

LSST Science Advisory Committee

Preliminary Report on the LSST Open Data Framework model

April 15, 2019

The LSST Science Advisory Committee (SAC) was asked by the LSST Project Director to comment on the "Open Data Framework" (ODF), in which the operations of LSST would be supported in full by the US funding agencies, obviating the need to raise money from international contributors, and allowing the data to be world-public. The ODF is in contrast to the current operations model, in which roughly 75% of the LSST operating funds would be contributed by the US funding agencies, with the remaining 25% (of order \$15M/year) to come from International Contributors in exchange for data rights, via agreements with the LSST Corporation.

The SAC was first briefed on the ODF concept in a 1.5 hour phone conference with Steve Kahn and Zeljko Ivezic (LSST Project Director and Deputy Director, respectively), Bob Blum (LSST Acting Operations Director) and Ed Ajhar (NSF Program Director for LSST) on Wednesday, April 3. Blum's slides from this discussion are available on the SAC website. We were asked to give a response in two weeks. Given this short amount of time, and the lack of specificity about many details of how ODF would work in practice, the opinions described here are almost entirely from the SAC members alone, without substantial input from the broader scientific community that we represent. Furthermore, these opinions may not be fully informed because, as we detail below, not all relevant information has been properly communicated to the SAC.

If the ODF is adopted, the SAC will continue to solicit and receive input from the community, and will explore the consequences of how the ODF will work in practice. That is, the current statement should be considered as preliminary, and will be followed up with a more in-depth and thoroughly considered report as we learn more and gather input from the community.

Feedback from the community about the ODF

We are very concerned about the way the ODF concept was communicated to the community, which led to rapidly propagating rumors and misinformation. We worry that decisions may be taken about LSST operations that will have a fundamental impact on the way the community will do science with LSST, without meaningful input from that community. Indeed, we were surprised and displeased that plans for such a fundamental change in the way LSST data would be made available to the science community was not explored with its Science Advisory Committee earlier on in the process. In its August 2018 meeting, the SAC reviewed and gave extensive comments on a pre-ODF draft data framework plan (document LPM-261 by Willman et al.). Having heard nothing in response to our report, and also nothing between the time of that meeting and early April about data-framework issues more generally, we are now being asked to respond very quickly to the ODF concept - for which no document describing how the ODF would work in practice yet exists. We hope that our response is not perceived simply as

pro forma feedback to provide political cover for a decision that has largely already been made. We are concerned that the funding agencies may be taking a decision before the project has developed a detailed reviewable implementation plan for the ODF.

We return to these themes at the end of this report.

Advantages of Open Data

In an ideal world, the SAC certainly supports the principle of open data, with world-wide accessibility; indeed, the SAC has argued for this in the past. Different scientific communities within LSST have different cultures with respect to the idea of open data, but the argument that total science is maximized when data are open and shareable is a convincing one, based on the experience with other major surveys such as SDSS, Gaia, 2MASS, and many others. In particular, the ODF would allow current LSST data rights holders to collaborate freely and share resources with those from the rest of the world on LSST projects. Indeed, making the LSST open will make it more likely that telescopes around the world will be used in following up LSST discoveries. In a decade where the most exciting problems in astronomy are likely to be best addressed by multiple observatories, open LSST data provides a foundation for building effective and open international collaborations.

An open data framework would also greatly simplify the data rights/data access rules. A specific example that the SAC commented on extensively in its August 2018 meeting concerns junior researchers; an open data framework would mean that they would not lose data rights if they are to move from, e.g., the US to another country at an institution that is not part of the LSST consortium.

How Data Access Would Work

The LSST Director told us that the total number of additional people likely to access the data in the open data framework over the current plans would be modest, especially those power users who will need significant resources from the data access centers (DACs). The SAC would like to see a more detailed description of the capabilities of the DACs in this context, together with a description of the number of users the DACs are likely to have and expectations for usage per user. That is, we would like to understand the differences between how data access would work in the current paradigm, and in the ODF paradigm. This information is key for the SAC to assess the impact of the ODF on the scientific community in the US, Chile, and around the world. The SAC requests a detailed description of how the ODF would be implemented in practice, especially if there are trade-offs in operations planning required to make the data available to a broader audience. These details will need to be specified in any updated operations proposal to the funding agencies; the SAC is eager to give feedback on the impact these would have on the scientific community.

On a related note, the SAC knows that a number of countries have expressed interest in developing DACs of their own, which would be welcome both in the current paradigm and if the

ODF is implemented. The SAC recommends that to the extent possible, these international DACs be a shared global resource open to the global community, with the compute resources to support a significant number of users.

The Impact of ODF on the US and Chilean Science Communities

The SAC is concerned about the effect of the ODF on the competitive edge that the US and Chilean communities would otherwise have in the LSST era. If the data are made world-public, scientists from the countries that have contributed almost all the construction and operations costs will have limited or no advantage in its scientific exploitation. That is, even though the world-wide science output of LSST may be maximized in the ODF, the scientific output of the US and Chilean communities may suffer.

A specific example of our concern is Euclid. Euclid and LSST are complementary because Euclid has the resolving power of a space mission, while LSST goes deeper and provides photometry in six filters. The combination is particularly powerful: LSST plus Euclid data allow the determination of precise photometric redshifts, and LSST will be able to calibrate its weak lensing measurements and improve its deblending with Euclid's superior resolution. With this in mind, it has long been anticipated that a data-sharing agreement that would benefit both experiments would be adopted. We are concerned that such an agreement may be untenable in the ODF. The large community of European scientists could combine the still-proprietary Euclid data with public LSST data to do high-profile science (in both cosmology and other fields), without involving US or Chilean astronomers (at least beyond the modest number of people in the US component of Euclid). European Euclid power-users could get a head start of a year or more on joint processing of the LSST and Euclid data. Indeed, because the Euclid data are initially proprietary, the bulk of the LSST community would not be *allowed* to be involved in such an analysis, at least until the Euclid data eventually become world-public. We recommend that the Project work with those science collaborations that are likely to benefit from Euclid data to prepare a written assessment of this issue, before the ODF decision is made.

While we have identified one such concrete example, there will likely be others, often from experiments that are still on the drawing boards. As a general rule, the bargaining power of US and Chilean communities is significantly reduced and foreign experiments with proprietary periods will enjoy unfair advantages if the LSST data become world-public. This is especially true because LSST is intrinsically amenable for follow-up observations: in many respects it is an ideal targeting survey for spectroscopic follow-up in all kinds of astronomy research.

The Role of Science Collaborations

We anticipate that the LSST Science Collaborations would continue to play a key role in LSST if the ODF goes forward; they help organize the community preparations for LSST and serve as a way of connecting researchers together to maximize LSST science. If LSST goes to an open data model, the project should formalize/acknowledge the role of the current science collaborations as one avenue of feedback for future policy decisions. The science

collaborations could also be very useful in giving feedback on decisions involving the DACs, especially policies governing international DACs. It is also worth mentioning that the Science Collaboration chairs submitted comments and questions to the Project about the ODF - during a period when the Chairs were not allowed to communicate with their Science Collaborations about ODF, so these comments and questions are not broadly representative. These comments and questions overlap in part with some of the points made here; we recommend that the Project provide a written response to their statement. It is of concern that the funding agencies have apparently instructed the Project not to respond to the comments and questions from the Science Collaboration Chairs, further impeding meaningful dialog.

The ODF and the financial impact on US funding agencies

The SAC is especially concerned that in this era of limited funding, increased costs for LSST operations will lead to a further squeeze on the NSF AST grants program and mid-scale innovative projects, making it yet more difficult to fund science analyses in the US with LSST. It is hard to see how, given the financial pressures, this will not be close to a “zero sum game”. There is a similar concern that, without a proprietary period for data access, there is less motivation to maintain the model of having cohesive, international science teams focused on comprehensive, full analyses of the data. Supported science teams are an approach that DOE, with international partners, uses for high energy and astroparticle physics experiments, and it is aligned with current support for the preparatory work in the LSST Dark Energy Science Collaboration (DESC). Increases in LSST operations funding has constrained DOE funding to support LSST DESC analysis pipeline development. A concern would be that increased DOE and SF operations commitments would take away from DOE and NSF support for the LSST science funding. The SAC recommends that the potential impact of additional operations cost commitments on NSF and DOE funding of LSST science analysis be fully understood before finalizing this decision. In order for US researchers to be leaders in analyzing the LSST data, and in realizing its scientific discovery potential, robust and sustained US federal science funding would be needed, in tandem with a commitment to meet operations costs.

If the AST grants program indeed becomes squeezed in supporting LSST science analyses and education, this may also have negative *differential* effects across the US. Wealthy institutions may have the capability to support their LSST research programs with institutional funds, developing accumulated advantages in LSST research/proposing/education, while less wealthy institutions will likely not be able to support such key groundwork independently. This could mean, considering the Matthew effect, that independently wealthy institutions reap the majority of the rewards from LSST, despite the fact that LSST was supported by all US taxpayers. This would not be consistent with fully sharing the research and educational opportunities of LSST broadly across the USA. For example, rural and other often underserved communities in particular may be at a strong disadvantage in reaping the scientific and educational opportunities offered by LSST.

Despite reassurances from Ed Ajar that the agencies would put together a reasonable operations budget, we similarly worry that the increased financial burden on the agencies will lead to possible descopes to LSST operations, with effects on the science efficacy of the data. This ties to the questions above about what data access would look like in the ODF paradigm: would all scientists have equal access to catalogs and images, and the compute resources needed to analyze the data?

The ODF and the International Community

It was argued to us that the Open Data Framework obviates the need for high-level agreements on data sharing between the US and foreign governments. It was further argued that this would allow the US agencies to control who gets the data and on what timescale. This second argument was unclear to the SAC; is the implication that world access would in fact be restricted for some countries or for some period of time?

On a related note, the SAC (which includes Chilean representation) is concerned about the status of the existing agreement between AURA and the Chilean government, which interprets the usual "10% of telescope time to the Chilean community" at least partially in terms of exclusive access to the LSST data. With the ODF framework, this agreement would presumably have to be renegotiated, which could be a challenge for LSST, and may increase the cost of the project to the US agencies. We are also concerned about the impact of the US agencies making a unilateral decision about ODF without thorough consultation with the Chilean community.

Finally, we emphasize that the way in which the information about the ODF possibility was discussed within the LSST community was far from ideal, as the project itself acknowledged to us. The decision to go to the ODF would have consequences for essentially all of the LSST constituents, but the information was distributed in a piecemeal fashion very late in the process, without request for input from any segment of the community until now. In addition, our understanding is that discussions about ODF with foreign partners have largely been only with the key representatives, and not with the broader scientific communities in those countries, so the impact on those communities may not yet be fully understood. Again, the Science Advisory Committee is being asked to respond on a very short timescale to the ODF idea with very few details in place. The value of feedback and input from the SAC would be much more meaningful if, in the future, it were consulted much earlier in the process. The SAC requests a plan for how the LSST Project will keep the SAC apprised of major decisions which would have significant effect on the scientific community. Communication is key, both for the SAC itself, and for the community that we represent.