

Relatório Final, Projeto N° 171006

# Sistemas Distribuídos para o controle de Instrumentos Astronômicos. Fase II

Guillermo Giménez de Castro (CRAAM/CAGE/EE)

São Paulo, 31 de março de 2018

## RESUMO

---

O presente projeto é uma continuação do projeto em execução durante o ano de 2016 (*projeto Mackpesquisa 161005*). O objetivo é participar no desenvolvimento de um sistema informático baseado em *objetos distribuídos* (ODs) para o controle do rádio telescópio *Large Latin American Millimeter Array* (LLAMA). LLAMA é um projeto internacional para instalar uma antena de 12 m de diâmetro no deserto de Atacama a 4800 m de altura. Durante a *fase II* do projeto Mackpesquisa pretendemos adaptar o *Data Model* originalmente desenvolvido pelo observatório ALMA, para as necessidades do telescópio LLAMA. O projeto prevê uma forte interação com outras instituições de e fora do Brasil.

## 1 INTRODUÇÃO

---

O objetivo principal do presente projeto é desenvolver um modelo de dados para o telescópio LLAMA. Para tanto se toma como base os modelos do observatório ALMA, chamados de ALMA Science Data Model (ASDM) e ALMA Project Data Model (APDM), para representar os dados astronômicos e a descrição do projeto de observação (alvos a observar, configuração instrumental, modos de observação, cronograma, etc) respectivamente.

O ASDM descreve conceitual e logicamente como os dados, meta-dados e dados auxiliares estão organizados. Dado aqui significa *dado astronômico* (no caso de LLAMA trata-se do *total power* ou seja o fluxo de energia que impacta na antena e é colhido por um receptor). Os meta-dados são as informações necessários para descrever os dados astronômicos, por exemplo dimensões, unidades, posição da antena, etc. E os dados auxiliares descrevem as condições de trabalho, como o estado atmosférico (pressão, temperatura e umidade), importantes no processo de calibração posterior.

O APDM é o modelo abstrato de informação necessária para descrever um Projeto de Observação desde que ele é proposto por um astrônomo até que ele é efetivamente executado pelo telescópio. Ele é usado para avaliar a proposta de observação em primeiro lugar (Fase I) e posteriormente para produzir o

*Scheduling* (cronograma de observações) do Observatório gerando *Scheduling Blocks* (SB), unidades mínimas de observação (Fase II). O APDM está construído em base a diagramas de classe UML exportados para *Schemas* em formato xml por meio de um gerador de código que também cria as classes na linguagem de desenvolvimento de ALMA. Durante a Fase 1, o astrônomo escreve a proposta observacional interagindo com um programa chamado de *Observing Tool* (OT).

## 2 SCIENCE DATA MODEL

Na proposta original do projeto iríamos trabalhar com o Dr. A. Bridger do *UK Astronomy Technology Centre*, em Edinburgh (Escóssia) no APDM por ser ele o líder do desenvolvimento dos códigos relacionados. No entanto, por problemas de agenda, decidimos ir ao Observatório de Paris para trabalhar durante uma semana com o Dr. F. Viallefond no SDM, o modelo de dados científico. Como informado no relatório (vide anexo) estudamos em detalhe o SDMv2 e discutimos estratégias para seu aproveitamento no telescópio LLAMA. No entanto, em razão da falta de definição de aspectos técnicos dos receptores de LLAMA foi proposto trabalhar com outro instrumento operado pelo CRAAM, o Telescópio Solar para Ondas Submilmétricas (SST por sua sigla em inglês) cuja atualização está em marcha e do qual temos pleno conhecimento porque é um instrumento em funcionamento.

Uma vez gerada a biblioteca de interfaces da aplicação (API) começou o desenvolvimento de um

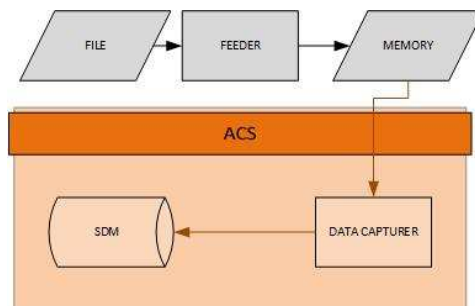


Figura 1. Diagrama do simulador de dados.

simulador do receptor. A ideia é utilizar dados do SST de nosso repositório e por meio de uma aplicação específica (*feeder*) criar uma sequência temporal que emula a aquisição de dados, desta forma um módulo (*data capturer*) rodando no ambiente ACS fará a coleta desses dados no formato SDM (Figura 1). O simulador está em processo de programação, e deve ser completado até o mês de maio, sendo que se trata do Trabalho de Conclusão de Curso dos dois alunos bolsistas deste projeto (Vide no anexo o póster do TCC-I que será apresentado em 10 de abril)

Apesar de não ter interagido pessoalmente com o Dr. Bridges está sendo analisado o APDM e o fluxo lógico desde a construção da proposta observacional até a geração do SB. Os SB são atômicos: eles não podem ser reiniciados se abortados. Eles contêm alvos (objetos a observar) e especificações técnicas. São usados *Observing Groups* para definir as configurações para os diferentes alvos e os grupos são executados em ordem. Grupo 0 é o inicial usado para calibrações, Grupos  $\geq 1$  são reservados para ciência. No entanto, estes Grupos podem conter calibrações misturadas com as observações de alvos astronômicos. Cada Grupo contém uma lista ordenada de ponteiros para *subSB*. Scripts específicos rodam os comandos de observação em uma ordem predefinida, embora os scripts têm liberdade

suficiente para decidir se uma calibração deve ser realizada antes que uma observação, por exemplo. Os alvos são definidos no SB e podem corresponder também a calibrações, não apenas a objetos astronômicos. Cada observação é dividida em

1. *Subintegration*. É a menor divisão temporal de aquisição de dados. Tipicamente tem duração de menos de um segundo (v.g. 50 ms)
2. *Integration*. Resulta da obtenção da média aritmética (em rádio astronomia conhecido como *integração*) dos dados *subintegrados* em intervalos regulares.
3. *Subscan*. Uma sequência de integrações de um mesmo alvo que não pode ser interrompida. Tipicamente cada *subscan* é armazenado em um BLOB (*binary large object*) do *dataset* do SDM.
4. *Scan*. É uma sequência de *subscans* que conformam a observação completa de um alvo e está formada por diferentes tipos de observação, incluindo calibrações, por exemplo.
5. *Scheduling Block*. É a unidade de agendamento das observações. Estão representados por arquivos xml.

Cada SB é interpretado por scripts (alguns estão escritos no próprio arquivo xml). A Figura 2 apresenta os diferentes alvos (ObsTarget) e sua hierarquia.

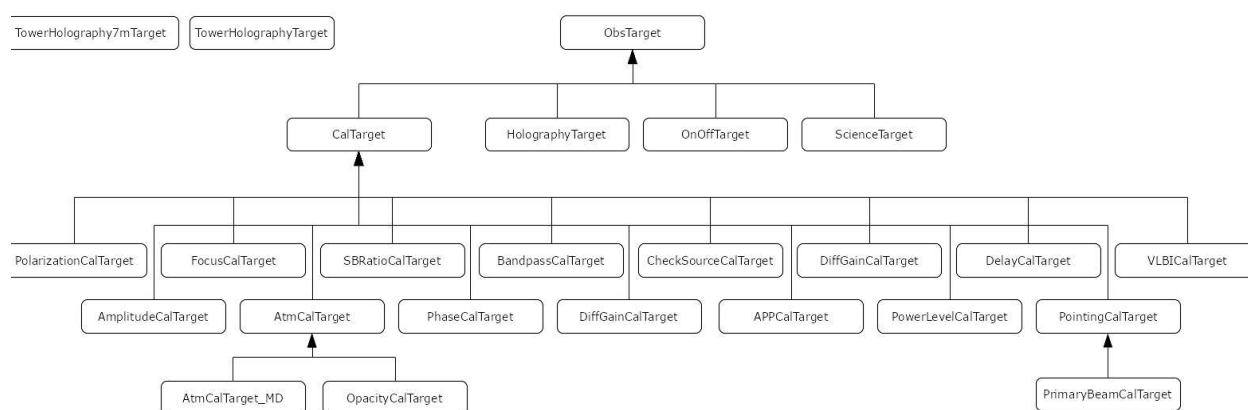


Figura 2. Hierarquia de alvos.

A análise deste esquema foi realizada por meio da leitura do código fonte do software de ALMA e ainda é incompleta. Mesmo assim já somos capazes de gerar um SB simples que é colocado no sistema de agendamento (*scheduling*) e produz arquivos de dados simulados (todos zeros) no formato SDM.

## 3 OUTRAS AÇÕES

### 3.1 ORGANIZAÇÃO DO INTEGRATED PROJECT TEAM DE COMPUTAÇÃO

Como parte de nossas atribuições no projeto LLAMA, temos a função de organizar e distribuir o trabalho do grupo de computação (C-IPT). Este grupo se encontra distribuído em vários países e instituições. Por esse motivo não podemos realizar reuniões *face-to-face* muito frequentemente. O nosso trabalho é desenvolvido por meio de:

1. Teleconferencias semanais usando o sistema Google Hangout,

2. Reuniões *face-to-face* quinzenais no Instituto de Astronomia, Geofísica e Ciências Atmosféricas (IAG) em São Paulo com os membros brasileiros da equipe.

## 3.2 ADAPTAÇÃO DO TMCDB PARA MYSQL

O *Telescope Monitoring and Control Data Base* (TMCDB) é uma parte central do sistema de controle do telescópio. Ele é um banco de dados onde está configurado tanto o hardware quanto o software. No caso do software, levando em consideração que o sistema é distribuído, ele também indica o processador onde um módulo específico roda. Inicialmente trabalhamos com uma versão construída por meio de arquivos xml distribuídos em uma hierarquia de diretórios. Este sistema muito simples de implementar e robusto de operar, é no entanto muito frágil e proclive a induzir erros. Por esse motivo seu substituto é um banco de dados relacional que conta com uma interface gráfica de edição: o TMCDB. Ele foi programado por ALMA originalmente em Oracle. Como LLAMA decidiu não utilizar produtos comerciais, sempre que possível, tivemos de fazer uma adaptação do código para MySQL que é distribuído sob licença LGPL. O procedimento foi bem sucedido tanto para o banco de dados quanto para sua ferramenta de edição TMCDBexplorer e ele faz parte do *branch master* de nosso repositório git.

## 3.3 ALMA OSS 2017-10

No fim do ano de 2017, o observatório ALMA decidiu publicar o código fonte de muitos de seus subsistemas usando uma licença LGPL (*Lesser General Public License*). O mesmo foi colocado em servidor git de acesso aberto (<https://bitbucket.sco.alma.cl/projects/OSS>) e contém a versão mais recente do código, a 2017-10. (A versão que nós estamos usando atualmente é a 2014-04.) A versão 2017-10 traz várias novidades, a mais importante é que roda em sistema RHEL (ou CentOS) 6 de 64 bits, exceto o subsistema CONTROL que continua sendo de 32 bits sobre RHEL 5. O repositório já foi clonado em nossos servidores e está sendo misturado (*merged*) com os nossos códigos. Já conseguimos fazer a compilação de mais do 90% deste código.

## 3.4 PARTICIPAÇÃO EM REUNIÕES

Durante o ano de 2017 participamos de várias reuniões de trabalho do projeto LLAMA. Destas reuniões participam os líderes de cada grupo (IPT) e os membros do Comitê Executivo e se discutem os avanços do projeto e o cronograma. Para cada reunião devemos apresentar um diagnóstico de nossa área. A seguir uma relação destas reuniões:

1. 28 Feb a 02 Mar, na sede de ALMA em Santiago, Chile.
2. 05 a 07 Jun, no Centro Científico Tecnológico do CONICET, La Plata, Argentina.
3. 28 a 29 Set, no Instituto de Astronomia, Geofísica e Ciências Atmosféricas (IAG), São Paulo.
4. 11 a 13 Dez, no Instituto Argentino de Radioastronomía, Villa Elisa, Argentina. Desta reunião participaram também os membros do Comitê Científico Externo formado pela Dra. Catherine Cesarsky (CEA, Fr), Dr. Félix Mirabel (IAFE, Arg), Dr. Thijs de Graauw (ALMA, Ch) e Dr. Lars Nyman (ALMA, Ch).

Além destas reuniões de trabalho, participamos também do “LLAMA Science Workshop II” realizado no IAG, na cidade de São Paulo, apresentando em sessão oral o trabalho: “A solar program for LLAMA early science”.

## 4 CONCLUSÃO E PERSPECTIVAS

---

O projeto LLAMA está muito avançado, embora não completou os objetivos no cronograma inicial. A antena, desmontada, se encontra no sítio de Alto Chorrillos (Salta, Argentina) onde deve ser instalada. Dentre outros eventos que atrasaram o desenvolvimento do projeto, aconteceu um acidente durante o transporte terrestre de um dos caminhões do convoio (eram mais de 20 em total) o que danificou uma peça de grande porte. No momento a firma fabricante da antena (Vertex da Alemanha) está realizando uma vistoria da qual deve resultar um plano de ação (substituição completa ou conserto parcial). Além disto há atraso na construção da infraestrutura do observatório. No novo cronograma, a antena será ensamblada durante o primeiro semestre de 2019, enquanto que os testes de integração e verificação (AIV por sua sigla em inglês) ficarão para a segunda metade do ano próximo. A *Primeira Luz* do instrumento deve acontecer no primeiro semestre de 2020.

Enquanto ao desenvolvimento do software ele também está atrasado, principalmente devido à falta de recursos humanos dedicados ao projeto. **Nesse sentido foi muito importante o aporte do MackPesquisa, ao permitir a participação de dois alunos.** Considerando a demora geral do projeto, no entanto, o desenvolvimento do software poderá cumprir com a meta e contar com os sistemas necessários quando a antena seja ensamblada e comecem os testes AIV.

Por último, achamos altamente produtivo o investimento realizado por MackPesquisa durante os dois últimos anos de apoio ao nosso projeto. O mesmo grupo de estudantes que no primeiro ano (2016) participou do curso de treinamento no *ALMA Common Software (ACS)* continuou trabalhando no segundo ano (2017) no desenvolvimento de software para LLAMA (e agora também do SST, um instrumento genuinamente de Mackenzie), neste ano de 2018 está completando a graduação e já está programando a continuação de seus estudos numa pósgraduação do Mackenzie na mesma linha que vem trabalhando há dois anos. Entedemos que essa é a melhor maneira de formar recursos e manter a continuidade das linhas de pesquisa e desenvolvimento.

# Anexo 1

## Visita ao Observatório de Paris. Relatório Científico

Guillermo Giménez de Castro (CRAAM/CAGE/EE)

São Paulo, 14 de julho de 2017

Durante a semana de 26 de junho a 1 de julho trabalhamos junto com o pesquisador François Viallefond no Observatório de Paris. O objetivo desse trabalho foi compreender em detalhe o modelo dados científico chamado de Science data Model version 2 (SDMv2) e como fazer a adaptação para o projeto LLAMA. Ambos os objetivos foram completados com sucesso. Nas linhas a seguir descrevemos o trabalho realizado.

## 1 O SDMv2

---

Os dados astronômicos registrados durante uma observação são agrupados em conjuntos conhecidos por seu nome em inglês como *datasets*. Um dataset consiste de *meta-dados*, tipicamente parâmetros de configurações instrumentais, junto com *dados auxiliares*, usados tipicamente para realizar calibrações dos dados, organizados na forma de tabelas. No caso do SDM essas tabelas usam o formato xml e contêm referências a arquivos de dados binários, estes contendo os dados astronômicos propriamente, por exemplo intensidades luminosas registradas. O formato dos dados binários é também predefinido, no caso de LLAMA, seguirá o padrão ALMA, que consiste em um contêiner de tipo MIME que pode conter diversos tipos de partes binárias referenciadas por subconjuntos de cabeçalhos xml, enquanto que a estrutura de dados destas partes é descrita no cabeçalho principal xml. De certa forma, estes dados binários conformam um pequeno banco de dados. Cada registro da *Tabela Principal* tem um identificador que referencia um arquivo binário que contém as *integrações* de um *subscan* para um determinado *processador*. Integração é um termo muito usado em rádio astronomia que corresponde à média dos dados durante um período de tempo determinado, *subscan* é um procedimento de observação que corresponde a uma configuração particular do sistema, já um *processador* é um equipamento inteligente utilizado para fazer a leitura dos dados, dependendo da complexidade do sistema.

Cada tabela tem uma estrutura bem definida que inclui as seguintes seções: 1) chave, que pode estar composta por mais de um campo; 2) palavras chave; 3) Descrição de dados: inclui o tamanho dos eixos e

coordenadas; 4) atributos não chave ; 5) dados, campo obrigatório e 6) *flags*. Ao todo são definidas 20 tabelas com as quais a configuração de uma observação é descrita.

## 5 O SOLAR SUBMILLIMETER TELESCOPE

---

Enquanto o telescópio LLAMA ainda tem vários aspectos por definir, propusemos utilizar um outro telescópio de propriedade de Mackenzie, o Solar Submillimeter Telescope (SST, Kaufmann et al, 2008) como exemplo para criar um *branch* (aplicação específica para um telescópio). O SST é um sistema multi-beam, o que resulta muito adequado para testar o SDMV2 porque ele foi desenvolvido com a intenção de implementar telescópios multi-beam. Por outra parte, como o SST é um instrumento que opera há mais de 15 anos, todas as configurações estão bem definidas e assim o *exercício* é completo. Na verdade, não se trata de um exercício porque um dos planos do CRAAM é fazer uma atualização do SST que incluirá necessariamente o sistema de aquisição de dados, pensando em utilizar o SDMV2.

Ao final da semana, o *branch* do SST foi criado, por meio da edição dos arquivos xml schema. A partir destes arquivos um *parser*, desenvolvido especificamente para o SDM, criou um conjunto de programas em C++ que definem as classes para manipulação das tabelas. Essas classes são incorporadas em um arquivo do tipo biblioteca dinâmica, conformando assim uma *Application Programming Interface* (API) para utilizar posteriormente no sistema de aquisição de dados (subsistema Control).

O SDMV2 está em desenvolvimento atualmente, e portanto muitos problemas foram identificados no esquema de geração de *branches*. Isto acabou resultando muito útil para a sua correção e aprimoramento, toda vez que se pretende que seja um padrão a utilizar em qualquer observatório astronômico.

Por último, é recomendável a confecção de um convênio entre o Observatório de Paris e a Universidade Presbiteriana Mackenzie para continuar a colaboração entre as partes. Atualmente existe um convênio entre a Universidade de São Paulo, líder do projeto LLAMA no Brasil, e o Observatório de Paris.

## REFERÊNCIAS BIBLIOGRÁFICAS

---

Kaufmann, P., et al, *New telescopes for ground-based solar observations at submillimeter and mid-infrared*, SPIE Conference Series, v. 7012, 2008.

# Anexo 2

Póster de apresentação do Trabalho de Conclusão (Fase I) do Curso de Ciências da Computação da Faculdade de Ciências Informáticas (UPM) dos alunos Rafael Buiat e Frederico Saar, sob orientação dos professores Luciano Silva e C. Guillermo Giménez de Castro.



# Modelo Digital de Dados para um radiotelescópio:

Frederico Saar Almeida Horta Barbosa,  
Rafael Buiat Cassiano Da Silva

Prof. Dr. Luciano Silva, Prof. Dr. Guillermo G. de Castro

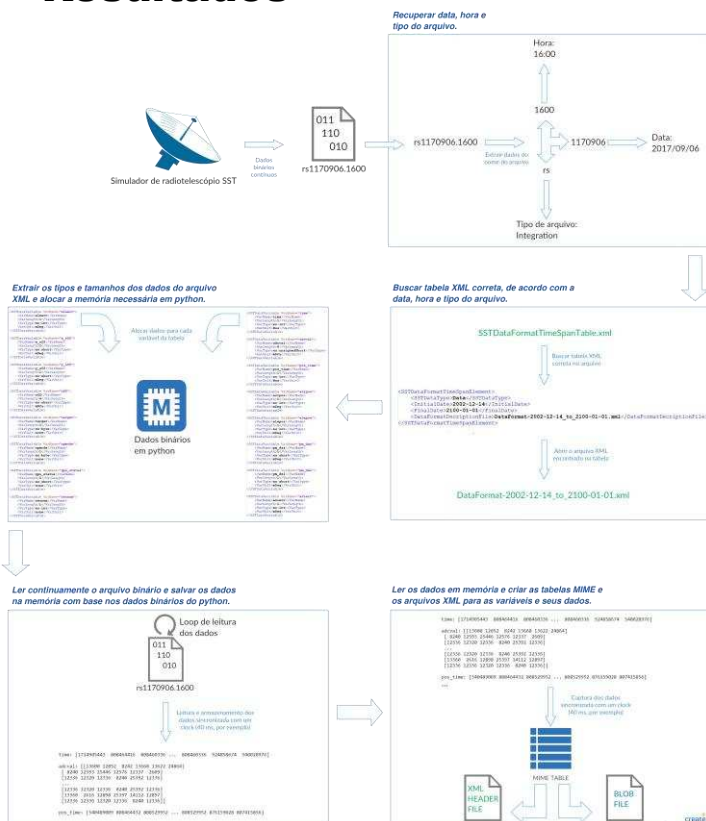
## Motivação e Objetivo

Desenvolver e implementar representação digital de dados para o radiotelescópio Large Latin American Millimeter Array (LLAMA), utilizando e criando um simulador de "Total Power", contribuindo para a comunidade científica.

## Metodologia

Foi necessário conhecer a parte técnica de LLAMA, como, a estrutura dos arquivos de armazenamento de dados, tipo e tamanho dos dados de leitura que cada equipamento gera e número de banda equivalente ao campo de frequência, para que então fosse realizada a leitura de dados e armazenamento dos mesmos em memória, utilizando Python.

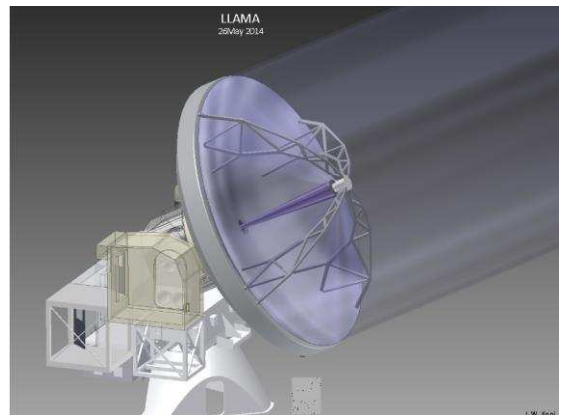
## Resultados



## Conclusões

Os arquivos com a estrutura original de dados foi lida e interpretada com sucesso por um programa desenvolvido em Python, que conseguiu representar os dados obtidos de um radiotelescópio em memória.

A tabela de dados SDM são guardadas como documentos XML, enquanto dados reais produzidos por processadores (correladores, radiômetros, etc.) por serem gerados em enormes quantidades, são guardados em blocos binários agrupados em BLOBs e uma espécie de etiqueta identificadora que poderá ser referenciada pelas tabelas SDM.



Representação em modelo 3D do radiotelescópio que será utilizado pelo grupo do LLAMA.

## Referências

E.M. Arnal et al, "LLAMA Observatory", IV LAPIS International Scholl – Millimeter/Submillimeter Astronomy with LLAMA. Facultad de Ciencias y Geofísicas. La Plata. 2012

Erich, G et al. Design Patterns: Elements of Reusable Object-Oriented Software. USA: Addison-Wesley. Pages:395.1994.

Chiozzi, H. Sommer et all (2009), "ALMA Common Software Architecture". Version: G COMP- 70.25.00.00-002-G-DSN.pdf. ALMA Internal Report.

Fischer, D (1998). "Basics of Radio Astronomy for the Goldstone-Apple Valley Radio Telescope". California Institute of Technology, Pasadena, California.