Fall 2023 ENSCI User Panel Report

November 29, 2023 held on Zoom

Members of User Panel

- **Dr. Eric Murphy** (UVA, NRAO; chair)
- **Dr. Antara Basu-Zych** (UMBC & NASA/GSFC)
- **Prof. Caitlin M. Casey** (UT Austin; EC)
- **Prof. Yicheng Guo** (U. Missouri)
- **Dr. Katarina (Dida) Markovic** (JPL; EC)
- **Prof. Michael Rutkowski** (Minnesota State University)
- **Prof. Michael Troxel** (Duke)
- **Prof. Gillian Wilson** (UC Merced)

Link to presentation material:
https://collaboration.ipac.caltech.edu/display/EUP/ENSCI+User+Panel+Meeting%3A+29+November+2023

Executive Summary:

The Euclid NASA Science Center at IPAC (ENSCI) User Panel convened via Zoom on November 29, 2023. Seven of the eight committee members were in attendance, with one having to leave slightly early. During the meeting, members of the ENSCI team provided the panel with 6 presentations covering a wide range of topics that included the general status of both Euclid and ENSCI, IRSA and public data, the Euclid science consortium, user support, and expected data products.

No additional feedback was solicited from community members (formally or informally) outside of the User Panel. In the future, it may make sense for ENSCI to solicit broader feedback through online surveys (or information obtained informally at AAS meetings, user emails etc.) throughout the year and present the ENSCI User Panel with this feedback ahead of meetings.

Below, we respond to each of the presentations made by the ENSCI staff. Brief summaries of each presentation are provided, along with a bulleted list of recommendations by the panel for ENSCI to consider. We would like to thank all of the speakers for their clear presentations.

The meeting was chaired by Eric Murphy, who replaced the previous chair Gillian Wilson earlier in 2023.

Finally, we would like to commend the entire ENSCI staff for all of the hard work that they have put into preparing the science center for the successful launch of Euclid. The team has done a truly remarkable job, which is clearly evident by the mature status for all of their activities to support US-based investigations of Euclid data.

Euclid General Status:

Summary:

Euclid is a European Space Agency (ESA) space mission with NASA participation, to study the geometry and nature of the dark Universe. Starting in early 2024, Euclid is carrying out a six-year survey of a third of the sky in the optical and near-infrared. It will measure the shapes and redshifts of galaxies out to distances corresponding to a look-back time of ~10 billion years. Euclid was launched on July 1, 2023, with first light images released on July 31, and the beautiful first scientific images in November as “Early Release Observations”.

During the commissioning phase in July 2023, several issues were discovered and had to be addressed. The problems seem to have been successfully solved, but caused an approximately 2 month slip in the schedule. Subsequent pre-mission verification activities seem to be going well and Euclid should be on track for the commencement of nominal survey observations. At the time of writing this report, the survey should be conducting Phase Diversity Characterization, with the nominal survey to start by February 2024.

Well-characterized and validated Euclid data will be publicly released within about 2 years of acquisition via the Euclid Science Archive System hosted by ESA, meaning in early 2026 as Data Release 1 (DR1). The total size of the data set will fall within the realm of “big data” at its petabyte scale. A subset of the data included in each of the three Euclid Data Releases (DR1: ~2026, DR2: ~2028, DR3: ~2031) will also be made available to the US community at ENSCI and will serve NASA archival science ADAP projects.

Panel Recommendation:

- *Euclid email exploder for US scientists*: The panel felt that it would be useful for the ENSCI website to provide a mechanism for the broader US community to sign up for a regular mailing list, which would provide and consolidate updates about the progress of the Euclid mission, data releases etc.

- *Survey buildup information*: The panel recommends that, as part of the above or otherwise, ENSCI should provide (or point to) information (i.e., a schedule) detailing the survey depth and sky coverage increase with current and future data releases.

ENSCI General Status:

Summary:

NASA has established the Euclid NASA Science Center at IPAC (ENSCI) in order to support US-based investigations using Euclid data, for both Euclid Consortium scientists, and US investigators at large. ENSCI is part of the Euclid Consortium's Science Ground Segment (SGS), providing algorithm and software development, participating in data quality assurance, and performing data processing. In addition, ENSCI supports the US research community by providing expert insight into the Euclid surveys, data processes, calibration, and products. The ENSCI staff team consists of 18 staff members, with contributions adding up to 11 FTE.

NASA has defined 6 primary tasks (T1-6) for ENSCI by which it measures ENSCI's success. For example, to provide a detector characterization data archive (T2), ENSCI stores 10 TB of data together with NASA's Test Reports - this archive has been in operation since 2016 using IPAC's open-source "Firefly" toolkit, planned to make available in the long-term crucial detector data. Notably, ENSCI passed their Operational Readiness Review (ORR) on May 31, this past year. The User Panel congratulates the ENSCI team for meeting their success criteria and for having impressed the ORR panel. Also in 2023, in time for the Euclid launch, ENSCI released a new website with a sleek and useful new design.

Panel Recommendations:

- *Information support for proposal and project planning*: The panel feels that it is critical for the US research community to have information about the timeline (especially leading up to DR1) of what information will be provided by and when. This would importantly support the preparation of proposals and project-planning by US scientists. In addition to this, it would be important to make clear how users may request additional information on this topic from ENSCI.

- *Collecting feedback from the US community at large*: The panel recommends that ENSCI develops and implements a way to formally solicit user input and feedback from the US research community (e.g., with an annual online survey). Furthermore, ENSCI should provide the resulting feedback to this User Panel in advance of any future User Panel meetings.

IRSA and Public Data:

Summary:

ENSCI has made excellent strides in preparing for the Euclid data, with plans underway to provide users access to the data via cloud access, IVOA-based application program interfaces (APIs), web applications, and the Fornax NASA science platform. The Euclid data products will be discoverable and accessible through Virtual Observatory protocols (e.g. ObsTAP, simple image and simple spectra access for catalogs, images and spectra).

The Euclid-specific web portal allows users to visualize the Euclid sky coverage, access data products and obtain cutout images. The catalog data will also provide easy-to-use photometry and other measured quantities. The web portal, in providing access to these catalogs, including image cutouts, quick look spectra, etc., will serve as a valuable resource to the community.

The Fornax initiative, which provides cloud-based, open-source data and software access in a unified science platform to all the NASA archives, will boost the multi-archive and multi-wavelength science potential for Euclid. However, it was not completely clear what specific Euclid data will be on the US (AWS) servers and how they would be maintained or kept synchronized with the other IRSA-served data.

Panel Recommendations:

- *Plan multiple methods for interacting with and downloading the data*: While many users may access the Euclid data using the Jupyter notebooks and cloud computing interfaces, the panel wants to stress the importance of allowing access to the data in a way that users can also analyze directly on their own machines. Therefore, it is also necessary to support straightforward and simple ways to download the images and spectral cutouts directly. In particular, the panel cautions against the need for installing special Python libraries or packages to access the data if/when possible.

- *Jupyter notebooks and establish best practices policies related to cloud-computing data/code sharing*: The panel hopes that some "starter" notebooks will be available for setting up basic analyses. It wasn't clear whether the Fornax platform allows users to share code and data products, such as user-generated catalogs. If this is allowed and enabled, it would be important for ENSCI to establish policies related to this, including best practices, how such code/data would be vetted or managed, who has access, and how that use would be credited or tracked.

Euclid Consortium Science:

Summary:

The primary science goals of the Euclid Consortium (EC) are to uncover the nature of dark energy via measurements of the large-scale mass density field via weak gravitational lensing (weak lensing, WL) and galaxy clustering (GC). These cosmology-based goals determine survey specifications, though there are several other working groups within the EC whose science interests span stars, planets, transients, galaxy clusters, and galaxy evolution.

The role of ENSCI is to support the US research community by providing expert insight into the Euclid surveys, data processes, calibration, and products. This includes North American EC members within the science working groups (SWGs). NASA has funded three US-based Euclid science teams whose members are part of the EC. These include:

- Constraining Dark Energy & Gravity with Euclid (PI: Rhodes)
- Infrared Background Anisotropies (PI: Kashlinsky)
- Precision Cosmology and Galaxy Growth via Nebular Emission (PI: Chary)

Several other US-based members of the EC exist who may not be members of these teams but who contributed in-kind data to enhance Euclid data products. The goal of ENSCI is to support all US members and integrate them into the EC SWGs and also support them with simulation data products that could help diagnose technical issues and challenges with Euclid datasets. Some of these products are only planned for internal release to the EC.

The EC publication plan calls for Key project papers and standard project papers. The key difference between these is the scope of the science, where key project papers are more central to the primary EC science goals or more broad, whereas standard project papers may deal with one-off interesting objects or a subset of analysis relevant to the primary science goals. They may also touch on topics of joint interest to the EC and outside science collaborations.

Panel Recommendations:

- *EC organization*: Beyond the science goals of the EC, it seems that the structure of the EC overall may be somewhat opaque to even North American members of the EC. ENSCI can play a pivotal role in clarifying structure; it currently seems as if most information is delivered to EC members via tag ups with fewer opportunities to interact asynchronously. Delivering more information directly via email lists or slack may facilitate more interaction amongst North American EC members.

- *Opportunities for non-EC members*: The panel also discussed some ambiguity as to the relationship between the EC SWG areas, the three funded teams, and the opportunities that non-EC members may have to pursue science interests that overlap with these

teams. While there may be opportunities to fund future archival research in calls like NASA ADAP related to Euclid science, will NASA preferentially fund science distinct from the goals of the existing three US-based teams? It would be beneficial for the public to also hear updates on the status of the three funded projects, where currently there is relatively little information (i.e. only the names of the team members).

ENSCI User Support:

Summary:

ENSCI user support prioritizes US-based researchers, but is open to all and part of the larger Euclid science platform ecosystem. Their support focuses on three main areas. First, they are supporting the US Science Teams for their routine operations as well as for providing advice or ancillary resources. Second, they will support US-based researchers. They expect active support of 200-500 US-based researchers in peak years, and will be working with IRSA to provide archival support for Euclid data. Third, ENCSI and IRSA will be the contact of the US-based archival community. They will provide documentation, data products, and additional data analysis tools to the community. Some tools for researchers (like L2 background estimator) are already available, but more will be hosted via IRSA in the coming years. ENSCI's support will be provided through multiple channels: help desk, Slack, tutorials, documents, etc.

The Panel was asked to provide feedback on three specific questions:

1) What types of documentation would be important to request from the Euclid Consortium? What can we provide in ENSCI?
2) Would a webinar for preparation (before Q1 release) be useful?
3) Considering the limitations, which data analysis tools are more important (useful) to provide?

Answers to these questions are provided in the panel recommendation section below.

Panel Recommendations:

We first summarize responses to the specific questions posed to the Users Panel about user support:

1) **What types of documentation would be important to request from the Euclid Consortium? What can we provide in ENSCI?**

    a) The panel feels that both the collaboration structure and the data access methods are opaque at this moment. Documents with sufficient information on

these two aspects are needed. The former should include team structure, observatory information, survey plans, some details of surveys (e.g., PSF, WCS, bandpass model, etc.). The latter should include definitional data storage and access information. The panel thinks that some sample queries (e.g., the SQL examples used in SDSS or Jupyter Notebook) would be necessary to help users to retrieve the data for their own sciences.

2) **Would a webinar for preparation (before Q1 release) be useful?**

    a) We strongly agree that such a pre-introduction to ENSCI and what will be available for the US research community to prepare for the use of Euclid data (especially in the case of tools for preparing e.g. simulations or funding proposals, and specific examples of how to access commonly needed information) would be highly useful.

    b) We recommend the webinar consider both live (to allow direct responses to specific questions/concerns raised from the community) and archived (to preserve that information for reference) use cases. It was suggested to possibly convert the live webinar into a cleaner, section-bookmarked youtube video, for ease of access.

    c) We think it is important to demonstrate some simple, specific science queries in the webinar (or youtube or documents). For example, finding Halpha emitters that meet these conditions and downloading their 1D spectra. This is also related to Point 1a above.

3) **Considering the limitations, which data analysis tools are more important (useful) to provide?**

    a) The highest priority identified was basic photometry and 1d spectra to support e.g., photo-z work and other basic studies of objects.

    b) The ability to use well-defined tools to perform forced (aperture or other) photometry and access 2d spectral cutouts were also identified as a next order need - i.e., basic ability to easily reprocess specific cutouts of the data, vs bulk data download. It would be ideal, for example, if line identifications and initial Flux, FWHM, EW or other object measurements could be provided and matched with cutouts. If these cannot be built-in tools, clear documentation and/or notebooks describing how one could do these would be beneficial to the community.

Additional Panel Recommendations:

- *Tools and documentation*: Some channels/tools may help provide efficient means for providing general user support such as: (1) youtube to share videos of webinars and small demonstrations; (2) an email listserv to broadcast big, high-priority efforts (which could be included in the general emails exploder suggested above), (3) a list of curated inquires from the community based on the existing helpdesk ticketing system – this could be especially useful considering the high demands during peak years.

Expected Data Products:

Summary:

Given the nature and large volume of the data to be obtained with Euclid over its lifetime, the community that will incorporate Euclid data products into their research includes (at least) two groups that will need to be supported by the ENSCI. Primary amongst these are members of the Euclid Consortium based in the US and their affiliated teams. A second group is the broader astrophysical research community who will use Euclid data products in support of other research programs (e.g., as a basis for sample selection, “big data”/machine learning activities, or as a complement with existing multi-wavelength surveys). ENSCI should provide these US community members with detailed descriptions of all available Euclid data products.

The Users Panel recognizes that the pipeline and public data products are still in development in this early stage of the active Euclid mission. Below, the Panel's recommendations focus largely on the user experience with the data products (which may reiterate some recommendations made in the IRSA and Public Data section above), as the specific data product decisions are made by the Euclid Consortium in concert with ENSCI and its support of the US community needs.

Panel Recommendations:

At present, the Users Panel recognizes that steps have already been taken by ENSCI to develop both components of the data products delivery strategy. We note specifically the use of a ``help desk'' to capture and address incoming comments and concerns from Euclid Consortium members on the access and availability of science products. Recommendations on this point are as follows:

- *Detailed Shear Calibration Strategy*: ENSCI should ensure that sufficient information is provided to the community to allow them to understand the methods of shear and redshift calibrations that could potentially allow users to validate these results themselves.

- *Delineate Public vs. Proprietary Data:* Documentation should be made available indicating which data products will be proprietary and which will be made available to the general public. ENSCI should continue to support the help desk to address users' requests for data products.

Regarding the data products and their accessibility by the community, the Users Panel recommends the following to be considered in advance of the Fall 2024 meeting of the Panel:

- *Incorporation within "Fornax":* ENSCI is monitoring the timescale for incorporation into Fornax. Including Euclid data products into this service should improve the visibility of ENSCI and the science that is possible with Euclid to the broader community. If/when incorporated, ENSCI should provide documentation specific to Euclid data products within the Fornax data structure.

- *Community-sourced data products:* ENSCI should help users understand how to meet the IRSA policy requirements necessary for creating user-contributed data products that could then be served by IRSA (e.g., multi-wavelength forced photometry).

- *Data product access:* The method(s) by which the community will access data products should be defined as soon as possible. Considering the ubiquity of the format, Jupyter notebooks could be provided to the user community that describe data access methods and example queries. At a minimum, data retrieval techniques should be provided (as with other IPAC-run query servers) online in an appropriate section of the ENSCI website for reference.

- *Functionality of online query pages:* IPAC has a long track record of providing interactive query and data manipulation websites for querying astrophysical databases (e.g., IRSA: WISE and NASA Exoplanet Archive) that provide quick, easy access to photometry, and reference data products and catalogs. The format for the initial data release should be decided and developed as soon as possible. ENSCI should work with IRSA to support functionality that will allow users to develop virtual tables that can be used to generate (interactive) spectral plots or plots of photometry, derived properties, or community-supported catalogs of the same.

- *Online compendium of substantive changes:* In addition to the advertisement of new data products to users via email, a succinct compendium of substantive changes/updates to data products should be presented on the ENSCI website or Euclid data products interface.

- *Citations for Technical Details/Data Methodology:* When publications from the Euclid consortium member parties themselves do not include sufficiently robust discussion of data reduction (e.g., methodology, systematics), data products (e.g., model assumptions), or value-added catalog analyses, ENSCI should suggest the appropriate documentation to be cited in publications that include such a discussion.