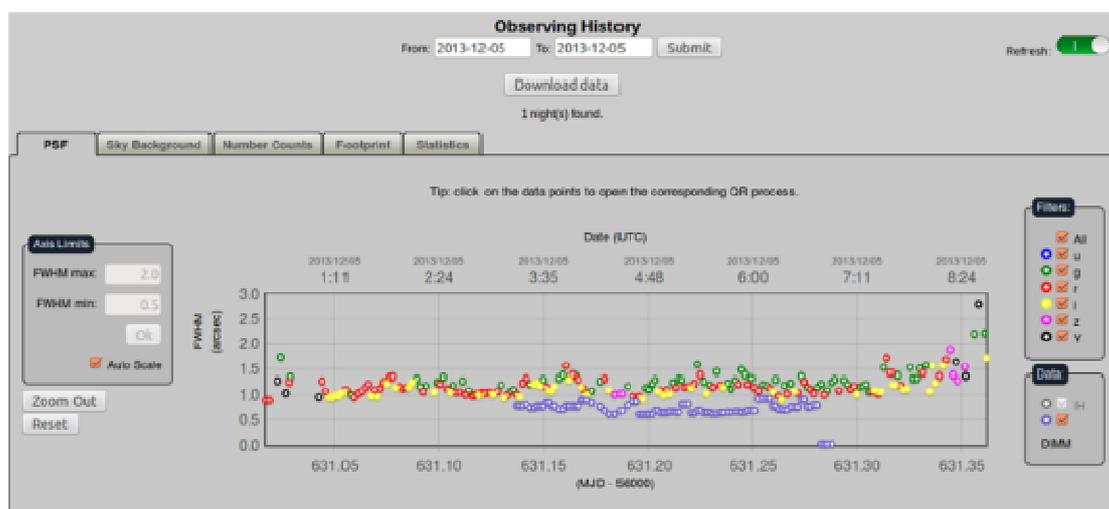
**Figura 9.4** – Tela do Quick Reduce mostrando os CCDs (4k x 2k) da câmera do DES monitorados pelo sistema.

Uma vez em funcionamento e configurado, o QR extrai de cada imagem produzida pela DECam os CCDs escolhidos, remove os efeitos instrumentais (Figura 9.5), identifica os objetos localizados na imagem, e calcula para os objetos classificados como estrelas parâmetros que são usados para caracterizar a qualidade da imagem.



**Figura 9.5** - A figura mostra o resultado da remoção dos efeitos instrumentais pelo QR. À esquerda a imagem original e à direita após o processamento.

A qualidade da imagem pode ser monitorada usando uma das interfaces do QR como mostra a Figura 9.6. A figura mostra a variação do tamanho de uma estrela em função do tempo (exposição) ao longo da noite. Em caso de rápidas variações um alerta é enviado ao observador, pois pode indicar a presença de nuvens ou problemas instrumentais.



**Figura 9.6** - Tela onde a qualidade das imagens obtidas pela DECam é monitorada em função do tempo durante uma noite de observações. As cores indicam os filtros utilizados em cada exposição.

O QR foi entregue a colaboração em 2012, durante o comissionamento da câmera, acompanhado de um manual e agora um artigo técnico sobre ele está em preparação. O sucesso do QR motivou a direção do CTIO a solicitar o apoio do LIneA em mantê-lo em funcionamento para qualquer usuário da DECam, inclusive para outros levantamentos como DECals, o que foi acordado com a direção do DES. O DECals está sendo

conduzido pelo projeto DESI que recentemente estabeleceu uma colaboração com o LIneA para desenvolver o sistema Quick Look Framework destinado a avaliar suas observações espectroscópicas. Isto pode ser interpretado como um reconhecimento da qualidade dos sistemas desenvolvidos pelo laboratório, em particular do QR. Como será discutido na seção 13 manter o QR funcionando durante as noites previstas para o DES é um dos importantes compromissos do LIneA.

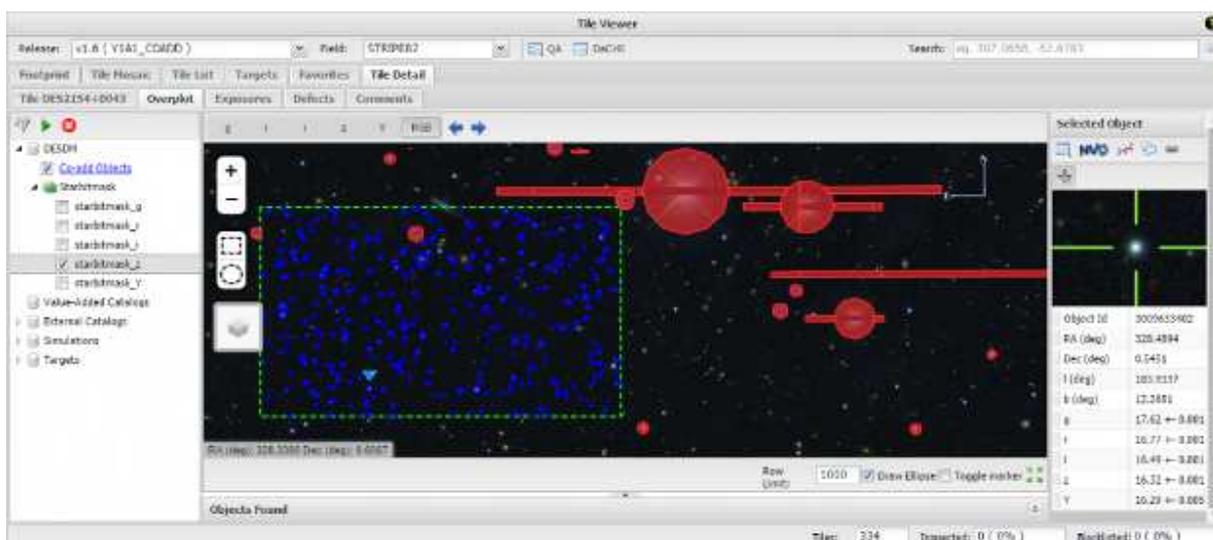
### 9.3 Data Server

Outro importante e complexo sistema entregue para a colaboração DES em Abril de 2014 foi o chamado *Data Server*. Este sistema é composto por um conjunto de ferramentas disponíveis no LIneA e no Fermilab que permitem:

- Acompanhar o progresso do levantamento DES.
- Verificar visualmente a qualidade das imagens, permitindo marcar defeitos, rejeitá-las e/ou fazer comentários.
- Sobrepor objetos de catálogos (internos e externos ao DES) às imagens do levantamento.
- Realizar buscas nos catálogos de objetos produzidos a partir destas imagens.
- Criar ou carregar listas de objetos de interesse.
- Visualizar estes objetos e criar “recortes de imagens” (*postage stamps*) para referência futura ou publicação.

A Figura 9.7 ilustra a ferramenta *Tile Viewer* usada na verificação dos dados do DES, uma das muitas disponíveis no sistema. Neste exemplo, em particular é possível não só ver a imagem, mas objetos detectados e as máscaras aplicadas para remover estrelas brilhantes. Esta ferramenta de visualização é fundamental para verificar o funcionamento de métodos automáticos.

O sucesso do *Data Server* com a colaboração foi grande e mostrou ao mesmo tempo que ao disponibilizar um produto complexo como este é fundamental prever que será necessário um bom tempo para a correção de problemas identificados pelos usuários durante operação e atender a demanda de novos casos de uso. Isto significa a necessidade de se ter pessoal suficiente para que seja possível conciliar as tarefas de correção e pequenas melhorias com o constante desenvolvimento. Outro problema identificado foi a dificuldade de preparar um manual para um sistema tão complexo. Uma experiência foi feita com a produção de pequenos vídeos mostrando diferentes casos de uso (Figura 9.8). Estes vídeos estão disponíveis no portal e no site do LIneA. Estes vídeos serão complementados por um artigo técnico em preparação.



**Figura 9.7** – Captura de tela do Tile Viewer, umas das ferramentas disponíveis na instância do Portal Científico instalado no Fermilab. No detalhe vemos uma imagem do DES com máscaras e objetos de um catálogo, sobrepostos.



**Figura 9.8** - Série de vídeos tutoriais produzidos para diferentes casos de uso. Em particular: como acessar a ferramenta Tile Viewer; uso do Tile Viewer para validação das imagens; demonstração da ferramenta Overplot.

O uso do *Data Server* tem mais de 180 usuários registrados, correspondendo a mais da metade de membros da colaboração DES na sua maioria pesquisadores internacionais. O sistema continua em desenvolvimento e, como será mencionado abaixo, está previsto para ser a face pública do DES quando os dados ficarem disponibilizados em 2017 (seção 13).

## 9.4 Infraestrutura para criação de catálogos científicos

A criação de catálogos para análises científicas específicas é talvez um dos maiores desafios de levantamentos fotométricos de grandes áreas do céu tendo em vista o volume de dados, o grande número de atributos associados a uma fonte (~700), o grande número de decisões que devem ser tomadas ao longo do caminho e a necessidade de preservar a memória de como estes catálogos foram produzidos. O Portal Científico foi desenhado como resposta a estes desafios, com as fases de instalação, preparação e criação de catálogos (veja Figura 9.1) na sua maior parte automatizada através de uma infraestrutura composta por 18 *pipelines* criando mais de 64 produtos que devem ser rastreáveis e reproduzíveis, caso necessário. A falta de homogeneidade dos dados devido



**Figura 9.9** – Pipelines usados na instalação e preparação dos dados, e na criação de catálogos no Portal Científico.

Release: Y1A1		Tag: COSMOS D04				
<b>Data Installation</b>						
Pipelines	Start	End	T	Nexc	Status	
Install Catalogs	2015-11-18 14:56:29	2015-11-18 14:57:34	0:01:05	<a href="#">1</a>		
Install Mangle Mask	2015-12-01 13:04:41	2015-12-01 13:37:17	0:32:36	<a href="#">7</a>		
Install Bright Mask	2015-12-01 10:37:28	2015-12-01 10:38:29	0:01:01	<a href="#">1</a>		
Install Images						
Install Depth Maps	2015-12-01 09:28:33	2015-12-01 09:45:59	0:17:26	<a href="#">5</a>		
Systematic Maps	2015-12-01 13:05:23	2015-12-01 13:21:42	0:16:19	<a href="#">24</a>		
Zeropoint Correction	2015-12-01 13:05:57	2015-12-01 13:16:27	0:10:30	<a href="#">19</a>		
Spectroscopic Sample	2015-12-01 13:28:13	2015-12-01 13:44:52	0:16:39	<a href="#">33</a>		
QA Coadd	2015-12-01 13:29:51	2015-12-01 15:13:09	1:43:18	<a href="#">6</a>		
<b>Data Preparation</b>						
Pipelines	Start	End	T	Nexc	Status	
SG Separation	2015-11-24 14:21:30	2015-11-24 14:24:15	0:02:45	<a href="#">4</a>		
Training Set Maker	2015-12-01 10:42:34	2015-12-01 10:43:23	0:00:49	<a href="#">20</a>		
Photoz Training	2015-12-01 09:57:48	2015-12-01 13:17:06	3:19:18	<a href="#">15</a>		
Photo-z Compute	2015-12-01 12:58:16	2015-12-01 12:59:41	0:01:25	<a href="#">5</a>		
Galaxy Properties						

**Figura 9.10** – Dashboard usado para monitorar o número de execuções, início, fim, duração e status do último processo de cada pipeline executado no portal.

A Figura 9.11 ilustra os testes de software que são realizados em cada pipeline para garantir que as últimas mudanças foram integradas com sucesso no ambiente de teste e podem ser levadas para produção.

Status of pipelines at Testing						
Stage	Pipeline	CD04	S82	Last change	Known issues	Pending
DI	Install Mangle Mask	<a href="#">10022397 (Feb 2)</a>	<a href="#">10022422 (Feb 2)</a>	- Added ra, dec columns on map table - Installation by field - registering PNG images		- Run on Y1A1 S82
DI	Install Bright Mask (Bad Regions Map)	<a href="#">10022398 (Feb 2)</a>	10022462 (Feb 3)	- Installation by field - Added ra, dec columns on map table	- missing data at ra < 0 - Rename pipeline - Maps show only a fraction of S82 at RIGHT	- we confirmed that it does not rem
DI	Install Depth Map	<a href="#">10022402 (Feb 2)</a>	10022108 (Jan 20)	- Installation by field - missing data at ra < 0		- Run on Y1A1 S82
DI	Systematic Maps	<a href="#">10022404 (Feb 2)</a>	10021526 (Nov 21)	- Installation by field	- Exptime map with constant value of 60s - missing data at ra < 0	
DI	Zeropoint Correction	<a href="#">10022405 (Feb 2)</a>	10022182 (Jan 28)	- Added parameters for provenance	- plot shows no correction on i-band because it applies only extinction correction	
DI	Spectroscopic Sample	<a href="#">10022408 (Feb 2)</a>	10022288 (Jan 28)		- Output class name (Training set spectroscopy)	
DI	QA Coadd	<a href="#">10022412 (Feb 2)</a>	10022299 (Jan 28)			
DP	SG Separation	<a href="#">10022413 (Feb 2)</a>		- Improved Notifications	Improve table that reports the number of star for each method (method   nstars)	- Run on Y1A1 S82
DP	Training Set Maker	<a href="#">10022414 (Feb 2)</a>	10022310 (Jan 29)	- SLR and Extinction correction		
DP	Photo-z Training	<a href="#">10022415 (Feb 2)</a>	10022345 (Jan 29)	- SLR and Extinction correction	- Cannot register class training?	
DP	Photo-z Compute	<a href="#">10022397 (Feb 2)</a>	10021995 (Jan 13) 10022027 (Jan 14)	- SLR and Extinction correction		- Release to testing - Run on Y1A1 S82
DP	Galaxy properties	<a href="#">10022397 (Feb 2)</a>	--	- Parallelization	- SLR and Extinction correction	- Release to testing - Run on Y1A1 S82
DP	Cluster Catalog	<a href="#">10022397 (Feb 2)</a>	--	- Added Bad Regions Map	- SLR and Extinction must be applied - special character on product log (& bitwise operator is being interpreted as html code)	- Release to testing - Run on Y1A1 S82

**Figura 9.11-** Testes de integração dos diferentes pipelines no portal de testes do LIneA.

## 9.5 Workflows científicos

O último estágio do sistema *end-to-end* que pode ser apreciado na Figura 9.1 são os *workflows* científicos que usam os catálogos (VACs) preparados no estágio anterior, para realizar diferentes análises científicas. No caso do DES os *workflows* desenvolvidos ou em fase de desenvolvimento permitem:

- Estudar a evolução das funções de massa e de luminosidade das galáxias.
- Calcular o grau de aglomeração das galáxias em pequena escala.
- Identificar aglomerados de galáxias.
- Determinar a função de correlação em grande escala para diferentes populações de galáxias.
- Modelar a estrutura da Via Láctea.
- Vincular parâmetros cosmológicos utilizando técnicas de MCMC levando em consideração diferentes hipóteses.
- Criar simulações da Via Láctea, e de populações de objetos no sistema solar.

Este sistema é usado, não só para analisar dados reais, mas também dados simulados para avaliar o desempenho de diferentes algoritmos, e o impacto dos critérios de seleção e de erros sistemáticos presente nos dados. Nesta fase a maior contribuição vem da equipe científica que provê os algoritmos de análise escritos em diferentes linguagens. Cabe a equipe técnica encapsular estes algoritmos de acordo com os requisitos da arquitetura do portal, resolver problemas de paralelização e ajudar na ingestão dos resultados no banco de dados ou sua disponibilização via uma interface específica.

Esta parte do sistema, mais orientada a ciência, é destinada principalmente a equipe brasileira e servirá de embrião para o que se imagina venha a ser um conjunto bem maior e mais sofisticado de *workflows* para utilização com os dados do LSST. O uso desses *workflows* com os dados do DES e de simulações numéricas serve, não só para obter resultados científicos de forma mais eficiente e testada, mas também como referência para estimar as necessidades de processamento na era do LSST.

## 9.6 Instâncias do portal em operação

O LIneA mantém as seguintes instâncias do portal:

- **Portal dos desenvolvedores (<username>.linea.gov.br)** - cada pesquisador ou tecnologista que desenvolve códigos e aplicações tem o seu próprio portal. Uma vez terminado o desenvolvimento e testado neste portal as mudanças são colocadas (*commits*) em um repositório central.
- **Portal de testes (testing.linea.gov.br)** - periodicamente estas mudanças são disponibilizadas neste portal que serve para testes de integração. Este é, em geral, o portal usado para validação de novas implementações. É um ambiente por definição instável mas que foi usado também para pesquisa pela dificuldade de se ter um ambiente de produção. Apenas 4 nós do cluster de processamento estão disponíveis para uso, o que limita os tipos de análises que podem ser feitas.
- **Portal de produção no LIneA (des-portal.linea.gov.br)** - este portal está sendo finalmente implantado. É um ambiente estável com acesso a 38 nós do cluster de processamento. Este portal será atualizado periodicamente com versões bem definidas e testadas da infraestrutura geral do portal, do sistema de criação de catálogos, e do *Data Server* acompanhados de documentação para benefício do usuário. Este portal deve servir de plataforma para atualização das instâncias mantidas nos centros internacionais.
- **Portal de produção no Fermilab (des-portal.fnal.gov)** – portal mantido no Fermilab contendo apenas o Data Server. Este portal é de uso de toda a colaboração e pode ser acessado por todos os pesquisadores afiliados ao DES registrados no *ldap* do projeto mantido pelo Fermilab.
- **Portal do NCSA (des-portal.ncsa.edu)** – em implantação.
- **Portal do CTIO** – disponível para todos os usuários da câmera DECam mas que só é acessível de dentro do observatório devido a restrições de segurança do *firewall* local.

A tarefa de manter estes vários portais sincronizados fica a cargo de um tecnologista sênior (*release manager*) e várias tentativas vem sendo feitas para simplificar e automatizar ao máximo o procedimento tanto para *builds* completos como para aplicação de *patches* para correções pontuais.

## 9.7 Avaliações internacionais

O desenvolvimento do Portal Científico tem sido avaliado periodicamente ao longo dos últimos anos por um painel internacional designado pelo diretor do DES seguindo o calendário e tópicos abaixo:

1. Outubro 2010 (Fermilab) – Introdução, *workflows* científicos.
2. Outubro 2011 (UPenn) – Precam, *Quick Reduce*, *workflows* científicos.
3. Maio 2012 (MPA) – *Quick Reduce*.
4. Julho 2013 (Fermilab) – Visão *end-to-end* e validação de dados.
5. Novembro 2013 (Fermilab) – Validação de dados e inspeção visual de imagens.
6. Agosto 2014 (Fermilab) – *Data Server*.
7. Novembro 2014 (Fermilab) – Criação de Catálogos
8. Maio 2015 (Fermilab) – Criação de catálogos e operação.

Seguem abaixo alguns comentários extraídos destas revisões:

- 2010 (Fermilab/EUA): *"The system is actually much more than a Science Portal, it is a comprehensive, web-based Science Analysis Center for the Dark Energy Survey Collaboration. **We don't know that anything quite like it exists yet it is forward looking and ambitious.**"*
- 2011 (UPenn/EUA): *"The vision of the Brazil portal is commendable and should go forward."*
- 2013 (Fermilab/EUA): *"The capabilities of the Science Portal have adapted directly the requests of users."*
- 2014 (Fermilab/EUA): *"The portal continues to steadily improve and offer more features that the Collaboration has requested and which are not provided in any other facility within DES."*
- 2015 (Fermilab/EUA): *"The overall conclusion is that the Data Server is providing important research facilities to the DES Collaboration. The evolution of the Portal capabilities is leading to a more operations-oriented view (Portal provides a controlled environment for the DESDM data release process). The Portal can and will still serve its other purposes, supporting down-stream science analysis (for example by creating Value-Added Catalogs and incorporating external data)"*

Os textos completos destas revisões que tiveram um relatório escrito podem ser encontrados nos apêndices relativos a esta seção.

Além destas revisões conduzidas pelo DES foi também solicitado a opinião de E. Bertin desenvolvedor de uns dos mais bem sucedidos softwares de astronomia (SExtractor) e de inumeras outras ferramentas para o projeto frances TERAPIX encarregado da redução de

dados do levantamento CFHT Legacy Survey. Sua avaliação está disponível no apêndice 9.7.

## 9.8 Novos projetos

Embora vários produtos já estejam no domínio público como os sistemas QR e o *Data Server*, ou num estágio avançado de desenvolvimento como a produção de catálogos científicos (*science-ready catalogs*) além de alguns dos *workflows* científicos, o trabalho de desenvolvimento continuará bastante intenso nos próximos cinco anos. A Tabela 9.1 lista os projetos necessários para o DES, que no momento, tem a maior prioridade tendo em vista que o levantamento acaba de concluir seu terceiro ano de observações dos cinco previstos.

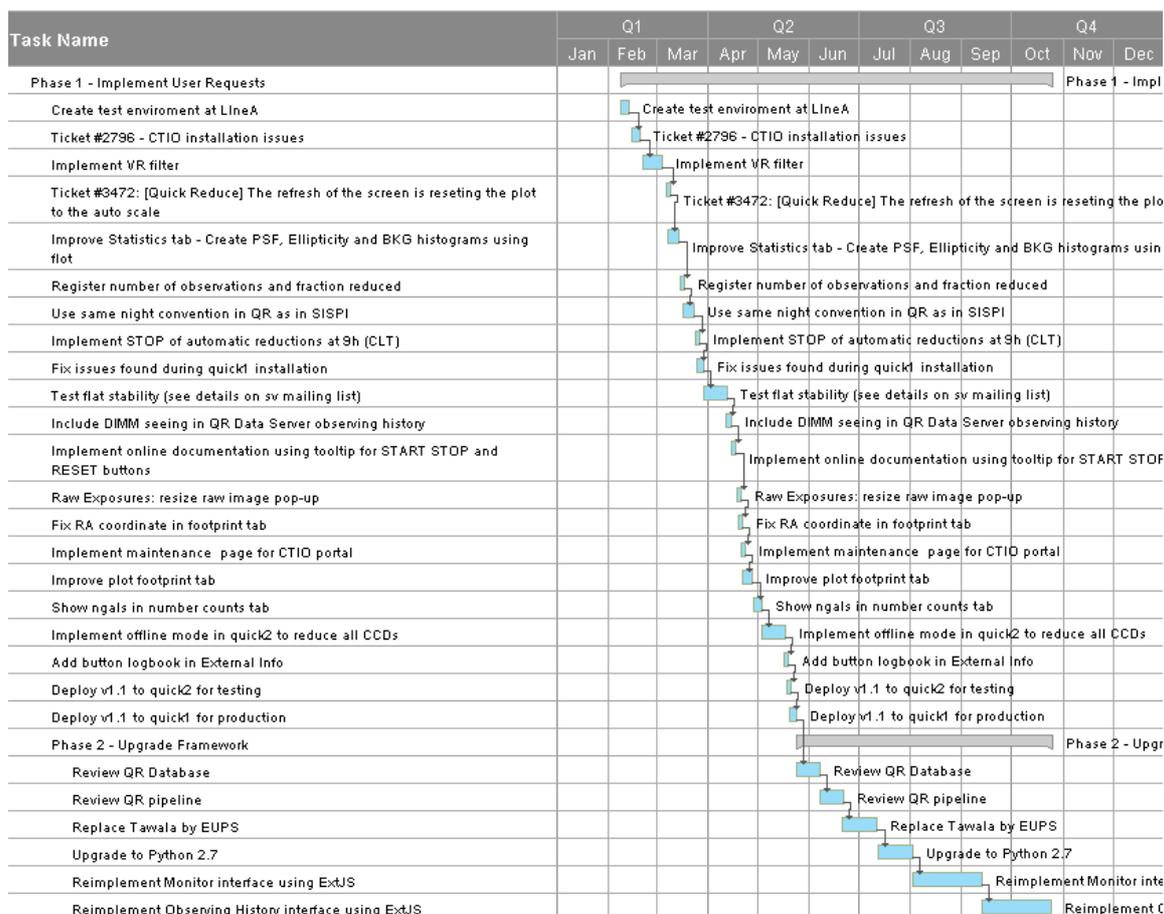
**Tabela 9.1** – Lista de Projetos de software para o DES.

<b>Software</b>	<b>Meses</b>	<b>Status</b>	<b>Projeto Alvo</b>
<i>Quick Reduce</i>	6	Em andamento	DES
<i>Portal infrastructure</i>	11	Em andamento	-
<i>Data Release interface</i>	10	Em andamento	DES
<i>Science-ready catalogs</i>	6	Em andamento	DES/LSST
<i>Science workflows (4)</i>	12	Iniciado	DES/LSST
<i>Parameter Estimation</i>	6	Iniciado	DES/LSST
Simulações	6	Iniciado	DES/LSST
<i>TNO pipeline</i>	6	Iniciado	DES/LSST
<b>Total</b>	<b>63</b>		

Cronogramas independentes para cada um desses projetos estão preparados evitando, numa primeira instância, paralelizar tarefas, como se apenas uma pessoa estivesse dedicada ao projeto. Os cronogramas incluem apenas o tempo estimado para o desenvolvimento. Geralmente, nesta fase não se inclui o tempo de validação e de correções de erros que podem significar pelo menos 30% do tempo de desenvolvimento. Além disso, surgem necessidades da inclusão de novos casos de uso que são difíceis de estimar.

A seguir são apresentados como ilustração alguns dos principais casos. Por exemplo, na Figura 9.12 mostramos o planejamento para implantar melhorias no *Quick Reduce* - Nela podem ser vistas duas fases distintas. A primeira estimada de 3 a 6 meses, procura atender pedidos de usuário e a segunda fase estimada em 6 meses para a modernização do sistema com novas tecnologias.

Outro projeto que é contínuo e tem seu próprio cronograma é o do desenvolvimento da infraestrutura do portal que inclui correções das mais variadas, além da adição de novas funcionalidades, algumas surgidas em função do modelo de operação adotado, outras simplesmente de caráter operacional. Um cronograma preliminar é apresentado na Figura 9.13 e requer da ordem de 11 meses para ser completado. Note que cada tarefa é acompanhada de explicações sobre o que precisa ser feito.

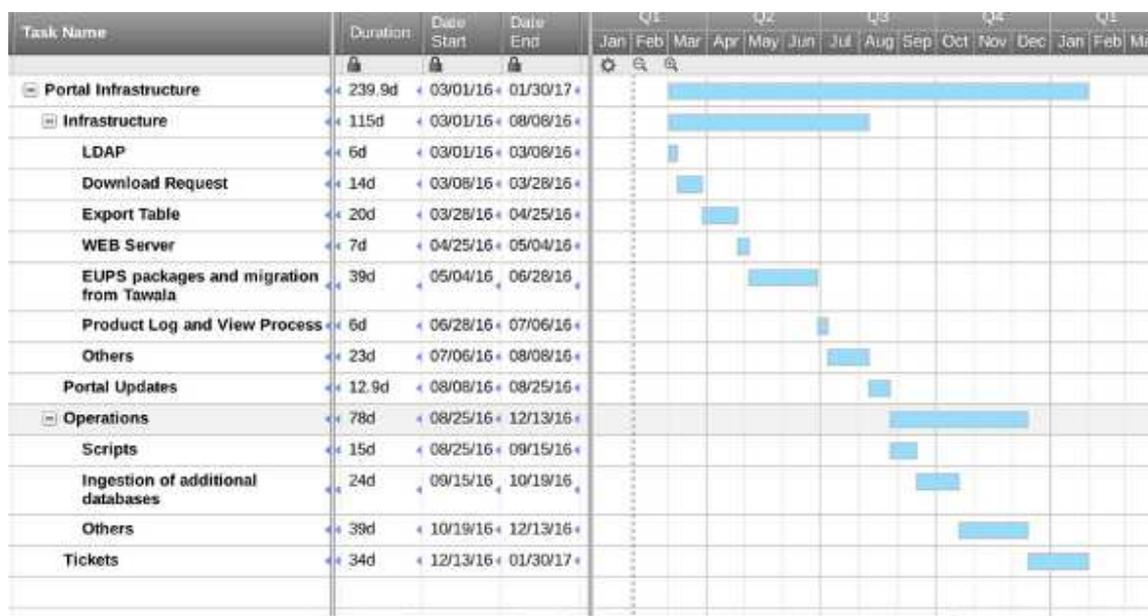


**Figura 9.12** – Visão detalhada do cronograma preliminar para o projeto Quick Reduce.

Um compromisso assumido com a colaboração DES e com uma proposta submetida ao NSF pelo NCSA é o de criar uma interface para prover acesso ao primeiro *Data Release* público, a ser feito pela colaboração em 2017. Esta interface é uma evolução do *Data Server* discutido acima em desenvolvimento há ~36 meses. O cronograma é apresentado na Figura 9.14, e seu desenvolvimento está estimado em 10 meses o que exige pelo menos a contribuição de 2 FTEs de tecnologistas para ser finalizado até 2017. Nesta estimativa não está incluído o período de *feedback* e *shakedown* que pode aumentar 30% a 50% do tempo estimado. Também não está incluído o tempo de migração para o NCSA. Estas estimativas são relativamente seguras tendo em vista a experiência adquirida nos últimos anos.

Com a entrada do grupo associado ao projeto de ocultação, novas demandas estão surgindo. A primeira já mencionada é a instalação do serviço de *Skybot* contendo previsões da localização de TNOs conhecidos no campo de uma dada imagem. A segunda demanda, é o *pipeline* a ser construído em colaboração com a equipe científica para identificar TNOs em imagens coletadas pela DECam provenientes do levantamento DES e DECals e prever futuras ocultações. Para isto foi construído o cronograma apresentado na Figura 9.15. O projeto está previsto para ser concluído em seis meses.

Manter este cronograma é extremamente importante tendo em vista a disponibilização pública dos dados do satélite GAIA no segundo semestre de 2016. A combinação do GAIA, que levará a um aumento significativo na precisão das posições das estrelas, e do DES, permitindo o acompanhamento de grandes áreas do céu por um longo período de tempo, levará a um aumento substancial no número de previsões de ocultações e isto pode revolucionar o estudo de objetos do sistema solar. A terceira demanda, ainda não analisada, é a necessidade de ingerir e analisar dados em diferentes formatos produzidos pela rede de observadores que participam do evento de ocultação que é o objetivo final deste projeto.



**Figura 9.13** – Cronograma preliminar para o projeto Portal Infrastructure.

Embora o desenvolvimento do Portal Científico tenha sido motivado pelas necessidades do DES a infraestrutura é genérica e adaptável a outros projetos e o conhecimento acumulado pela equipe de TI do LIneA permite alavancar o desenvolvimento de soluções para outros projetos.

A Tabela 9.2 lista outros projetos de interesse, alguns deles por razões estratégicas já foram até iniciados. A Figura 9.16 apresenta uma visão de alto nível do cronograma acordado entre o LIneA e o projeto *Dark Energy Spectroscopic Instrument* (DESI) para o desenvolvimento do sistema *Quick Look Framework* (QLF), a ser entregue em 2018. Este sistema será usado para avaliar a qualidade dos 15.000 espectros que serão obtidos em cada exposição a uma cadência de 20 minutos aproximadamente. O sucesso do *Quick Reduce* levou a colaboração DESI a adotar a arquitetura desenvolvida pelo LIneA. Um acordo entre o LIneA e o projeto DESI foi recentemente assinado e deverá permitir a participação de alguns pesquisadores vinculados ao laboratório nesse projeto.

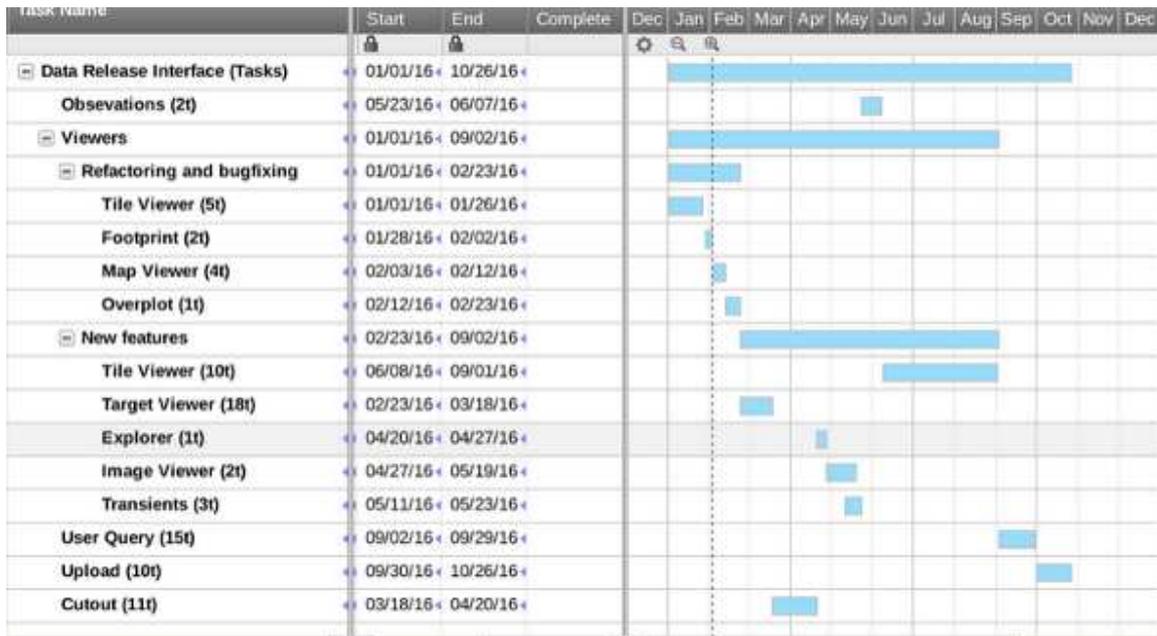


Figura 9.14 – Cronograma preliminar para o projeto Data Release Interface.

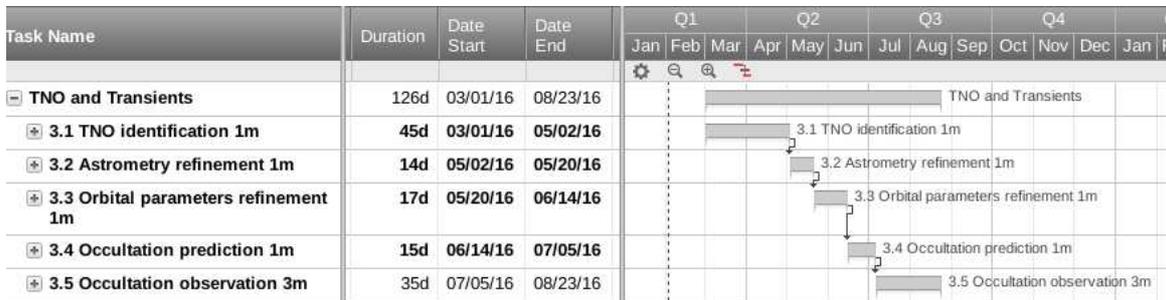
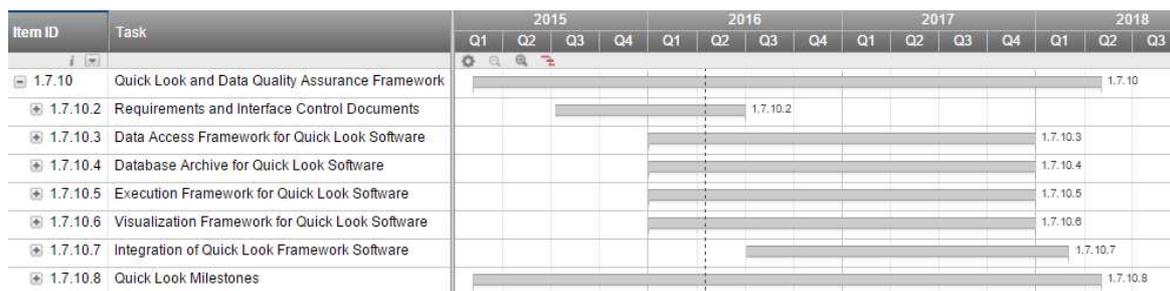


Figura 9.15 - Cronograma preliminar para o projeto TNO pipeline.

Tabela 9.2- Projetos futuros.

Software	Meses	Status	Projeto Alvo
Quick Look Framework	24	Em andamento	DESI
LSST Stack	12	Em andamento	-
Transients	12	iniciado	DES/LSST
Integração SDSS	6	Não iniciado	-
MaNGA pipeline	12	Não iniciado	SDSS
Qserv	24	Não iniciado	LSST
SOAR Archive	3	Não iniciado	LNA
<b>Total</b>	<b>87</b>		



**Figura 9.16** – Cronograma acordado com a colaboração DESI para o desenvolvimento do Quick Look Framework (QLF).

O portal também terá um papel fundamental para o *Brazilian Participation Group* do LSST, pois servirá para o desenvolvimento de algoritmos que permitam a eficiente exploração científica do grande volume de dados que serão gerados continuamente pelo projeto, ao longo de 10 anos. O uso de simulações e a experiência adquirida com os dados do DES serão fundamentais para garantir a participação pró-ativa da equipe brasileira e para este fim colaborações técnicas com o NCSA e o LSST foram estabelecidas.

A colaboração técnica com o grupo de *Data Management (DM)* do LSST incluiu a ida de um dos membros do time de TI do LIneA para realizar um estágio em Tucson/AZ. Existe grande interesse por parte do LSST nas ferramentas de *Quality Assessment (QA)* e exploração de dados que o LIneA desenvolveu para o DES e no *know-how* acumulado pelo time de TI do LIneA nos diferentes aspectos de gerenciamento e análise de dados.

A relação de trabalho já estabelecida entre o LIneA e o NCSA através do DES é muito favorável para esta colaboração tendo em vista que os dados do LSST serão processados no NCSA. Atualmente percebe-se uma transição natural do time de DM do DES no NCSA para o time de DM do LSST no NCSA e o LIneA é um braço importante do desenvolvimento de software no DES, como estabelecido nos acordos com NCSA e em propostas para o NSF. O LSST precisa de ferramentas de QA em vários níveis, para garantir a qualidade do software de redução de dados, para verificar a qualidade das imagens processadas (Nível 1) e para verificar a qualidade dos produtos dos releases anuais (Nível 2). Também foi estabelecido que as ferramentas de QA devem ficar disponíveis para o usuário final (Nível 3), o que requer interfaces que facilitam a utilização das mesmas. Um dos objetivos desta colaboração técnica é a construção de um protótipo para QA de um conjunto de *datasets* precursores com dados do DES, CFHT e HSC sendo processados com o *LSST software stack*. O *LSST software stack* foi desenvolvido para analisar dados de diferentes câmeras imageadoras e por isso o sistema de redução está sendo desenhado para ser genérico. Um passo nessa direção foi o modelamento de um *database* que cumpre com este requisito. Além do *database* de QA, os componentes de visualização e de workflow são igualmente importantes. A experiência anterior com o DES tem sido fundamental para avaliar e empregar novas tecnologias como é o caso da biblioteca *Python Bokeh* e *Datashader* que permitem a visualização e interatividade com 1 bilhão de *data points* no browser. O sucesso de tecnologias adotadas pelo LIneA no DES

também foi fundamental ao definir o uso do *Django* como *backend* e *ExtJS* como *frontend*. Na área de *workflows* científicos estão sendo investigadas alternativas para concatenar as diferentes tarefas de QA a fim de executá-las facilmente em diferentes ambientes de processamento. Com a extensão do acordo que formou o LIneA, será possível dar continuidade a este trabalho contando com o apoio do time de TI acelerando de sobremaneira o desenvolvimento de protótipos e convergindo nas soluções empregadas em produção, marcando definitivamente a contribuição brasileira para o LSST.

## 9.9 Metodologia de desenvolvimento

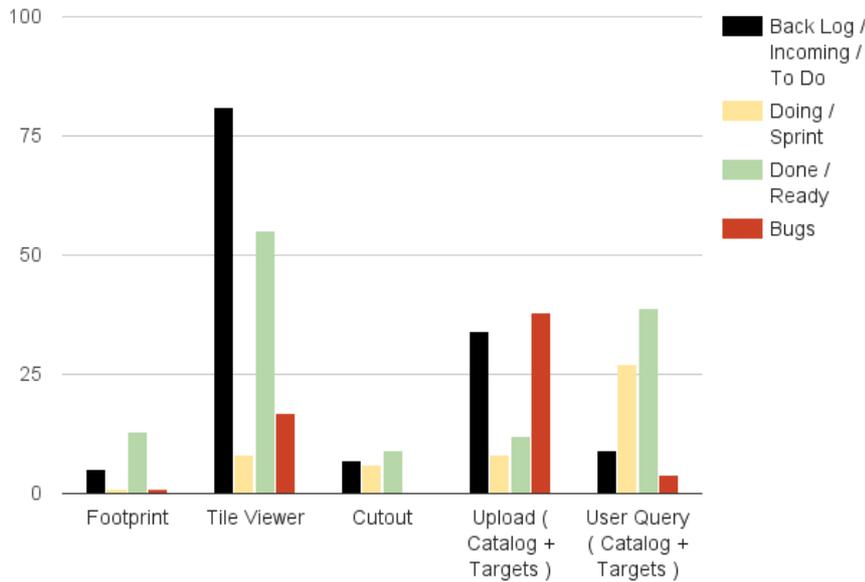
No início da construção do Portal, o esforço de desenvolvimento de software foi extremamente subestimado e, ao longo dos anos, várias tentativas foram feitas por um time sem experiência para formalizar o processo de desenvolvimento. Inúmeras metodologias e ferramentas foram testadas e descartadas. Apesar de um longo aprendizado com erros e acertos, os processos agora estabelecidos, foram adotados com algum grau de sucesso.

Os projetos técnicos e científicos são identificados e o escopo de cada um deles discutido em reuniões periódicas da equipe de coordenadores de TI, mais os pesquisadores interessados e o coordenador do laboratório. Quando um caso é considerado apropriado, solicitam-se maiores detalhes aos proponentes. Então, são identificados membros do time de TI cujos perfis mais se adaptam ao projeto. Embora o objetivo principal do portal seja atender a demanda científica, outros fatores devem ser levados em consideração. Portanto, para definir a prioridade dos projetos a serem desenvolvidos são considerados: 1) a importância científica e alinhamento com o status do projeto internacional e interesse da equipe brasileira; 2) os compromissos assumidos com os parceiros internacionais e o cronograma com as datas de entrega; 3) a criticidade operacional; 4) a disponibilidade do time de TI; 5) o tempo de duração do projeto.

Uma vez decidido o desenvolvimento de um projeto de software no LIneA, o processo segue alguns conceitos das metodologias ágeis com reuniões diárias no formato *stand-up* onde os membros dos times técnico e científico fazem um breve relato sobre as atividades e identificam interdependências nas suas tarefas. A partir de reuniões semanais por projeto, as tarefas são identificadas e planejadas ao longo de *sprints*. Ao término de cada *sprint* é realizada uma reunião de validação com os interessados quando novas tarefas podem ser definidas. O detalhamento e acompanhamento das tarefas é feito através do software *Trello* e o planejamento a médio e longo prazo através da ferramenta *SmartSheet* usada na criação dos cronogramas apresentados acima. Para melhorar a comunicação entre os membros do time é utilizada a ferramenta *Slack* que permite criação de canais de conversação permanentes. Documentos de requisitos são compartilhados no *Google Drive* e versões finais são mantidas no repositório de documentos *doc-db* desenvolvido pelo Fermilab.

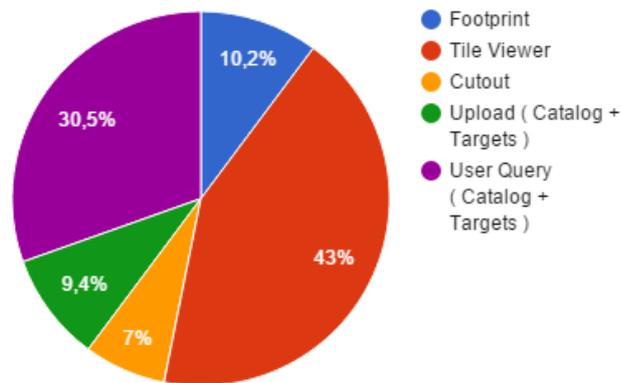
A Figura 9.17 dá uma ideia do número de tarefas envolvidas no desenvolvimento do *Data Server* por subsistema. Neste exemplo pode ser visto o número total de tarefas é da

ordem de 300 e como comentado neste documento a fração de correções não é desprezível, facilmente chagando a 30% do total de tarefas.



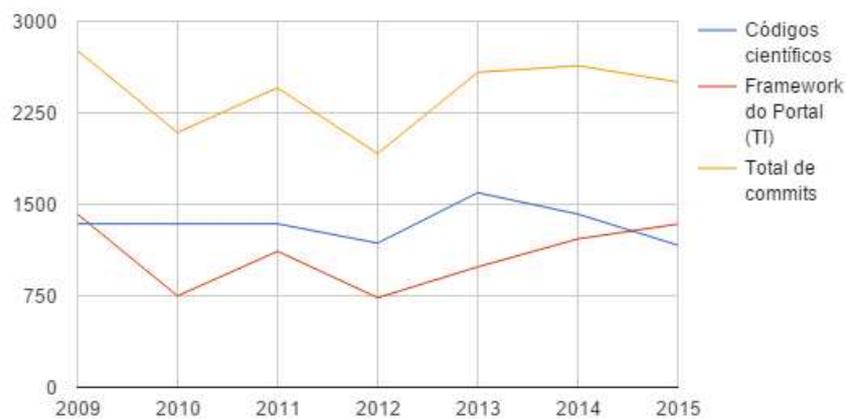
**Figura 9.17-** Distribuição de tarefas a serem feitas, em andamento, e feitas, além de correções para cada subsistema do Data Server a ser entregue para a colaboração no NCSA em 2017.

O uso do *Trello* ilustra áreas críticas no desenvolvimento de software. No exemplo acima, dois subsistemas correspondem a 70% das tarefas e a alocação de pessoas deve levar isto em consideração (ver Figura 9.18).

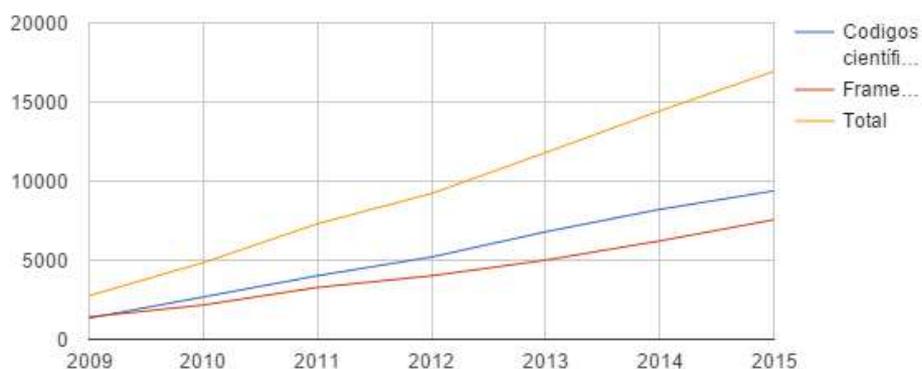


**Figura 9.18-** Distribuição das tarefas associadas aos subsistemas do Data Server.

A dimensão de um projeto e a produtividade da equipe de TI pode ser medida aproximadamente pelo número de mudanças no código (*commits*) efetuadas periodicamente pelos membros do time no repositório oficial. Isto é mostrado na Figura 9.19 em função do ano, ou o acumulado (ver Figura 9.20) onde também é indicada a natureza do código se trata de uma mudança "de infraestrutura" ou "de aplicação", embora a fronteira entre estas categorias seja bastante cinza. O importante é que o número de *commits* chega a mais de 2.250 por ano, é estável e corresponde a aproximadamente 6 por dia. Como eles são feitos principalmente pelos programadores, e o time atual conta com aproximadamente de 4 a 6 tecnologistas que fazem *commits* isto em princípio implica que em média um desenvolvedor é capaz de realizar da ordem de 5 tarefas por semana. Por outro lado, a experiência adquirida no monitoramento dos *sprints* sugere que um número mais razoável é de não mais de 4 tarefas/semana/programador. É este valor que é usado nas estimativas de duração de um projeto.



**Figura 9.19** - Número de submissões (*commits*) ao repositório representando modificações nos códigos do portal devido a novas implementações ou correções.



**Figura 9.20** - Número acumulado de submissões (*commits*) ao repositório representando modificações nos códigos do portal devido a novas implementações ou correções ao longo dos anos.

Com a estabilização da infraestrutura do portal e o crescimento do número de projetos atendendo a uma grande variedade de objetivos, é importante considerar a criação de um Comitê de Controle de Mudanças (*Change Control Board*) a exemplo do que é feito em outros grandes projetos para, não só discutir prioridades, como para minimizar efeitos colaterais de mudanças e garantir a aderência ao cronograma de atividades. Uma tentativa neste sentido já foi feita mas de forma prematura considerando a velocidade das mudanças do portal na época e a ausência da figura de um gerente que pudesse coordenar as reuniões e ter a visão global e atualizada do estado do desenvolvimento.

## 9.10 Documentação

Um importante legado do LIneA é a base de conhecimento sendo criada que na medida do possível é registrada em documentos para futura referência. A listagem a seguir mostra uma compilação da documentação técnica que inclui documentos de especificação da infraestrutura do Portal, ferramentas, pipelines e procedimentos de operação. Estima-se que a documentação existente cobre 75% das funcionalidades implantadas no portal e procedimentos de operação sendo de grande utilidade para o time técnico. No entanto, esta documentação precisa ser mantida e atualizada e o time atual não tem condições de fazê-la. Outros aspectos de documentação que são deficientes no projeto incluem a documentação de código e documentação voltada ao usuário. Isso precisa ser corrigido, tendo em vista a necessidade de tornar o código do Portal *open source*.

### Portal Infrastructure

1. Orchestration Layer
2. Garbage Collector
3. Configuration Manager
4. Search Engine
5. Data Sources
6. Export Infrastructure
7. Virtual Environment
8. Migration Tawala EUPS
9. Improvements on Portal Loading
10. Provenance Implementation
11. Publish Products and External Catalogs
12. Portal Notifications
13. Plano de trabalho RNP-LIneA
14. Implementacao do Science DMZ
15. Production migration plan
16. Refactoring Plan

### Quick Reduce

1. Quick Reduce Design
2. QR user requests

#### Catalog Infrastructure

1. Data Discovery
2. Data Installation
3. Data Installation Refactoring
4. Data Preparation
5. Building Science Ready Catalogs
6. DESDM Data Release tags
7. Catalog database design
8. Products database design
9. Pipeline inputs and outputs
10. E2E framework
11. Integrating simulations in the E2E framework
12. Status of pipelines
13. QA Coadd Design
14. QA Single Epoch Design
15. External Catalog Comparison
16. VAC infrastructure
17. VAC creation design

#### Data Release Interface

1. Portal viewers
2. Run Query Interface
3. Catalog Server Interface
4. Target Properties
5. Transients Viewer
6. Upload Targets
7. Upload Targets Refactoring
8. Upload Catalog
9. Data Release Products Interface

#### Science Workflows

1. Galaxy Archeology pipelines
2. LSS pipelines
3. WAZP cluster Finder pipeline
4. Galaxy Evolution pipelines

#### DESI Quick Look

1. DESI Quick Look Design

#### Operations Procedure

1. Roteiro de instalacao do Portal
2. Software development process at LIneA

3. Creating a new pipeline
4. Updating and creating new code
5. QR operations procedure
6. Updating production environment at LIneA
7. Updating production environment at FNAL
8. Adding new packages to EUPS
9. Entrada e saída de membros do LIneA
10. LIneA Newcomers Guide
11. Glossary
12. Ticket System
13. Submitting jobs in the cluster

Além dos 62 documentos listados acima estão sendo preparados cinco artigos sobre diferentes aspectos do portal entre os quais:

1. Analyzing Large Data Sets: the Science Portal Solution
2. The DES Data Release Infrastructure
3. DES Science Portal: Building Science Ready Catalogs
4. Quick Reduce: Quality Assessment of DECam Images in Real-time
5. DES Science Portal: Photo-z for Y1 data

Finalmente, um dos objetivos para 2016 é de se obter uma licença ou mesmo uma patente para o trabalho. Algumas alternativas já foram exploradas, mas a falta de pessoal impediu que este projeto fosse levado adiante.

## 9.11 A equipe técnica

Um dos maiores desafios enfrentados pelo LIneA tem sido a intensidade do esforço do desenvolvimento de software ditado pelas necessidades da área científica, os compromissos assumidos com as diferentes colaborações, e o amplo escopo de conhecimento exigido para o desenvolvimento do portal. Isto se reflete nos requisitos para a formação da equipe de TI do LIneA. As principais áreas de atividades do time de TI incluem programação web, desenvolvimento de ferramentas para a visualização de dados, gerenciamento de dados em bancos relacionais e em sistemas de arquivos distribuídos, processamento de dados em paralelo, desenvolvimento de *workflows* científicos, além de atividades de correção de erros, testes e operação.

Ao longo dos cinco anos a equipe de tecnologistas tem oscilado entre 8 e 11 membros, complementada por 3 a 4 pesquisadores que vem contribuindo também para a infraestrutura do portal principalmente na área mais próxima da ciência que é a criação de catálogos científicos onde existe uma área cinza do que é TI e do que é ciência. A Tabela 9.3 mostra o perfil da equipe que contribuiu para o desenvolvimento do portal ao fim de 2015, incluindo nove membros da equipe de TI (todos terceirizados) quatro cientistas de dados envolvendo um aluno de doutorado, um pós-doutorando com bolsa do CNPq, e dois pesquisadores, estes, os únicos contratados. Infelizmente, o número de analistas foi reduzido a partir deste ano em 1,5 FTEs pelo desligamento de uma analista (maternidade) e 0,5 FTE pela ida de um dos analistas mais ativos para Tucson para trabalhar metade do

tempo no projeto LSST. Isto foi um arranjo feito pelo LIneA por considerar importante a abertura de uma canal entre os projetos DES e LSST e o intercâmbio de conhecimento. Não só o número é pequeno, mas, como mostrado na Tabela 9.4, o número de membros com um determinado conhecimento é bastante reduzido tornando todo o esforço vulnerável a mudanças. Neste ponto é importante frisar que apenas quatro membros estão há mais de três anos no time, sendo que o mais antigo está há 7 anos. A partida de membros da equipe de TI ao longo dos anos teve um impacto no desenvolvimento do portal implicando na necessidade de longos períodos de aprendizado. Por outro lado, novos membros também trouxeram benefícios agregando experiência em novas tecnologias.

**Tabela 9.3** - Participantes no desenvolvimento do Portal (ano de 2015).

<b>Perfil</b>	<b>#</b>
Programadores	4
Ferramentas de visualização	2
Analistas de sistema	3
Cientistas de dados	4
<b>Total</b>	<b>13</b>

**Tabela 9.4** - Perfil de conhecimento dos membros da equipe de TI.

<b>Função</b>	<b>#</b>
Programação web	2
Ferramentas de visualização	2
Banco de dados	2
<i>Workflows</i>	4
Paralelização	2

Com estes números é possível estimar a capacidade de produção do LIneA. Por exemplo, no caso do *Data Server* 40% das tarefas exigem o conhecimento de ferramentas de visualização. Como o *Data Server* está estimado em 300 tarefas, e só a parte de visualização corresponde a 120 tarefas, isto implica que são necessárias 30 FTE-semanas, isto é 30 semanas para uma pessoa ou 15 semanas (4 meses) se para este projeto forem alocadas as 2 FTEs com o conhecimento necessário. Infelizmente, caso isto seja feito, será impossível levar adiante os projetos QR, QLF e TNOs que também exigem este tipo de conhecimento. Isto é o dilema de coordenar um time pequeno para o escopo do projeto.

Um ponto importante a ser frisado, é a necessidade urgente de um especialista no gerenciamento de projetos de software. Nos últimos anos isto tem sido feito desviando

pessoas de sua função, o que causa impacto no desenvolvimento, mesmo em regime de tempo parcial.

Embora o time de TI já tenha a experiência de construir cronogramas realistas para um projeto, o grande desafio tem sido montar um cronograma integrado para os 15 projetos listados acima de forma a cumprir os prazos estabelecidos pelos diferentes atores.

Usando o caso do *Data Server* como exemplo um projeto requer a execução da ordem de 15 tarefas por mês. Portanto, os projetos listados na Tabela 9.1 envolvem da ordem de 900 tarefas ou o equivalente a 60 FTE-meses. Para isto ser feito numa escala de seis meses seriam necessários 10 programadores com as necessárias características. São essas considerações que formam a base das necessidades apresentadas abaixo. Este tipo de análise também é fundamental para avaliar novos projetos e compromissos, sendo uma das funções da comissão de e-ciência proposta para o LIneA.

Novamente, neste contexto é que as vantagens de um centro de e-ciência devem ser consideradas. Em primeiro lugar, o centro poderia se encarregar de avaliar as necessidades de cada projeto, ajudar na escolha de profissionais, mantendo um corpo de colaboradores individuais ou de firmas de onde contratar, preparar e definir cronogramas, levando em consideração os vínculos do projeto. Por exemplo, cada projeto poderia incorporar pelo menos um cientista de dados e de alguma forma contratar o número de programadores adequados que seriam integrados ao time de TI do centro, no caso o LIneA. Haveria o benefício do aproveitamento da infraestrutura já montada, do compartilhando ferramentas, da experiência do time como um todo, e da orientação de membros seniores estáveis que preservassem o conhecimento em longo prazo. Desta forma o dimensionamento da equipe seria função da demanda científica, implicando numa racionalização dos recursos financeiros e humanos. Tentativas neste sentido estão sendo feitas com parcerias com firmas de desenvolvimento de software, mas os custos ainda são relativamente altos pela necessidade de terceirização.

## 10 Suporte aos usuários e outros serviços

Tendo em vista a complexidade da operação do LIneA ao longo do anos, várias ferramentas para trabalho colaborativo foram testadas e implementadas com o propósito de aumentar a eficiência do trabalho e para servir de documentação do trabalho sendo realizado. Atualmente os afiliados do laboratório tem acesso a um grande conjunto de ferramentas colaborativas, entre elas:

- **GoToMeeting** - sistema de web conferência a ser usado em *webinars*, reuniões de comitês, de subgrupos, e até reuniões virtuais do laboratório como um todo. Com esta facilidade, mantemos no âmbito do LIneA uma série de *webinars* de frequência semanal com convidados de grandes centros de pesquisa, ajudando a manter os corpos técnico e científico atualizados sobre as novidades internacionais.
- **Doodle** - mecanismo de agendamento de reuniões.
- **LIneA calendar** - todas as atividades do LIneA são mantidas num calendário compartilhado por seus membros.
- **Mailchimp** - sistema de gerenciamento de envio de e-mails para lista de assinantes e acompanhar os resultados de campanhas.
- **SurveyMonkey** - ferramenta usada para a realização de diferente enquetes.
- **Sistema de e-mail próprio** - o *webmail* mantido pelo LIneA tem se mostrado robusto e eficiente, possuindo sistema de backup diário e está disponibilizado aos integrantes do laboratório.
- **Servidor de listas** - usado para atender os diversos subgrupos, mantendo o histórico de comunicações (ex.: linhas de pesquisa, setor administrativo, comitês). Atualmente, mais de uma centena de listas são mantidas para atender as necessidades de comunicação.
- **Slack** - ferramenta para o envio de mensagens em tempo real, armazenamento e busca para comunicação entre membros de diferentes grupos de trabalho e o time de TI.
- **Overleaf** - para a redação de artigos em colaboração.
- **Wiki** - sistema de gerenciamento de conteúdo disponível para que os membros de cada projeto e subprojeto possam colocar informações relevantes e formar a base de conhecimento de fácil acesso aos membros (ex.: relatórios, resultados de pesquisa, minutas de reuniões, links importantes, pró-formas de compras).
- **Banco de documentos** - ferramenta desenvolvida pelo Fermilab que permite o gerenciamento de documentos (rascunhos de artigos, apresentações, relatórios técnicos e científicos, teses). Atualmente, mais de 300 documentos estão armazenados no sistema mantido pelo LIneA.
- **Repositório de códigos (GIT)** - um repositório de códigos centralizado. Este é um importante legado, pois além de permitir o desenvolvimento colaborativo, mantém o histórico dos algoritmos desenvolvidos por alunos e pós-doutorandos.
- **Smartsheet** – ferramenta para o gerenciamento de projetos e de recursos. Usada principalmente pela equipe de desenvolvimento de software para registrar e

acompanhar as atividades do laboratório a médio e longo prazo, indicando datas e metas a serem alcançadas.

- **Trello** - Um “quadro de avisos” virtual que permite o acompanhamento de subprojetos, suas tarefas, prazos, e membros envolvidos. O mesmo é útil tanto para atividades de pesquisa e desenvolvimento como administrativa, e vem sendo usado com sucesso pelos membros do LIneA.
- **Centro de Dados do LIneA** - Os afiliados ao LIneA tem acesso a esta facilidade para o processamento, armazenamento, e desenvolvimento de software.
- **Portal Científico no CTIO** – utilizado para acessar o pipeline *Quick Reduce* que avalia em tempo real a qualidade das imagens obtidas pela câmera DECam construída pela colaboração DES. A pedido do CTIO o software é usado não só pelo levantamento DES mas também por qualquer observador com a DECam. O QR já analisou mais de 500 mil exposições ou mais de 1,2 milhões de CCDs, ao longo dos últimos três anos. O software se mostrou extremamente robusto apresentando um número reduzido de defeitos e que foram resolvidos logo após sua instalação no início das operações em 10/2013.
- **Portal Científico no Fermilab** – um sofisticado subsistema do portal utilizado para a validação de imagens produzidas anualmente pelo NCSA combinando todas as imagens acumuladas nos anos anteriores. A interface também é usada para extrair informações do banco de dados, fazer a inspeção dos objetos selecionados usando as imagens disponíveis, e para a distribuição de catálogos e de simulações. Este sistema é o protótipo da interface que está sendo desenvolvida para instalação no NCSA, que deverá fazer a primeira distribuição pública dos dados do DES em 2017. Este é um compromisso assumido pelo LIneA num projeto apresentado ao NSF.
- **Portal Científico no LIneA** – o sistema completo que além das ferramentas disponíveis nos outros sites, possui ainda a infraestrutura para a construção de catálogos prontos para diferentes tipos de análise, inclui os *workflows* científicos e os *pipelines* para o cálculo dos parâmetros cosmológicos. No LIneA são mantidos um portal de testes e um de produção, além dos portais de cada desenvolvedor.
- **Helpdesk** - sistema de comunicação entre usuários e mantenedores da infraestrutura computacional (provida pelo LIneA) para atender as demandas do time científico, minimizando as dificuldades dos pesquisadores no uso dos sistemas.

Para apoiar as atividades do laboratório um membro do time de TI é responsável por vários aspectos de operação. Suas funções incluem:

- O registro de novos membros.
- A abertura de contas para uso dos computadores, no portal dando acesso ao repositório de códigos e na wiki.
- A inclusão de novos membros nas listas de e-mail apropriadas a administração destas listas.
- Compilação de novas bibliotecas requisitadas por usuários externos ou pelo time de TI.

- Criação de pacotes de software (EUPs).
- Manutenção do *web site* do LIneA.
- Manutenção de livro de registros de eventos com os equipamentos.
- Manutenção da lista de patrimônio e movimentação de máquinas.
- Servir de interface com firma de manutenção.
- Monitoramento independente do status das máquinas, temperaturas dos centros, e estado da rede.
- Monitoramento do estado de *backup* realizado diariamente.
- Monitoramento do estado dos portais no CTIO, Fermilab e NCSA (Figuras 10.1 a 10.4).
- Avisos sobre paralisações programadas do sistema.
- Reinicialização das máquinas e dos principais portais.
- Recebimento de novos equipamentos.
- Suporte para os *webinars*.
- Distribuição do *LIneA News*.
- Preparação de relatórios periódicos.
- Reunião semanal para revisão do funcionamento do centro de dados.
- Triagem de tickets submetidos ao *helpdesk*.

De ctio-portal@quick1.ctio.noao.edu

Assunto CTIO portal status (quick1) 2016-02-06

Para ctio-portal-notify@linea.gov.br

Current time: Sat Feb 6 08:00:01 CLT 2016

Status of CTIO portal running at quick1:

```
Science Portal is running...
Maestro is running...
Pardal is running...
SISPI monitor is running...
QR daemon is running...
QR monitor is running...
QR test environment is not running...
```

QR Reduction Mode: AUTOMATIC

**Figura 10.1** - Captura de tela de um e-mail enviado automaticamente do CTIO a cada 12 horas reportando o status do portal e o grau de ocupação do disco usado para a análise das imagens.

```

De Science Portal <noreply@portal.linea.gov.br>☆
Assunto [QR Data Server] Transferring data from to FNAL
Para Science Portal★

Data taken: 2016-02-06UT
Route: (quick1) > FNAL(fnal)

Method: UDT

Status: success
Number of files: 180
Number of failures: 0

Total volume: 1.19 GBytes
Transfer rate: 6.60 Mbps

```

**Figura 10.2** – Captura de tela de e-mail enviado automaticamente do Fermilab indicando o recebimento das informações do Quick Reduce que são disponibilizadas no Data Server.

```

De des-portal@fnal.gov.☆
Assunto Data Server portal @ Fermilab status (des31) 2016-02-06
Para fnal-portal-notify@linea.gov.br☆

Local time: Sat Feb 6 06:00:01 CST 2016

Status of Data Server at Fermilab:

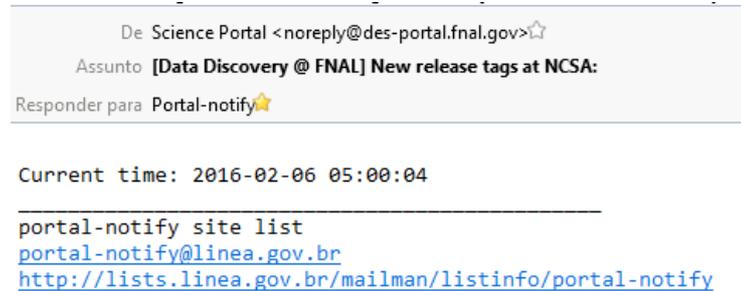
Science Portal is running...
Maestro is running...
Pardal is running...
Release is running...
Gavo is not running...
Register is not running...
Backup info:
The files are being stored at ms01:/mnt/backup/fnal

```

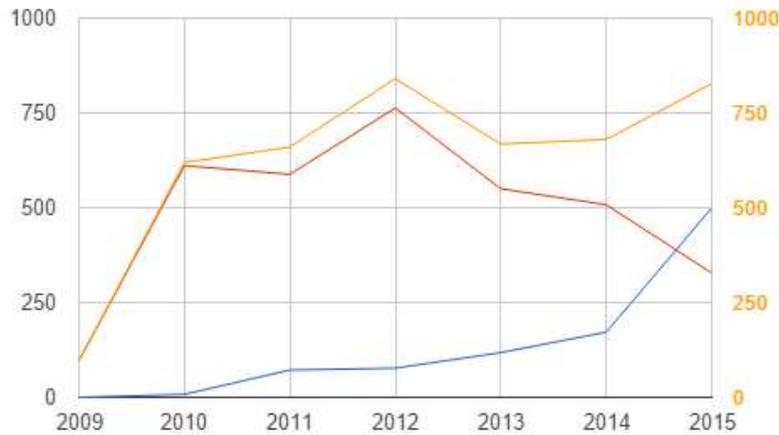
**Figura 10.3** – Screenshot de e-mail enviado automaticamente do Fermilab a cada 6 horas reportando o status do portal e os últimos acessos.

Existe a orientação de que todos os pedidos de serviço ou relato de problemas com as máquinas ou software sejam submetidos ao *helpdesk* que é monitorado não só pelo operador, mas também pelos coordenadores de TI, garantindo desta forma que alguma ação será tomada. A Figura 10.5 mostra o número de tickets submetidos por ano ao *helpdesk* bem como o número de tickets fechados e ainda aguardando resposta, enquanto a Figura 10.6 mostra as mesmas quantidades acumuladas ao longo dos anos.

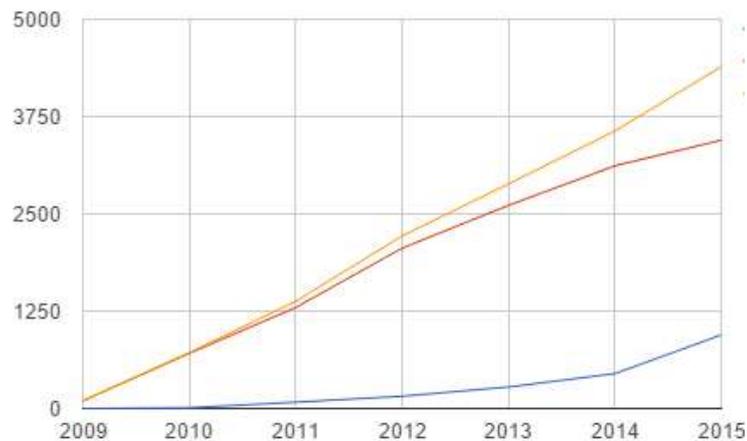
Destas figuras já foram eliminados os spams. O número total de tickets reais desde 2009 é ~4.000. Em média são recebidos 60 tickets por mês com variada complexidade.



**Figura 10.4** – Captura de tela de e-mail enviado automaticamente do NCSA (diariamente) notificando se algum novo release de dados está disponível.



**Figura 10.5** - Número de tickets abertos (azul), fechados (vermelho) e total (amarelo) por ano.



**Figura 10.6** - Número acumulado de tickets abertos (azul), fechados (vermelho) e total (amarelo).

## 11 Eventos

### 11.2 Participação

Um aspecto importante para assegurar a formação de jovens pesquisadores e para a permanente atualização dos pesquisadores mais seniores é a participação nos encontros internacionais realizados pelas colaborações. A frequência dessas reuniões gerais é de pelo menos uma (SDSS, DESI, LSST) ou duas (DES) vezes por ano, dependendo da colaboração, além de outras reuniões temáticas ou aquelas necessárias para discutir assuntos específicos. Por exemplo, reuniões sobre o portal como as revisões feitas anualmente por um painel indicado pelo Diretor do DES, a instalação de novos sistemas e a reunião anual do **Advisory Committee do Astrophysical Research Council** do SDSS onde é necessário a participação do representante institucional.

No período 2011-2015 pesquisadores e tecnologistas associados ao LIneA participaram de mais de 30 reuniões internacionais relacionadas aos projetos apoiados (DES, SDSS, DESI, LSST) sendo que em várias com a participação de mais de um membro tendo em vista a variedade de tópicos cobertos nestas reuniões.

Além das reuniões internacionais pesquisadores do LIneA também participam de reuniões nacionais entre elas da SAB, SBF, SBC e apresentações ao Conselho diretor do LIneA.

### 11.2 Organização

No período 2011-2015 o LIneA organizou as seguintes reuniões internacionais:

[XLDB Workshop South America 2014](#) – reunião organizada em colaboração com a Universidade de Stanford e o SLAC em 3-4 de Junho de 2014. Este *workshop* fez parte de uma série de encontros que procura reunir grupos interessados em encontrar soluções para os desafios de gerenciar grandes volumes de dados reunindo, acadêmicos e empresas da área de TI, e cientistas envolvidos em projetos de Big Data (Figura 11.1).

• [SDSS III International Collaboration Meeting](#) – reunião internacional da colaboração SDSS realizada no período 26-29 de Junho de 2012 (Figura 11.2).



HOME PROGRAM PARTICIPANTS VENUE TRAVEL ORGANIZATION PHOTOS CONTACT

## XLDB Workshop June 3-4, 2014 Observatório Nacional Rio de Janeiro, Brazil

Extremely Large Databases (XLDB) event series attempts to tackle challenges related to extreme scale data sets. Main activities include identifying trends, commonalities and roadblocks related to managing and analyzing extreme scale data sets, and facilitating development and growth of appropriate technologies including (but not limited to) databases. XLDB attendees include a mix of data-intensive industrial and scientific users, as well as big data academic researchers, and vendors.

Since 2007 the community has been meeting annually each fall at Stanford University, in California. In addition, satellite events are organized on different continents to connect with, and engage communities world-wide working on data-intensive challenges. Such events were organized in Europe and Asia in the past (Edinburgh/UK in 2011, Beijing/China in 2012, Geneva/Switzerland in 2013). Organizing next satellite event in Brazil would allow to engage the South America data-intensive community in the XLDB community.

Both the European and Asian communities found XLDB events very valuable, and expressed strong interest in organizing similar XLDB events again in the near future.

For more information about past events access [this link](#).

Sponsors:




**Tweets** Follow

**XLDB Rio 2014** @XLBDRio2014 5 Jun 14  
#xldb South America 2014 photos xldb-rio2014.linea.gov.br/photos/@XLBDBConf  
Expand

**XLDB Rio 2014** @XLBDRio2014 4 Jun 14  
Thanks to all participants and speakers. It was a great workshop! #xldb

**XLDB Rio 2014** @XLBDRio2014 4 Jun 14  
#XLDB South America is finished... or is it? The data must flow...

**XLDB Rio 2014** @XLBDRio2014 4 Jun 14  
global.gotomeeting.com/meeting/join #xldb is online. Broadcasting live!  
Expand

**XLDB Rio 2014** @XLBDRio2014 3 Jun 14

Tweet to @XLBDRio2014

**Figura 11.1** – Chamada para o workshop XLDB organizado em colaboração com o SLAC para discutir assuntos relacionados à Big Data. A reunião contou com o patrocínio da Huawei e SGI e teve a participação de vários grupos brasileiros que poderiam se beneficiar de um Centro de e-Ciência.

**SDSS-III**  
Collaboration Meeting  
June 26-29, 2012

HOME | GENERAL INFORMATION | VENUE | PROGRAM | REGISTRATION | PARTICIPANTS | GROUP PHOTO | CONTACT

**Observatório Nacional**  
Rio de Janeiro - Brazil

have not yet registered, do it ASAP.

**03/26/2012** - Registration is now open.

**03/15/2012** - If you want to receive regular updates about the meeting, please follow us on Twitter @sdss3rio

**03/15/2012** - Welcome to the SDSS-III collaboration meeting website! Registration will be opened in 1-2 weeks.

**02/13/2012** - We invite the members of the CoCo to send comments about the site.

**02/13/2012** - The Brazilian Participation Group is proud to host the fourth collaboration meeting of the Sloan Digital Sky Survey III at the Observatório Nacional.

**APOGEE**

**BOSS**

**MARVELS**

Star	Mass ( $M_{\odot}$ )
Sun	1.0
MCD001	25 $M_{\odot}$
MCD002	30 $M_{\odot}$
MCD004	40 $M_{\odot}$
MCD006	60 $M_{\odot}$
MCD007	91 $M_{\odot}$
MCD011	62 $M_{\odot}$
MCD012	39 $M_{\odot}$

**SEGUE-II**

Ministerio de Ciência, Tecnologia e Inovação  
**BRASIL**  
PAÍS SECO E PAÍS SEM FOME

LInEA  
Laboratório Interdisciplinar de Astrofísica

CBPF  
Centro Brasileiro de Pesquisas Físicas

Laboratório Nacional de Computação Científica

Observatório Nacional

CNPq  
Conselho Nacional de Desenvolvimento Científico e Tecnológico

FAPERJ  
Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro

FINEP  
Financiadora de Estudos e Projetos  
Ministério de Ciência, Tecnologia e Inovação

**Figura 11.2** – Chamada para a reunião internacional da colaboração SDSS-III organizada pelo BPG apoiado pelo LInEA.

Outras reuniões organizadas pelo mesmo grupo antes da criação oficial do LInEA foram:

- [DES International Collaboration Meeting](#) – reunião internacional da colaboração DES realizada no período 26-29 de Maio de 2009 (ver Figura 11.3).
- [New Astronomy: The Data Challenge](#) – reunião internacional em 29 de Maio de 2008, cujo principal enfoque foi de Big Data prevista pelos grandes projetos astronômicos como os levantamentos PANN-STARS, DES e LSST entre outros, e organizações como AMPATH signatária do recente acordo entre LInEA, LNA, RNP, ANSP e LSST. Isto

demonstra que o interesse do LIneA no LSST e sua recente entrada vem de longa data assim como as negociações.

• A Glimpse into the Future of Astronomy: Celebrating 180 Years of Astronomy in Brazil – reunião internacional realizada nos dias 27-28 de Maio de 2008, para discutir os principais projetos da década. Na reunião foram discutidos projetos como os grandes telescópios óticos (GMT, TMT, E-ELT), de rádio (ALMA, SKA), X-ray (Constellation-x) e os levantamentos incluindo o SDSS, PAN-STARRS, DES e LSST. Esta reunião serviu para despertar o interesse da comunidade brasileira nesses grandes projetos e iniciar a discussão sobre alternativas de participação neles. Além destas reuniões internacionais o LIneA organizou mais de 15 reuniões com seus membros entre reuniões gerais (pelo menos uma ao ano) e reuniões com temas científicos ou técnicos mais específicos. Estas reuniões complementam aquelas semanais ou quinzenais organizadas por diferentes grupos de trabalho. Uma lista destas reuniões é apresentada na web site do LIneA.

Membros do LIneA também participaram e ajudaram na organização da trilha de astronomia do BRESCI - Brazilian e-Science Workshop do Congresso da Sociedade Brasileira de Computação (CSBC) nas seguintes reuniões:

- 2012 Curitiba, PR
- 2013 Maceió, AL (Palestrante convidado: Jacek Becla, SLAC)
- 2015 Brasília, DF

The Dark Energy Survey

International Collaboration Meeting  
Rio de Janeiro, Brazil, May 26-29, 2009

DES-Brazil  
Dark Energy Survey

Home  
General Information  
Program  
Registration  
Travel  
Accommodations  
Maps  
Announcements  
Outreach

**The Dark Energy Survey International Collaboration Meeting**  
Rio de Janeiro, Brazil, May 26-29, 2009

Welcome to the DES International Collaboration Meeting

Local Organizing Committee

Marcio Maia	ON (Chair)	maia@on.br
Martin Makler	CBPF	martin@cbpf.br
Paulo Cabral Filho	LNCC	cabral@lncc.br
Paolo S. Pellegrini	ON	pssp@on.br
Simone Franco	LNCC (Secretary)	simone@lncc.br

Ministério da  
Ciência e Tecnologia

BRASIL  
CENTRO BRASILEIRO DE RECURSOS COMPUTACIONAIS

**Figura 11.3** - Portal da reunião internacional da colaboração do DES realizado no Rio de Janeiro.

## 12 Divulgação

Como parte do esforço de atrair jovens talentos e colaborar na formação científica da população o LIneA procura divulgar seu trabalho de várias formas entre as quais:

- Blogs - o site [www.linea.gov.br](http://www.linea.gov.br) é usado para a publicação periódica sobre eventos ou resultados de destaque relacionados as atividades do laboratório e de seus associados. O formato é propositalmente feito livre para despertar o interesse do leigo que é o público alvo.
- Anúncio de *Webinars* - enviado para uma lista de mais de 700 recipientes anunciando a realização de um seminário e disponibilizando um link para o sistema *GoToMeeting*. Um exemplo do conteúdo desses anúncios enviados pelo sistema de mala direta *mailchimp* é apresentado na Figura 12.1.
- LIneA News (Notícias) - um compilação de blogs e da agenda de *webinars* enviados periodicamente aos assinantes da mala direta. Como exemplo, a Figura 12.2 mostra o último número de 2015.
- LIneA na mídia – compilação feita de todas as referências ao laboratório na mídia.
- Galeria de Imagens – para atrair o grande público um esforço também é feito de colocar no site do LIneA imagens astronômicas selecionadas do levantamento fotométrico DES (<http://www.linea.gov.br/060-divulgacao/2-galeria-de-imagens/>).
- Apresentações públicas – Palestras públicas organizadas pelo LIneA como a proferida por Michael Wood-Vasey do SDDS-III, no Planetário do Rio de Janeiro, incluindo tradução simultânea com ampla participação do público leigo.
- *Press Releases* – ocasionalmente também são preparados *press releases* como recentemente feito com a assinatura do acordo com o LSST.

Um resumo em números destas atividades acima é apresentado na Tabela 12.1. Além disso, o LIneA mantém uma presença nas redes sociais com contas no Twitter para o público em geral e para profissionais, integrado aos portais de produção e Femilab, e no Facebook.

O LIneA tem colaborado com o CEFET no contexto da Semana Nacional de Ciência e Tecnologia participando dos seguintes eventos de divulgação:

- Semana de Extensão, 2013, CEFET RJ: Aplicações do portal científico do DES.
- II Workshop de Computação Aplicada, 2014, CEFET/RJ: Surfando os tsunamis do e-Universo.
- III Workshop da Escola de Informática & Computação, 2015, CEFET/RJ: Luz, Câmera, 5 anos.

O LIneA também faz um esforço para atrair novos talentos através de um programa de divulgação sendo estabelecido em colaboração com o grupo de *Education and Public Outreach* (EPO) do LSST em conjunto com planetários internacionais. Neste sentido o Planetário-RJ está aderindo ao projeto de INCT submetido pelo laboratório (apêndice 15.1) e deverá participar ativamente do desenvolvimento deste programa. Uma

ferramenta desenvolvida pelo LineA (target viewer) já despertou interesse por parte do grupo de EPO do LSST.




**Friday, Dec 04, 2015** ⌚ **4:00 PM - BRT** WATCH WEBINAR

Webinar will take place exceptionally on Friday

## Astrostatistics: Opening the Black Box



**Jake VanderPlas**  
*University of Washington eScience  
Institute*

The large datasets being generated by current and future astronomical surveys give us the ability to answer questions at a breadth and depth that was previously unimaginable. Yet datasets which strive to be generally useful are rarely ideal for any particular science case: measurements are often sparser, noisier, or more heterogeneous than one might hope. To adapt tried-and-true statistical methods to this new milieu of large-scale, noisy, heterogeneous data often requires us to re-examine these methods: to pry off the lid of the black box and consider the assumptions they are built on, and how these assumptions can be relaxed for use in this new context. In this talk I'll explore a case study of such an exercise: our extension of the Lomb-Scargle Periodogram for use with the sparse, multi-color photometry expected from LSST. For studies involving RR-Lyrae-type variable stars, we expect this multiband algorithm to push the effective depth of LSST two magnitudes deeper than for previously used methods.

*The seminar will be broadcasted via web and is limited to the 20 first remote attendees. Room will be available 15 minutes before the talk, when you can click WATCH WEBINAR to attend it or <http://goo.gl/oaLc8Z>. System requirements are [here](#). ON locals can gather at the conference room in LineA/GPA building.*

**Figura 12.1** – Exemplo do aviso do webinar da semana enviado para uma lista de mais de 700 pessoas atuando nas áreas de astronomia, física e TI.



LineA News N° 16, 2015

## Notícias

---



**LineA assina acordo com o revolucionário projeto LSST**  
 14 de Dezembro  
 O Large Synoptic Survey Telescope (LSST) é um telescópio sendo construído em Cerro Pachón situado nos Andes chilenos. Previsto para entrar em operação no início da próxima década, o LSST irá... [Leia mais →](#)



**LineA coloca no ar nova Home Page**  
 14 de Dezembro  
 A Home Page do LineA está de visual novo. Além de ampliar o conteúdo sobre os levantamentos apoiados e as atividades científicas e de Tecnologia de Inovação, temos uma Galeria de imagens e Notícias. Venham nos visitar em [www.linea.gov.br](http://www.linea.gov.br). [Leia mais →](#)

**Figura 12.2-** Exemplo das notícias enviadas a uma lista de mais de 700 pessoas divulgando os blogs e os convidados para a série de webinars patrocinados pelo LineA.

**Tabela 12.1** – Divulgação em Números.

Atividade	#
Blogs	89
LineA News	16
LineA na mídia	66

Embora importante, a divulgação científica é uma tarefa que consome tempo e exige não só uma coordenação, mas a colaboração de todos os membros do LIneA sendo um exemplo das responsabilidades administrativas que devem ser assumidas pelos participantes.

## 13 Acordos e compromissos

### 13.1 Acordos Projetos Internacionais

Ao longo dos anos o LIneA e o projeto estruturante Astrosoft do ON tiveram que negociar e assinar vários Memorandos de Entendimento para viabilizar a entrada de pesquisadores nos projetos internacionais entre os quais:

- DES – memorando para entrada do consórcio DES-Brazil em 2007 envolvendo contribuição financeira e de trabalho. Isto deu origem ao projeto do Portal Científico que agora inclui o *pipeline* de avaliação dos dados QR, uma interface para a distribuição pública dos dados em 2017 e ao fim do projeto e a infraestrutura para a criação de catálogos para uso científico. O fim do projeto está previsto para 2020-2021 e até lá o LIneA tem o compromisso de dar sustentação a equipe científica brasileira, como também dar manutenção aos softwares desenvolvidos pela nossa equipe para a colaboração.
- SDSS-III – memorando para entrada do consórcio *Brazilian Participation Group* (BPG) em 2008 envolvendo contribuição financeira e de trabalho. Os recursos financeiros foram provenientes do ON e a contribuição em trabalho significou a implantação de um centro de distribuição de dados específico para o projeto complementando aquele existente na Universidade de *Johns Hopkins*. A idéia é a de atender as necessidades dos pesquisadores locais e servir como alternativa ao site principal. O grupo BPG foi formado por pessoas que manifestaram interesse após consulta pública, inclusive com uma apresentação a todos os astrônomos do ON. O projeto terminou oficialmente em dezembro de 2015, um ano após a conclusão das observações.
- SDSS-IV - memorando para a continuação da participação do BPG em 2014. Novamente este acordo foi uma combinação entre trabalho e recursos financeiros o que tornou possível uma redução de US\$ 400 mil na taxa de entrada. A contribuição do LIneA inclui a melhoria do desempenho do sistema de distribuição de dados e a instalação e monitoramento de computadores no telescópio do *Las Campanas Observatory*. Uma chamada pública foi feita a comunidade astronômica brasileira com diversos pesquisadores manifestando interesse (vide apêndice 14.3). O projeto está previsto de terminar em 2021 um ano após a finalização das observações.
- LSST – memorando assinado em 2015 cuja responsabilidade é definir, em conjunto com o LNA, um processo para a escolha de pesquisadores para formar um grupo de participação brasileiro e dar suporte a este time em seu trabalho de se preparar para as observações que se iniciam a partir de 2021 e terminarão não antes de 2032.
- DESI – carta de intenções do diretor do projeto propondo os termos para a entrada de pesquisadores brasileiros. No momento um MOA está sendo negociado em troca do desenvolvimento do sistema *Quick Look*.

A documentação relativa a estes vários memorandos está disponível no apêndice 13.1. Como pode ser visto existe uma continuidade de projetos e o aprendizado científico e técnico nos projetos atuais, somados a experiência operacional adquirida pelo LIneA nos últimos cinco anos, serão de enorme importância para a formação de novas lideranças científicas e o melhor aproveitamento científico dos dados usando novas tecnologias e expandindo a infraestrutura disponível

### 13.2 Acordos Técnicos

Além desses memorandos de entendimento outros acordos técnicos foram estabelecidos com os projetos DES, DESI, LSST e com o NCSA e a Universidade de Illinois. Estes acordos de colaboração técnica são fundamentais para consolidar as conquistas do LIneA e acelerar sua preparação para a era do LSST. Em particular, a criação de um Centro Regional de Dados do LSST no Brasil é um passo essencial nesta direção e agora é o momento de definir uma estratégia para atingir este objetivo. É bom lembrar que a implantação deste centro é uma condição necessária para uma possível expansão do número de participantes brasileiros. Ao todo no apêndice 13.2 seis acordos são apresentados, em ordem alfabética, inclusive um detalhamento do que os projetos DES e DESI esperam do LIneA nos próximos anos.

### 13.3 Solicitações de Serviços

Além dos memorandos de entendimento e os acordos técnicos mencionados acima, ao longo dos anos o LIneA vem acumulando vários pedidos de apoio e solicitações relativas a manutenção e implementação de novas funcionalidades nos produtos já entregues que devem ser encarados como compromissos do laboratório. No apêndice 13.3 são apresentados 12 dessas solicitações. Estas incluem cartas:

- Do diretor da AURA no Chile, Diretor do CTIO e *DECam Instrument Scientist* solicitando o uso do QR por toda a comunidade
- Do diretor do DES, e líderes de diferentes grupos de trabalho (*Strong lensing, Clusters, Transientes and Moving Objects*) solicitando apoio das ferramentas desenvolvidas.
- Do diretor do SDSS-IV sobre o serviço *SkyServer/CasJobs* mantido pelo LIneA.
- Do diretor do DESI sobre o software *Quick Look Framework*.
- Da direção do LNA para apoio ao armazenamento de dados do SOAR.
- Do líder do Grupo de Ocultação solicitando apoio para a identificação de TNOs nas imagens da DECam.

O simples número e diversidade de solicitações mostra a relevância e o impacto que os produtos de software desenvolvidos pelo LIneA tem. É importante salientar que isso só foi

possível graças à criação de um grupo especializado e que teria sido impossível ser feito no âmbito de uma UP tradicional, pois requer conhecimentos de astronomia, de TI e de cientistas com a vivência necessária e interesse científico para definir os objetivos e coordenar os trabalhos. Esta é a singularidade do modelo do LIneA que só é possível criando uma saudável mistura de talentos e o que o torna tão essencial na era da ciência de grandes volumes de dados.

## **14 Manifestação de Apoio**

### **14.1 Cartas Institucionais de apoio ao LIneA**

Em várias oportunidades ao longo dos anos de funcionamento do LIneA, expressões de apoio a suas atividades como um todo ou em relação a algum projeto do laboratório foram feitas por representantes institucionais. No apêndice 14.1 podem ser achadas estas manifestações. Alguns exemplos são:

- Comitê Executivo do LIneA.
- Conselho do Instituto de Física da UFRGS.
- Instituto de Física da UFRGS.
- Instituto de Física da USP.
- Pós-Graduação do Observatório do Valongo/UFRJ.
- Pró-reitoria de Pós-graduação e Pesquisa da UFSM.
- Pró-reitoria de Pesquisa da UNESP.

### **14.2 Cartas Individuais de apoio ao LIneA**

Além das cartas institucionais 9 cartas de pesquisadores afiliados a sete diferentes instituições nacionais e estrangeiras (AIP, ON, UFF, UFRJ, UFRGS, UNESP, USP) e em diferentes estágios da carreira foram enviadas ao LIneA manifestando apoio as suas atividades e ao trabalho sendo realizado.

### **14.3 Outras cartas de apoio**

Cartas de apoio ao projeto LSST do LNA/LIneA e da Sociedade Astronômica Brasileira e à Secretaria Executiva do MCTI (apêndice 14.3)

Cartas de Apoio ao projeto submetido pelo LIneA ao CNPq pleiteando um INCT para o qual contribuíram dez instituições nacionais e internacionais entre as quais: CEFET; LNCC; LSST Data Management; LSST Education and Public Outreach; ON; Planetário do Rio de Janeiro; RNP; SLAC UFRGS e UNESP. Estas cartas estão incluídas no apêndice 14.4.

Cartas de apoio ao projeto Pronex submetido à FAPERJ dos quais fazem parte as firmas INFAX, SGI e SLACAM. Estas cartas estão incluídas no apêndice 14.5.

## 15 Finanças

### 15.1 Orçamentos Consolidado

Com relação à preparação do orçamento anual do LIneA o acordo de sua criação especificava:

***“A previsão orçamentária anual do LIneA será elaborada pelo Comitê de Gestão e submetido à aprovação do Conselho.***

***Parágrafo Primeiro*** – O orçamento deverá especificar os recursos que poderão ser providos pelas agências usuais de fomento (CNPq, FAPERJ, FINEP, etc.), o aporte que deverá ser solicitado diretamente a cada uma das UPs signatárias deste Acordo e recursos extra orçamentários, a serem solicitados diretamente ao MCT.

***Parágrafo Segundo*** – Os recursos orçamentários provenientes de cada uma das UPs signatárias a serem destinadas ao LIneA concorrerão em igualdade de condições com os demais projetos desenvolvidos nas três instituições.

***Parágrafo Terceiro*** – Caberá aos pesquisadores participantes do LIneA elaborar os projetos acordados para as agências de fomento

Infelizmente, a realidade se mostrou bem diferente. Em geral, os signatários apoiaram fortemente os pedidos para as agências de fomento e solicitações extra orçamentárias (TED) ao MCTI. Entretanto, a participação do LIneA no orçamento das respectivas instituições foi pequeno até mesmo no ON, cuja missão é a mais próxima das atividades fim do LIneA, e aonde o laboratório participa de um projeto estruturante.

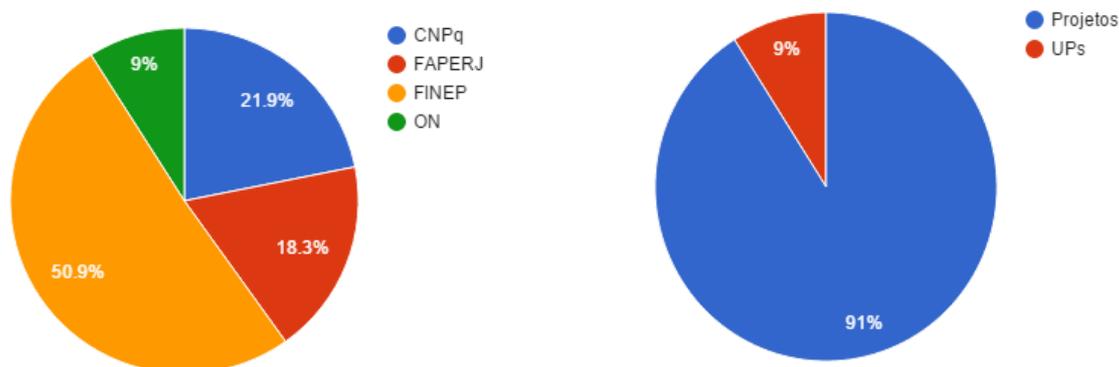
Por outro lado, na discussão abaixo também será considerada, na medida em que seja possível quantificar, a contribuição não financeira das instituições que proporcionaram ao laboratório apoio administrativo ou logístico e importantes oportunidades. Também importante é separar o custo da operação do laboratório do apoio à participação de pesquisadores brasileiros no projeto.

A Tabela 15.1 resume a contribuição financeira dada pelas diferentes organizações envolvidas para a operação do laboratório. A contribuição do CNPq inclui uma ação transversal e bolsa de produtividade, mas não inclui bolsas de pesquisadores associados que foram usadas para pagamento de despesas em suas respectivas universidades nem viagens suas e de seus alunos a reuniões das colaborações internacionais. A contribuição da FAPERJ inclui bolsa de Cientista do Estado e auxílios de pesquisa como PRONEX e Apoio a Instituições Sediadas no Estado. Também não estão incluídas contribuições da FAPESP apoiando os pesquisadores paulistas associados.

Tabela 15.1 – Contribuição Financeira para a operação do LIneA.

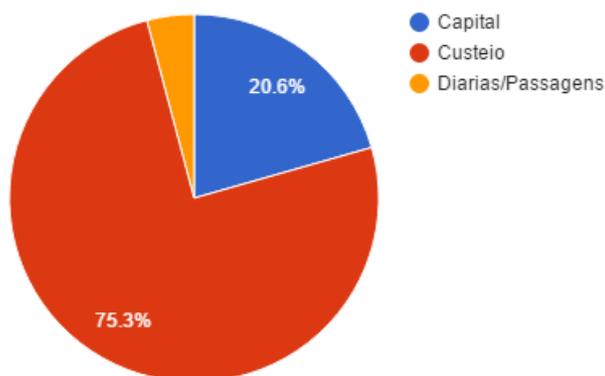
Organização	Reais
CNPq	1.928.147,50
FAPERJ	1.611.964,39
FINEP	4.487.266,02
ON	794.793,49
<b>Total</b>	<b>8.822.171,40</b>

Considerando apenas as dotações financeiras para operação, o total investido pelo LIneA no período foi de **R\$ 8,8 milhões**, sendo que a fração contribuída por cada agência ou UP é apresentada na Figura 15.1 (painel esquerdo). Nesta figura pode ser visto que a maior contribuição ao longo dos cinco anos considerados neste relatório vem da Finep (50%) seguida pelo CNPq (22%). FAPERJ (18%) e ON (9%). O painel direito da Figura 15.1 mostra que 91% do custo de operação do laboratório (pessoal e equipamento) é coberto por projetos e não com recursos orçamentários.



**Figura 15.1** – Distribuição das contribuições financeiras para a operação do LIneA mostrando a contribuição por organização (painel esquerdo) e separando projetos e orçamento.

A Figura 15.2 mostra como estes recursos foram gastos. Como pode ser visto a partir da Figura 95% dos recursos foram utilizados para as atividades técnicas do laboratório e infraestrutura com 20% indo para a compra de equipamentos e 75% para o pagamento de pessoal envolvido no desenvolvimento de software e instalação e manutenção da infraestrutura física. Isto é feito através das firmas contratadas para este fim, em alguns casos com considerável overhead.

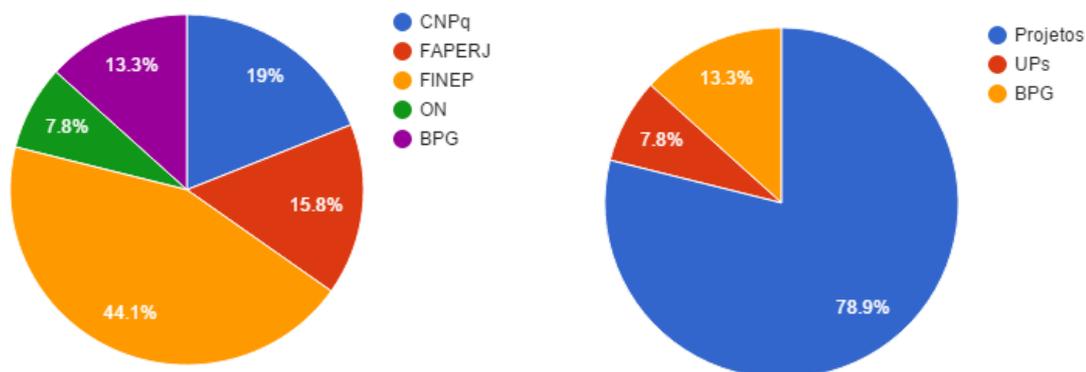


**Figura 15.2** – Destino do orçamento do LIneA nos cinco anos de operação.

Isto não inclui todas as contribuições. Além das mencionadas acima deve ser incluída a contribuição para o pagamento da participação brasileira no projeto SDSS que no período 2011-2015 totalizou **R\$ 1.349.173,60**. Apesar de ser pago pelo ON este valor tem sido obtido por meio de TEDs ao MCTI e, portanto é mostrado separadamente na Figura 15.3. Incluindo este pagamento o total do custo financeiro do LIneA é de **R\$10,2 milhões**, dos quais o pagamento ao SDSS corresponde a 13% do total. Mesmo incluindo este pagamento a contribuição vinda de projetos ainda corresponde a 80% de todo o custo financeiro do LIneA e demonstra mais uma vez que o orçamento do LIneA é destinado a infraestrutura do laboratório.

Também deve ser reconhecida a contribuição não financeira dos outros parceiros do LIneA. Por exemplo, extremamente valiosa tem sido a contribuição do LNCC/CBPF pela cessão do espaço no POP-RJ, e agora em suas instalações em Petrópolis. Este custo foi orçado comercialmente em R\$ 85 mil por mês (R\$ 35 mil no caso do LNCC) quando esta solução foi considerada pelo LIneA como alternativa para a instalação de suas máquinas. Isto corresponde a uma contribuição efetiva de **R\$ 2,1 milhões a R\$ 5,1 milhões** no período de cinco anos sendo considerado. Também importante tem sido o apoio administrativo dado pelo LNCC no gerenciamento do projeto FINEP e consultas feitas ao seu pessoal técnico.

Igualmente importante tem sido a contribuição da RNP que tem tomado diferentes formas. Em primeiro lugar, seu time técnico tem apoiado o LIneA na tentativa de melhorar o desempenho das transferência de dados com seus vários parceiros o que inclui a configuração da rede, do software utilizado para a transferência e a configuração a ser utilizada para a máquina que recebe os dados. Este trabalho foi feito em 2013 e agora está sendo repetido devido à mudança do LIneA para Petrópolis. A RNP também ofereceu para hospedar as máquinas do LIneA em seu *Internet Data Center* (IDC) em Brasília. Finalmente, a RNP viabilizou a assinatura do acordo com o LSST, que é de suma importância para o LIneA já que expande o horizonte do uso do Portal Científico. Sua contribuição, em conjunto com a da ANSP, foi avaliada para o LSST como um benefício equivalente a US\$ 2 milhões.



**Figura 15.3** – Distribuição das contribuições financeiras para o LIneA mostrando a contribuição por cada organização (painel esquerdo) e separando a contribuição vinda de projetos e de orçamento. A contribuição do BPG vem do MCTI através de TEDs do MCTI.

## 15.2 Projetos Submetidos

Além dos projetos acima mencionados, durante o ano de 2015 vários pedidos de auxílio foram submetidos ou preparados entre os quais se destacam:

- FAPERJ – em resposta ao edital Pronex E46/2014 foi submetido o projeto intitulado *Núcleo de e-Astronomia: do Sistema Solar à Energia Escura* com a participação de 37 pesquisadores e estudantes, 10 tecnologistas de oito instituições nacionais (CEFET, LNCC, ON, UFF, UFRGS, UNESP, UFTPR, USP) e internacionais (apêndice 15.1). O pedido totaliza ~R\$ 900 mil, para a compra de um cluster de processamento com 33 nós e de estações de trabalho, para pagamento diárias e passagens. Aguardando julgamento.
- CNPq – em resposta ao edital INCT foi submetido o projeto intitulado *INCT do e-Universo (INCT e-U)* envolvendo ~40 pesquisadores e estudantes, mais a equipe de TI do LIneA. O pedido é de R\$ 10 milhões, sendo R\$ 4,5 milhões em capital, R\$ 2,5 milhões em bolsas e R\$ 3 milhões em custeio a maior parte em diárias e passagens tendo em vista a natureza internacional dos projetos apoiados (apêndice 15.2).
- *National Science Foundation* – submetido em conjunto com o NCSA para o desenvolvimento da interface que será utilizada para a distribuição dos dados do DES em 2017, um dos importantes compromissos assumidos pelo LIneA. A parte da proposta que se refere ao LIneA está incluída no apêndice 15.3.
- FAPESP – projeto temático intitulado *Estudo de Modelos de Negócios para Federação de Serviços para Suporte a e-Ciência* visando à criação de um centro de dados compartilhado, em parceria com UFCG, UFSCAR, EMBRAPA, RNP. Um total de R\$ 2 milhões foi solicitado, grande parte em bolsas de capacitação técnica (apêndice 15.4).

Em resumo, atualmente o LIneA busca recursos extra orçamentários da ordem de R\$11 milhões. Em paralelo, alternativas de financiamento para a criação de um Centro de Dados para o LSST para a fase pré-levantamento estão sendo feitas com a FAPERJ, FAPESP, LNCC e RNP. Entretanto, nenhum desses projetos resolve o grande problema do LIneA que é financiar os tecnologistas.

## 16 Dificuldades e riscos

Após cinco anos de operação é possível identificar as principais dificuldades e riscos para o sucesso do trabalho sendo desenvolvido pelo laboratório em suas diferentes atividades. Estas questões certamente devem ser abordadas no processo de avaliação. Abaixo apresentamos alguns destes fatores.

### 16.1 Científicos

- Dificuldades de atrair pós-doutores devido principalmente ao baixo valor das bolsas disponíveis, o que dificulta a sobrevivência dos pesquisadores nos grandes centros. Embora com projetos atraentes, isto dificulta o recrutamento deste tipo profissional.
- A falta de experiência do pesquisador brasileiro, e por consequência de seus estudantes, de trabalhar em grandes projetos de colaboração o que pode levar a algumas situações de conflito.
- A falta de experiência de trabalhar em projetos internacionais, o que leva a certo grau de intimidação para os mais jovens e que dificulta a sua inserção.
- Falta de uma maior dedicação do pesquisador e estudante brasileiro para torná-los mais competitivos internacionalmente.

### 16.2 Técnicos

- O envelhecimento das máquinas, a impossibilidade de manter contratos de manutenção, a falta de agilidade na compra ou reparo de equipamentos de informática através das UPs. A impossibilidade de poder planejar a substituição e expansão do centro de dados em médio prazo. Tudo isto coloca risco a operação e o aproveitamento científico do investimento feito.
- Dado o grande volume de dados e a necessidade de se utilizar o portal para a análise, a arquitetura das máquinas a serem utilizadas tem que se adequar a forma de operação prevista para o Portal.
- Editais das agências pouco adequados, com alguns sendo dirigidos apenas a ciência e outros para pesquisadores na área de TI sem reconhecer a natureza singular do LIneA.
- Dependência extrema em alguns tecnólogos não contratados.
- Limitada participação de contratados no projeto de software com quase todo o time de TI sendo terceirizado. Portanto, não existe no momento nenhuma garantia que será possível preservar o conhecimento acumulado sobre o desenvolvimento do Portal.
- O tamanho do time e o cronograma agressivo dificulta a manutenção do software, uso de novas tecnologias e manutenção de documentação adequada.
- A falta de pessoas qualificadas para atender as necessidades dos projetos de software tendo vista o enorme escopo do mesmo.

### 16.3 Gestão e financiamento

- Acordo formalizado entre as UPs é apenas parcialmente cumprido, em particular, a participação financeira foi pequena.
- O desinteresse demonstrado pelo ON em relação ao LIneA.
- A falta de um orçamento anual próprio e estável.
- Enorme ônus na preparação de projetos e prestação de contas para inúmeros editais para manter a operação do laboratório.
- A falta de perspectivas para a contratação de pessoas chave para o LIneA.
- Instabilidade causada pela incerteza de recursos para pagamento de pessoal.
- Dificuldade de contratação de serviços especializados.
- Dificuldade no recrutamento de tecnologistas devido a baixos salários.
- Falta de um plano de progressão funcional para retenção de tecnologistas.
- Excessivo custo de terceirização dos serviços especializados.
- Falta de agilidade e de recursos para a execução de reparos emergenciais ou reposição de equipamentos de computação.
- Situação institucional do laboratório mal definida.
- Convívio difícil dentro do ON.
- Falta de representatividade junto ao MCTI.

## 17 Perspectivas

Graças ao enorme esforço realizado, as perspectivas do LIneA em relação ao seu futuro são as melhores possíveis.

Do ponto de vista técnico durante este período o LIneA desenvolveu, implantou e manteve dois complexos subsistemas do Portal que não só validaram o conceito proposto mas consolidou a posição do laboratório como um parceiro confiável na entrega de produtos úteis e de qualidade a colaboração internacional do DES. Estes subsistemas já estão em uso pela comunidade internacional no CTIO e no Fermilab e, em breve, no NCSA. A qualidade do produto motivou, por exemplo, a colaboração do DESI em convidar o laboratório para participar do desenvolvimento do *Quick Look Framework* para validar os 15.000 espectros a serem obtidos durante uma única exposição colocando o trabalho do grupo em destaque em várias comunidades astronômicas.

Em longo prazo, a assinatura do acordo com o LSST garante que o investimento feito pelo LIneA para o DES poderá ser estendido e melhorado, integrando novas tecnologias nos próximos 5 anos, dando a oportunidade única da equipe brasileira estar altamente preparada para capitalizar nos dados do LSST, já no período de verificação científica, previsto para começar em 2020-2021. Até lá, será possível usar o próprio DES para desenvolver os vários *workflows* científicos inclusive como o *pipeline* para identificação nas imagens de TNOs, e desta forma prever com mais precisão ocultações. Neste contexto é bom lembrar que vários membros do projeto TON apoiado pelo LIneA acabam de ser aceitos como colaboradores externos do DES.

No momento, uma das preocupações é, junto com o LNA e a RNP, estabelecer o processo de montagem do *Brazilian Participation Group* do LSST e dar início aos trabalhos de construir a infraestrutura de software, montar a base do Centro Regional de dados do LSST e se for o desejo da comunidade trabalhar para estender o acordo para incluir toda a comunidade astronômica e cosmólogos interessados em desvendar o mistério da energia escura.

Ainda nos planos está se explorando: i) como integrar o uso do supercomputador Santos Dumont no modelo de operação do LIneA; ii) como integrar o portal num ambiente de nuvens privadas atualmente sendo previsto para o LSST; iii) participar do desenvolvimento e testes do sistema de execução de buscas distribuídos paralelamente para atender as demandas de mineração de dados na era do LSST; iv) liderar a implantação de um Centro de Dados regional para o LSST; e v) tomar parte da criação de um Centro de e-Ciência no Brasil. Este Centro é fundamental para dar suporte não só ao LIneA mas a outros laboratórios semelhantes cujas necessidades de infraestrutura e de expertise poderiam ser atendidas e compartilhadas com uma consequente redução nos custos.

Além das necessidades dos projetos apoiados o LIneA, dependendo dos recursos disponíveis, temos ainda o interesse de apoiar outras iniciativas como a criação de centro de dados para o SOAR não só para dados brutos mas na medida do possível de dados reduzidos. Numa primeira fase isso seria feito para imagens utilizando o software genérico de redução de dados desenvolvido pelo LSST aplicado a diferentes imageadores.

## 18 Necessidades

### 18.1 Pessoal

Para consolidar o trabalho sendo desenvolvido pelo LIneA alguns ajustes devem ser considerados.

Do ponto de vista científico seria fundamental criar mecanismos para atrair um maior número de pós-doutorandos que fazem a diferença em qualquer projeto. É uma fase da carreira importante para afirmação profissional e as oportunidades oferecidas pelos projetos apoiados pelo LIneA são imensas. Atrair pós-doutorandos brasileiros e estrangeiros compromissados é o grande desafio que não depende apenas do LIneA mas de uma política nacional.

Os pesquisadores seniores devem colaborar no treinamento de seus estudantes para trabalhar em grandes colaborações exigindo deles uma participação pró-ativa nas atividades científicas patrocinadas pelo LIneA como teleconferências, *webinars* e o uso de ferramentas colaborativas. É neste contexto que eles terão o treinamento necessário para atuar de uma forma assertiva no contexto internacional.

É necessário reconhecer o enorme escopo de atividades do LIneA e a necessidade de se ter pelo menos um encarregado para gerenciar as inúmeras demandas administrativas, e um encarregado para o gerenciamento do processo de desenvolvimento de software que atualmente onera os mais talentosos tecnologistas. Alguém que seja responsável pela organização dos *stand-ups*, reuniões de planejamento e validação, supervisão do time, supervisão do cronograma, documentação e interface com o responsável do hardware.

A equipe de TI tem que ser expandida e fortalecida para atender satisfatoriamente o grande número de projetos que procuram atender demandas científicas com prazos bem definidos e estabelecidos pelos projetos internacionais. Em particular, adotando práticas estabelecidas no setor é fundamental. Ter pelo menos dois tecnologistas para cada tecnologia empregada no portal. Isto garante que as atividades de desenvolvimento, de correções e de documentação podem ocorrer paralelamente, promovendo a qualidade do código e sua preservação, e assegura continuidade em caso de saídas de elementos do time. Criar salários competitivos e uma política de progressão funcional para vencer o alto nível de pressão e a grande volatilidade do mercado de trabalho na área de TI. Em particular, pelo menos dois tecnologistas devem reter a visão global do sistema que se pretende usar para o LSST.

Levando-se em conta a experiência adquirida ao longo dos últimos cinco anos, o aumento do número de projetos e subprojetos, os prazos de entrega exigidos para o aproveitamento científico dos dados sendo produzidos pelos levantamentos em andamento e o apoio a ser dado ao BPG do LSST, implicam que o aumento do time de TI é inadiável. Uma estimativa do que se imagina seja o mínimo de pessoal necessário para

atingir as metas descritas acima é apresentada na Tabela 18.1, levando em conta os diferentes perfis técnicos. É importante enfatizar que para o LIneA seria extremamente benéfico compartilhar com outros projetos profissionais cuja atuação seria pontual ou apenas por uma fração de tempo. Esse é o caso do perfil de analistas e de banco de dados que também são os profissionais com maior remuneração. A possibilidade de haver este tipo de compartilhamento é uma das vantagens da criação de um Centro de e-Ciência.

Tabela 18.1 - Equipe de TI do LIneA.

Perfil	FTEs (Atual)	FTEs (necessárias)
Operador	1	2
Operador Científico	0	0.5
Programador	6	9
Analista	1	2
Database	0.05	0.5
Cientista de Dados	1.25	3
Gerente de Projetos	0	1
<b>Total</b>	<b>9.3</b>	<b>18</b>

Além disso, tendo em vista o grande número de atividades administrativas como compras, reuniões, prestação de contas, relatórios, organização de eventos e viagens, preparação e armazenamento documentação, trato com fornecedores, ficou claro a necessidade de um gerente administrativo.

No início de sua operação o LIneA contratou um serviço de consultoria para ajudar na formulação de processos administrativos focado principalmente para a área de desenvolvimento de software. Para implantar esta organização foi contratada uma gerente de projetos. Infelizmente, após três anos ficou claro que a tentativa de ter apenas uma pessoa para gerenciar todas as atividades do LIneA era insustentável mostrando a necessidade de dois profissionais com perfis distintos. Entretanto, devido às limitações orçamentárias foi dada prioridade a contratação de programadores. Por outro lado, ficou evidente que sem profissionais adequados a gestão do laboratório, com a complexidade de um instituto, não pode ser conduzida somente contando com a boa vontade de pesquisadores.

## 18.2 Equipamento

As necessidades de equipamento são ditadas em curto prazo pela demanda criada pelos levantamentos científicos em andamento e as necessidades para processar e distribuir os

dados destes projetos ao público, compromissos assumidos com os pesquisadores brasileiros envolvidos no DES-Brazil e BPG do SDSS, e com as colaborações DES e SDSS. Em particular, a grande demanda de espaço de armazenamento e de processamento vem do DES que acaba de completar metade do levantamento e deve liberar em breve os dados dos três primeiros anos de observação. Isto exige a imediata expansão do sistema de armazenamento (sendo feita emergencialmente com recursos da FAPERJ) e a compra de um novo cluster duplicando o número de nós disponíveis. Lembramos que alguns processos exigem mais de 42 horas de processamento, e deve-se atender, pelo menos, a cinco grupos de trabalhos, simultaneamente. Neste mesmo período está também previsto o uso de nossos sistemas pelo projeto MaNGA cuja a demanda deve ser alta considerando o tipo de processamento a ser feito em cubos de dados, embora esta ainda não tenha sido avaliada objetivamente.

Para atender a demanda do DES foi proposta, em diversas oportunidades, a compra de um cluster da SGI (proposta BRMC 808042) com 48 nós (588 núcleos) o mais cedo possível e daqui a dois anos uma complementação com mais 48 nós. Isto garantiria o funcionamento de um cluster robusto até o final do DES. Com a compra do segundo cluster, o atual seria usado a partir de 2018 para a implantação da base de dados distribuída seguindo o modelo a ser adotado pelo LSST. Isto poderia ser testado na distribuição pública dos dados dos três primeiros anos do DES em 2017 e ao final do projeto em 2020.

Em paralelo, está previsto: 1) o crescimento gradual da capacidade de armazenamento do sistema ALTIX XE1300 para ter da ordem de pelo menos 4 PB para o início do período de verificação científica do LSST, deixando para comprar os 2 PB finais o mais próximo desta data, desta forma garantindo tecnologias mais modernas; 2) aumento do número de servidores para atender a distribuição de dados do SDSS e DES; 3) Desktops/laptops para a equipe de TI.

### **18.3 Projetos Científicos**

No momento, o único compromisso do LInEA é com a taxa de participação para o BPG no SDSS-IV. Faltam três parcelas de U\$ 130 mil, além dos recursos necessários para a instalação e manutenção de uma máquina operando no LCO.

Um projeto que deveria ser considerado pela comunidade astronômica brasileira é o DESI que, em conjunto com o LSST, serão os projetos mais importantes para desvendar a natureza da energia escura. Atualmente, a nossa participação é mínima, mas um esforço deve ser feito para colocar jovens pós-doutores neste projeto de grande impacto científico. O custo para um grupo de três pesquisadores é de US\$ 200 mil e a forma de pagamento pode ser negociada.

## 19 Previsão Orçamentária

Levando em consideração a discussão acima as previsões de recursos necessários para a compra de equipamentos, participação em projetos científicos, contratação de pessoal e de diárias e passagens para os próximos cinco anos são apresentadas nas Tabelas 19.1, 19.2 e 19.3, respectivamente.

Na Tabela 19.1 a ênfase é o cluster de processamento tendo em vista que recursos da FAPERJ estão sendo usados para adicionar mais 240 TB de armazenamento em curto prazo. Outro investimento constante é a compra de servidores para atender a crescente demanda dos dados do SDSS principalmente pelo projeto MaNGA, e a necessidade de se preparar para a liberação dos dados do DES. Também está previsto uma reserva técnica para poder responder as diversa falhas que ocorrem anualmente (e.g., discos, fontes, computadores).

**Tabela 19.1** - Previsão para a compra de equipamentos (US\$ mil).

Equipamento	2016	2017	2018	2019	2020
Banco de dados SDSS/DES	60	60	60		
Cluster	400		400		
Armazenamento	0	520		520	520
Estação de trabalho	6		6		6
Reserva técnica	20	20	20	20	20
<b>Total em US\$</b>	<b>486</b>	<b>600</b>	<b>486</b>	<b>540</b>	<b>546</b>

Na Tabela 19.2 estão incluídos recursos para o pagamento das três parcelas restantes da participação do BPG no SDSS-IV e proposta para a compra de duas participações (6 pesquisadores) no revolucionário projeto DESI, no qual uma posição (3 participantes) estão sendo negociados em troca do desenvolvimento do sistema *Quick Look Framework*.

**Tabela 19.2** - Participação em projetos internacionais (US\$ mil).

Projeto	2016	2017	2018	2019	2020
SDSS	130	130	130	0	0
DESI	80	80	80	80	80
<b>Total em US\$ (mil)</b>	<b>210</b>	<b>210</b>	<b>210</b>	<b>80</b>	<b>80</b>

Finalmente, na Tabela 19.3 é feita uma estimativa para o custeio do LIneA ao longo de cinco anos incluindo a contratação das firmas para a instalação e manutenção dos equipamentos, incluindo o banco de dados do SDSS e a manutenção dos equipamentos no *Las Campanas Observatory*, e a firma de software. A questão salarial tem que ser tratada com cuidado. Vários problemas foram diagnosticados como: 1) falta de

governança sobre alguns salários; 2) imprevisibilidade dos recursos disponíveis; 3) necessidade de realinhamento com o mercado; 4) definição de uma política de progressão funcional. Um sério problema é a existência de desajustes internos onde a criticidade do trabalho do tecnologista, sua experiência e tempo de casa não são devidamente levados em consideração.

**Tabela 19.3** - Custeio de pessoal, diárias e passagens (R\$ mil).

<b>Custeio</b>	<b>2016</b>	<b>2017</b>	<b>2018</b>	<b>2019</b>	<b>2020</b>
Pessoal	3.700	3.900	4.100	4.300	4.500
Passagens e diárias	50	50	50	50	50
<b>Total</b>	<b>3.750</b>	<b>3.950</b>	<b>4.150</b>	<b>4.350</b>	<b>4.550</b>

Em relação ao orçamento é bom lembrar que no modelo atual de financiamento existe um enorme *overhead* com a taxa de administração da fundação, mais a taxa de administração cobrada pelas firmas prestadoras de serviço. A eliminação destes intermediários representaria uma grande economia para o LIneA, em particular tendo em vista que da ordem de 60% dos recursos vão para o pagamento de pessoal.

Considerando o dólar a R\$ 4,0 na Tabela 19.4 mostramos o orçamento estimado para o LIneA que é inferior a R\$ 7 milhões lembrando, que ao comparar com o orçamento das UPs estas não incluem o custo de pessoal. Uma comparação mais justa seria com um orçamento anual de ~R\$ 2,4 milhões, comparável ao orçamento da COAA do ON nos últimos anos, com a importante diferença do envolvimento da comunidade e a alta produtividade científica e visibilidade patrocinada pelo LIneA.

**Tabela 19.4** - Orçamento Estimado do LIneA para os próximos cinco anos (R\$ mil).

<b>Ano</b>	<b>2016</b>	<b>2017</b>	<b>2018</b>	<b>2019</b>	<b>2020</b>
Orçamento (R\$ mil)	6534	7190	6934	6830	6734

Também é importante considerar como este planejamento deve ser integrado àquele que previsto para implantação do Centro Regional de dados do LSST. As características de um centro como imaginado para o Brasil onde o foco será a distribuição de catálogos para mineração e para alimentação de *workflows* científicos foram discutidas com a equipe técnica do LSST e as necessidades são apresentadas abaixo.

Os produtos e serviços do LSST são classificados em três níveis:

- Nível 1 – inclui imagens, imagens de diferença, catálogos de fontes detectados nas imagens de diferença e catálogos de objetos do Sistema Solar. O principal objetivo do nível 1 é a identificação eficiente de transientes em tempo real. A cada noite de observação serão obtidas cerca de 2000 exposições (ou 15 TB de dados) totalizando 5,5 milhões de exposições ao longo dos 10 anos de levantamento.

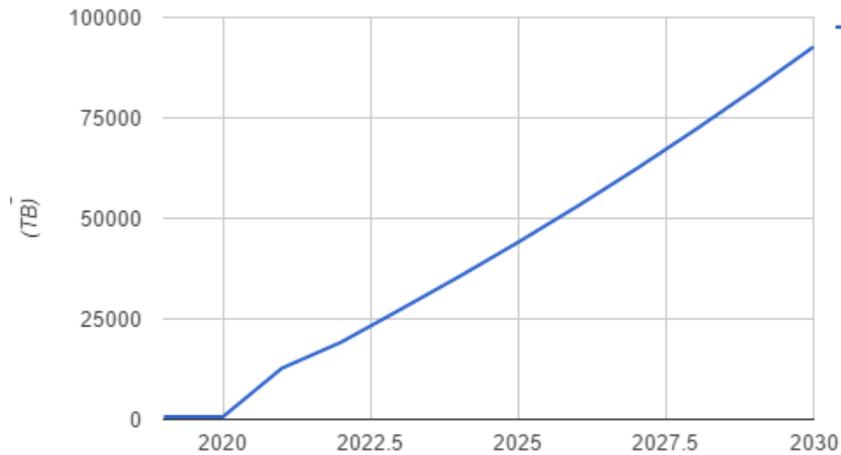
- Nível 2 - corresponde ao *release* de dados anual e que inclui exposições individuais calibradas, fontes detectadas nessas exposições e, um conjunto final de imagens co-adicionadas e objetos identificados nessas imagens. São previstos 11 *releases* anuais e as estimativas para o último *relesse* (DR11) incluem um banco de dados com 37 bilhões de objetos (15 PB), identificados em 24 PB de imagens co-adicionadas.
- Nível 3 - inclui produtos criados pela comunidade usando os serviços para os níveis 1 e 2 possibilitando a ciência com este enorme volume de dados.

O modelo de crescimento de um Centro de Dados Regional do LSST planejado para servir apenas catálogos de objetos dos níveis 2 e 3 é detalhado na Tabela 19.5. O dimensionamento do centro tem como base o fluxo de dados anual e a capacidade de processamento necessária. As estimativas de armazenamento até o final do levantamento (DR11) incluem não apenas o volume do banco de dados final (15 PB), mas pelo menos uma réplica para tolerância a falhas, um *data release* anterior bem como os *overheads* de redundância e indexação dos dados, e uma área de *scratch* para usuários do Nível 3. A estimativa do número de cores e nós de processamento é feita com base na projeção do número de cores por processador ao longo dos anos. O custo de aquisição de armazenamento, processamento e de servidores de apoio é estimado na última coluna da tabela e não inclui o custo de operação do centro. Os valores são aqueles estimados no mercado americano.

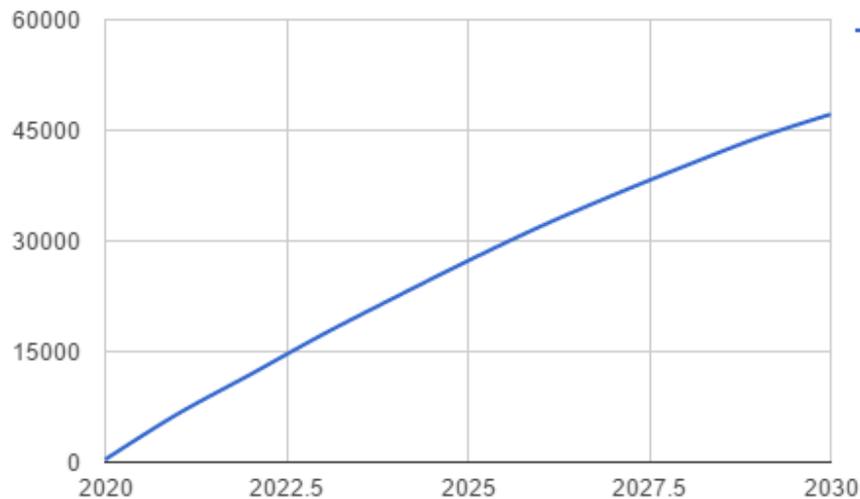
Tabela 19.5 - Modelo de crescimento do centro de dados regional do LSST.

Ano do levantamento	Ano calendário	Storage (TB)	Processamento (TeraFlops)	# Nós	Processamento (# Cores)	Servidores de apoio	Custo (US\$1k)
	2019	618	1,5	9	365	1	44
	2020	618	1,5	9	365	1	0
1	2021	12.735	29,0	104	6.543	16	474
2	2022	19.180	55,4	178	11.888	16	318
3	2023	27.286	83,2	241	17.326	16	281
4	2024	35.518	111,9	287	22.372	16	284
5	2025	44.075	141,6	332	27.300	16	231
6	2026	53.006	172,1	307	31.965	16	373
7	2027	62.332	203,4	289	36.200	16	275
8	2028	72.076	235,2	273	40.193	16	229
9	2029	82.189	266,9	258	43.993	16	203
10	2030	92.691	299,0	245	47.172	16	173
							<b>2886</b>

Examinando a tabela fica claro o rápido crescimento da demanda já a partir de 2021 principalmente no que diz respeito ao volume de armazenamento necessário. As Figuras 19.1 e 19.2 mostram o crescimento estimado de ser necessário na capacidade de armazenamento e número de núcleos, respectivamente, para um centro trabalhando apenas com os catálogos. Note que nestas estimativas existem considerações sobre a evolução da tecnologia. Por exemplo, o número de nós tem um pico em 2025 caindo para 245 em 2030, quando está prevista a necessidade de 300 TFlops de processamento.

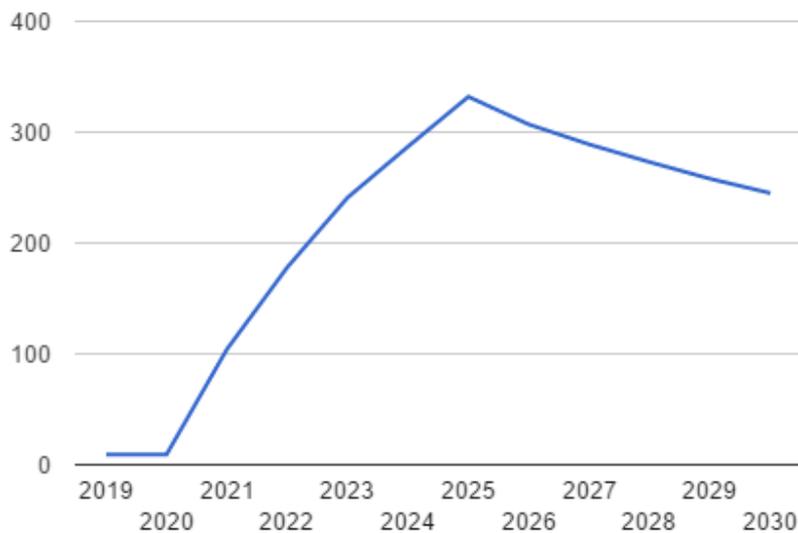


**Figura 19.1** – Crescimento em Terabytes da capacidade de armazenamento exigidas para o Centro Regional de Dados do LSST.



**Figura 19.2** – Crescimento do número de núcleos de processamento exigido para o Centro Regional de Dados do LSST.

Para responder a este desafio, uma das recomendações é a criação imediata do comitê de e-Ciência do LIneA proposto na seção 5 deste documento. Esta comissão seria encarregada de avaliar o uso sendo feito da infraestrutura, principalmente pelo DES, precursor do LSST, escalonar para o número de pesquisadores a serem atendidos na era do LSST. Ela também deveria investigar, em maiores detalhes, como os números apresentados acima foram obtidos, e propor um perfil de investimentos para os próximos anos levando em conta as necessidades do DES e dos outros projetos já em andamento e o período de transição para o LSST. Os dados levantados por esta comissão devem ser usados pelo Conselho Diretor para articular com o MCTI, MEC e as agências de fomento como a Finep e as FAPs os recursos e a cadência necessária para garantir uma operação contínua e estável.



**Figura 19.3** - Número de nós previstos para um centro dando suporte a catálogos dos níveis 2 e 3.

Um fato é claro, um centro como o LIneA focado nas necessidades da comunidade astronômica será vital para garantir o sucesso da participação brasileira no LSST. No momento a solução institucional mais adequada para atender as necessidades do LIneA seria este ser uma ação especial da RNP cuja participação foi fundamental para a assinatura do acordo.

## 20 Resumo

A experiência do LIneA foi um enorme sucesso como pode ser visto por qualquer métrica que seja utilizada. Não só ele cumpriu a missão estabelecida originalmente em relação aos levantamentos DES, SDSS-III, mas usou o trabalho realizado para entrar em novos projetos como SDSS-IV, DESI e LSST expandindo o horizonte deste tipo de atividades até pelo menos 2032, e se envolvendo em projetos de grande impacto procurando desvendar o mistério da energia escura com grandes implicações para física fundamental.

Ao mesmo tempo gerou produtos que estão sendo usados em centros internacionais e criou uma rede de colaboração técnica com os grandes laboratórios americanos. Finalmente, vem contribuindo na formação de jovens pesquisadores na ciência de grandes levantamentos e na ciência de dados uma tendência universal da pesquisa aplicada.

O LIneA ajudou a criar as condições para inserção de pesquisadores brasileiros interessados em grandes levantamentos montando uma infraestrutura computacional adequada e ferramentas para enfrentar aos desafios impostos por este tipo de projetos geradores de grandes volumes de dados. Os resultados científicos são claros como pode ser constatado pela produção científica e impacto, gerando mais de 20 artigos por ano e mais de 6000 citações em cinco anos. Do ponto de vista do desenvolvimento de software o destaque foi a produção de três importantes sistemas sendo utilizados pela colaboração DES, que serve de base para a participação brasileira no LSST, e a implantação e operação de um centro de processamento e distribuição de dados de médio porte atendendo as necessidades dos projetos apoiados.

O escopo do trabalho exigido para o bom funcionamento do laboratório é muito grande e complexo tendo em vista a operação de um centro de dados de médio porte e em expansão para atender a crescente demanda, um agressivo programa de desenvolvimento de software, e todas as outras atividades científicas e administrativas necessárias para sua operação. Estas atividades transcendem barreiras geográficas e as atividades de qualquer dos institutos envolvidos sendo, portanto vital encontrar soluções que deem sustentação a natureza interinstitucional do laboratório. Na verdade, a principal singularidade da organização é o de preencher uma importante lacuna que existe em várias áreas de atividades que é lidar com o gerenciamento de grandes volumes de dados. Por este ângulo o LIneA pode ser visto como um precursor de algo necessário que é a criação de Centro de e-Ciência combinando as expertises de organizações como o LNCC e RNP na infraestrutura física e de conhecimento na área de TI complementadas por times que entendam as especificidades das diferentes áreas a serem atendidas.

Refletindo sobre os riscos do projeto fica claro que alguns elementos fundamentais para o sucesso do laboratório em sua missão são:

- 1) O funcionamento do conselho e do comitê gestor, que definam o escopo do trabalho nos próximos cinco anos, proponham ao MCTI um orçamento adequado e que supervisionem o funcionamento do mesmo a cada 6 meses.
- 2) A aprovação de um plano plurianual que garanta a continuidade dos projetos dando certo grau de segurança ao pessoal contratado.
- 3) A alocação de um orçamento próprio que leve em conta as necessidades de pessoal de TI e gestão, e de equipamentos.
- 4) Mecanismos que garantam agilidade na compra de equipamentos e contratação de pessoal especializado minimizando o *overhead*.
- 5) Mecanismos de retenção do conhecimento como a contratação de pessoas chaves do processo criativo.

A experiência mostrou que muitas partes do acordo não foram cumpridas como se imaginava, gerou uma enorme resistência dentro do ON que favorece desproporcionalmente outros projetos da astronomia, e mostrou as sérias dificuldades impostas pelo regime jurídico das UPs na operação de um laboratório multiusuário como o LIneA com um amplo escopo de atuação.

O futuro do LIneA depende em primeiro lugar do reconhecimento de sua singularidade e o papel crucial que tem para garantir o aproveitamento científico dos projetos apoiados e montar o arcabouço necessário para o bom aproveitamento desta oportunidade única que a astronomia brasileira tem de participar do LSST. Sua missão para o futuro deve ser expandida para liderar a criação de um Cento Regional de Dados e continuar o desenvolvimento do portal científico que será fundamental para o bom e eficiente aproveitamento científico dos dados do LSST.

Para isso, no entanto, será necessário rever sua situação institucional e, em conjunto com o MCTI, explorar alternativas que levem em consideração os pontos levantados acima. A solução ideal seria o reconhecimento da necessidade da criação de um Centro de e-Ciência, mencionado algumas vezes neste documento, com orçamento próprio e com a flexibilidade de uma organização social. A experiência do LIneA mostra que para a operação de um laboratório multiusuário é fundamental ter agilidade na compra e reposição de equipamentos de informática, poder contratar de maneira estável e sem intermediários um time de tecnólogos, e poder fazer uso de consultores de acordo com as necessidades. Acredito que um dos desafios para o futuro do laboratório é encontrar uma solução que atenda a estas necessidades

### **Agradecimentos**

Este documento teve contribuições dos coordenadores dos vários grupos de trabalho, da equipe de TI, da secretaria do LIneA e de seus fornecedores.

## Apêndices

### 1 Histórico

- 1.1 Portaria Acordo LIneA - DOU 213, 08.11.2010
- 1.2 Acordo de Criação do LIneA
- 1.3 Aditivo LIneA-RNP
- 1.4 CONVÊNIO MCT OBSERVATORIO NACIONAL-FAPERJ (rascunho)

### 4 Projetos Apoiados

- 4.1 Protocolo Criação DES-Brazil
- 4.2 Anúncio de Oportunidade SDSS-III
- 4.3 Resposta da Comunidade ao AO sobre SDSS IV APOGEE
- 4.4 Resposta da Comunidade ao AO sobre SDSS-IV MaNGA
- 4.5 Carta para SCUP de pesquisadores interessados no Apogee2 (SDSS-IV)
- 4.6 Carta para SCUP de pesquisadores interessados no Manga (SDSS-IV)
- 4.7 Termo de Compromisso BPG-members
- 4.8 Carta Convite Joint DES-LSST Workshop, Mar.2014
- 4.9 Proposta de Regimento Interno BPG- LSST

### 5 Organização Interna

- 5.1 Proposta de Estatuto para o LIneA 02122013
- 5.2 Carta Conselho LIneA a SECEX

### 6 Afiliação

- 6.1 Participantes Institucionais do DES
- 6.2 Participantes Institucionais do DESI
- 6.3 Participantes Institucionais do SDSS-III
- 6.4 Participantes Institucionais do SDSS-IV

### 7 Ciência

- 7.1 TNO colaboradores externos do DES

### 8 Centro de Processamento e Distribuição de Dados

- 8.1 Serviços SLACAM
- 8.2 Enrich Your Research Proposal

## 9 Desenvolvimento de Software

- 9.1 Portal Review 2010
- 9.2 Portal Review 2011
- 9.3 Portal Review 2013a
- 9.4 Portal Review 2013b
- 9.5 Portal Review 2014
- 9.6 Portal Review 2015
- 9.7 Cientista de Dados Terapix IAP France E. Bertin

## 12 Divulgação

- 12.1 LSST Education and Public Outreach (EPO)

## 13 Acordos e Compromissos

### 13.1 Acordos Projetos Internacionais

- MOU DES - DES-Brazil
- MOU SDSS-III BPGI
- MOU SDSS-IV BPG
- MOA LSSTC-Brazil
- Carta de Interesse DESI

### 13.2 Acordos Técnicos

- LSST-LIneA
- MOA Draft V1 LIneA LSST Tech Collaboration.doc
- MOA NCSA-LIneA
- Statement of Work DESI-LIneA
- University of Illinois Urbana –LIneA
- Statement of Work DES-LIneA

### 13.3 Solicitações de Serviços

- AURA Director C. Smith (QR)
- CTIO Director N. van de Bliet (QR)
- DECam Instrument Scientist A. Walker (QR) 2016
- DECam Instrument Scientist A. Walker (QR) 2013
- DES Cluster WG coordinator K. Romer (Viewer, WAZP)
- DES Director J. Frieman (DES Portal)
- DESI Director M Levi (QLF)
- DES SL WG coordinator E. Buckley (Viewer)
- DES TNO WG Coordinator D. Gerdes (Exposure viewer)
- Grupo de Ocultação (Identificação TNOs in DECam images)
- LNA (SOAR Archive)

SDSS-IV Director M. Blanton (CASJOB)

#### 14 Manifestações de Apoio

##### 14.1 Cartas Institucionais de Apoio ao LIneA

Comitê Executivo LIneA  
Conselho IF da UFRGS item 4  
Instituto de Física da UFRGS, 08.01.2016  
Instituto de Física da USP, 08.01.2016  
Pós-graduação OV-UFRJ 09.08.13  
Pró-Reitoria de Pesquisa da UNESP 15.01.2016  
Pró-Reitoria de Pós-graduação e Pesquisa UFSM 12.2015

##### 14.2 Cartas Individuais de Apoio ao LIneA

Apoio Comunidade ao LIneA  
Carta Apoio Basílio Santiago, UFRGS 09.08.13  
Carta Apoio Cristina Chiappini, AIP, 12.08.13  
Carta Apoio Cristina Furlanetto, UFRGS, 09.08.13  
Carta Apoio Fernando Saliby, UFF, 12.08.13  
Carta Apoio Hélio Jaques, UFRJ Pós Astronomia, 09.08.13  
Carta Apoio Marcos Lima, USP, 10.08.13  
Carta Apoio Rogério Riffel, UFRGS, 12.08.13  
Carta Apoio Rogério Rosenfeld, UNESP Instituto Física, 09.08.13

##### 14.3 Cartas Apoio ao Projeto LSST

Carta da SAB ao MCTI 12.02.2015  
Carta LNA-LIneA ao MCTI

##### 14.4 Cartas Apoio ao Projeto INCT

CEFET  
LNCC  
LSST DM  
LSST EPO  
ON  
Planetário RJ  
RNP  
SLAC  
UFRGS  
UNESP

##### 14.5 Cartas Apoio ao Projeto PRONEX

Infax

SGi  
Slacam

15 Finanças

- 15.1 Projeto Pronex
- 15.2 Projeto INCT
- | 15.3 Projeto NSF
- 15.4 Projeto FAPESP