PROCEEDINGS OF SPIE

Euclid's US Science Data Center: lessons learned from building a small part of a big system

Teplitz, Harry, Rusholme, Benjamin, Aussel, Hervé, Dabin, Christophe, Helou, George, et al.

SPIE.

Euclid's US Science Data Center : Lessons Learned from Building a Small Part of a Big System

Harry I. Teplitz[a*], Benjamin Rusholme[a], Hervé Aussel[b,c], Christophe Dabin[d], George Helou[a], Mark Holliman[e], Andrea Zacchei[f] on behalf of the Euclid Consortium

[a]Caltech/IPAC, Pasadena CA 91125; [b]AIM, CEA, CNRS, Université Paris-Saclay, Université Paris Diderot, Sorbonne Paris Cité, F-91191 Gif-sur-Yvette, France; [c]Institut d'Astrophysique de Paris, 98bis Boulevard Arago, F-75014, Paris, France; [d]Centre National d'Etudes Spatiales (CNES), Toulouse, France; [e]University of Edinburgh, Edinburgh, United Kingdom; [f]INAF-Osservatorio Astronomico di Trieste, Via G. B. Tiepolo 11, I-34131 Trieste, Italy

ABSTRACT

Euclid is an ESA M-class mission to study the geometry and nature of the dark universe, slated for launch in mid-2022. NASA is participating in the mission through the contribution of the near-infrared detectors and associated electronics, the nomination of scientists for membership in the Euclid Consortium, and by establishing the Euclid NASA Science Center at IPAC (ENSCI) to support the US community. As part of ENSCI's work, we will participate in the Euclid Science Ground Segment (SGS) and build and operate the US Science Data Center (SDC-US), which will be a node in the distributed data processing system for the mission. SDC-US is one of 10 data centers, and will contribute about 5% of the computing and data storage for the distributed system. We discuss lessons learned in developing a node in a distributed system. For example, there is a significant advantage to SDC-US development in sharing of knowledge, problem solving, and resource burden with other parts of the system. On the other hand, fitting into a system that is distributed geographically and relies on diverse computing environments results in added complexity in constructing SDC-US.

Keywords: Euclid, distributed data processing, lessons learned, US Science Data Center

1. INTRODUCTION

Euclid is an ESA M-class mission to study the geometry and nature of the dark universe. The mission is the second medium-class mission (M2) in the Cosmic Vision program (2015-2025). *Euclid* was adopted by ESA's Science Program Committee (SPC) in June 2012, and is slated for launch in 2022. The *Euclid* mission is described in detail by Laurijs et al. [1].

The *Euclid* design is optimized for measurement of two of the primary probes of Dark Energy: (1) *Weak Gravitational Lensing*, which uses observations of tiny distortions in the images of ~1.5 billion galaxies, to infer the presence of foreground mass concentrations that cause the distortions via gravitational lensing; and (2) *Galaxy Clustering*, which uses the positions and redshifts of ~30 million galaxies to measure the cosmic expansion history through baryonic acoustic oscillations (BAO), which originated from primordial sound waves frozen when the Universe first became transparent. *Euclid* will also enable unprecedented advances in many areas of astrophysics, from objects in the Milky Way halo to the light of the first stars detected in background fluctuations. *Euclid* will deliver high-quality morphologies, masses, and star-formation rates for billions of galaxies, looking back over 75% of the age of the Universe.

Euclid will consist of a 1.2m Korsch space telescope operating at L2 with two instruments: visible light instrument (VIS; [2]) with wide band (550 to 900 nm) imaging with thirty-six 4k by 4k CCDs, covering ~0.5 deg^2 with 0.1 arcsecond plate-scale pixels; and the Near Infrared Spectrometer and Photometer (NISP; e.g. [3] and the references therein), designed for near-infrared (NIR) slitless spectroscopy and imaging, with sixteen 2k by 2k detectors, covering ~0.5 deg^2 with 0.3

---

*[*] Caltech/IPAC, MC 314-6, Caltech, Pasadena CA 91125; hit@ipac.caltech.edu

arcsecond pixels. *Euclid* will survey fifteen thousand deg^2 of the extra-galactic sky (Galactic latitude |b| > 30 deg), and observe Deep Fields encompassing 40 deg^2.

Euclid data will be made public within about two years of acquisition. *Euclid* plans four "quick" releases (possibly 50 deg^2, which preview the data) and three major releases, each including thousands of square degrees.

1.1 The Euclid Science Ground Segment

ESA provided the spacecraft and telescope through a selected industrial contractor, as well as the CCD detectors. The Euclid Project at ESA has responsibility for operating the Mission Operations Center (MOC) and the Science Operations Center (SOC). The Euclid Consortium is tasked with: delivering the VIS and NISP instruments to ESA, developing and implementing the Science Ground Segment (SGS) to analyze *Euclid* data, and performing the science analysis enabled by *Euclid*.

The SGS has been described in detail in previous conference proceedings (e.g. [4] and [5]). Briefly, science requirements for *Euclid* surveys and data processing are determined by Science Working Groups (SWGs). These requirements drive algorithms to be developed by Organizational Units (OUs) within the SGS. Algorithms are used as the basis for production software developed by the Science Data Centers (SDCs), which are also responsible for running the analysis production pipelines. The data processing will be distributed between SDCs in participating countries.

The EC and ESA are working together on the Euclid Archive System (EAS). The EAS will serve as both the real time archive for storage of intermediate products during data processing and as the repository of final science products for the EC and for the public.

1.2 NASA contribution to *Euclid*

In January 2013, NASA joined the *Euclid* mission. A Memorandum of Understanding (MOU) between NASA Headquarters and ESA establishes the high-level roles and responsibilities of each party regarding their collaboration on Euclid. The Explorers Program Office located at the Goddard Space Flight Center (GSFC) is the Program Office for the Euclid project. JPL manages the Euclid Project Office for NASA.

The US has a three-pronged contribution to *Euclid*:

1. NASA delivered to ESA 16 flight and 4 spare Sensor Chip Subsystems for the NISP instrument.

2. The Euclid NASA Science Center at IPAC (ENSCI) has been established to support the NASA-selected *Euclid* investigators and the broader community interested in *Euclid* archival research and to serve as a U.S. node of the Euclid Science Ground Segment.

3. NASA funds three competitively-selected science investigations. These selected science investigations are: (1) "Precision studies of galaxy growth and cosmology enabled through a physical model for nebular emission" (PI R. Chary, Caltech); (2) "Looking at infrared background radiation anisotropies with *Euclid*" (PI A. Kashlinsky, GSFC); and (3) Constraining dark energy and gravity with *Euclid*" (PI J. Rhodes, JPL). Members of the NASA teams have been admitted to the Euclid Consortium.

In addition, there are several US based scientists and engineers who are EC members but are not associated with one of these investigations. Currently, there are about 100 US members of the EC.

In addition to the participation of the NASA-selected teams, many US scientists will use *Euclid* data when it is publicly released. *Euclid* was mentioned in more than 60 Astro2020 Science White Papers. The amount of *Euclid* archival research was discussed in an Astro2020 APC White Paper [6].

2. *EUCLID*'S SCIENCE DATA CENTERS

The processing for the Euclid SGS will be distributed between national SDCs. The system as a whole, comprised of the individual SDCs working together, is a highly heterogeneous collection of resources, varying from supercomputing facilities to university departments, representing a range of hardware architectures, software systems, security models, and

administrative policies. The Euclid SGS pipeline thus assumes little about the computing environment underlying the SDCs, borrowing in part from the CERN distributed processing model.

The overarching pipeline agent issues processing orders to individual SDCs. Within each SDC, local software interprets the orders and runs jobs on a local cluster. The data flow is shown in Figure 1. Raw data from the spacecraft will pass through the Mission Operation Centre in Darmstadt, Germany to the Science Operations Center in Madrid, Spain before being split between the SDCs.
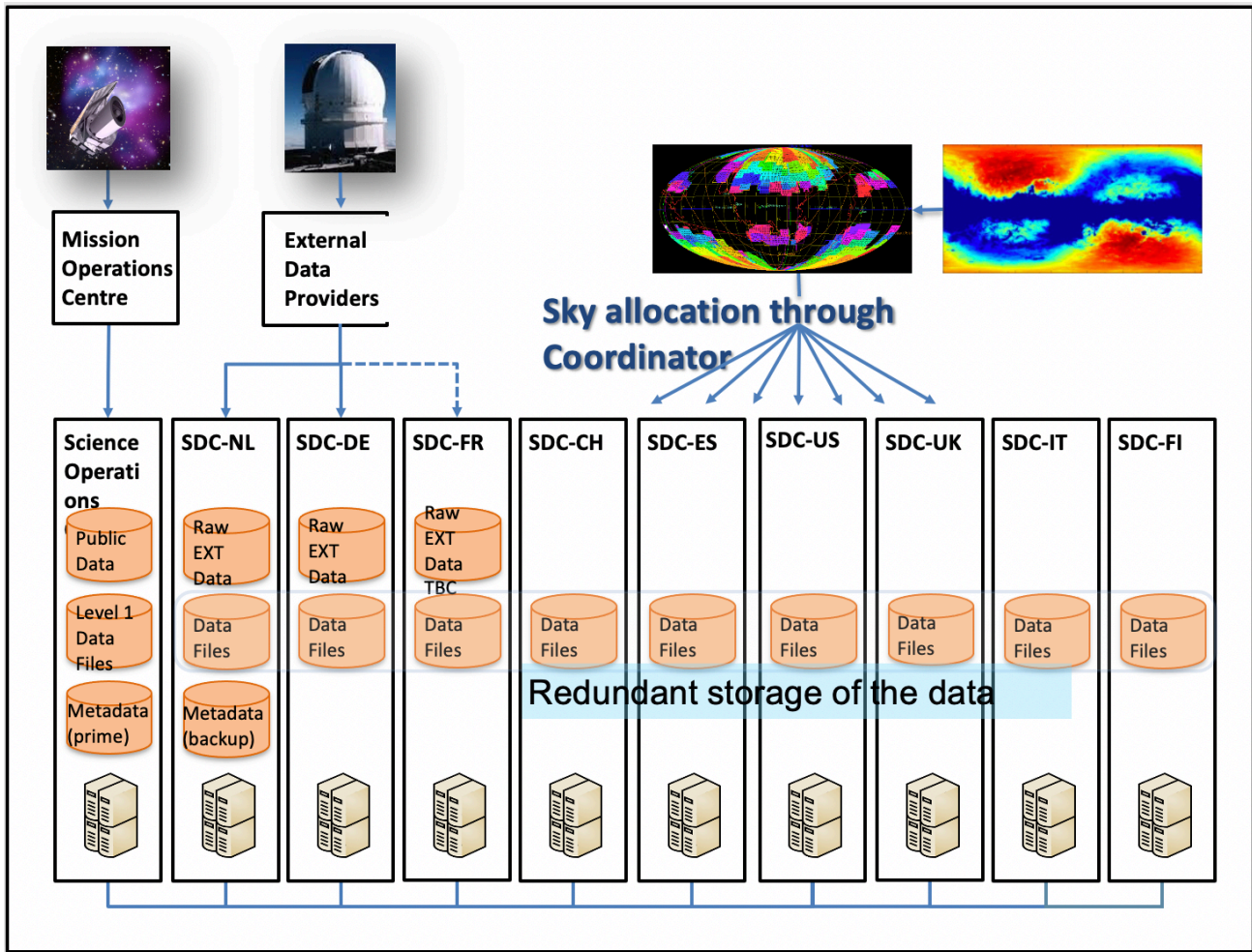

Figure 1. Data flow for the Euclid SGS distributed processing. The figure is used with permission from the Euclid Consortium's SGS Project Office, who provided it.

2.1 SGS development process

SGS development is organized around a series of "Science Challenges" (SCs) of increasing complexity, scope, and integration of different ground-segment components. These challenges are roughly yearly, commencing with a detailed planning exercise, and ending in a documented review process.

Additionally, the System Team (ST) which runs the national SDCs have separate "Infrastructure Tests" (ITs) to investigate performance and robustness of the hardware and software underlying the pipelines. These ITs have historically been scheduled during the development period between the SCs for most efficient use of schedule.

2.2 The US Science Data Center

NASA established ENSCI to support US-based investigations with *Euclid* data. ENSCI is part of IPAC at Caltech, which partners with NASA, NSF, JPL and the world-wide research community on a range of astrophysics and planetary science projects. IPAC fills the science and/or data center role for multiple NASA missions, including Spitzer, WISE, and SPHEREx.

ENSCI participates in the SGS, including software development and operating a node in the distributed processing system, SDC-US. For code development SDC-US is an auxiliary to SDC-Italy. Specifically, SDC-US develops software modules for specific steps in the pipeline, integrates them into the full pipeline, and delivers them to SDC-Italy for final verification.

The SGS distributed organization allows for varying national commitments. There will be only limited replication between the SDCs for load balancing and redundancy. SDC-US is anticipated to provide 5% of the total data processing and storage. Hardware choice and configuration will follow IPAC best practices, ensuring that ENSCI computing is supportable within the IPAC environment and benefits from lessons learned by other big data projects.

3. LESSONS LEARNED IN BUILDING SDC-US

SDC-US is one of the nodes in the Euclid SGS distributed processing system. The system as a whole has been designed to provide both uniform processing and the flexibility to accommodate local requirements. The design has had incorporate the resource planning windows under which the participating national centers operate.

The SGS has overcome a number of challenges in developing this extensive design. Here, we discuss the decisions that have particularly affected SDC-US and the lessons drawn from them.

3.1 Use an interface layer for support local design constraints

A key part of the successful SGS design to support distributed processing is creating a robust interface layer that allows individual SDC's, including SDC-US, to optimize the design of the low-level compute and storage systems in line with local standards. Uniformity of commanding is needed to support the distribution of computing and to meet the goal that almost any part of the processing can be done at any of the SDCs. Ideally, the central processing coordinator "sees" the distributed system as a unified compute cluster on which jobs can be assigned in an efficient way, without regard to exceptions for particular SDC implementations.

Thus, there needs to be a robust interface between this centralized view and the reality of varying infrastructure choices between SDCs. Some centers choose to implement their computing on shared resources, while others procure dedicated infrastructure. The SGS design supports these various choices. SDC-US, in particular, uses dedicated hardware for SGS processing.
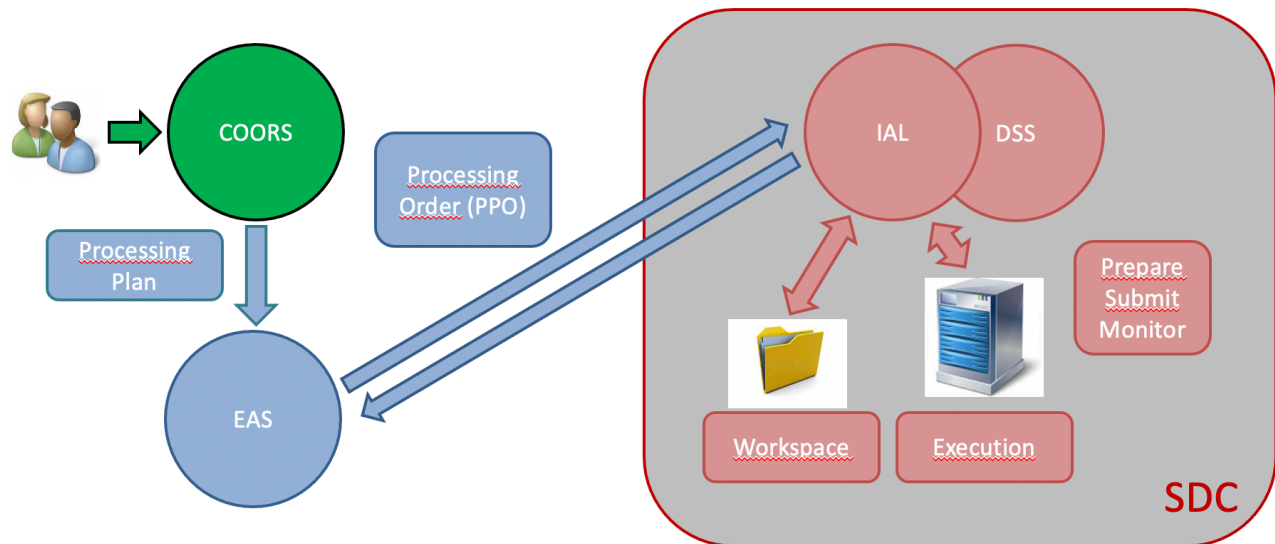
Figure 2. Overview of the components of SGS distributed processing.

Figure 2 shows an overview of the components that the SGS design uses to ensure the separation of system design from local implantation choices. The major components are:

- Central executive software (Common Coordination System; COORS) creates processing plan
- Euclid Archive System (EAS) sends pipeline processing order (PPO) to individual SDCs
- Infrastructure Abstraction Layer (IAL) submits & monitors processing tasks to local high-performance cluster
- Distributed Storage Server (DSS) handles data transfer
- IAL updates EAS with results & metadata as specified in PPO

3.2 Ensure frequent and inclusive communication

In building the distributed SGS system, numerous interests had to be balanced and diverse inputs considered. Some parts of the system have significantly more resources than others, but smaller centers often offer unique expertise or resources. SDC-US in particular is a small part of the system, but strives to offer the benefit of IPAC's experience with other space missions, past and ongoing. Similarly, information flow from the SGS to national centers has been crucial to keep all parts of the project working towards the same goal, both in general and in detail.

Frequent and inclusive communication between SGS stake holders has been key to allowing SDC-US and other smaller parts of the system to keep up with the evolving design. The SGS has ensured frequent organization meetings between sub-system managers, together with regular meetings of domain experts. For example, routine meetings of the System Team, which handles development of the computing infrastructure and challenges, have been essential to the continuing flow of information. Equally important has been the openness of the SGS management (the Project Office) to dedicated meetings with SDC-US to address items of particular interest.

On the SDC-US side, the lesson to take away has been the need for full participation in the available channels of communication. One particular challenge for SDC-US has been the extensive travel required (except during the current work-from-home situation) to attend SGS meetings. We have found that support for this travel is well justified, and the benefit of in person meetings substantially outweighed the additional cost.

To illustrate the benefits of expansive communication channels, it is important to note that information needs to be exchanged on both large- and small-scale problems. For example, we can look to the NISP Interfaces meetings that SDC-

US helped to organize. In these meetings, the SGS brought together the members of several sub-systems that have related concerns but do not always have enough time for communication between the teams. SGS-wide meetings often have too many objectives to dedicate multiple days to NISP-only concerns. On the other hand, each individual team's weekly meetings have to deal with lower level details, during which it can be difficult to include other organizational units. So, the SGS scheduled annual, multi-day meetings dedicated to only NISP imaging and spectroscopy, and including all relevant teams. These meetings included scientists and developers from the pipeline teams, together with representatives of the NISP instrument team and the group responsible for simulating pixel-level data.

3.3 Forecast long-term resource needs

The Euclid SGS is a large system, and it relies on the effort of scientists and developers from all the countries within the EC. Securing funding of algorithm and code development is the responsibility of each contributing center, because the EC does not have centralized budget authority. It is thus necessary for the SGS to provide the information that centers need in order to communicate with their own national funding agencies.

A particularly important aspect of this challenge is the need for long-term forecasting for resource requirements. In the case of SDC-US, effort is funded by NASA, following a plan that encompasses all phases of the mission. This long-term budget plan is standard for NASA projects, both those that are managed solely by NASA itself and those that are contributions to international projects. Our ability to forecast resource needs has been key success in the budget process. These forecasts have included effort by scientists and developers during the pre-launch, operations, and post-mission phases, as well as computing/storage needs anticipated by the System Team.

3.4 Plan proportional distribution of resource expectations

As discussed above, the contribution to the SGS from various countries is not distributed equally. Nonetheless, national centers that are a small part of the system can still play an important role in the project. A key lesson learned from the development planning is to set realistic expectations for the requirements from each center. Proportional distribution of resource expectations has successfully kept all parts of the system involved without being either overwhelmed or ignored.

3.5 Schedule time for iteration

Together with realistic forecasting for resource allocation, SDC-US has benefited from scheduling substantial time for iteration in long-term plans. The distributed nature of SGS development has necessarily increased the time needed for interaction between development teams. In particular, validation of formats, meta-data, and features in simulations have required significant time for iteration. The complexity of the data model required for *Euclid* processing requires buy-in from many stakeholders at diverse locations, naturally increasing the time it takes to reach consensus. SDC-US incorporated this time into our long-term schedules, and experience has supported that decision.

3.6 Enable local testing in advance of system-wide tests

The diversity of the individual SDCs comprising the SGS is a problem for configuring and testing the infrastructure. Software must be flexible enough to be configured for different job schedulers and storage systems. There is no single standard against which to develop and test.

Early science challenges demonstrated the difficulty of separating issues with the underlying infrastructure from those in the processing functions, when everything is tested simultaneously. The system team realized we needed a dedicated infrastructure tests to check for issues. These tests (which approximate the processing with simplified imitation data) have allowed us to continuously investigate the infrastructure without having to coordinate with all the groups that develop the pipeline. These tests, when combined with well-defined discrete updates, have allowed SDC-US to provide *verified* service to the SGS.

3.7 Ensure system-wide tests have the scope and coverage to detect problems

System Team tests have evolved significantly since they were first introduced, as new infrastructure issues have been discovered. This is similar to the process of unit testing software, where tests are written as bugs in the code are fixed. It has been essential to test as many aspects of the SDC-US as possible. Initial tests were designed for basic functionality; receiving and successfully completing commanded PPO instructions. The tests are now becoming more complex and probing the resources provided by the SDC, such as memory per node and storage performance. This provides guidance for both tuning code development for efficiency, and to evaluate local performance and inform future hardware purchases.

The distributed nature of the development adds an additional challenge to the development of validation tests that must run in the full system. Test scripts need to be fully automated and cannot rely on operator intervention. In locally focused testing, it is tempting to assume a knowledgeable operator will be part of the process of running and evaluating validation scripts. In the distributed SGS system, it is imperative that automated scripts be self-contained. The success of the science challenges has relied, in part, on careful planning of validation tests.

3.8 Bandwidth planning and testing

The distributed Euclid SGS processing requires exchanging data between the SDCs, and thus depends on sufficient wide-area network performance. This is a particular concern for SDC-US as a geographical outlier far from all other SDCs. TCP networks constantly tune to prevent congestion, which results in myriad complex issues affecting performance where the round-trip time (RTT) of the synchronize/acknowledge control loop is large.

It can be hard to gain an accurate picture of network performance due the interaction of these issues, especially packet loss in the intermediate hops, which is out of the control of the endpoints. It is thus easy to underestimate the available bandwidth without careful attention to all details. We have found it is critical to measure transfer speed between all relevant locations, in both directions, sampling both at different times of day and over long periods of time, and for a variety of transfer times (or file sizes).

The Euclid SGS is currently using the scripts around the iPerf network measurement tool[†] to monitor bandwidth between all SDCs. Initial tests to/from SDC-US appeared to indicate a problem, but this was determined to be a result of the test design. Specifically, the bandwidth measurement was low because the tests only ran for 2 seconds (compared to approximately 200ms RTT), which was too short to allow the TCP window to scale. Now, the iPerf tests run for 30 seconds and use 5 parallel streams, and current results demonstrate that sufficient bandwidth is available for processing needs.

The Distributed Storage Server which manages the inter-SDC transfer encountered the same issue in handling small files. Originally files were authorized and transferred independently, which never allowed TCP to scale. The DSS has been subsequently improved to use daily token authentication and to batch files together into larger blocks, resulting in much improved transfer performance.

We continue to monitor network performance and in parallel investigate more advanced possibilities such as the perfSONAR network measurement toolkit and Science DMZ architectures[‡].

4. CONCLUSION

As astrophysics increasingly relies on very large collaborations to study increasingly complex questions, *Euclid*'s SGS development provides valuable lessons for the design of future projects. Development of the SDC-US, in particular, has required addressing a number of issues. The distributed nature of the system has been both a challenge and an opportunity. Overall, the process had greatly strengthened the ability of SDC-US to perform as a node within the processing system, as well as the ability to contribute to the system. Many of the lessons learned during the development to date will be useful to keep in mind when new systems are designed.

[†] https://iperf.fr
[‡] https://fasterdata.es.net and https://pacificresearchplatform.org

REFERENCES

[1] Laureijs, R. et al., "Euclid Definition Study Report", 2011arXiv1110.3193L

[2] Cropper, M. et al., "VIS: the visible imager for Euclid", 2018SPIE10698E..28C

[3] Maciaszek, T. et al., "Euclid Near Infrared Spectrometer and Photometer instrument concept and first test results obtained for different breadboards models at the end of phase C", 2016SPIE.9904E..0TM

[4] Zacchei, A. et al., "Management of the science ground segment for the Euclid mission", 2016SPIE.9913E..1BZ

[5] Frailis, M. et al., "The Euclid Science Ground Segment Distributed Infrastructure: System Integration and Challenges", Astronomical Data Analysis Software and Systems XXVI, 2019ASPC..521..612F

[6] Teplitz, H. et al., "Supporting Archival Research with Euclid and SPHEREx Data", 2019BAAS...51g..36T