

# LSST LineA BPG visit report

**William O'Mullane and Ranpal Gill October 2018**

## Introduction

LSST DM and Communications representatives were invited to LineA from 24 to 28 of September 2018 to present LSST DM, communications and EPO efforts and to review the LineA science and technical activities. This was a meeting of the LSST Brazil Participation Group (BPG). There was a packed agenda but with adequate time for open discussion.

### September 24, 2018 - Monday morning

09:30 am to 09:45 am	<b>Introduction</b> - Luiz da Costa
09:45 am to 10:00 am	<b>Clusters of Galaxies</b> - Michel Agüena
10:00 am to 10:15 am	<b>Galaxy Evolution</b> – Julia Gschwend
10:15 am to 10:30 am	Coffee Break
10:30 am to 10:45 am	<b>MW Fitting</b> - Adriano Pieres
10:45 am to 11:00 am	<b>Large Scale Structure</b> – Flávia Sobreira
11:00 am to 11:15 am	<b>Theory-Joint-Probes</b> – Rogerio Rosenfeld
11:15 am to 11:30 am	<b>TransNeptunian Occultation</b> – Julio Camargo

### September 24, 2018 - Monday afternoon

02:00 pm	<b>Demo Infrastructure</b> – William O'Mullane (room 7)
02:00 pm	<b>Ranpal Gill/ LineA Managers</b> - Ranpal Gill (auditorium)

### September 25, 2018 - Tuesday morning

09:30 am to 09:40 am	<b>Introduction</b> - Luiz da Costa
09:40 am to 10:00 am	<b>Data Transfer</b> – Carlos Adean
10:00 am to 10:20 am	<b>Science Portal</b> – Felipe Machado
10:20 am to 10:40 am	<b>Observational Properties Maps</b> – Ricardo Ogando
10:40 am to 11:00 am	Coffee Break
11:00 am to 11:20 am	<b>Data Preparation &amp; VACs</b> – Julia Gschwend
11:20 am to 11:50 am	<b>Network Future &amp; e-Science Data Center</b> – A. Moura and A. Hazin

### September 25, 2018 - Tuesday afternoon

01:30 pm to 02:15 pm	<b>Data Management Status</b> – William O'Mullane
02:15 pm to 02:30 pm	<b>Q&amp;A</b>
02:30 pm to 02:45 pm	<b>LineA Science Server</b> – Ricardo Ogando
02:45 pm to 03:00 pm	<b>Jupyter Notebook</b> – Julia Gschwend
03:00 pm to 03:15 pm	<b>Cloud Initiatives</b> – Jeferson Souza and Carlos Adean
03:15 pm to 03:30 pm	Coffee Break
03:30 pm to 03:45 pm	<b>Santos Dumont Overview</b> - Wagner Leonardo

### September 26, 2018 - Wednesday morning

09:30 am to 09:40 am	<b>Introduction</b> - Luiz da Costa
09:40 am to 10:00 am	<b>Visualization Tool Firefly</b> – Glauber Costa
10:00 am to 10:20 am	<b>LineA Science Server</b> – Ricardo Ogando (Discussion)
10:20 am to 10:40 am	<b>Jupyter Notebook</b> – Julia Gschwend (Discussion)

### September 26, 2018 - Wednesday afternoon

01:00 pm to 02:00 pm	<b>LSST Communications Office</b> – Ranpal Gill
02:00 pm to 03:00 pm	<b>LSST Education and Public Outreach (EPO)</b> – Amanda Bauer
03:00 pm	<b>Q&amp;A</b>

### September 27, 2018 - Thursday morning

09:30 am to 11:30 am	<b>Science Platform Demo</b> - William O'Mullane (room 7)
----------------------	---

### September 28, 2018 - Friday

09:30 am	<b>Unconference (R. Ogando)</b>
----------	---------------------------------

## General Findings

Though we covered Science, Technology, Management and Communications the focus was on Technology and Communications as will be born out below.

## Science presentations

On Monday science topics were presented articulating a broad range of astronomy topics. These ranged from searching the solar system for minor body occultations, through to Galaxy evolution and Large Scale Structure. Model fitting at milky way scale and theory probes were also discussed. Studies beyond the milky way with Galaxy Clusters were also presented.

Some presentations were remote and the infrastructure worked very well - the audio quality was good and there were no dropouts.

Some further discussion was held with Julio on Asteroid codes and GAVITEA<sup>1</sup> was pointed out as a potentially interesting project which wrapped various asteroid modeling tools in python and made them available as a web application.

The science cases are supported by the Science Portal which allows for very easy execution of a range of pipelines. Scientist may also develop new pipelines.

**We did not quite understand the need to develop a web front end for every pipeline especially since they are chained queries with temporary tables - is this overhead needed ?** CasJobs did not provide this facility and pretty complex queries were made frequently.

## Technical topics

### Science Portal

The Brazilian science portal is backed by python and postgres. It is good to see this is developed on top of available GAVO code for data access. Visualization is provided at a large scale by Aladin Lite and at pixel level by Visiomatic which seems an efficient approach - again is good to see reuse of existing tools and the resulting efficiency.

It is pretty clear the pipeline system will not scale to LSST volume. Currently queries take many hours on DES data which is a fraction of the size of LSST data. Even importing LSST data to the current postgres would probably be a challenge - one could consider alternatives like Hadoop or QSERV but either requires some dedicated effort (DESC are putting someone in SLAC for six months to learn about QSERV) .

### Database

Postgres seems an appropriate system for the data volume currently in house. There were specific worries about the efficiency of some queries, query plans had been checked and optimizations made. We did not go into these in detail.

---

<sup>1</sup> <https://www.sciencedirect.com/science/article/pii/S0273117718303673>

**Investigate why the database server does not use all available memory on the 256GB machine. There is probably some setting limiting memory use and thus slowing queries down.**

**Consider, in addition, getting some SSDs in the database server to hold the index files.**

We did look at the Gaia Archive system which is also built on Postgres, that system has a large memory machine making queries very fast.

Some of the queries seem to be full table scans and the database may be a bottleneck.

**Consider DASK/Hadoop or Spark for the pipelines which are touching every row in a table.**

We clarified the nature and use of Qserv. It is a distributed database loaded at each data release to allow fast queries on the Object and Source Catalogs. It is publicly available and installable with Kubernetes. The use case is not quite the same as currently being done with Postgres in LIneA. In principle Qserv can be put on top of other databases, it is currently run on MariaDB.

## Infrastructure

There seems to be a lot of difficulty with Openstack on the project. On the other hand it was not clear what problem Openstack was trying to solve. Apparently this was chosen since LSST were using it - but LSST DM no longer use Openstack. One should take care to achieve Astronomy with technology and not start open ended research on technology in the name of Astronomy.

Many institutes run pipeline steps in containers, LSST is providing containers for each release and for jupyter. Even if one uses virtual machines, containers inside the VMs make life much simpler.

**Invest more effort in containerizing the pipelines and web services with Docker.**

We discussed the merits of Kubernetes and containers which LSST are using for deployment. For example the JupyterLab container with the latest stack build was deployed on the LIneA jupyter server for the stack tutorial. With the addition of a small dataset the tutorial ran without problems. We discussed composing a localised container with additional packages if needed. We looked into the google credits program but it is not available in Brazil.

**Consider asking for Brazil to be added to the google academic credit program<sup>2</sup> to investigate Kubernetes**

## Visualization

We have been discussing the need for a focal plane quick viewer on LSST Camera Diagnostic Cluster. This is an area of potential collaboration. A viewer would need to be easy to deploy (e.g. docker image) and clean and simple. Firefly currently can display a full camera image but all agree a hierarchical image would be better. Considering each CCD is 4K and we have 4K screens we discussed 3 levels of hierarchy CCD, Raft and Full focal plane. Tiles based on the physical devices would make more sense than a resampling to something like HEALPix. Though HiPs<sup>3</sup> is synonymous with HEALpix it is a generalised schema and it should be possible to load the raft and CCD (even amplifier) polygons in a HiPs format. This would allow many viewers such as AadainLite or Firefly to quickly load a focal plane image. This would be a good avenue of investigation.

---

<sup>2</sup> <https://lp.google-mkto.com/CloudEduGrants.html>

<sup>3</sup> <http://aladin.u-strasbg.fr/hips/hipsdoc.pdf>

## Management topics

Use of smart sheets for sprint planning seems appropriate. We were impressed that developers report actuals in Slack but worry about the overhead of the scrum master copying the daily actuals to the sprint sheet.

**Consider the google sheet<sup>4</sup> approach (could possibly be done in smartsheet) for sprint planning and tracking.**

We noted requirements are captured in google docs with one doc per requirement, this seems unstructured and does not lead to good traceability (e.g. there are no id numbers on the requirements). Trac is used for bug reporting which is ok (LSST moved of Trac a few years ago for Jira).

**Use a tool that enables end to end traceability of requirements<sup>5</sup> - testing - problem reports - change control such as Jira (perhaps Trac can do this also).**

It is good that there is a weekly meeting for focused and miscellaneous discussions - it is incorrectly called a CCB. It is not clear CCB items get discussed in any forum. There is no evidence of a risk management process - risks are reported in the status report but there is no risk management plan or risk register. We fear then that risks are not regularly assessed and mitigated.

**Consider changing the name of the current meeting to strategy/status meeting and having a small CCB meeting at least monthly.**

**Consider organising risks in a tool so that a risk register can be extracted into a report (this can be done in Jira and possibly in Trac).**

There is little evidence of general management documentation for structures and processes, implying an CMMI<sup>6</sup> organisational maturity of level 1 (informal).

**Strive to become at least CMMI level 2 (Documented), have a dedicated change control board meeting with a documented workflow process, have a light risk management plan with quarterly review.**

We noted google docs are used for minutes - this is ok even good.

## Communications Discussions

Branding is important for brand reputation and integrity. It is noted that the LSST brand, specifically the logo, is already being used in various mediums.

**Follow the brand usage guidelines<sup>7</sup> and use the provided logo. If in doubt contact [communications-team@lists.lsst.org](mailto:communications-team@lists.lsst.org) if in the future an alteration needs to be made the guideline<sup>8</sup> should be followed.**



Logo provided:

---

<sup>4</sup> <http://adsabs.harvard.edu/abs/2014SPIE.9150E..1EG>

<sup>5</sup> <http://aspbooks.org/custom/publications/paper/495-0371.html>

<sup>6</sup> [https://en.wikipedia.org/wiki/Capability\\_Maturity\\_Model\\_Integration](https://en.wikipedia.org/wiki/Capability_Maturity_Model_Integration)

<sup>7</sup> <https://docushare.lsst.org/docushare/dsweb/Get/Document-28976>

<sup>8</sup> <https://docushare.lsst.org/docushare/dsweb/Get/Document-28699>

Meetings carrying the LSST name have taken place in the past and are planned for the future. This links back to brand reputation. Although meetings may not be funded by LSST (it may be funded by LSST Corporation) by including “LSST” in the meeting name means it falls under the LSST umbrella potentially affecting the brand integrity.

**Consider using the Meeting Code of Conduct template<sup>9</sup> which can be adapted to suit local needs and norms.**

Communications should have performance indicators, there is some level of communications that needs to happen to keep stakeholders updated but beyond that any additional effort and resource usage should have metrics to measure efficacy. LSST has not yet defined these metrics so it is a potential area for collaboration.

The meeting had a good mix of attendees but not all BPG members were represented.

**Consider surveying attendees for feedback on LSST BPG 2018 meeting and it’s contents and also survey BPG members that did not attend to find out why and how the meeting could be made more applicable or accessible for them.**

## Conclusion

It was a busy week in Rio. The Brazil Participation Group seems active and engaged if perhaps lacking the latest LSST information. We hope to have provided much up to date information on LSST and were happy to hear about all the BPG and LIneA projects.

---

<sup>9</sup> <https://docushare.lsst.org/docushare/dsweb/Get/Document-28973>