

1 May 2023

This document reports on the IRSA User Panel meeting held virtually on April 17th and 22nd, 2023. The role of the User Panel is to evaluate IRSA's content and services from the user's perspective, advise on the priorities for ongoing/planned activities (given current funding level), and to suggest improvements or new services not covered in the current work plan. The members of the Panel attending the meeting were: Federica Spoto (Harvard & Smithsonian), Michael Blanton (New York University), Rose Finn (Siena College), Bahram Mobasher (UC Riverside), David Kaplan (University of Wisconsin-Milwaukee), Alina Kiessling (JPL) and Remy Indebetouw (University of Virginia and NRAO; Chair). The meeting was scheduled for 4 hours split over two days and consisted of presentations from Vandana Desai, Andreas Faisst, Shoubaneh Hemati, Steven Groom, and Luisa Rebull on the various activities of IRSA and live discussion time among the panel. Some material was provided beforehand for the panel to review asynchronously, which we think was effective. In this letter, we summarize comments and suggestions from the User Panel.

The panel is pleased to see progress on the numerous and diverse projects that IRSA is undertaking. However, we believe that we could give more appropriate suggestions and recommendations if we had greater clarity about the current resources (financial, staff, compute), and how much of those resources are already earmarked for specific missions, as opposed to potentially being available for new projects. We understand that much of IRSA's work is necessarily tied to a specific mission, and appreciate the chart shown with the mission timelines overlaid. It would clarify our understanding to add to such a chart for the broader deliverables, e.g., refactoring of "vintage" applications that have reached end of life, and implementing prototype cloud platform applications. For the next meeting, **we request an organization chart, a summary of staff and their time commitments to projects, and a summary of the ongoing maintenance and near-term milestones for each project; we would also appreciate hearing which projects are under-resourced vs. well-resourced.**

We remain concerned about whether there is sufficient staffing to carry out all of the existing work (in addition to having sufficient staff to take on new, strategically valuable efforts). It would be best to avoid single-individual points of expertise, such as only a single person providing most of the helpdesk support, but we understand that redundancy may not be easy within the current organizational and funding structure. We also wonder whether an internship or mechanism to hire post-graduate students would help. That would both provide training and familiarity to the next generation, and potentially help carrying out some projects.

The committee is pleased to see successful hiring of staff with mission-specific expertise in SPHEREx and Euclid, and that plans for those missions are progressing. We think that the planned interactive user tools for those missions look appropriate and should be effective. In discussing the general needs for SPHEREx and Euclid, the IRSA team pointed out the need for large-scale server-side analysis tools, for example for machine learning applications; we note that the presented user tools did not, however, address this less interactive use case.

For the next meeting, we would like to **request more detail on the plans for NEOWISE and NEO Surveyor** - we would like to understand if handling the data will require any additional resources or infrastructure, and what plans IRSA has for improving the solar system user tools.

The transition from modest-sized data and interactive data examination to “big” data, cloud and platform processing will be a continuing evolution of archives over the next few years. The committee commends IRSA’s leadership and successful efforts with NAVO and metadata standards, which are increasing discoverability and inter-archive collaboration.

The jupyter notebooks developed by IRSA will be an indispensable resource for the community. The notebooks demonstrate commonly-used analysis procedures, and the demonstration of various multiprocessing options will help scientists effectively use the resources available within the science platform. The documentation within the notebooks is not yet fully developed, and we suggest that the team makes the documentation standard, for example by using consistent section names and section headers. (The link for the convolution notebook didn't work. [https://github.com/IPAC-SW/ipac-sp-notebooks/blob/main/parallelize/Parallelize\\_Convolution.ipynb](https://github.com/IPAC-SW/ipac-sp-notebooks/blob/main/parallelize/Parallelize_Convolution.ipynb))

The scope of the transition is large, and represents a fundamental change in ways that astronomers work. As such, we understand that the details will evolve, and a fully comprehensive plan is not yet possible. We also understand the constraints of NASA requiring a joint archive platform solution in NASA-cloud. However, we do have some concern that the timescale is too long to make this product competitive.

We **encourage IRSA to pursue limited-scope shorter-timeline prototypes**, for example: Perhaps allowing a small group of users to propose to obtain access to the existing internal platform, for beta-testing, early feedback, and facilitating interesting science outcomes that would be challenging as an “outside” user. Another specific suggestion would be to use NASA’s request for putting Euclid data in the cloud as an opportunity to demonstrate limited-scope platform-like functionality, building upon the demonstrator Jupyter-notebooks that you’ve shared, but allowing some additional access to compute resources. IRSA could also consider deploying a prototype in the SciServer framework which has existed since 2015.

We strongly **support the concept of requiring user registration for larger queries**, queries that may not easily be able to be made near-synchronous, and platform work. In our experience, it is already commonplace for archives and data centers to request an email for such larger queries, and to register for the use of CPU-significant resources. This does not conflict with the requirement for open access.

Finally, we **recommend that IRSA assess the relative use level today of programmatic vs interactive UIs**, and insofar as possible, the trend in recent years. Server statistics should already exist, albeit likely needing some curation. Acknowledging that those metrics are imperfect, they should still be useful for planning purposes, and prioritizing interactive UI vs programmatic and cloud UIs in the future.

Thank you,

Federica Spoto, Michael Blanton, Rose Finn, Bahram Mobasher, David Kaplan, Alina Kiessling and Remy Indebetouw