

LAMBDA Advisory Group FY 2024 meeting report

| | |
|--|----------|
| LAMBDA Advisory Group FY 2024 meeting report..... | 1 |
| Executive Summary..... | 1 |
| Data Capacity and accessibility..... | 1 |
| Archive review..... | 2 |
| ADAP organisation/engagement etc..... | 3 |
| Review of the LAMBDA Charter..... | 3 |
| General tools/software..... | 4 |
| The functioning of LAG..... | 4 |

Executive Summary

The LAMBDA Advisory Group (LAG) has been reinstated after a hiatus, and has had two productive meetings with LAMBDA staff on a range of issues. In this report we have a list of recommendations related to LAMBDA's data capacity and accessibility, archiving, and engagement with NASA's ADAP program. On request we have also provided input on the LAMBDA charter, and the LAG itself including the nomination of new members, suggested term limits and processes for communication between LAMBDA and the LAG.

Our overarching recommendation to LAMBDA is to focus on increasing data capacity and enabling data accessibility as a top and urgent priority, to ensure that LAMBDA remains relevant. We also urge LAMBDA to argue for increased levels of funding to enable the other valued services of LAMBDA including educational resources/tools.

As users of LAMBDA, we look forward to continued engagement with the LAMBDA team on this excellent community resource, ways to improve it and ways we can ensure sustainability and effectiveness of the platform.

Data Capacity and accessibility

Providing long-term archival of data associated with microwave frequency experiments has been LAMBDA's most important contribution to the community. We recommend this continue to be the main focus of efforts since there is no other program that guarantees archival of large volumes of survey data. However, given the rapid expansion of data volume and breadth of science goals of ground-based observatories, storage requirements have shot well past what LAMBDA currently offers.

Capacity: Experience with recent data releases from the Atacama Cosmology Telescope, for example, shows that surveys are now limited in what they can release through LAMBDA.

Specifically, recent releases have only included the main data products (~ few TB) and not the associated simulations (~dozens of TB) that are essential to scientific analysis. The latter are instead released at large supercomputing facilities such as NERSC, which do not guarantee long term archival of these products. This dual release strategy is far from ideal and will get much worse as new surveys like Simons Observatory and CMB-S4 begin data releases. The storage capacity at LAMBDA further lags behind other NASA archives such as MAST, or LAMBDA's parent HEASARC. **We recommend expanding LAMBDA's storage capacity to the ~100 TB level with a path to petabyte-scale expansion for this service to continue being relevant.**

Accessibility: A second equally important inadequacy identified by the advisory group was the interface to data products provided by LAMBDA. Once again, the scale and volume of products makes it such that online URLs for download of a data product (or zipped products) have become impractical. Full-scale deployment of analysis of high-resolution data almost always happens on a computing facility with tens to hundreds of nodes. This makes LAMBDA's online link-based interface insufficient for seamless access to dozens or hundreds of files. **To correct for this, we strongly recommend that LAMBDA implement a Globus endpoint (e.g. through JHU SciServer) that allows large computing facilities to selectively and rapidly sync data products.** Failure to do so will relegate LAMBDA to more of a 'vault' than a living and accessible archive facility.

Archive review

The LAMBDA Advisory Group is pleased with the praise that the LAMBDA archive received from the 2024 NASA Astrophysics Archives Programmatic Review (AAR2024), a once every 5-year assessment. As the Review correctly recognized, LAMBDA is and will continue to operate as, "the main hub of the CMB community." The review recognizes and endorses that LAMBDA continues to serve as a source for all major CMB experiments, including LiteBIRD, SO, CMB-S4, SO, etc. These include Decadal Priorities. The LAMBDA Advisory Group strongly endorses the current and future holdings as the scientific use of the data is increasingly done using a multi-experiment approach; it is important to have straightforward access to all of this data.

In addition to its holdings, the Review encouraged LAMBDA to continue to invest in software: (1) to keep a modern website operation; (2) to create value-added tools; (3) to ensure interoperability (by taking over maintenance of the General Coordinates Network and enhancing interoperability through the VO standards and through integration with the Fornax platform).

While the LAMBDA Advisory Group congratulates the LAMBDA archive team and applauds its future goals and excellent Review, we also recognize that its budget has been largely flat for several years in the face of increasing costs. With continued flat budgets from NASA (effectively cuts when correcting for inflation) this path is clearly not sustainable. As experimental datasets become larger and as labor costs continue to grow, NASA must recognize that a budget bump-up to the LAMBDA budget is a necessity. We are sympathetic to the real budget issues that NASA faces, but we also recognize that the LAMBDA budget is objectively very small (even at the level of only NASA archives) and a bump-up even smaller.

We urge the leadership of LAMBDA to seek a moderate budget increase.

ADAP organisation/engagement etc.

LAMBDA plays an important role in the Astrophysical Data Analysis Program (ADAP) under the ROSES NRA. The LAG noted that NASA missions are required to provide public data but are running into storage size limitations on LAMBDA. (One example is the Taurus Ballon program, which forecasts data volumes larger than currently available on LAMBDA.) Similarly, ADAP has requirements to make products publically available. The ability of LAMBDA to support the ADAP program has been limited by the LAMBDA data capacity.

The LAG therefore recommends improved support for ADAP activities by expanding the hosting of NASA mission data and the hosting of ADAP created data products.

These activities are well aligned with the core mission of LAMBDA, and since LAMBDA is already a central provider of data for ADAP efforts, an expansion of LAMBDA storage space available would have a relatively large impact on the ADAP program and its recipients.

When hosting ADAP products, the LAG recommends that they be organized by ADAP paper/program, separate from the underlying source data. Further structure could be added in the form of tags or links to the source datasets.

Review of the LAMBDA Charter

A review of LAMBDA's mission statement was requested in this advisory report. The mission statement is:

LAMBDA exists to serve the CMB research community, and the greater cosmological research community. In particular, LAMBDA:

- Develops and maintains data archives.
- Develops and maintains data access and analysis tools.
- Offers scientific expertise on NASA's CMB missions
- Carries out data-intensive processing of vital importance to NASA's CMB community
- Conducts education and outreach efforts aimed at early career researchers and the general public

While the LAMBDA Advisory Group supports all of these activities, it does not find that they all accurately reflect the current role of LAMBDA in the CMB community, nor its current resources. To the extent that limited resources must be allocated between different activities, the LAG advises that they should not all be weighted equally.

LAMBDA's primary role is as a data archive, and this is how it provides the most value to the CMB community. **The LAG advises that developing and maintaining the data archive be**

the highest priority, including improvements to the data capacity and accessibility as outlined earlier in this report. Developing and maintaining data access tools further enhances this primary function.

The development and maintenance of other data analysis tools are largely outside the scope of what the community expects from LAMBDA. Similarly, the computational requirements of CMB data processing have grown significantly, to a point where it is difficult for LAMBDA to contribute on this front. Beyond high priority software for data archiving, data access, or outreach, LAMBDA has some potential to leverage its position as an archive with data from multiple experiments. The footprint comparison tool is a simple and effective example of this.

The LAG commends LAMBDA on its educational and outreach resources, and recommends continued development of these resources. These are essential services to the community and the public, and LAMBDA's impact here cannot be overstated. The importance of these activities is second only to LAMBDA's primary role as an archive.

The LAG recommends that the charter be updated to reflect LAMBDA's current priorities. We would be happy to consult with stakeholders and make specific suggestions in a future report.

General tools/software

The online tool for simulating the universe that replaces a previous CAMB interface is very useful and informative, and we thank the LAMBDA team for developing the software for hosting on the site. One suggestion may be to have an option to overlay some experimental data at different scales to highlight why different experiments can constrain combinations of parameters with different relative precision.

LAMBDA provides an excellent repository for linking to software developed related to CMB experiments. However, given the fast-changing software landscape and budgetary pressure, **we recommend that the primary service of LAMBDA should be linking to external software repositories, rather than in developing a suite of new LAMBDA-specific software tools.**

The functioning of LAG

The LAMBDA Advisory Group was reconstituted after a long hiatus and was tasked with defining the term lengths of the members, and its selection procedure. We make the following recommendations:

- Current members of the LAG can make suggestions to LAMBDA for new members who might serve on the group, but are not involved in the recruitment or selection of new members.

- The suggested term length for membership on the LAG is 2 years, with a subgroup of current LAG members serving 3 years to provide the institutional memory, with half the group rotating every two years. Membership on the LAG could be open to renewal, subject to the approval of the LAMBDA executive. The size of the LAG could be reduced to 4 members to ensure that the number of rotating members is constant yearly, but this is not essential.
- The LAG should select their own chairperson
- The LAG should meet twice a year, and prepare a formal report once bi-annually, with informal recommendations occurring after the other bi-annual meeting. At the start of each LAG meeting, the LAMBDA executive can report/respond to the reports from previous meetings.