Multi-Mission Data Processing System (MDPS) Study Review Report

October 25-26, 2023

Cerese Albers, Ph.D., Chair of MDPS Study Review Board

Program Executive, Earth Science Data Systems

Introduction and Background

The purpose of the Earth Science Data Systems (ESDS) program is to oversee the life cycle of NASA's Earth science data- from acquisition through processing and distribution. That includes: processing of instrument data to create Earth System Data Records; active management of NASA's Earth science data as a national asset; upholding NASA's policy of free, full, and open sharing of all data, tools, software, and ancillary information for all users; engaging members of the Earth science community in the evolution of data systems; and developing data system capabilities optimized to support rigorous science investigations and the unique needs of various scientific disciplines. As such, the ESDS program is leading the research and development of technology for management and analysis of complex Earth science data and recently convened two studies aimed at addressing the data processing needs of the upcoming Earth System Observatory (ESO) missions.

The first of the two studies was the ESO Data Processing Study conducted in 2022-2023 which identified two candidate architectures for further exploration. These included architectures named, "Type Two Variant 3" (T2V3) and "Type 2 Variant 4" (T2V4). Type 2 refers to a general classification of "Managed Services" for infrastructure, data, catalog, and analysis services. The Variants are differentiated by factors within Type 2 Managed Services. Variant 3 takes core Science Data System (SDS) infrastructure services and data and catalog services, and adds analysis services to it (e.g., interactive visualization for algorithm development, calibration / validation, and product validation). Variant 4 adds to Variant 3 and includes processing (batch execution only). In addition to further exploration of those candidate architectures, the ESO Data Processing Study made three specific recommendations for follow-on efforts. The first recommendation was that NASA consider a service-based data processing architecture for the ESO missions, requiring a multi-mission organization (MMO) to develop the architecture and standards across missions. It encouraged that such an organization develop and deliver common infrastructure, data, catalog, and analysis services, and consider a generic processing service. These became part of the second study, the Multi-mission Data Processing System (MDPS) Study and gave rise to deliverables that were subject to this review. A second recommendation was that NASA should sponsor a follow-on study to develop the concept of operations, requirements, preliminary design, and a rough cost estimate for service-based architecture. This was partially executed through the MDPS Study, but without addressing cost. The third recommendation was that NASA sponsor an additional study into how the data system could support integrated science use cases that are being developed as part of the ESO mission concepts. This third recommendation is considered forward work by ESDS and not part of the MDPS Study.

MDPS Study Inputs

The inputs to the MDPS Study included the ESO Study Phase 1 findings, SPD-41a requirements, prior architecture standards, interfaces, cloud-based architectures, and cross-center implementations and

workshop presentations, as well as specific ESO use cases, and ESO project requirements (as of July 2023). Inputs were gathered from the Earth Science Data and Information System (ESDIS) project's "SIPS-In-The-Cloud" requirements derived from a collection of existing Joint Polar Satellite System (JPSS) Science Investigator-led Processing Systems (SIPS) requirements. From the ESO missions: Mass Change (MC) comprehensive Science Data System (SDS) Level 3 requirements, the Surface Biology and Geology (SBG) Level 3 and preliminary Level 4 requirements, and a use case from the Atmosphere Observing System (AOS) mission that included a Level 2A Aerosols product with single input and output. The use case from MC included a Level 2 product with complex processing and infrastructure requirements. The use case from SBG included a Level 2 product with multiple inputs and processing steps.

Purpose of Convening an MDPS Study Review

The MDPS Study was well-underway when an official review was convened by Earth Data Officer Katie Baynes for 25-26 October 2023 to review and assess the success of the study from the perspectives of potential key stakeholders of such a system in the future. The review was held as an open meeting for observation and participation both live at NASA Langley Research Center (LaRC) and virtually. The purpose included reviewing the MDPS architecture systems engineering materials, mission use cases, and the contributed mission requirements against the T2V4 candidate architecture recommended by the first study. Review materials, presentations and the Agenda can be found publicly accessible here.

Review Board Representation

The MDPS Study review board was chaired by a Program Executive from the NASA Earth Science Data Systems (ESDS) Program. MDPS Review Board members included representatives from each of the three ESO missions: AOS, SBG, and MC, the ESD Flight Program with an Operating and Other Missions representative, Earth Science center representatives from Ames Research Center (ARC), Goddard Space Flight Center (GSFC), the Jet Propulsion Laboratory (JPL), Langley Research Center (LaRC), and Marshall Space Flight Center (MSFC) as well as the Planetary Data Officer (PDO) and a representative from the Office of the Chief Science Data Officer (OCSDO). A list of official MDPS Study review board members is available in Appendix A.

The review board was briefed by the chair on the protocols, but also on the focus of the review, its objectives, and success criteria.

Review Objectives

- Demonstrate the maturity of the architecture and system design aligns with validating the concept architecture;
- Demonstrate architecture and system design responds to the specified requirements of the T2V4 concept recommendation;
- Demonstrate stakeholder inputs and objectives have been incorporated into requirements and validation objectives;
- Demonstrate the architecture and system definition is sufficient to proceed into a development and prototyping phase.

Review Success Criteria

1. Assess if the maturity of the architecture is responsive and credible in supporting the study objectives;
2. Affirm stakeholder inputs have been documented and incorporated into validation objectives;

3. Ensure that the architecture as documented responds to the specified requirements;
4. Evaluate the degree of definition of requirements, system design, and concept of operations are sufficient to support the prototype.

The Chair charged the review board to consider the objectives and success criteria from their various perspectives and expertise. The charge was to evaluate: 1) if the study's approach was feasible to architect a common data processing system appropriate for use by a mission or project, 2) what risks must the study specifically address in order to make successful, implementable recommendations, 3) can the candidate architecture meet as-designed project mission data processing requirements, 4) if there were any specific concerns about the concept of operations compared to existing concepts and processes underlying current mission data processing pipelines, infrastructure, interfaces, etc., and 5) in a separate closed session at the end, if any programmatic concerns or risks about the overall approach might make it difficult to implement and might need additional exploration. Since programmatic concerns such as cost, support for interdisciplinary science, structure of system reviews, types of service agreements, requirements at the NASA level, organization structure, etc. were out of scope, the review board's programmatic recommendations resulted in a separate report that the Chair authored based on collective review board feedback and sent to the Earth Data Officer for separate consideration. Review board members were encouraged to ask questions throughout the review and to submit Request For Action (RFA) forms on all of these points.

Review Progression and Summary of Discussions

The agenda for the MDPS Study review is publicly available here.

The MDPS Study introduction began by outlining the differences between the proposed architectural types and variants, and managed services. It discussed key aspects and attributes, including how the design includes software capabilities and applications that would be provided and managed by a third party, a Multi-Mission Organization (MMO), to take care of the technical aspects. Applications and capabilities would be intentionally designed and developed from the outset to be provided and managed this way, and delivered over the internet so missions and projects could be accessed from anywhere via secure connections. The MMO would take care of provisioning server infrastructure, maintenance, and updates. Services could be made multi- or single-tenant so different versions could meet different requirements. The key attributes include scalability, multi-tenancy, automation of updates (with coordination), security and compliance, built-in collaboration potential, and integration of systems. The MDPS could have deployment flexibility focused on open-source compliance with NASA Science Policy Directive 41a (SPD-41a). A positive impact on future projects from this could be reduced concern on the core capabilities of a processing system, their related workforce staffing, infrastructure, software, and security compliance, the enablement of rapid development of science algorithms with a validation environment. However, some attendees viewed this as a source of potential concern in handing all of that over to a single organization. Additional potential impacts that were viewed positively included that the baseline deployment of production-like project MDPS could be deferred until required since configuration could be done in mere weeks, and it would be inherently compliant with open source science objectives. However, it was noted that successful implementation would require components that were out of scope for the review such as Service Level Agreements (SLA), operational level agreements, and a high degree of project coordination. For the rest of the first day, the MDPS

Study review focused on detailed presentations of the T2V4 Concept of Operations, the T2V4 System Requirements Review, and the T2V4 System Design Review.

The second day of the review began with the remaining presentations from the System Architecture Advisory Team (SAAT) on MDPS architecture and a design assessment, and a presentation on Open Science Alignment of the proposed MDPS by the ESDS Headquarters program scientist, Dr. Katherine Saad, who was also a team member of the SAAT. The SAAT was charged with evaluating the proposed architecture of the MDPS. Based on the outcomes of the workshops held during the ESO Data Processing Study and resulting data synthesis, four evaluation criteria were defined. This included mission development, enabling data system efficiencies, supporting Earth systems and science applications, and promoting open science principles. The SAAT took the review objectives and coupled them to the four evaluation criteria, and then evaluated the MDPS Study artifacts in order to inform the architecture and design of the MDPS. The artifacts the SAAT evaluated included the MDPS Concept of Operations document, operational scenarios, MDPS Level 1 through Level 3 requirements and their traces, architecture and design diagrams and presentations, mapping of requirements to architecture components, mapping the T2V4 recommended architecture to the MDPS proposed architecture, and mapping of evaluation criteria to requirements. For each of the four objectives, the SAAT presented its assessment, a summary of its findings, and indicated any concerns and risks.

The SAAT also provided valuable additional considerations and suggestions. It recommended that the MDPS focus on simple, foundational services to bound the MDPS scope. It also recommended that multi-organization managed services will require more stringent technical scope and interface agreements. The SAAT recommended the MDPS support a hybrid infrastructure (on-prem, cloud, and high-end computing) since several ESD missions were already moving in that direction. The SAAT also pointed out that fluidity in what constitutes a mission-deployed service or managed service made the architecture difficult to assess. It recommended up-front attention to efficiently build a service capable of being shared among multiple missions. The SAAT noted that even excellent examples of tenant-deployed services tend to be significant sources of expense and operational overhead, which needs to be considered for anything tenant-deployed on a NASA system. Finally, the SAAT indicated that T2V3 architecture was more achievable, and that the T2V4 which adds generic processing services was indeed a stretch goal. Overall, the SAAT deemed that the artifacts they assessed were sufficiently developed and the implementation team had demonstrated sufficient readiness to move into the prototype development phase.

The Open Source Science evaluation of MDPS requirements against SPD-41a was conducted as the final piece of the review. The ESDS HQ team had developed an initial compliance matrix to trace which SPD-41a requirements would apply to the MDPS. With that as a starting point, Dr. Saad assessed whether the plan for the MDPS aligned with the ability to advance open science, evaluated ways in which open science principles were incorporated into the system architecture, and identified the ways that the architecture considered relevant elements of SPD-41a. Also included were the guardrails of Findable, Accessible, Interoperable, and Reusable (FAIR) principles highly endorsed by NASA SMD and the Desirable Characteristics of Data Repositories for Federally Funded Research released in May 2022 by the National Science and Technology Council's Subcommittee On Open Science. Dr. Saad concluded with remarks that to what extent the elements of MDPS might be considered a repository is an ongoing discussion, but it still might have implications for enabling capabilities and/or be applicable in niche cases. Dr. Saad recommended further exploration using federal guidance on the creation of accessible

digital content (software, websites, documents, applications) and ideally with input from accessibility subject matter experts.

After these sessions, the review board was briefed again by the Chair before a closed session discussing in-scope feedback. After the closed review, the Chair presented a generalized debrief of the discussion to the open audience. Immediately following, the Chair and the Convener conducted an impact assessment open discussion of overall themes that had arisen across the two-day review. The review concluded with mention of next steps, and an expression of gratitude to the reviewers and participants. Instructions were clearly provided that Requests For Action (RFA) forms would be accepted until 31 October 2023 via the online form and could only be submitted by official board members. The board members were reminded that they could submit an RFA on behalf of anyone so any participant could ensure their feedback was included by contacting a board member.

Review Board Feedback

Requests for Action

The review board was given until 31 October 2023 to submit Request For Action (RFA) forms, and fifty-nine were submitted. Of those fifty-nine, forty-six were accepted and thirteen were deferred. The RFAs were categorized into five affinity groups: additional use cases (eleven), clarification (fourteen), requirements update (thirteen), programmatic (thirteen), and further study (eight). Specific RFAs and their dispositions are openly available here.

A key finding from the review from the first affinity group of RFAs was to solicit additional use cases for inclusion in the study. The Earth Data Officer Katie Baynes commenced this action on 27 November 2023, and asked key stakeholders to respond with additional use cases by 12 January 2024. Another key finding from the review revolved around getting clarity on scope and deliverables, and around adding specificity or impact of specific features of the proposed architecture. RFAs on requirements suggested the inclusion of additional requirements, removal of some existing requirements and suggested changes to existing requirements. Programmatic RFAs were suggestions, as they were out of scope for the formal review, but generally included recommendations on implementation and operational team staffing and planning, on infrastructure capacity planning and scalability, on cost modeling support to teams in implementation phase (commissioning), and on scope increase beyond the T2V4 architecture. The final set of RFAs focused on topics recommended for further study. There were multiple recommendations on evaluating duplicative capabilities and potential reuse of existing software components and capabilities. There were multiple suggestions to gain a fuller understanding of the relationship and interfaces between MDPS and infrastructures of SMD Core Services as well as ESDIS enterprise offerings. Finally, there were recommendations to further study aspects of project needs for inclusion within the MDPS.

Selected Topics and Themes from Requests For Action and Review Board Sessions

- *Architecture Requirements:* Several RFAs submitted focused on the T2V4 system architecture requirements. There were suggestions to separate the requirements of the system from those of any third-party service provider. In this way, the requirements of a particular cloud provider, a high-end computing center, or an on-premises system would be clearly delineated from those of the MDPS.

- *Down Time Requirement:* There was a request that zero planned down time be removed as a requirement, as the provided rationale was insufficient. The board noted that there should still be some down-time requirement.
- *Version Roll-back Requirement:* The board recommended modifying the requirement on managed services providing a common rollback mechanism to a requirement to just performing rollbacks as needed.
- *Open Source Science:* The board noted that the proposed open development of the MDPS might benefit the open science community more than any individual mission could. It encouraged investigating further how the MDPS would help missions meet the requirements in SPD-41a and map its support accordingly for potential stakeholder consideration. Regarding system requirements for open source science, the board offered that rather than listing specific licenses it would be more appropriate to say, "an open source license will be chosen compliant with SPD-41a."
- *Reuse of enterprise tools and services:* Several RFAs encouraged the MDPS Study team to explore the feasibility of reusing existing ESDS programmatic tools and services in lieu of custom development, in order to reduce the potential risk of spending development resources where existing services and architectures could be reused that already have stakeholder familiarity. Examples of reuse that were offered included GIBS and Worldview as well as a recommendation to explore the feasibility of instantiating a Cumulus instance for data storage and distribution and linking directly to a DAAC instance of it, which could prove useful to mission tenants. In a similar vein, the MDPS Study team was encouraged by the board to explore ways to hook into existing Integrated Development Environments (IDEs) for stakeholders used to their own pre-existing IDEs. Therefore, the board recommended overall that the ESDS program perform a gap analysis on what infrastructure and services already exist and are possibly open sourced already that the MDPS could leverage.
- *Benefits of Additional Systems Engineering Support:* The ability for an MMO to estimate the compute resources needed for scientific data processing for a given mission tenant was questioned by the review board, which recommended additional systems engineering be put in place to manage the uncertainty in scoping resource requirements. Also, the board discussed that systems engineering support to help prospective tenants with cost estimations would be a wise approach. Accurate estimations would require empirical data, and then wise use of it to make good estimates of costs before deployment. The board asserted that an MDPS MMO would need to be sufficiently staffed to support encountering multiple new issues simultaneously in order to satisfy data processing requirements. Additionally, specific mention was made that reprocessing rates as listed in the requirements are less than the current state of art. A recommendation by the board was to provide revised reprocessing capability to the expected volumes and code complexity of ESO missions, such as SBG.
- *Analysis Use Cases:* The review board also recommended that there be a collection and review of analysis use cases beyond the ESO missions to enable robust data transformation and data analysis capabilities in the MDPS.
- *MMO Staffing:* The board recommended increasing operational support for missions during the commissioning phase activities, including assigning MMO liaisons. The board requested investigation into what the minimum number of concurrent missions/tenants would need to be

in the MDPS to support adequate Mission Management Organization (MMO) staff for all tenants, and additionally, what the maximum number would be.

- *Load Testing:* It was noted by the review board that a requirement that the MDPS support multiple concurrent tenants implies the necessary establishment of a minimum number of tenants to load test the system, but a maximum should be established as well.
- *Benchmarking run time performance and summary statistics:* The board also expressed that the commonly available application catalog should have a live catalog of benchmarked aggregate run time performances and useful summary statistics to increase the value to both current and prospective tenants of the MDPS.
- *Supporting Applied Sciences, Suborbital, and other communities in the MDPS:* It was noted that the architecting within SMD is more geared toward orbital missions, but that suborbital missions could be supported by this system as well if it were architected to support those use cases, including appropriate MMO staffing support. There was also discussion that Earth Action / Applied Sciences community interests were not well-represented at the MDPS review. The board did discuss that any ultimate MDPS construct should be scoped appropriately and that it discouraged crafting MDPS to be "everything to everyone."
- *Expanding Prototyping Activities to Include ESO Missions and additional platforms:* Prototyping for missions that do not yet exist is inherently difficult. The board discussed that there should be a feedback-loop during prototyping activities with ESO mission teams. The board also encouraged ESDS to consider what other types of missions would benefit from MDPS based on requirements, including additional prototyping such as a case study for a mission using the high-end computing capability of the MDPS instead of cloud. The board indicated that this would more fairly allow the MDPS system to be evaluated on its ability to support missions across multiple processing platforms.
- *Cost Benefit Analysis and Business Case Development:* Regarding prototyping, the board strongly encouraged ESDS to conduct a cost benefit analysis to elucidate and quantify potential efficiency gains, and to make a business case for funding the full-scale development of an MDPS. Such a business case could prove beneficial in overcoming current inertia for individual data processing pipelines. Furthermore, the board recommended that ESDS identify a potential stakeholder that would want to adopt the MDPS paradigm and encourage that stakeholder to provide ongoing feedback during the prototyping process. This in turn could result in stakeholder advocacy and could drive increasing return on investments.
- *Architecting with Ability to have Tenants from Other SMD Divisions:* The review board members that were external to Earth Science advocated that the MDPS concept offers an opportunity for tenants in other SMD divisions, as long as that potential is architected in. The review board discussed that missions external to ESD would incur technical debt from the MDPS being developed with too much of a geo-centric system and that other missions would then have added cost to adapt to be able to use the MDPS. A science-discipline-agnostic system could potentially allow for funded support to come from other missions as tenants outside of Earth Science. Furthermore, building the MDPS as an open source system will readily enable contributions from other divisions which would increase the applicability of the MDPS to support external missions. The review board encouraged broad participation across SMD to determine what SMD wants versus what they have requested.

- *Taking a Gradual Approach to Rolling Out MDPS Capabilities:* The board had a lot of discussion on the need to build incrementally, not as a holistic system. Instead of developing the entire proposed architecture at once, the review board discussed the value of developing a la carte services that could be available to missions sooner than an entire MPDS. Services that support calibration and validation that are in line with the benefits of the T2V3 architecture were recommended. Building stepping stones that build up the capabilities gradually and permit tenants of all classes early could be highly beneficial as long as limitations are clearly expressed.
- *Concerns with the human element of a multi-tenant MDPS Architecture:* Multiple teams from different organizations with differing priorities are going to need to have their development schedules and operations closely coupled with one another as they work on their respective systems. Beyond that, the deployment model ensures a need for significant coordination and knowledge sharing. The recommendation was to consider studying the human factors through the prototyping activities, noting that realistic operations simulations need to occur ultimately to know if the recommended architecture is truly feasible.

Conclusion by the MDPS Study Review Board

In response to evaluating the success criteria for the review, the review board concluded that: The maturity of the architecture is responsive and credible in supporting the study objectives; Stakeholder inputs have indeed been documented and incorporated into validation objectives; The architecture as documented responds to the specified requirements; and the board evaluated that the degree of definition of requirements, system design, and concept of operations are sufficient to support moving to prototyping. However, the board noted that the proposed prototyping activity was solely cloud-based. Thus, it recommended the ESDS program collect additional use cases for the MDPS Study team to analyze and that ESDS consider commissioning additional non-cloud platform tenant prototype development to fairly assess the MDPS capacity to support both types of tenants.

Summary

The MDPS Study review was convened by the Earth Data Officer on 25-26 October 2023 as an open review for observation and participation, with the goal of assessing the success of the Earth System Observatory (ESO) Multi-Mission Data Processing System (MDPS) study. The purpose included reviewing the MDPS architecture systems engineering materials, mission use cases, and the contributed mission requirements against the T2V4 candidate architecture recommendation from the prior study. The MDPS Study aimed to respond to recommendations from that prior study, called the ESO Data Processing Study. The first study was driven to investigate whether a common MDPS architecture could be used across the ESO projects to process mission science data, could it promote open science principles, could it enable efficiencies, and could it advance Earth system science and applications. The second study essentially aims to further refine the first study's recommendations and to validate the architecture with ESO use cases and requirements in a prototype deployment.

The objectives of the MDPS Study Review were to demonstrate the maturity of the architecture and system design, its responsiveness to the specified requirements of the concept recommendation, that the stakeholder inputs and objectives had been incorporated, and that the architecture and system definition is sufficient to proceed to prototyping.

The review board representation was inclusive of various kinds of stakeholders. Though the review was held openly and participation in discussion was encouraged by members and non-members of the MDPS Study review board alike, the official board included representatives from all of the ESO missions, the NASA Centers' Earth science programs, the mission program offices of the Earth Science Division, a representative from the Office of the Chief Science Data Officer (OCSDO), and the perspective from another SMD division: the Planetary Science Data Officer. These review board members submitted a total of fifty-nine Requests For Action (RFAs), which were dispositioned by the MDPS Study team and the Review Board Chair. Forty-six were accepted and thirteen were deferred. The RFAs fell into five general affinity groupings: Additional Use Cases (11), Clarification (14), Requirements Update (13), Programmatic (13), and Further Study (8). Review board discussion themes, general outcomes, and key findings from these RFAs have been summarized in this report.

Overall, the review board concluded that the MDPS Study review met the four success criteria set by the ESDS Program that were sufficient to evaluate how well the MDPS Study had achieved the objectives and whether it was ready to proceed to prototyping. In addition, the Earth Data Officer provided the review board with the opportunity to provide programmatic feedback, which was outside the scope of the review, but resulted in thirteen of the RFAs. These RFA themes and the review board's closed programmatic recommendations were summarized and shared with the Earth Data Officer for future consideration.

Appendix A: Review Board Membership and Role

Official Review Board Member Name	**Representing**
Cerese Albers	ESDS / Chair of the Review Board
Mike Obland	ESD Flight, Operating, & Other ESD Missions
Curt Tilmes	AOS
Felix Landerer	GRACE-C
Peter Xaypraseuth	SBG
Jon Jenkins	ARC
Sean Bailey	GSFC
Jennifer Cruz	JPL
Jeff Walter	LaRC
Brent Roberts	MSFC
Andrew Mitchell	OCSDO
Robin Fergason	Other SMD Division (Planetary Science Division)