Final Report of the Open Source Science for Earth System Observatory Mission Data Processing Architecture Study

Steering Committee

*Andrew Bingham, JPL**

Andrew Mitchell, GSFC

Chelle Gentemann, Farallon Institute

*Luke Dahl, JPL**

System Architecture Working Group

Natasha E. Stavros, U. of Colorado

*Elias Sayfi, JPL**

Andrew Michaelis, ARC

Evelyn Ho, GSFC

Leslie Ott, GSFC

Curt Tilmes, GSFC

*Bernie Bienstock, JPL**

*Hook Hua, JPL**

*Qing Yue, JPL**

Wenying Su, LaRC

Sean Harkin, MSFC

Adrian Parker, NOAA

Christopher Engebretson, USGS

Study System Engineering

Sara Lubkin, GSFC

*Karen Yuen, JPL**

National Aeronautics and Space Administration

*Jet Propulsion Laboratory
California Institute of Technology

January 2023

CL#23-0918

This research was carried out under a contract with the National Aeronautics and Space Administration.

CONTENTS

APPENDIX

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

The Open Source Science for Earth System Observatory (ESO) Mission Data Processing Architecture Study was sponsored by Kevin Murphy, Chief Science Data Officer of NASA's Science Mission Directorate (SMD) and Program Manager for the Earth Science Division (ESD) Data Systems.

The study purpose is to assess if a common Mission Data Processing System (MDPS) architecture can be used across the ESO projects to process the mission science data, while promoting open science principles, enabling efficiencies, and advancing Earth system science and applications. To make this assessment, the study formed two teams consisting of a diverse set of experts in the field of science data processing systems: a Steering Committee (SC) responsible for the leadership of the study and making programmatic assessments, and a System Architecture Working Group (SAWG), responsible for the technical assessments. The study held two open workshops to understand the ESO mission science data processing needs, identify stakeholder objectives and constraints, and understand the state of the art in mission science data processing systems. This information was used by the SAWG to identify three types of architectures that could meet the objectives of the study and improve on the current approach to implementing the MDPS across projects. In a Type 1 architecture, each project develops an MDPS independently, but conforms to varying degrees of development and interface standards and policies. In a Type 2 architecture, each mission develops a MDPS using common services provided and managed by a multimission organization. Variations within this architecture are the number of services that are available to use. In a Type 3 architecture each mission develops its MDPS on a multi-tenant platform that is provided and managed by a multimission organization.

To select the optimal architecture, we evaluated the desirability of an architecture against its complexity. The Desirability Score (SC) measures the degree to which an architecture responds to the study objectives. This score was developed and scored jointly by the SC and SAWG. The Complexity Score (CS) combined the assessment of the Technical Complexity (TC) provided by the SAWG with the assessment of the programmatic complexity provided by the SC. By maximizing the desirability and minimizing the complexity, we identified that a Type 2 architecture (managed services) with some infrastructure, data, catalog, and analysis services provided by a multimission organization is the optimal architecture for the ESO missions.

We recommend that NASA conduct a follow-on study to establish a preliminary design and implementation approach for the Type 2 architecture. The study should include strong engagement from the ESO missions and provide an assessment of any impacts to the project. This study should commence as quickly as possible to reduce delays in support of active development ESO missions. Throughout the study, we identified a strong desire by the science community for a data system to support cross-mission analysis, since this was out of scope for this particular study, we recommend an additional study that addresses these use cases and builds on this study.

1 STUDY INTRODUCTION AND OBJECTIVES

A critical component of an Earth observing mission is the ground-based system that transforms raw data collected by the instrument into archive ready, scientifically valid data products. This system, which we refer to as the Mission Data Processing System (MDPS), consists of geophysical retrieval algorithms, operational software, computing infrastructure, documentation, and team procedures. It also includes the suite of software tools that support the development of the processing algorithms, data validation and analysis, and data reprocessing.

To date, each flight project builds a mission-specific MDPS that is tightly coupled with the instrument system and is responsive to the needs of the project science team during the implementation and operations phase of the project. This approach ensures mission success, but has long-term challenges. A dedicated staff with deep knowledge of all aspects of the MDPS must be maintained throughout the project lifecycle; intersystem dependencies limit the infusion of new technology; the project on its own is unable to negotiate on commodity costs such as compute resources; and silos of knowledge are formed that are difficult for external scientists to penetrate. Moreover, this approach may also make it challenging for projects to meet the objectives of NASA's SMD Policy Directive-41 (SPD-41) that calls for openness in mission data and software.

NASA's development of the Earth System Observatory (ESO) provided a unique opportunity to study whether a shared approach to building an MDPS across missions could overcome these challenges. The ESO consists of four missions intended to obtain coordinated measurements to study the Earth as a system and enable applications.

This study was sponsored by Kevin Murphy, Chief Science Data Officer (CSDO) of NASA's Science Mission Directorate (SMD) and Program Manager for the Earth Science Division's (ESD) Data Systems. The objective of the study was to identify and assess potential architectures that can:

- Meet the science processing objectives of the Surface Deformation and Change (SDC), Surface Biology and Geology (SBG), Mass Change (MC), and Atmosphere Observing System (AOS) missions
- Promote open science principles
- Enable data system efficiencies, and
- Seek opportunities that support Earth system science and applications

This study was largely guided by the following definitions and constraints:

- Open science is defined as a collaborative culture enabled by technology that empowers the open sharing of data, information and knowledge within the scientific community and the wider public to accelerate scientific research and understanding (Ramachandran et. al, 2021).

- The MDPS will adhere to open data, software, algorithm, and publication policies (i.e., SMD Data and Information Policy currently in draft form).
- Missions will develop algorithms (Level 1–4), software, and documentation in open systems from inception.
- Representatives from each of the ESO missions will participate as members of the study team.
- The study must solicit input from a broad and diverse set of flight project teams, industry partners, open science experts, and stakeholders across a diverse spectrum of the science mission data systems community.
- The study must be conducted in an open manner, through public workshops and open Requests for Information (RFI), and include broad social media and outreach efforts. The public workshops were designed to enable community participation, engage key stakeholders and promote diverse and inclusive discussions. In all, the study held two of the workshops and one hybrid meeting.

2 STUDY APPROACH

The study consisted of two teams: the Steering Committee (SC) and the System Architecture Working Group (SAWG). The Steering Committee had overall responsibility for the study; they interfaced with the external stakeholders, evaluated programmatic related considerations, and reported to the CSDO. The SAWG had the responsibility to identify common MDPS architectures that could be shared across the missions and assessed if each could meet the objectives of this study. The membership of the SAWG was selected to ensure that a diverse set of insights and opinions were considered. The members of each team, their qualifications, and organizational affiliations are listed in the Appendix.

The workflow for the study is highlighted in Figure 2-1. First, the ESO stakeholder objectives, ESO mission requirements, and community state of the practice were collected through two public workshops. The NASA stakeholders (Program Directors) set the top-level objectives for the overall ESO program with respect to the need and use of NASA's science data. The ESO mission science and MDPS requirements, objectives, and constraints provided the SAWG with an understanding of the scope, purpose, and performance of each MDPS. The community state of the practice provides insight into the current and future MDPS architectures that are used within and outside of NASA.

The SAWG analyzed and studied the information provided from these sources to identify the potential architectures that could satisfy the objectives of this study and assess the merits of each architecture. Next, the SC assessed the architectures from a programmatic perspective The results of these assessments were combined to make a final recommendation.

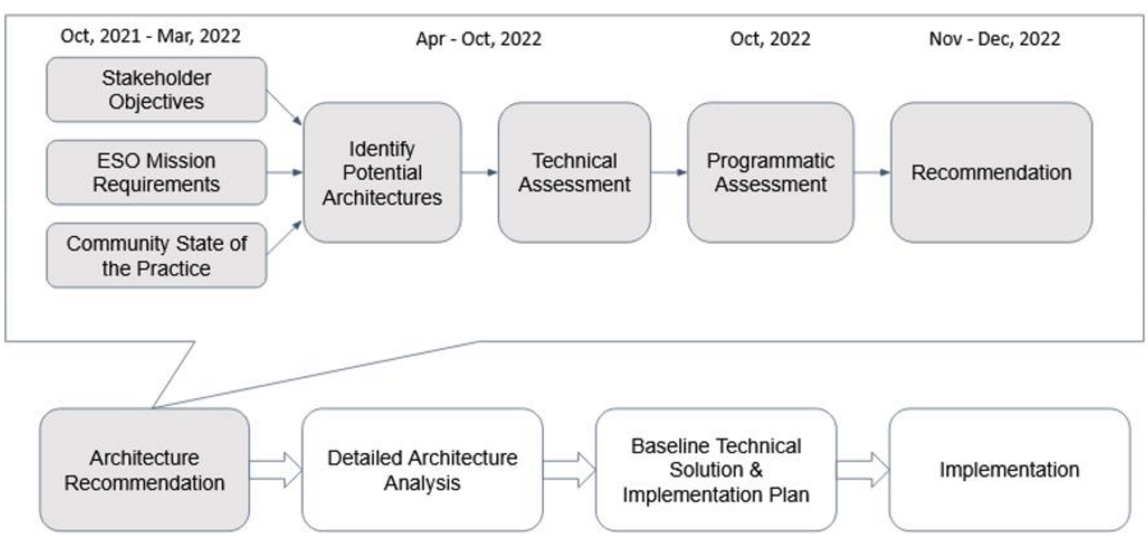


Figure 2-1. Summary of the approach used to conduct the study.

3 SUMMARY OF WORKSHOP FINDINGS

The study was conducted via two publicly attended virtual workshops, as well as weekly meetings held by members of both the SC and SAWG, along with separate focused sessions as necessary. Workshop #1 (October 1-20, 2021) focused on gathering needs and considerations for evaluating different open science data system architectures to support Earth system sciences and mission science data system efficiencies with the explicit goal of informing both qualitative and quantitative evaluation criteria. Workshop #2 (March 1-4, 2022) aimed to understand the state of practice in MDPS and open science, and sought community input on data system architectures. The weekly sessions conducted by the SAWG analyzed workshop findings to inform both evaluation metrics and development of MDPS architectures.

The findings of Workshop #1, with registered participants totaled at 141, were evaluated as stakeholder priorities, stakeholder considerations and constraints, and a synthesis of common themes by study objective. For a detailed description of stakeholder priorities, considerations and constraints, see the Workshop #1 Final Report (Stavros et al. 2021, https://hdl.handle.net/2014/53042). Important to the methodology for recommending an architecture is the synthesis of common themes by study objective, which were distilled into evaluation criteria of “desirability” or the degree to which the architecture responds to the study objectives, against which each candidate architecture was evaluated (See Technical Assessment). These themes were:

1. ESO MDPS locations in the cloud and/or on-premises (on-prem), support forward, on-demand and low latency processing, interfaces with external systems, and are cost-constrained;
2. Availability of efficiency opportunities, including NASA Distributed Active Archive Center (DAAC) co-location, flexibility/scalability to adapt to varying data volumes/compute needs, and common data formats;
3. Advancement of Earth system science by sharing of data/algorithms, supporting multidisciplinary research, and a common architecture that enables cross-ESO science objectives; and
4. Promotion of open science needs are publicly accessible, implemented on an extensible analysis platform with access controls that can track metrics both for cost accounting and adoption encouragement.

The findings of Workshop #2 (Stavros et al. 2022, http://hdl.handle.net/2014/54626), with registered participants totaled at 134, helped inform the design and development of candidate architectures for evaluation. NASA/non-NASA and mission/non-mission MDPS contained common functionalities, and were built on a wide variety of implementations. The main common components, at the highest level, were the data component and the processing component, which were the main drivers for the architectural decisions and costs. With science data processing data volumes surpassing petabyte scales, processing nodes from a few to thousands, the various MDPS were deployed either solely in an on-premises facility, wholly in a commercial cloud platform such as Amazon Web Services (AWS), or hybrid (on-premises and cloud). Some took advantage of NASA High End Computing Capability (HECC) for overflow and reprocessing. Both software and hardware heritage were prominent across missions implemented within an organization and evaluated as a leading factor in making subsequent missions cheaper and more efficient.

Workshop #2 identified three main MDPS architectures as: 1) Single Instance: one system for one mission, 2) Multimission System: one instance to process multiple missions; and 3) Co-located MDPS and DAAC: one system for one mission but sharing functions with the DAAC.

A common theme among the non-NASA systems e.g., National Oceanic and Atmospheric Administration (NOAA), Japan Aerospace Exploration Agency (JAXA), Indian Space Research Organisation (ISRO), Agenzia Spaziale Italiana (ASI), German Aerospace Center (DLR), was their multimission designs (one instantiation to support multiple missions), as well as their implementation of a data lake. The NASA missions were generally single instantiations per mission. NASA's Earth Science Data and Information Systems (ESDIS) migration of all DAAC data to AWS is a foundational step in constructing a data lake, around which efficient processing and access services are under development.

Discussions from Workshop #2 highlighted some open issues. While NASA Scientific Information Policy (SPD-41) mandates the policy, the challenges between open science and cybersecurity remain unaddressed, especially at the organizational level. Better processes for community contributions, interoperability, quality assurance, data and metadata standards, and the advancement of capabilities such as Analysis Ready Data (ARD) remain a challenge. When examining multimission systems, the increased efficiency and support for system science must be weighed against cost management (particularly if it's opened for public use) and interdependency complications. Movement to the commercial cloud is a prominent, yet open debate on the benefits and limitations of on-prem versus cloud remain, especially with respect to cost and capabilities.

With the criteria defined from Workshop # 1 and the state of practice in MDPS survey completed in Workshop #2, the SAWG designed architectures (See System Architecture Identification (SAWG)) and performed a trade study (See Technical Assessment) that establishes viable architectures and implementation approaches. The findings of the trade study and the recommended architecture options are presented in this report.

4 SYSTEM ARCHITECTURE IDENTIFICATION (SAWG)

The SAWG employed an approach that mapped community feedback from Workshop #1 and Workshop #2, the latter of which collected the state of science data processing systems. From the feedback, a set of desirements and complexity factors, which consisted of feasibility and maturity (see Appendix for details) were obtained and determined to be relevant for evaluating candidate architectures. The approach assessed various architecture options and how each could meet the community's desirements against complexity factors. The foundational desirements are:

- ESO Mission Science Processing Objectives
- Support Earth System and Science Applications
- Promote Open Science Principles
- Enable Data System Efficiencies

The ESO Mission Science Processing Objectives included baseline needs for ESO missions such as forward keep-up processing, bulk reprocessing, supporting hybrid processing capabilities to leverage on-premise and cloud capabilities, and latency needs. In support of Earth System and Science Applications, the key emphasis was on characteristics that enable cross-ESO mission science goals. Under Open Science Principles, architectures were evaluated based on how well they facilitate openness, support non-project access, enable collaborative platforms, etc. On Data System Efficiencies, characteristics were assessed on how different architectures improve upon data production processing costs for both large and small missions, varying latency needs, and how commensurate the architecture is to the NASA DAAC's data lake (see Glossary) concept.

Workshop #2 provided the SAWG with the state of the practice and state of the art in MDPS architectures across the community. Beginning with the gathered information, the SAWG took on the task of defining architectures for evaluation, using these guiding principles:

1. Start where NASA is today,
2. Build incrementally, and
3. Push beyond the realm of feasibility.

To ensure the full breadth of possibilities were considered, starting where we are today grounded the team in beginning in a known state. Building incrementally, using differentiating factors, provided a systematic approach to developing architectures, and pushing beyond the realm of feasibility allowed the team to explore well beyond the known and existing architectures. Three main architecture types (see Glossary for definition of Architecture Type) were identified, each more ambitious than the previous:

1. Type 1: Independent MDPS

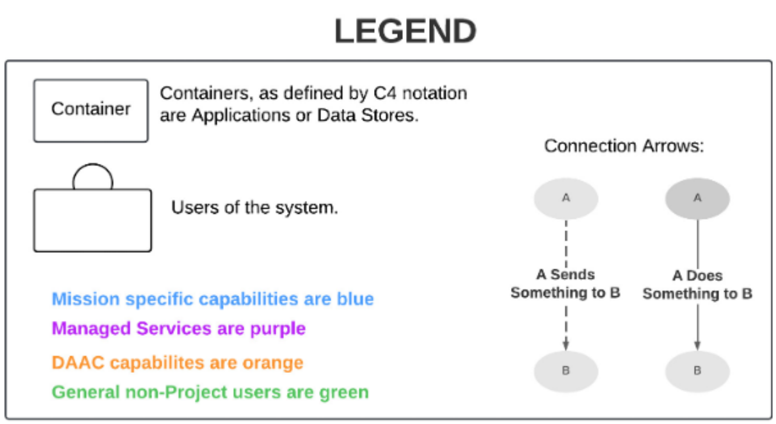
2. Type 2: MDPS with partially managed services shared across the ESO missions
3. Type 3: MDPS as a fully managed system, where each ESO mission is essentially a tenant on the fully managed system

In addition, the SAWG established a common understanding of managed or shared services. Managed services are defined as services that provide a functional capability with well-defined interfaces and are intended to be used by more than a single project. Managed Services are owned and operated by an organization team. The organization team develops, operates, maintains, and evolves the service over time. A service level agreement (SLA) is provided for any managed service. The managed services approach is intended to consolidate and improve cost efficiencies across the service's development, operations, evolution, and workforce.

Beginning with the three identified architecture types, as well as all other stakeholder inputs, the SAWG defined an MDPS by common functionalities using a Block Definition Diagram (BDD) (Stavros et al. 2022, http://hdl.handle.net/2014/54626). The BDD blocks represent all MDPS components relevant to the study, with each component containing a set of properties, constraints, and high-level relationships.

Each candidate MDPS architecture was modeled using C4-like notation[1] to assess behavioral characteristics. The C4 diagrams were at the container level of detail as defined in the BDD, and were presented to the steering committee. The full set of tabulated and diagrammed architecture variants (See Glossary for the definition of Architecture Variant) were evaluated based on inputs from the stakeholders, steering committee, and SMEs (see Technical Assessment section).

The following figures show the high-level architecture types to the C4 container level.



[1] https://c4model.com/

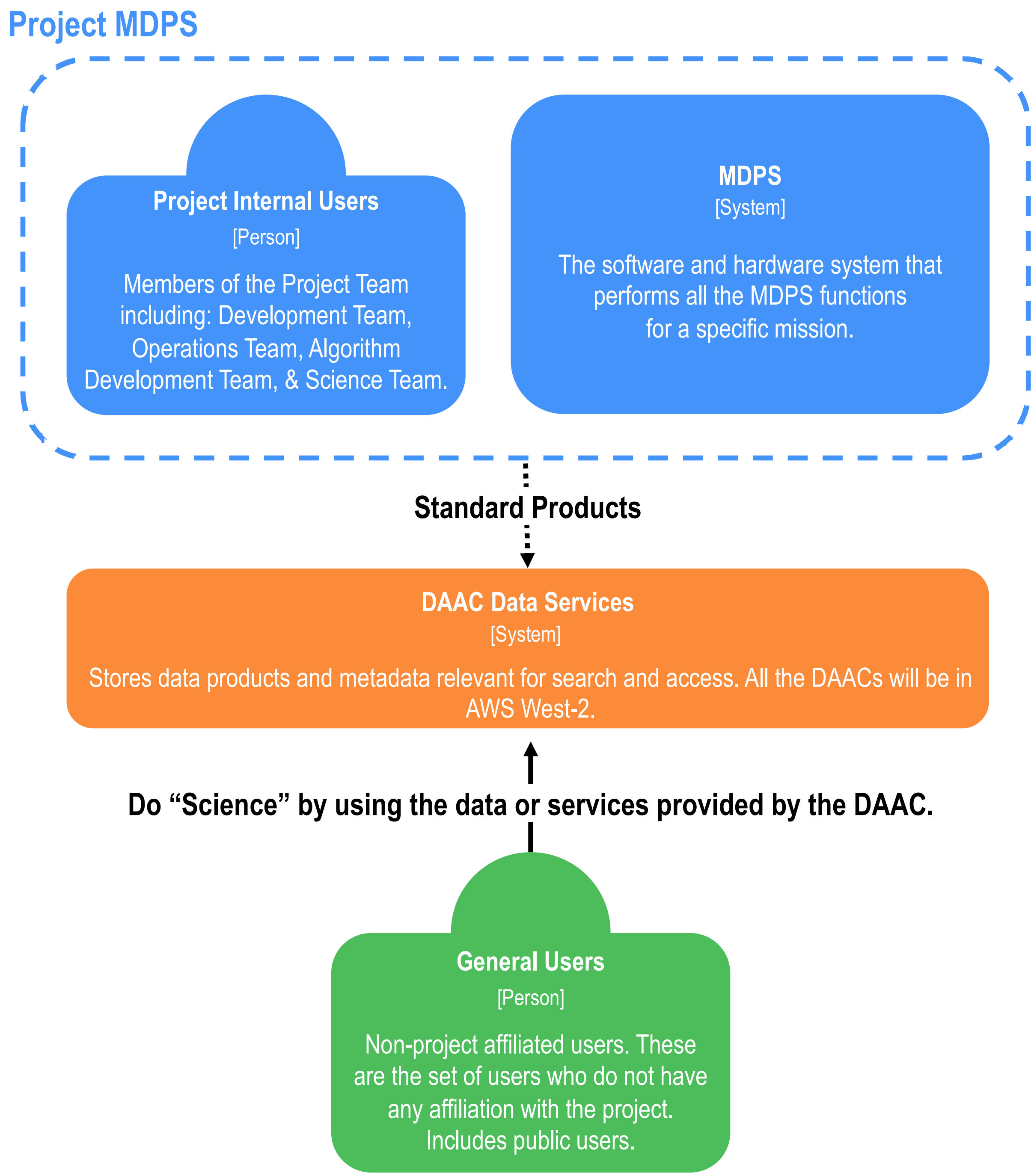


Figure 4-1. Type 1: Independent MDPS.

Figure 4-1 illustrates a Type 1 MDPS that is independently developed, deployed, operated and maintained by a specific mission. All requirements are owned by the mission project team.

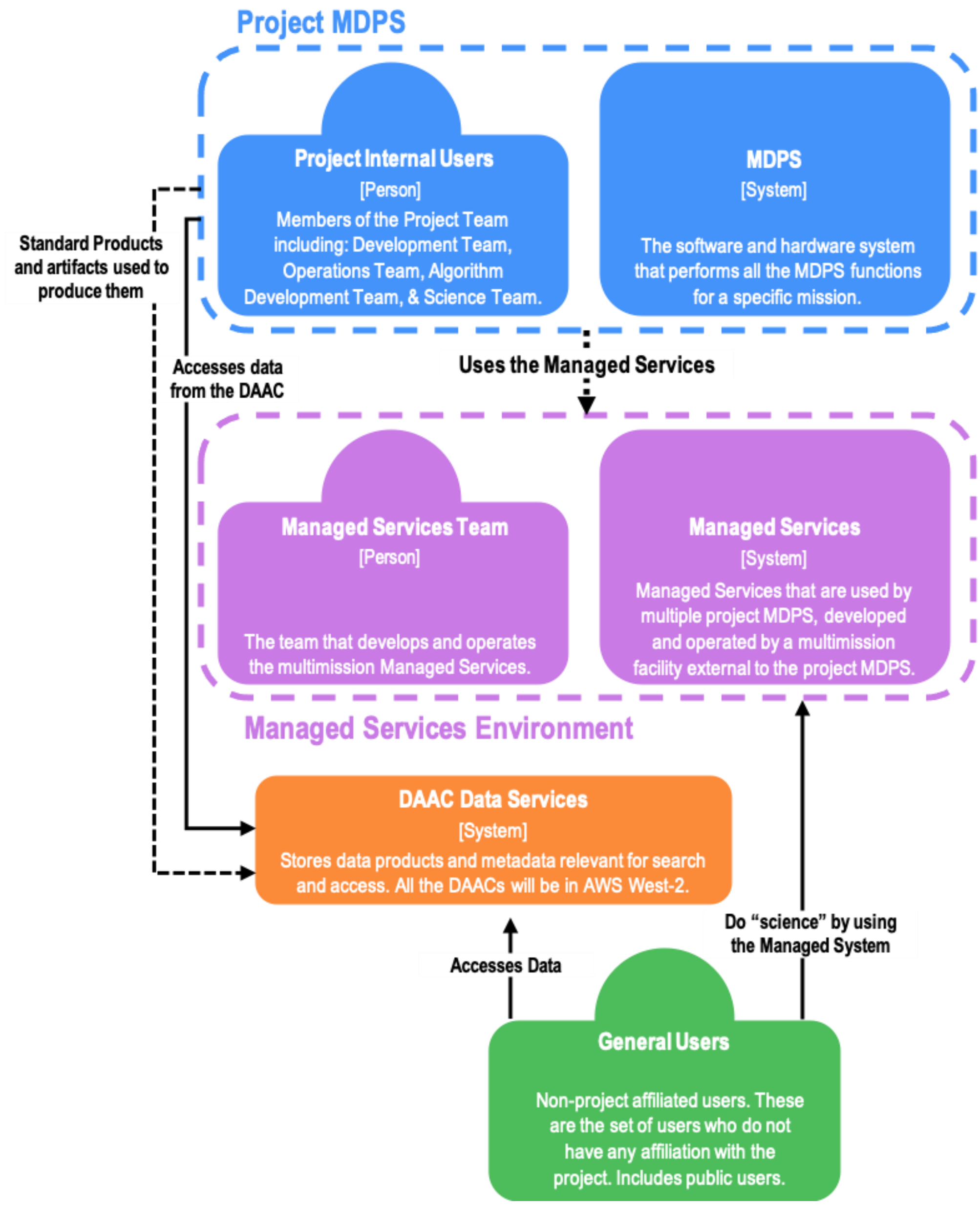


Figure 4-2. Type 2: Managed services.

Figure 4-2 illustrates the introduction of managed or shared services in a Type 2 MDPS. The intent is that several MDPS use a suite of centrally managed services developed, deployed, operated, and maintained by a team external to any specific mission. A mission retains the responsibility to develop, deploy, operate, and maintain some core MDPS functionality, but uses shared services as available.

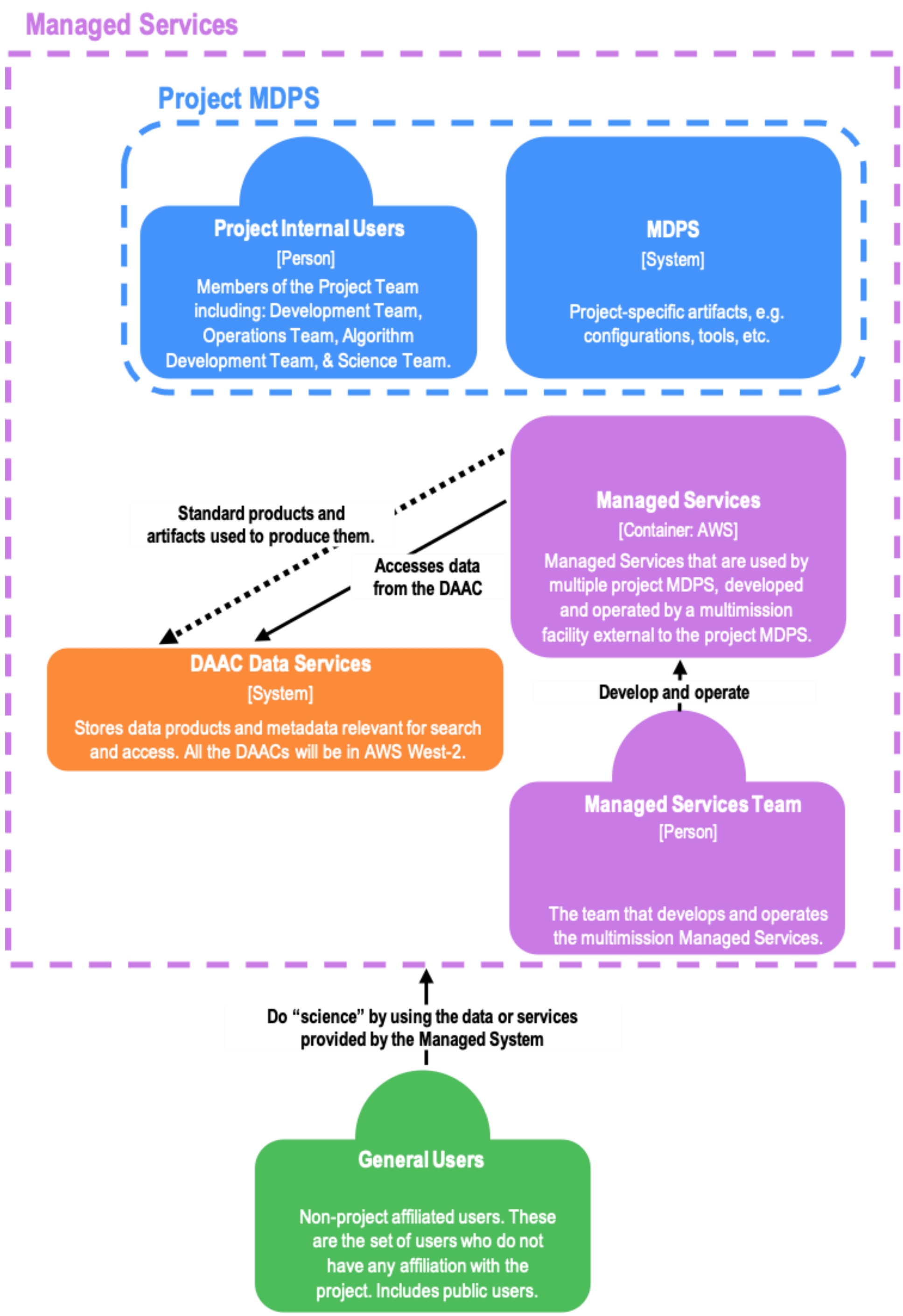


Figure 4-3. Type 3: Fully managed system.

Figure 4-3 illustrates a Type 3 MDPS, which is a system of managed shared services upon which several MDPS may be built. Missions are provided with a complete set of managed services for mission data processing, such as a managed infrastructure service, an artifacts catalog, and a generic data storage, processing, and job execution service. Additionally, algorithm development and analysis environment services may also be provided. Services are developed, deployed, operated, maintained, and evolved by a team external to the missions.

Within these three architecture types, variants were identified that adhere to the same topology and behavior of the architecture type, but are distinguished by differentiating factors. Type 1 includes six variations, Type 2 five variations, and Type 3 two variations. Table 4-1 shows the architecture groups, encodings, and high-level differentiating factors, with more detailed characteristics; general diagrams are given in General Diagrams for the Collection of Architectures and Variants. A high level description of the variants is presented in Table 4-1.

Table 4-1. MDPS Architecture Types and Variants.

Architecture TYPE 1 - Independent MDPS		
Variant Coding	**Independent MDPS**	**Differentiating Factors**
T1V1	Single Instance	Status quo
T1V2	Single Instance with Auto-deployment	Some General User support
T1V3	Co-located with DAAC plus Analysis Environment	Lowered Storage Cost
T1V4	Co-located with DAAC and Algorithm Interoperability	Improved General User support as you go down this list
T1V5	Co-located with DAAC and System Interoperability	
T1V6	Co-located with DAAC and Full System Interoperability	
Architecture TYPE 2 - Managed Services		
Variant Coding	**Managed Services**	**Differentiating Factors**
T2V1	Infrastructure Services	Core infrastructure services only
T2V2	Infrastructure, Data, and Catalog Services	Adds Data and Catalog services
T2V3	Infrastructure, Data, Catalog, and Analysis Services	Adds Analysis services (e.g. interactive visualization for algorithm development, cal/val, product validation)
T2V4	Infrastructure, Data, Catalog, Analysis, and Generic Processing Services	Adds Processing (batch execution only)
T2V5	Infrastructure, Data, Catalog, Analysis, and Full Processing Services	Adds full processing (batch execution and workflow orchestration) As you go down the list: • More managed services • Increased support for the general user • Increased data system efficiencies • Increased System Science
Architecture TYPE 3 - Fully Managed System		
Variant Coding	**Managed System**	**Differentiating Factors**
T3V1	Multi-Project MDPS	MDPS as a managed system
T3V2	Multi-Project MDPS and DAAC	MDPS & DAAC combined

It is worth pointing out that Table 4-1 indicates a progression among the variants within each type as the level of sharedness or interoperability increases. Type 1 variants show an increase in the degree of interoperability, Type 2 shows an increase in the degree of managed services shared by multiple missions or projects, and Type 3 shows an increase in sharedness with DAAC capabilities. All types show an increasing level of complexity and potential cost and risk corresponding to the progression of variants, but also present increases in opportunities such as efficiency, openness, and ability to engage the open source science community.

All candidate architectures under consideration, i.e., the set of architecture types with several variants each, were discussed as a concept and tabulated into an architecture variant table (Table 4-1 and Architecture Variant Tables). The SAWG then drilled down on each tabulated candidate architecture, considering strengths, weaknesses, architecture-specific attributes, and any relevant implications for each stakeholder group (multimission ESO, non-NASA stakeholders, and NASA Headquarters). Items included in the discussion were the ability to deploy the MDPS quickly and consistently, various levels of process orchestration, low-level processing capabilities, computing infrastructure, and shared or common services. Shared services will be shared across all ESO MDPS to some extent, such as authentication and authorization, artifacts catalogs, and other infrastructure services.

Other novel concepts considered were shared "short-term" and "long-term" storage between the MDPS and DAACs, co-location strategies for data and data lakes, processing and analysis environments and services for activities such as Calibration/Validation, algorithm development, and open source science community participation. Additionally, cross-mission system component interoperability with conventions and standards, e.g., Open Geospatial Consortium (OGC) (https://www.ogc.org/standards), were considered.

5 TECHNICAL ASSESSMENT

5.1 GENERAL METHODOLOGY

After identifying the architecture types and variants, the SAWG used the NASA Systems Engineering Handbook (NASA/SP-2007-6105 Rev 1) to conduct a trade study evaluating desirability (D) against technical complexity (TC), which included Feasibility Factors (FF) and Maturity Factors (MF). Desirability criteria came directly from common themes synthesized from Workshop 1 (Appendix Desirability Evaluation Criteria) and aligned with the four objectives of the study: 1) the Data System shall support mission needs, be portable, have well defined interfaces, be relatively mature before use, and be able to be developed within existing budgets; 2) the Data System shall support a data lake, be flexible and efficient, accommodate varying compute needs, and encourage standard data formats; 3) the Data System shall enable data/algorithm/tools sharing to facilitate the advancement of cross-ESO science goals; and 4) the Data System shall provide a community-based, publicly accessible analysis platform that is cybersecurity compliant.

Feasibility Factors were defined to represent management complexity (Appendix Feasibility Factors)—e.g., cost, team management, requirements burden, cybersecurity conformance, staffing, etc. Maturity Factors (Appendix Maturity Factors) represented development, operations and maintenance complexity. Within the system engineering framework, the trade space places both Desirability (D) (x-axis) and Technical Complexity (TC) (y-axis) in a two-dimensional space and assigns scores to each criteria weighted (w) by the value of importance of meeting/accommodating that criteria:

$$\text{DS} = \sum_{i=1}^{28} w_i * D_i \qquad \text{(Equation 1)}$$

$$\text{TC} = \sum_{j=1}^{7} w_j * MF_j + \sum_{k=1}^{4} w_k * FF_k \qquad \text{(Equation 2)}$$

where DS is the Desirability Score with i denoting Design Criteria 1 to 28 (9 for mission development, 5 for data system efficiencies, 6 for Earth Science and Applications, and 8 for open source science), j denoting MF Criteria 1 to 7, and k denoting FF Criteria 1 to 4.

Each architecture type and variant were assessed by the SAWG for their ability to meet desirability evaluation criteria on a score from 0 to 3 for each criterion. A score of 0 represents a heavy level of effort (LOE) to accommodate that criterion with that architecture variant, 1 represents a moderate LOE, 2 represents a low LOE, and 3 represents no adaptations necessary. LOE was defined by complexity to achieve the criterion, which includes resources required(cost) and schedule for development—Technology Readiness Level (TRL), etc.—i.e., "If this architecture exists, what is the LOE required to meet this criterion?" These scores increase in value to optimize a cumulative DS that is high. Figure 5-1 depicts the median score for each architecture type and variant meeting the different objectives of the study.

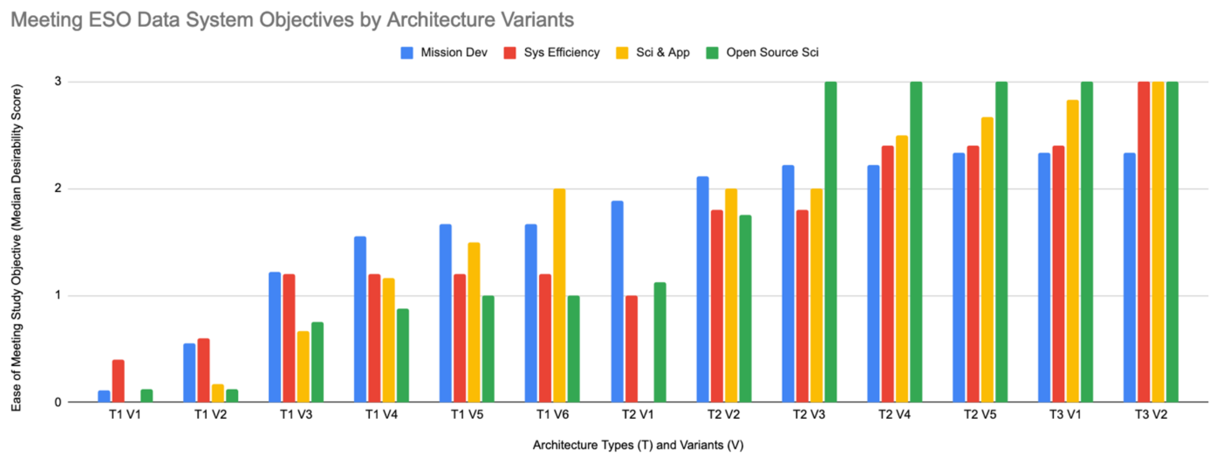


Figure 5-1. The median Desirability Score per study objective for each Architecture Type (T) and Variant (V). A score of 0 = heavy level of Effort (LOE), 1 = moderate LOE, 2 = low LOE, and 3 = no adaptation needed. LOE denotes the complexity to achieve that criterion based on the architecture.

Weights (w) for Desirability Scores were assigned by polling the stakeholders of the study: the sponsors, the SC, and the ESO MDPS representatives. Each criterion was a parameter in the survey. Stakeholders individually assigned a score of 0 as "not important/necessary," 1 as "nice to have (bonus)," 2 as "good to have (needed)", and 3 as "must have (essential)". These weightings increased in value to weigh more heavily via increased value of the cumulative DS score. After polling all stakeholders independently, meetings were held among stakeholders to discuss divergent perspectives (Figure 5-2) and allow stakeholders to recast their priorities after discussion. The mean weight was then used to calculate the cumulative DS (Equation 1).

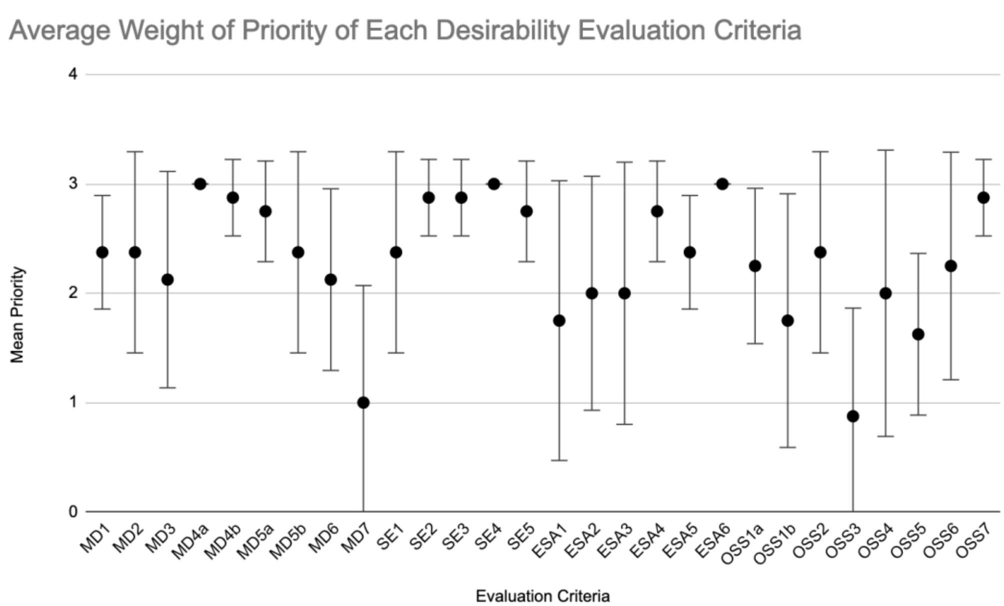


Figure 5-2. The mean and standard deviation of the priority weighting assigned across stakeholders.

The desirement component of scoring primarily focused on evaluating positive attributes of the architecture candidates. The SAWG also factored in negative attributes in the scoring of feasibility and maturity risk factors. Within the feasibility component, the SAWG attributed larger weighting factors to the cost of implementation and technical complexity of each architecture candidate. The SAWG agreed that these would weigh more heavily on feasibility than cybersecurity compliance. For the maturity component, the SAWG also prioritized TRL and readiness of a development process for managed services at a slightly higher weight than concepts maturity for operations and maintenance.

The SAWG assigned Technical Complexity (TC, Equation 2) on a score from 0 to 5 for each criterion of risks, where a score of 0 indicates the greatest maturity/feasibility and a score of 5 represents least maturity/feasibility. Different from DS, a 5-level scale was used for MF and FF because a score of 0 to 3 did not provide enough granularity to differentiate between different architectures regarding their risk postures. Feasibility and maturity were rated from most to least, with most being 0 and least being a 5, so that architectures that were the hardest to achieve would get the highest score which would be used to mitigate desirability. These scores decrease in value to optimize a cumulative TC that is low, thereby a better score would be a lower TC, i.e. with reduced risks.

The scores (D, MF, and FF) were assigned based on the detailed characteristics of each architecture type and variants, including both their strength and weakness for supporting the multimission aspect of ESO, external stake-holder participation, and management / programmatic considerations. Then following Equations 1 and 2, DS and TC were calculated for each system, which were used by the SAWG and SC to determine the recommended MDPS architectures.

5.2 HIGHLIGHTED ARCHITECTURE VARIANTS

The SAWG engaged in in-depth discussions of each architecture as scores were assigned and tabulated. Architecture highlights were conducted, including dominant and conflicting needs, implications, and detailed characteristics. Detailed scoring tables were developed (see Technical Scoring Tables in the Appendix). A brief narrative of a subset of architecture variants is provided below to give the reader insight into the debated variants. The reader can refer to the Architecture Variant Tables in the Appendix for a brief description of each architecture variant. These specific architectures are highlighted here because they represent the most desirable variants, within each type, that were within an acceptable TC range. Correspondingly, these were the architectures that generated the most discussion and debate.

T1V5 (See Appendix Diagrams of Select Architecture Variants, Figure A-1.: T1V5) is an independent MDPS assumed to be co-located with a DAAC, and, where relevant, components are interoperable via standards (e.g., OGC). Algorithms, processing, and data access would be interoperable across the collocated MDPS. Implications for multimission ESO with this approach are an increased burden on the missions to meet any interoperability standards. However, there may be opportunities for improved Earth Systems Science support with this approach. Implications for external stakeholder participation are that any product creation and analysis must still be done on their infrastructure.

A few strengths of this approach are (1) the potential to facilitate the integrated ESO mission system science, and (2) enabling open source science (OSS) via interoperability with MDPS and external systems, allowing algorithms to be developed outside and be used in MDPS and vice-versa. A weakness of this approach is the potential need to enforce standardized interface constraints on the MDPS.

T2V2 (See Appendix Diagrams of Select Architecture Variants, Figure A-2.: T2V2), MDPS are assumed to use a set of managed infrastructure, data, and cataloging services, and to maintain some level of interoperability on these components. Infrastructure-level services include an artifacts catalog, authentication and authorization, system metrics, logging, costing, and cybersecurity apparatus.

Implications of this approach for multimission ESO are less burdensome for user and infrastructure management and provide shared access to data across the missions, with some increased dependency and gains from sharing. Implications for external stakeholder participation are that this approach would allow an on-demand creation of an instance of some MDPS, and external stakeholders may have a "view" into all data products from all the missions using the shared infrastructure services. Programmatic implications are that an external office would be required to provide managed services. The office must manage and account for any financing across the missions and potentially external stakeholders.

A strength of this variant is the consolidation of common shared services managed by a core team, allowing the reuse of mission-specific components. However, a mission's MDPS still allows for significant customization for specific mission needs. This may reduce development costs and time for additional missions and reduce overhead, where mission teams must be trained in infrastructure-level technology and cloud services. Additionally, instead of each MDPS handling data, a shared data service provides access to short-term data that are not intended to be delivered to the archives. A few weaknesses are (1) the possibility of effort duplication in separate MDPS development and operations, (2) the potential for difficulties in prioritization (de-conflicting and merging of changes) between the different needs of all MDPS, and (3) the reliance on an external multimission organization for the managed services.

T2V3 (See Appendix Diagrams of Select Architecture Variants, Figure A-3.: T2V3) includes all T2V2 characteristics, but a managed analysis environment, or service, is added. This architecture provides the capability for a team to instantiate one or more managed analysis environments with minimal effort, based on needs that may vary through the mission lifecycle. The analysis environment supports multiple development languages and tooling, and may interface with the artifacts catalog and is generally used. It is still assumed that MDPS will use the other managed infrastructure-level services, such as an artifacts catalog, authentication and authorization, system metrics, logging, costing, and some cybersecurity apparatus, as with T2V2, and many of the T2V2 strengths and weaknesses still hold.

T2V4 (See Appendix Diagrams of Select Architecture Variants, Figure A-4.: T2V4), many T2V2 and T2V3 characteristics are carried forward, but a generic processing service is added. The generic batch-oriented processing service provides the capability to schedule work on a set of computing resources with minimal effort, but this service falls short of any mission-specific workflow orchestration. The processing service essentially abstracts away scaling up or out any computational work, i.e., an MDPS-specific orchestrator submits jobs to the batch service, but the team is not responsible for process scaling, only for specific MDPS workflows. It is up to the tenant users to use this managed processing capability based on their needs, which may be variable. Teams can consider latency-optimized workloads (forward processing, urgent response processing, near real-time [NRT] processing) or cost-optimized workloads (bulk reprocessing campaigns), with the former being more expensive and the latter less expensive. Again, it is still assumed that MDPS will use other managed infrastructure-level services, such as an artifacts catalog, authentication and authorization services, system metrics, logging, costing, and some cybersecurity apparatus, as with T2V2 and T2V3.

The T2V2 and T2v3 implications for a multimission ESO are still true with the T2V4 architecture. However, with the addition of the cross-mission analysis environment and generic processing services, teams must integrate with them. Implications for external stakeholder participation with this approach will continue as is with T2V3, but there may be an opportunity for an external stakeholder to tighten the development to production cycle for higher-level community-driven products.

An additional strength above T2V3 is that with the T2V4 approach, each mission may move closer towards a shared development and operation process, which can positively impact cross-mission efficiencies. Each mission can still choose a specific workflow orchestration mechanism that best meets its needs, allowing for significant customizations, with the result that teams are no longer burdened with managing at-scale processing. A weakness above T2V3 with this approach is that the large-scale processing for the shared processing service may be non-trivial for a shared services team to stand up and maintain.

T3V1 (See Appendix Diagrams of Select Architecture Variants, Figure A5.: T3V1) architecture, all MDPS are built on a fully managed and shared system. Any infrastructure, processing, data, artifact catalog, or analysis environments are developed, deployed, operated, and maintained by an external entity, and funded externally to the mission. Mission-specific teams are only responsible for their own algorithms and workflow templates, which would be registered in the fully managed artifact catalog. All data services support multimission data products, which presents an opportunity for cross-mission, data-rich analysis environments. Any deployed analysis environment within the managed system has full access (read-only) to data produced by the mission. All public non-project users may deploy a copy of any MDPS on the shared system using templates provided by an official mission team, depending on available funds and approvals.

The T3V1 implications for multimission ESO are that the level of effort to establish a basic MDPS may be significantly reduced compared to the other architectural approaches.

A few other notable strengths of the T3V1 architecture are that multiple-discipline products may be generated together on the same multimission system, and there are efficiency gains with the economy of scale due to the "share everything" nature of the system. The approach lowers the infrastructure sustainment responsibilities of MDPS so teams can focus on producing the science products instead of managing low-level system components and infrastructure.

There is the potential for improved support for public users, as they can be on-boarded with an account and funding, instantiate, and use the templates in the managed artifact catalog, and run prototype high-level community-driven products with a limited effort.

Several weaknesses worth noting with the T3V1 approach are the "one size fits all" managed services model may be challenging to design, develop and maintain to address all mission needs. In addition, the approach may limit innovation, as each MDPS may not be able to replace any components to meet mission needs. Furthermore, large-scale processing for the shared processing service for the multi-tenancy system may be non-trivial.

5.3 SAWG RECOMMENDATION: T2V3 ARCHITECTURE, POTENTIAL STRETCH GOAL OF T2V4

The SAWG identified and evaluated the architectures based on how well they can demonstrably meet the ESO data system objectives while balancing out technical complexity. As evidenced in Figure 5-3 below, the T2V3 architecture achieves relatively high desirability with a commensurate level of technical complexity. T2V3 requires just slightly more technical complexity than T2V2, yet provides a disproportionately higher level of desirability.

Because each subsequent variant within a type builds on the previous variant, the SAWG explored the viability of T2V4 as a stretch-goal architecture based on what could realistically be achieved. T2V4 builds on T2V3 by adding “generic processing”. Generic processing was defined to include job processing of algorithms to create data products (i.e., bulk [re-]processing), but excludes workflow processing (i.e. “full processing”), which can get complicated and is nuanced for each mission. We determined that T2V4 incrementally adds value to the SAWG-recommended T2V3 architecture, and is at the edge of achievability. As such, T2V4 should be considered as a stretch-goal architecture for further analysis in follow-on efforts.

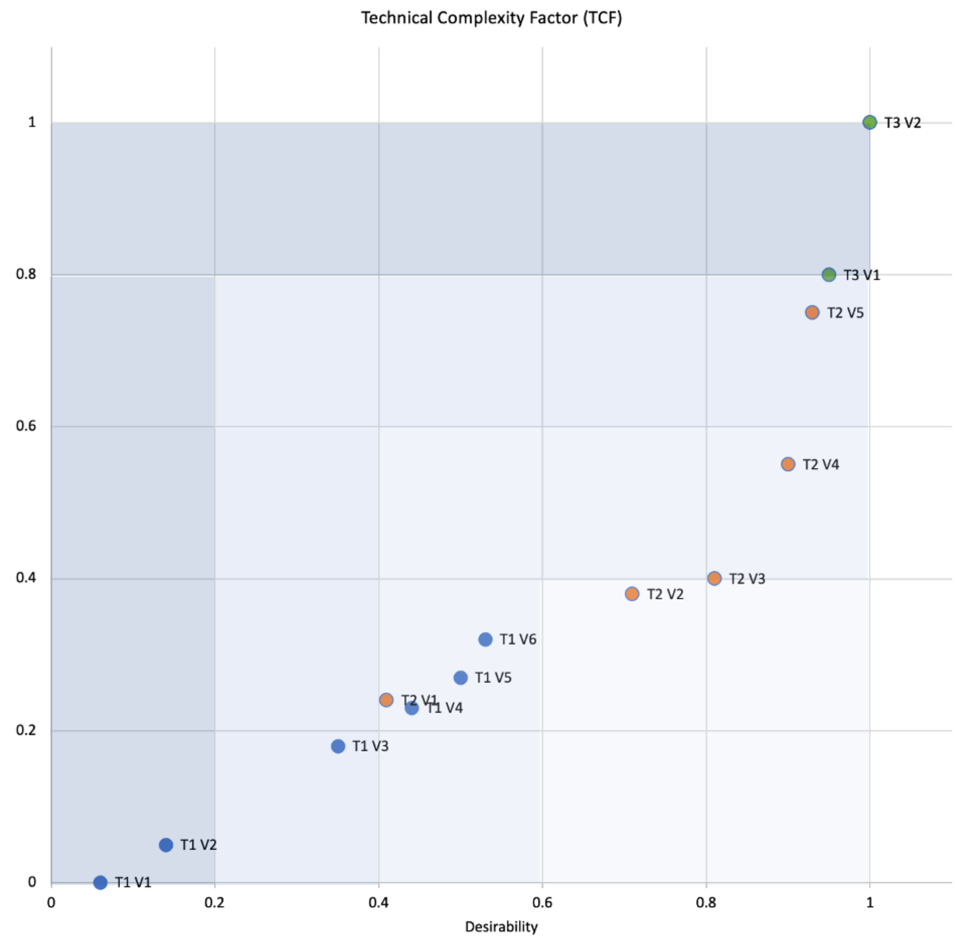


Figure 5-3. Independent SAWG assessment of architectures against the DS and TC values by architecture variant. Note that the bottom right denotes the area of highest desirability and lowest technical complexity.

6 PROGRAMMATIC ASSESSMENT

6.1 DESCRIPTION & RESULTS

The programmatic assessment considers the non-technical aspects of implementing and maintaining each of the candidate architectures. This assessment is intended to balance the technical assessment by considering the complexities of levying additional requirements or interfaces on projects, creating and allocating responsibilities to a new Multimission Organization (MMO) that is needed to manage the architecture and provide specifications and services as necessary, and the potential impact on the ESO flight projects. The programmatic assessment was intentionally divorced from the technical assessment provided by the SAWG, to not introduce a bias into either recommendation based on perceived or real impacts.

Over the course of the study, the steering committee had discussions and received feedback from the ESO missions and program management. In order to evaluate programmatic complexity, the committee distilled the gathered information into four factors to assess the potential barriers and impacts related to: cultural, cost, schedule, and resource complexities. These factors and their definitions are described below.

- Cultural Complexity (CC) pertains to the challenges of coordinating and integrating the requirements and processes of multiple missions within a centralized service framework. Architectures are considered culturally complex when the multi-mission organization has a broad supervisory span that can result in bureaucratic entanglements, impeding the efficiency and autonomy of individual missions.
- Cost Complexity (CoC) pertains to the challenges of estimating and securing an adequate budget for implementing an architecture, from both a service provider and service user standpoint. The complexity of cost is contingent upon various aspects, such as the number of services included in the architecture, the level of maturity of these services, and the clarity around funding sources and mechanisms.
- Schedule Complexity (SC) pertains to the expected timeline for implementing an architecture and its alignment with the ESO flight projects' implementation schedule. At the time of conducting this analysis, the ESO flight projects were anticipated to conclude Phase-B and transition to Phase C/D in the 2024 timeframe.
- Resource Complexity (RC) pertains to the availability of talent and level of skill required to implement a particular architecture. This assessment was conducted during a time when NASA was experiencing significant employee attrition due to the post-pandemic changes in the employment landscape.

In order to assess the programmatic complexity of each candidate architecture, we created a metric termed the Programmatic Complexity Factor (PCF) which is the normalized value (0-1) of the overall complexity to implement a candidate architecture. A PCF = 0 represents the easiest architecture to implement from a programmatic perspective, while a PCF = 1 represents the most difficult architecture to implement from a programmatic perspective.

Each of the four members, m, of the SC scored the four complexity attributes (cultural, cost, schedule, resource) of every architecture between 1 and 10, to produce an average Programmatic Complexity score for each architecture, a, ($\underline{}$PC(a)):

$$\underline{PC}(a) = \sum_{m=1}^{4} \quad (CC(a,m) + CoC(a,m) + SC(a,m) + RC(a,m))/16 \qquad \text{(Equation 3)}$$

The average Programmatic Complexity was then normalized to create a PCF for each architecture:

$$PCF(a) = \frac{(\underline{PC}(a) - \underline{PC}(a)_{min})}{(\underline{PC}(a)_{max} - \underline{PC}(a)_{min})} \qquad \text{(Equation 4)}$$

The scores from each member and resulting calculations are documented in the Appendix. It should be noted that each member of the SC created their scoring independent of each other. They formed their score based on broad and diverse experience working on similar projects and activities, input received during Workshops #1 and #2, feedback from discussions with the ESO flight projects, as well as diverse stakeholders.

The results are summarized in Figure 6-1 below, which shows the PCF of each architecture. The T1V1 architecture is considered the easiest architecture to implement from a programmatic perspective, with T3V2 as the most difficult. Within each architecture type, the programmatic complexity increases as a function of the variant; this is because each variant contains more functionality than the previous variant and therefore is more difficult to implement. T2V1, T2V4 and T2V5 stand out as having high variance. This indicates that the members of the SC have differing opinions on the complexity of a service-based architecture.

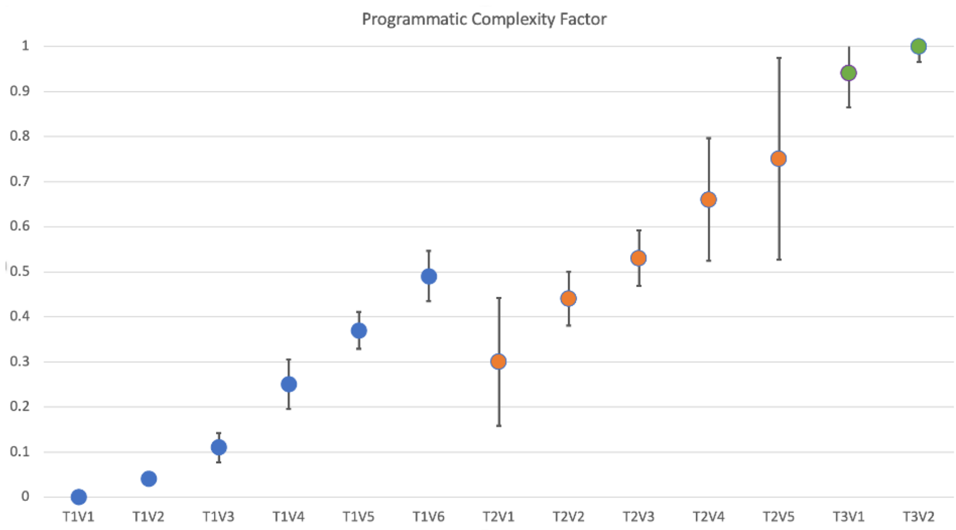


Figure 6-1. The Steering Committee scored the four complexity attributes (cultural, cost, schedule, resource) of each architecture to produce an average Programmatic Complexity score, which is normalized along with the variance, to create the Programmatic Complexity Factor.

In Figure 6-2 we show a scatter plot of the PCF relative to the Desirability of each architecture. This chart suggests that if we were to select an architecture purely from a programmatic perspective, we would select architecture T2V2 because it optimizes for a high Desirability and low PCF.

Figure 6-2. Programmatic complexity of each architecture relative to the architecture desirability. This chart suggests that Architecture T2V2 is the optimal choice from a programmatic perspective because it is mathematically closer to the point of minimum complexity and maximum desirability.

6.2 PROGRAMMATIC DRIVERS AND TAKEAWAYS

Projects are driven to minimize risk, adhere to budget and schedule constraints, and achieve mission objectives. One of the most common approaches to achieving programmatic adherence is to reuse technology, processes, and procedures that are demonstrably proven to reduce risk and maximize success. Not surprisingly, the potential to introduce a new external dependency and required interface to support mission data processing needs introduces some measure of uncertainty to the missions underway. Furthermore, the uncertainty of meeting scheduled review milestones and implications to development and operations costs cannot be ignored when evaluating new approaches to providing MDPS capabilities. Another factor is alignment with the strategic goals of this study. The SC noted a number of programmatic drivers and findings that are summarized below to characterize the feedback received.

- All ESO projects recognize the value of improving access to science data, supporting reproducibility and transparency, and engaging a more diverse and broad community. They are committed to supporting open sourced science, but are weary about how it should be implemented.
- The ESO projects have existing cost plans based on strong inheritance from existing architectures and implementations. Any new approaches imposed on them will therefore require a reevaluation of their existing cost estimates.
- The ESO projects are entering Phase-A and will require the maturation and selection of the best architecture to support the development of their mission requirements in Phase B. It is uncertain if a new architectural approach can be flushed out to the level of maturity within the timeframe required.
- The ESO missions are concerned about controlling costs and overall budget adherence. The introduction of interfaces and a MMO that will be delivering capabilities are perceived to potentially increase costs. Missions are concerned with expectations that they can reduce budgets if another provider is delivering capabilities.
- Developing a dependency on a provider outside of the core MDPS team will require not only tight coordination, but a clear understanding of managed service requirements and a documented SLA that establishes clear expectations for both the MMO and the project community.
- Enabling access to mission data for non-NASA personnel will be important to enabling broad participation from a distributed science community.
- There were a number of indications of applicability of the Type 2 architecture enabling missions beyond the current ESO scope. The upcoming Earth System Explorer (ESE) announcement of opportunity as well as ongoing Earth Venture Instrument (EVI) and Earth Venture Missions (EVM) opportunities would all benefit from the value of these shared services. These will be particularly important to support interoperability and analysis as these new science products are generated.
- Dependencies on the program of record data (e.g. NOAA provided on AOS, and DLR-provided on MC) and complementary products (ESA products on SBG) requires non-NASA data integration to support mission science objectives.
- There is a strong demand for cross-data interoperability at Level 3 and beyond. Meeting decadal survey objectives will require cross-mission and cross-instrument access, integration, and analysis to enable the science community.

6.3 COMBINED TECHNICAL AND PROGRAMMATIC ASSESSMENT

In Figure 6-3 we show the combined Technical and Programmatic Complexity Factor, which was computed by averaging the two scores, plotted relative to the Desirability Score. Based on this figure, we conclude that the Type 2 architectures generally provide the optimal balance of meeting a high desirability, while minimizing the technical and programmatic complexity. Specifically, the T2V3 architecture yields the most optimum solution.

Figure 6-3. Combined Technical and Programmatic complexity of each architecture relative to the architecture desirability. With this chart we conclude that Architecture T2V3 is the optimal choice because it is mathematically closer to the point of minimum complexity and maximum desirability.

7 CONCLUSIONS AND RECOMMENDATIONS

The SC and SAWG established and followed a thorough process to identify candidate architectures that could meet the study objectives to satisfy the ESO mission processing requirements, promote open science, enable efficiencies, and advance Earth system science and applications. This process resulted in the identification of three architecture types, each of which have architecture variants:

1. Independent MDPS (Type T1) is an architecture type where each mission implements a mission-specific MDPS, but has varying degree of interoperability requirements levied on the approach.
2. Managed-Services (Type T2) is an architecture type where each mission implements a MDPS using a common architecture, but using a varying number of services provided by a multimission organization.
3. Fully Managed Services (Type T3) is an architecture type where each mission implements its MDPS on a multi-tenant platform provided by a multimission organization.

Together, the SC and SAWG assigned a Desirability Score to each architecture, which represents the degree to which the architecture responds to the study objectives. Independently, the SC assessed the programmatic complexity of each architecture and the SAWG assessed the technical complexity of each architecture. By combining these assessments and comparing with the desirability we conclude that the T2V3 (managed infrastructure, data, catalog, and analysis services) is the optimal architecture. However, we also note that a T2V4 architecture (managed infrastructure, data, catalog, analysis, and processing services) provides a significant increase in desirability over T2V3 with a modest increase in complexity, and therefore should not be fully discounted.

We have three key recommendations:

1. NASA should consider a service-based architecture for the ESO mission data processing. This approach will require a multi-mission organization to develop the architecture and standards across the ESO missions, and develop and deliver the services that support the architecture. These services include the infrastructure service, data management and catalog service, and analysis service, and potentially generic processing service.
2. NASA should sponsor a follow-up study to thoroughly examine the recommended architecture. This study should involve a review of the functional and performance requirements at the system level, as well as a review of the system-level architecture and design. The study should include strong engagement from the ESO missions and provide an assessment of any mission impacts. This study should commence as quickly as possible to reduce delays in support of the ESO missions that are in development now.

3. NASA should sponsor another study on how the architecture can facilitate integration and processing of data from across the entire observatory. Various science and applications based use cases are emerging that require cross-mission integration, which the recommended architecture has the potential to support.

The entire study committee concluded that addressing the goals of meeting ESO mission objectives, enabling efficiencies, supporting Earth system science and applications, and promoting open science principles may be accomplished through the advancement of a managed services architecture. Historical barriers to this approach are largely mitigated through technology advancements, broad support and adoption of open source software, and a cultural recognition of the value of managed services.

8 REFERENCES

Mell, P. and Grance, T., (2011) “The NIST Definition of Cloud Computing – Recommendation of the National Institute of Standards and Technology”, Special Publication 800-145, https://nvlpubs.nist.gov/nistpubs/legacy/sp/nistspecialpublication800-145.pdf

Ramachandran, R., Bugbee, K., Murphy, K., (2021) “From Open Data to Open Science” *Earth and Data Science*, 8, Issue 5 https://doi.org/10.1029/2020EA001562

Stavros, E. et al., (2021) “ESO Mission Data Processing Study - Summary of NASA Program Offices and ESO Mission Requirements, Constraints, Recommendations, and Opportunities”, http://hdl.handle.net/2014/53042

Stavros, E. et al., (2022) “Workshop #2 Report; ESO Mission Data Processing Study: Summary of State-of-the Practice and State-of-the-Art Mission Data Processing System Architectures”, http://hdl.handle.net/2014/54626

Thriving on a Changing Planet – A Decadal Strategy for Earth Observation from Space (2018) https://doi.org/10.17226/24938

Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

A. APPENDIX

A.1 Steering Committee

Dr. Andrew Bingham has more than 20 years of experience managing and leading teams responsible for delivering science processing, archive, distribution, and analysis systems for Earth and planetary missions. He is the manager for the Science Data and Instrument Operations System Section at NASA's Jet Propulsion Laboratory (JPL).

Luke Dahl has more than 20 years of experience as an architect and systems engineer supporting diverse activities ranging from enterprise infrastructure and applications, multimission instrument operations, and Earth and planetary science data systems at NASA's Jet Propulsion Laboratory (JPL).

Dr. Chelle Gentemann is a senior scientist at the Farallon Institute leading research on open science, cloud computing, remote sensing, and physical oceanography. For more than 20 years, she has worked on every aspect of passive microwave satellite missions, both domestically and internationally, from launch through decommission, including calibration, algorithm development, validation, data distribution, and science applications. She is also serving as Project Scientist for NASA’s Transform to Open Science (TOPS).

Dr. Sara Lubkin has a broad background in Earth science. She has worked with NASA since 2015 and is a Distributed Active Archive Center (DAAC) Operations Engineer with NASA’s Earth Science Data and Information System (ESDIS) Project at NASA's Goddard Space Flight Center.

Andrew Mitchell has multiple years of experience executing the technical and financial management of the science systems of NASA's Earth Observing System Data and Information System (EOSDIS). He is responsible for managing the processing, archiving, and distribution of Earth science data while ensuring scientists and the public have access to these data to enable the study of Earth. He is the project manager for ESDIS at Goddard.

Karen Yuen has more than 20 years of technical and management experience in science applications, project formulation, system engineering, and communications for multiple NASA missions and programs. She is the Science Data Applications Lead for the Orbiting Carbon Observatory-2 (OCO-2) and OCO-3 missions.

A.2 System Architecture Working Group (WG) Participants

Study Architecture Working Group Members 10/5/21			
#	**Focus Area**	**Name**	**Org**
1	Co-Chair	Elias Sayfi	JPL
2	Co-Chair	Natasha Stavros	UC Boulder
3	SBG & NISAR Science Processing Rep.	Hook Hua	JPL

Study Architecture Working Group Members 10/5/21			
#	**Focus Area**	**Name**	**Org**
4	AOS Science Processing Rep.	Curt Tilmes	GSFC
5	MC Science Processing Rep.	Bernie Bienstock	JPL
6	Instrument Algorithm Developer	Qing Yue	JPL
7	Applied Science Algorithm Developer	Wenying Su	Langley
8	Data Processing Workflow Community-focused Developer	Andy Michaelis	Ames
9	Geophysical Numerical Modeler	Lesley Ott	GSFC
10	Science processing architect from a comparable big-data science-based organization or project	Evelyn Ho	GSFC
11	Science processing architect from a comparable big-data science-based organization or project	Chris Engebretson	USGS
12	Science processing architect from a comparable big-data science-based organization or project	Adrian Parker	NOAA
13	Science processing architect from a comparable big-data science-based organization or project	Sean Harkin	MSFC

A.3 Desirability Evaluation Criteria

ESO Mission Science Processing Objectives ["Mission Dev"]	
The Data System should provide an accessible platform for the community to develop L3+ products that adhere to NASA metadata and provenance standards	MD1
The Data System should support on-demand product generation as soon as data is available from the ground data system.	MD2
The Data System should be portable to support deployment on-prem, in-cloud, multi-cloud, and hybrid infrastructure.	MD3
The Data system external interfaces should go through Authentication and Authorization.	MD4a
The Data system external interfaces should use standardized access protocols.	MD4b
The Data System should have the ability for forward-stream and bulk [re-]processing.	MD5a
The Data System should be compliant with DAAC archive retrieval.	MD5b
The Data System should be demonstrated (TRL6+) by the earliest ESO mission launch	MD6
The Data System is cost-constrained by ESO mission budget capacity	MD7
Enable Data System Efficiencies	
The Data System should support a data lake.	SE1
The Data System should be flexible to efficiently (cost, bandwidth, processing capability) support large and small data volumes.	SE2

The Data system should accommodate variable compute needs over time, which is crucial to reducing costs.	SE3
The Data System should support services to create standard data formats (ESDIS Standards)	SE4
The Data System should keep up with forward-stream processing demand.	SE5
Support Earth System and Science Applications	
The Data System should enable ESO data sharing before the data archive.	ESA1
The Data System should enable data access from non-ESO missions.	ESA2
The Data System should enable users to share algorithms.	ESA3
The Data System should enable development and sharing of data tools (e.g., software, code libraries, etc.).	ESA4
The Data System should enable on-demand processing.	ESA5
The Data System should meet cross-ESO mission science goals.	ESA6
Promote Open Science Principles	
The Data System should provide an analysis platform.	OSS1a
The Analysis Platform should be accessible by NASA and Non-NASA users while supporting algorithm sharing and on-demand batch processing.	OSS1b
The Analysis Platform should enable users to control user-generated resources: private and public code repositories, containers, binaries, data, etc.	OSS2
The Analysis Platform should enable public access.	OSS3
The Analysis Platform should enable user authentication and authorization for provenance of contributions.	OSS4
The Analysis Platform should automate standardized open source science guidelines (e.g., metadata standards, provenance, etc.).	OSS5
The Data System should allocate cost accounting by user activity.	OSS6
The Data System should be compliant with cybersecurity policies (e.g., authenticate & authorize, user management, etc.). *NOTE: Intellectual Property compliance is covered by other evaluation criteria.*	OSS7

A.4 Maturity Factors

Each of the architecture variants and types were evaluated against the four maturity factors listed in the table below. SAWG scoring of each in the “Scoring” worksheet was based on a 5-point scale, with 0 score indicating the greatest maturity and a 5 score indicating the least maturity, multiplied by the “Weighting” factor.

Maturity Factors (MF)	
High TRL at ESO mission implement need	D8
Development process	D9
Operations model	D10
Maintenance model	D11

A.5 Feasibility Factors

Each of the architecture variants and types were evaluated against the seven feasibility factors listed in the table below. SAWG scoring of each in the "Scoring" worksheet was based on a 5-point scale, with a 0 score indicating the greatest feasibility and a 5 score indicating the least feasibility, multiplied by the "Weighting" factor.

Feasibility Factors (FF)	
Cost of implementation to get to this Architecture from where we are now	D1
Technical Complexity	D2
Cost Tracking Complexity	D3
Requirements Complexity	D4
Team Complexity	D5
Schedule Complexity	D6
Cybersecurity Conformance	D7

A.6 Architecture Variant Tables

Type	Type 1 - Independent MDPS					
Variant #	**Variant 1**	**Variant 2**	**Variant 3**	**Variant 4**	**Variant 5**	**Variant 6**
Short Name	Single Instance	Single Instance with Auto-deployment	Collocated with DAAC + Analysis Environment	Co-located with DAAC and Algorithm Interoperability	Co-located with DAAC and System Interoperability	Co-located with DAAC and Full System Interoperability
Common standards and interfaces	No assumed interoperability	Interoperable deployment standards	Common cloud storage mechanisms	Interoperability for algorithm representation for portability	Interoperability of algorithms, processing, data access across collocated MDPSs.	T1V5 standards + cloud storage standards + notebook interfaces
Concept	Each mission independently developed and deployed. Each MDPS sends data to a remote long-term data storage entity. This can be on-premise and in the cloud. This is how most MDPS's are implemented now.	MDPS is containerized/VMized so as to enable automated deployment of MDPS services.Enables reproducible services and deployment by project and non-project users.	Independently developed and deployed MDPSes into the same cloud region.Some components of MDPSes are collocated with DAAC archives.Multiple MDPSes are deployed into the same cloud region as the data storage entity, can share data, using data directly from the long term data store, so all input data aren't needed to be stored in the MDPS.	Just like the "Co-located with DAAC + Analysis Environment" but additionally all MDPS agree to use a standard interface for executing PGEs.	Independent MDPS where relevant system components are interoperable via standards (e.g. OGC). Enables algorithms , processing, data access to be interoperable across MDPSes that are collocated.	Independent MDPS where relevant system components are interoperable via standards (e.g. OGC). Enables algorithms , processing, data access, analysis services, and common services to be interoperable across MDPSes that are collocated.
Implications for multi-missions ESO	Each mission is implemented independently with no dependency or efficiencies gained from other missions.	Simplifies experimentation and deployment by Project users	Lower storage cost and lowers the bar to perform analysis tasks.	Increase work to meet interoperability, improved Earth System Science support	Increase work to meet interoperability standards, improved Earth System Science support	Increase work to meet interoperability standards, improved Earth System Science support
Implications for external	External users have no access or influence on the MDPS	Deploy their own instance of the MDPS.	Perform analysis of a specific mission's products.	Produce and analyze a specific mission's products, but must still	Produce and analyze all the missions' products, but must still	Produce and analyze all the missions' products, share analysis

Type	Type 1 - Independent MDPS					
Variant #	**Variant 1**	**Variant 2**	**Variant 3**	**Variant 4**	**Variant 5**	**Variant 6**
stakeholder participation				have their own infrastructure.	have their own infrastructure.	environment, but must still have their own infrastructure.
Implications to management / programmatic	Status quo.	Cost may go up with added functionality and support	Cost may go up with added functionality and support	Cost may go up with added functionality and support	Cost may go up with added functionality and support	Cost may go up with added functionality and support
Strengths	• Maximum flexibility to build a custom system for the mission needs.	• Enables MDPS to be easily deployable to both on-prem and in cloud by project and non-project users.	• Collocated and minimizes egress• Avoid duplicate storage costs by allowing MDPS to use input data directly from long term storage •Independent MDPS allow customization for mission needs	• Allows reuse of "PGE Processing" component • Independent MDPS allow customization for mission needs • Compliance of algorithm sharing facilitates OSS interoperability and reproducibility.	• Facilitates integrated ESO mission system science. • Enables OSS via interoperability with MDPS and external systems. Allows algorithms developed outside to be used in MDPS, and vice-versa.	• Collocated and minimizes egress. • Avoid duplicate storage costs. • Independent MDPS allow customization for mission needs.• More reuse of software components. • Allows reuse of "PGE Processing" component • Independent MDPS allow customization for mission needs • Compliance of algorithm sharing facilitates OSS interoperability and reproducibility.
Weakness	• High cost to develop and operate each entire system independently. • Additional time/cost to transfer data from MDPS to remote data storage.	• High cost to develop and operate each entire system independently. • Additional time/cost to transfer data from MDPS to remote data storage.	• High cost to independently develop and operate each MDPS	• Does not imply multi-mission • High cost to independently develop and operate each MDPS • Requiring common PGE interface may restrict PGE run scenario	• Enforcement of standardized interfaces constrains MDPS approach to those interfaces	* May be difficult to share internal MDPS data store catalog. * High cost to synchronize internal MDPS data stores.• Requiring common PGE interface may restrict PGE run scenario

Type	Type 2 - Managed Services					Type 3 - Fully Managed System	
Variant #	**Variant 1**	**Variant 2**	**Variant 3**	**Variant 4**	**Variant 5**	**Variant 1**	**Variant 2**
Short Name	Infrastructure Services	Infrastructure, Data, and Catalog Services	Infrastructure, Data, Catalog, and Analysis Services	Infrastructure, Data, Catalog, Analysis, and Processing Services	Infrastructure, Data, Catalog, Analysis, and Full Processing Services	Multi-project MDPS (the "gmail" analogy)	Multi-project MDPS and DAAC
Common standards and interfaces	Interoperability on Infrastructure Services	Interoperability from T2V1 + data and catalog services	Interoperability from T2V2 + Analysis Services	Interoperability from T2V3 + Processing Services	Interoperability from T2V4 + Full Processing Services		
Concept	The MDPSes share on a managed core set of services. The multi-mission component is only for core services. Core Services includes infrastructure-level services. (artifacts catalog, A&A, metrics, logging, costing, cyber - i.e. MCP & NGAP now)	Shared MDPS service providing access to private MDPS store. The data access is more stable throughout the life cycle of MDPS.	The MDPSes share a managed core set of services. The multi-mission component is only for core services. Multi-Project Core - Services includes infrastructure-level services. (metrics, logging, costing, cyber - i.e. MCP&NGAP now) MDPS Services - Generic generic data storage services + Analysis Environment (this is both for Algorithm Development and Analysis)	The MDPSes share on a managed core set of services. The multi-mission component is only for core services. Multi-Project Core - Services includes infrastructure-level services. (metrics, logging, costing, cyber - i.e. MCP&NGAP now) MDPS Services - Generic data storage services + Generic processing services (which implies that Algorithm Interoperability is also present)	The MDPSes share on a managed core set of services. The multi-mission component is only for core services. Multi-Project Core - Services includes infrastructure-level services. (metrics, logging, costing, cyber - i.e. MCP&NGAP now) MDPS Services - Generic data storage services + Full processing services (which implies that Algorithm Interoperability is also present)	The entire MDPS is managed, not just the core.	The entire MDPS is managed, not just the core. But the core services are shared with DAACs.

Type	Type 2 - Managed Services					Type 3 - Fully Managed System	
Variant #	Variant 1	Variant 2	Variant 3	Variant 4	Variant 5	Variant 1	Variant 2
Implications for multi-missions ESO	Easier user and infrastructure management.	Shared access to data across the missions. Increased dependency and gains from sharing.	Shared analysis across the missions. Increased dependency and gains from data sharing and analysis.	Shared analysis and processing across the missions. Increased dependency and gains from sharing.	Shared analysis and processing across the missions. Increased dependency and gains from sharing.	Easy to stand up basic MDPS. Lots of dependency on external services	Easy to stand up basic MDPS and DAAC. Lots of dependency on external services.
Implications for external stakeholder participation	On-demand creation of an instance of the MDPS.	View products from all the missions.	View and analyze products from all the missions.	Produce and analyze all the missions' products.	Produce and analyze all the missions' products.	Access to all MDPS functions	Access to all MDPS and DAAC functions.
Implications to management / programmatic	External office required to provide managed services. Color of money	External office required to provide managed services. Color of money	External office required to provide managed services	External office required to provide managed services. Color of money	External office required to provide managed services. Color of money	External office required to provide managed services. Difficult to produce one-size fits all. Color of money	External office required to provide managed services. Difficult to produce one-size fits all. Color of money
Strengths	* Consolidates common core and managed by core team allowing reuse of the very reusable non-mission specific components. * Separate MDPS allow customization for mission needs. * Reduces development cost and time. * Reduces overhead where MDPS need to be trained in specific cloud technology/ services.	• Instead of each MDPS handling data, a shared project office provides access to non-DAAC SIPS data.	• A managed data and analysis services will enable more multi-mission system science. • The missions' data management is interoperable. • Each ESO mission is able to develop and deploy their algorithms in different ways.	" Each missions data management is interoperable. " Shared development and operation of common pieces. " The missions' processing algorithms are interoperable. " Each mission can choose the workflow mechanism that best meets their needs.	" Each missions data management is interoperable. " Shared development and operation of common pieces. " The missions' processing algorithms are interoperable. " Workflow mechanism no longer has to be built individually"	• Multiple discipline products are generated together on same multi-mission system. • Larger efficiency and economy of scale • Support public users, they can be on-boarded with an account and funding. * Lowers infrastructure sustainment responsibilities of MDPS.	* Increasing efficiency with shared core services between MDPS and DAAC.

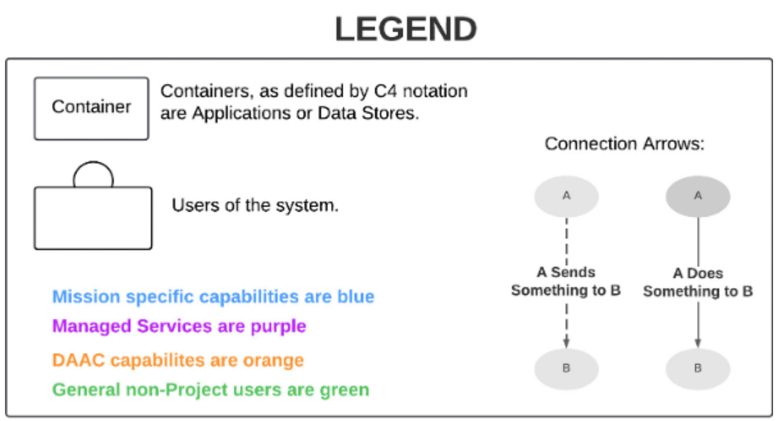
Type	Type 2 - Managed Services					Type 3 - Fully Managed System	
Variant #	Variant 1	Variant 2	Variant 3	Variant 4	Variant 5	Variant 1	Variant 2
Weakness	* Some possible duplication of effort in separate MDPS development and operation * Possible difficulties in prioritization (de-conflicting and merging of changes) between the different needs of MDPSes.	• Public non-project users cannot access the Project MDPS services. They will have to deploy their own copy to use analysis environment and processing.	• Each ESO mission may still have to develop and deploy their algorithms in different ways.	* Requires large-scaling of job processing from shared Processing Service	* Requires large-scaling of job processing from shared Processing Service * Workflow has to meet the needs of all ESO missions which increased Technical Complexity	* Requires large-scaling, support of multi-tenancy, higher risks. * Very difficult to design/develop/ maintain " One Size Fits All" software. " Limit innovation where each MDPS can replace components to better meet mission needs.	* Cross color of money lines. * Very difficult to design/develop/maint ain "One Size Fits All" software. " Limit innovation where each MDPS can replace components to better meet mission needs.

A.7 Programmatic Scores

		T1V1	T1V2	T1V3	T1V4	T1V5	T1V6	T2V1	T2V2	T2V3	T2V4	T2V5	T3V1	T3V2
Cultural complexity	ST Member #1	1	1	1	3	5	6	3	4	5	7	9	10	10
	ST Member #2	1	1	3	4	5	5	6	7	8	9	9	10	10
	ST Member #3	1	1	1	2	4	4	4	5	6	6	7	8	10
	ST Member #4	1	1	2	4	4	5	3	4	7	8	8	10	10
	Ave	1.0	1.0	1.8	3.3	4.5	5.0	4.0	5.0	6.5	7.5	8.3	9.5	10.0
	Var	0.0	0.0	0.9	0.9	0.3	0.7	**2.0**	**2.0**	**1.7**	**1.7**	0.9	1.0	0.0
Cost complexity	ST Member #1	1	1	1	3	4	5	3	4	5	7	9	10	10
	ST Member #2	1	1	1	3	4	5	5	6	7	8	9	9	9
	ST Member #3	1	1	2	2	3	4	5	6	6	8	10	10	10
	ST Member #4	1	2	3	3	4	5	5	6	6	7	8	10	10
	Ave	1.0	1.3	1.8	2.8	3.8	4.8	4.5	5.5	6.0	7.5	9.0	9.8	9.8
	Var	0.0	0.3	0.9	0.3	0.3	0.3	1.0	1.0	0.7	0.3	0.7	0.3	0.3
Resource complexity	ST Member #1	1	2	3	4	5	6	4	5	6	8	9	10	10
	ST Member #2	1	1	2	3	4	5	6	7	8	9	10	10	10
	ST Member #3	1	1	2	3	4	5	5	5	6	6	7	9	10
	ST Member #4	1	2	3	3	5	6	6	6	6	7	8	9	10
	Ave	1.0	1.5	2.5	3.3	4.5	5.5	5.3	5.8	6.5	7.5	8.5	9.5	10.0
	Var	0.0	0.3	0.3	0.3	0.3	0.3	0.9	0.9	1.0	**1.7**	**1.7**	0.3	0.0

		T1V1	T1V2	T1V3	T1V4	T1V5	T1V6	T2V1	T2V2	T2V3	T2V4	T2V5	T3V1	T3V2
Schedule complexity	ST Member #1	1	2	2	4	5	6	3	5	7	9	9	10	10
	ST Member #2	1	1	1	2	3	5	5	5	6	7	8	9	10
	ST Member #3	1	1	2	3	4	5	4	5	5	7	7	8	8
	ST Member #4	1	2	2	5	5	7	5	6	6	7	8	9	10
	Ave	1.0	1.5	1.8	3.5	4.3	5.8	4.3	5.3	6.0	7.5	8.0	9.0	9.5
	Var	0.0	0.3	0.3	**1.7**	0.9	0.9	0.9	0.3	0.7	1.0	0.7	0.7	1.0
		T1V1	T1V2	T1V3	T1V4	T1V5	T1V6	T2V1	T2V2	T2V3	T2V4	T2V5	T3V1	T3V2
Average Programmatic Complexity (1-10)		1.00	1.31	1.94	3.19	4.25	5.25	4.50	5.38	6.25	7.50	8.44	9.44	9.81
Programmatic Complexity Factor (PCF) (Scaled 0-1)		0.00	0.04	0.11	0.25	0.37	0.48	0.40	0.50	0.60	0.74	0.84	0.96	1.00

A.8 Diagrams of Select Architecture Variants



Architecture diagram components, connectors, and color coding.

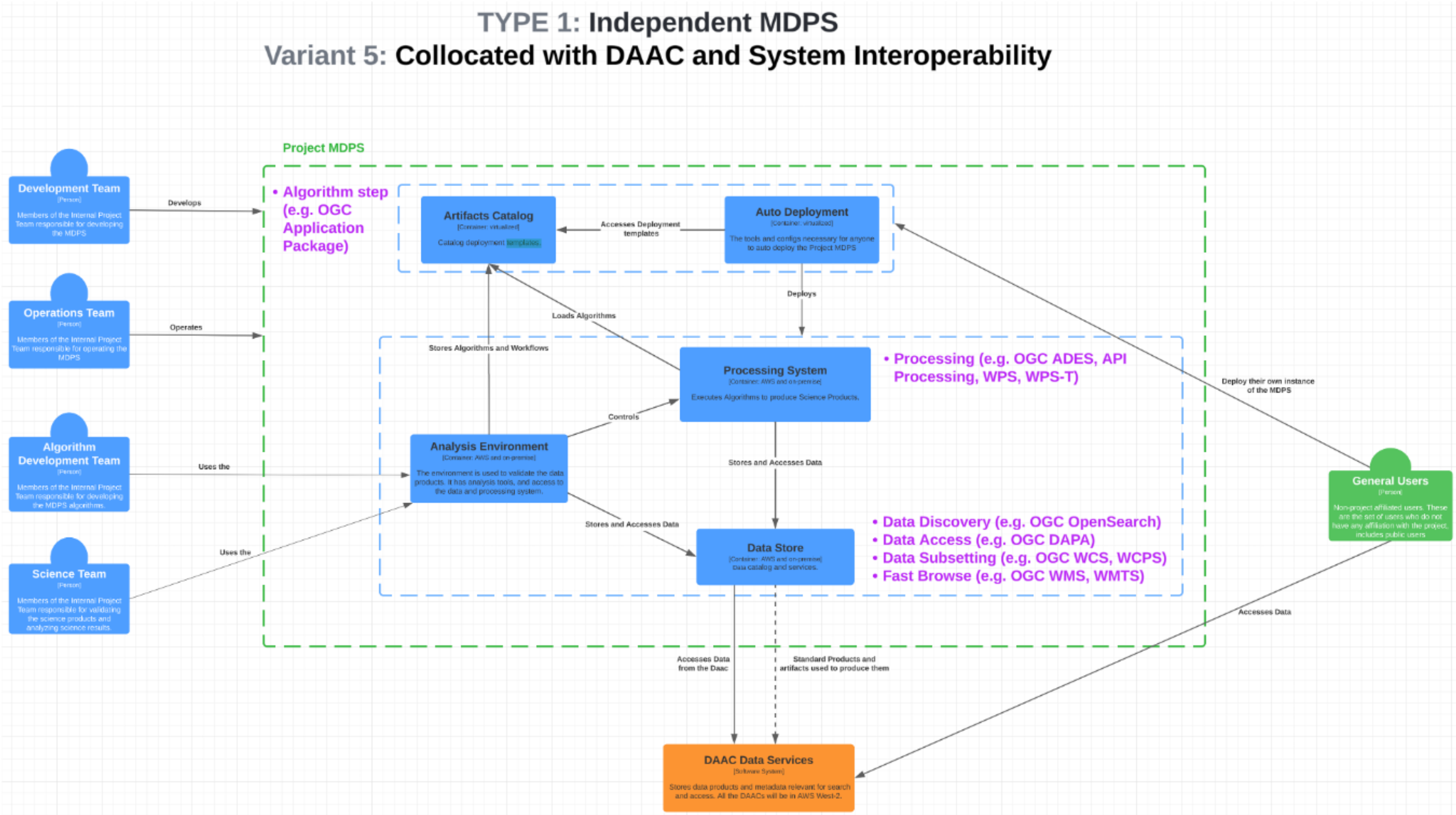


Figure A-1. T1V5: TYPE 1, Variant 5 - Independent MDPS, collocated with DAAC with system interoperability.

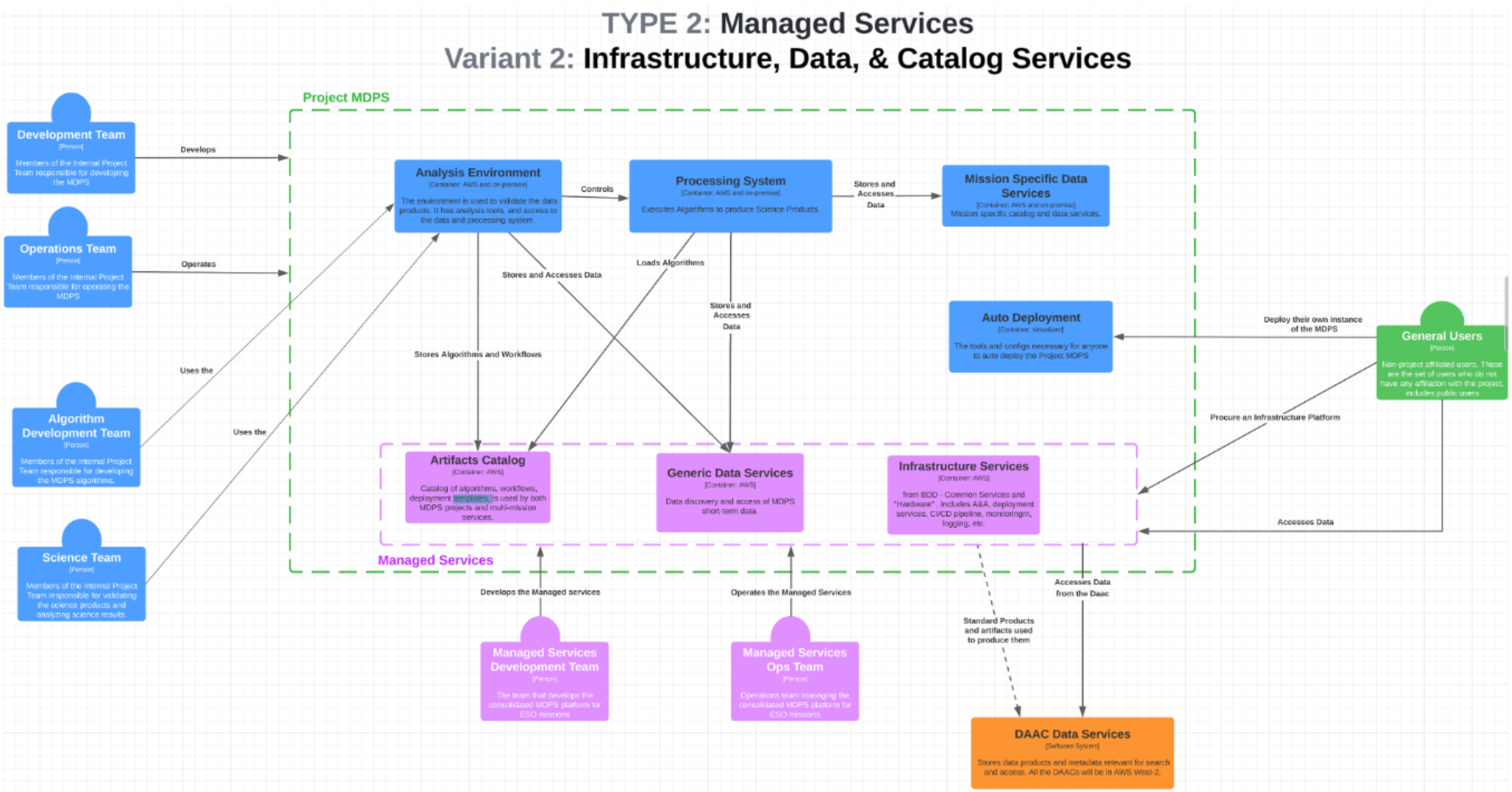


Figure A-2. T2V2: TYPE 2, Variant 2 - Managed infrastructure, data, and catalog services.

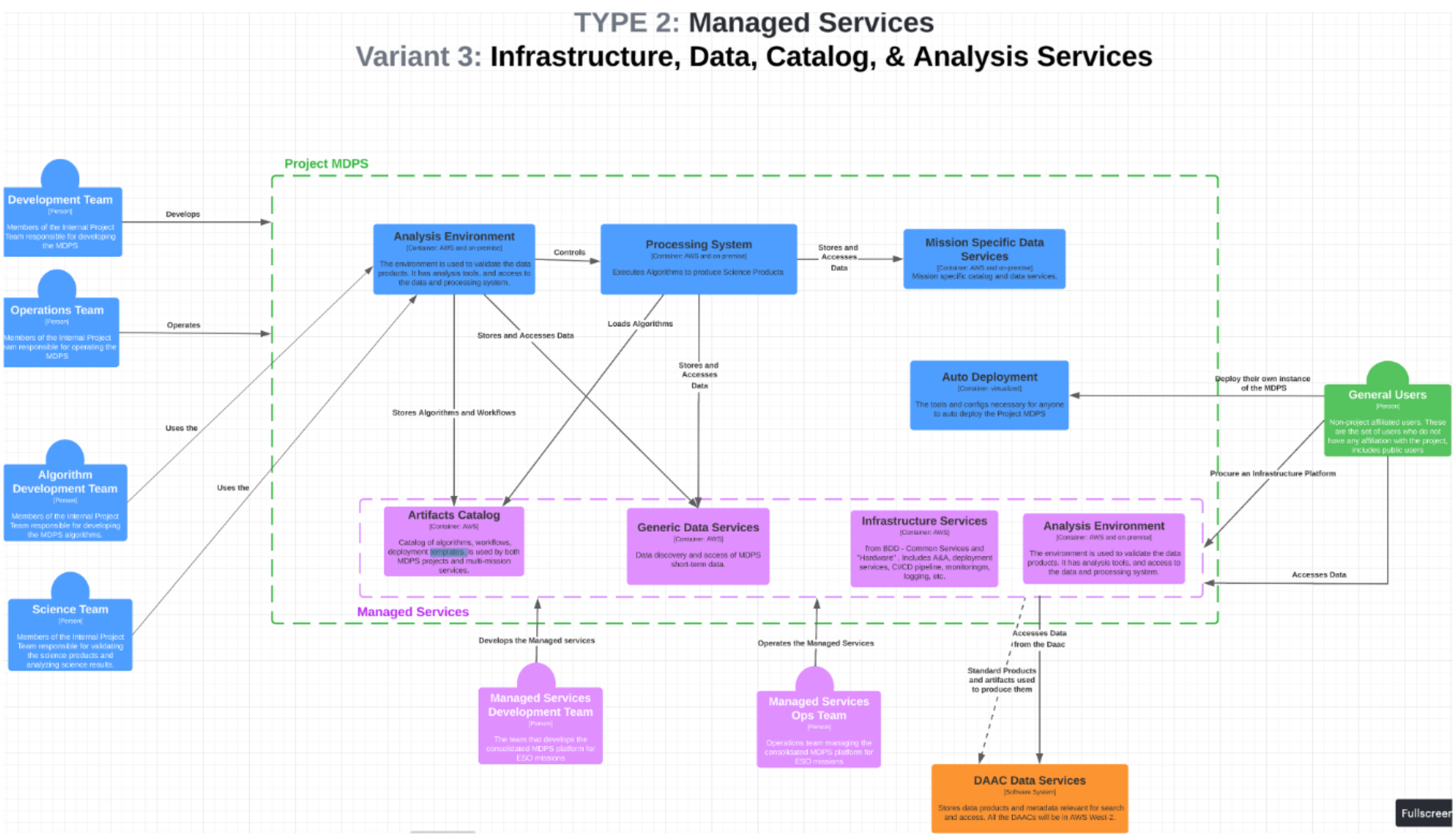


Figure A-3. T2V3: TYPE 2, Variant 3 - Managed infrastructure, data, catalog and analysis services.

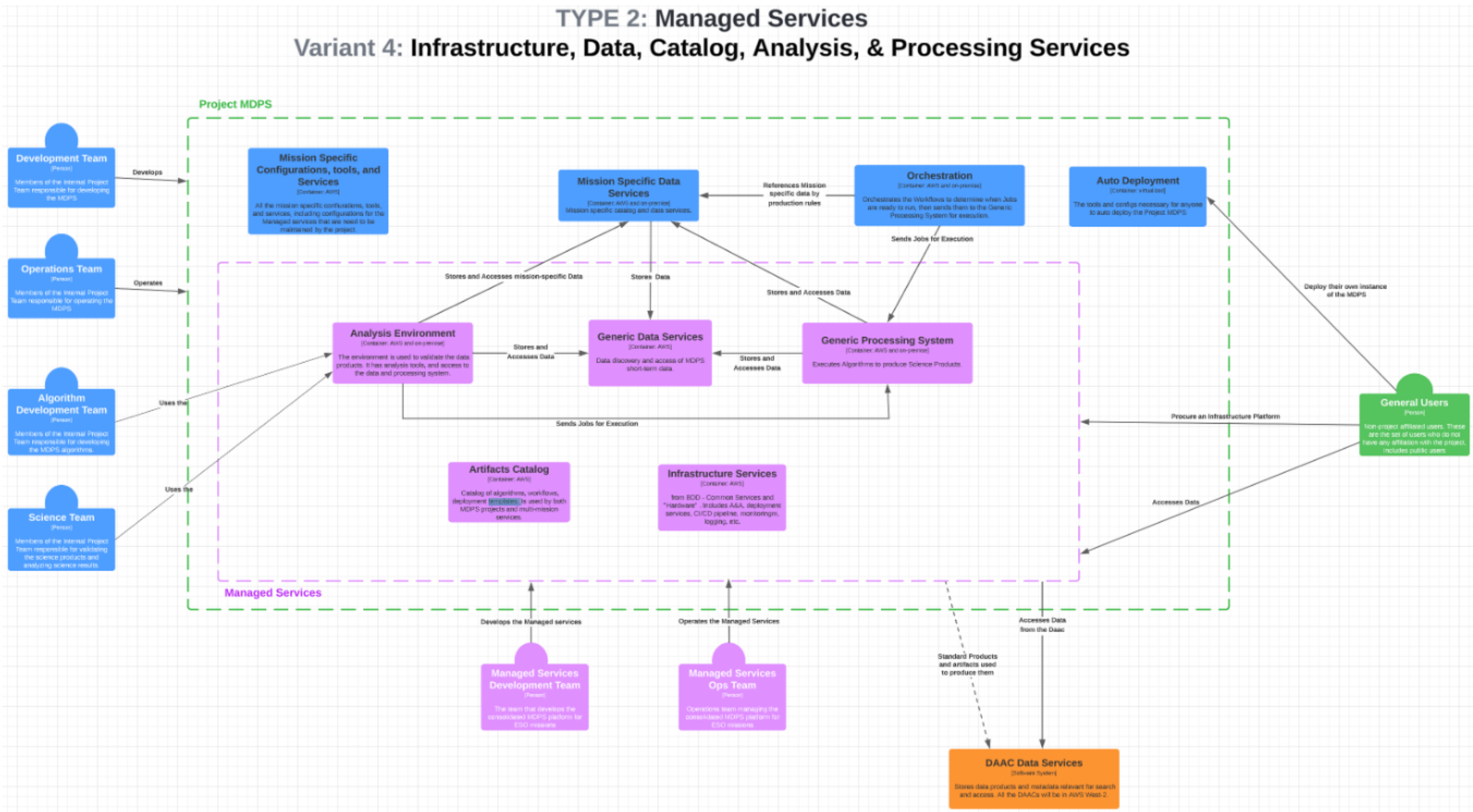


Figure A-4. T2V4: TYPE 2, Variant 4 - Managed infrastructure, data, catalog, analysis, and processing services.

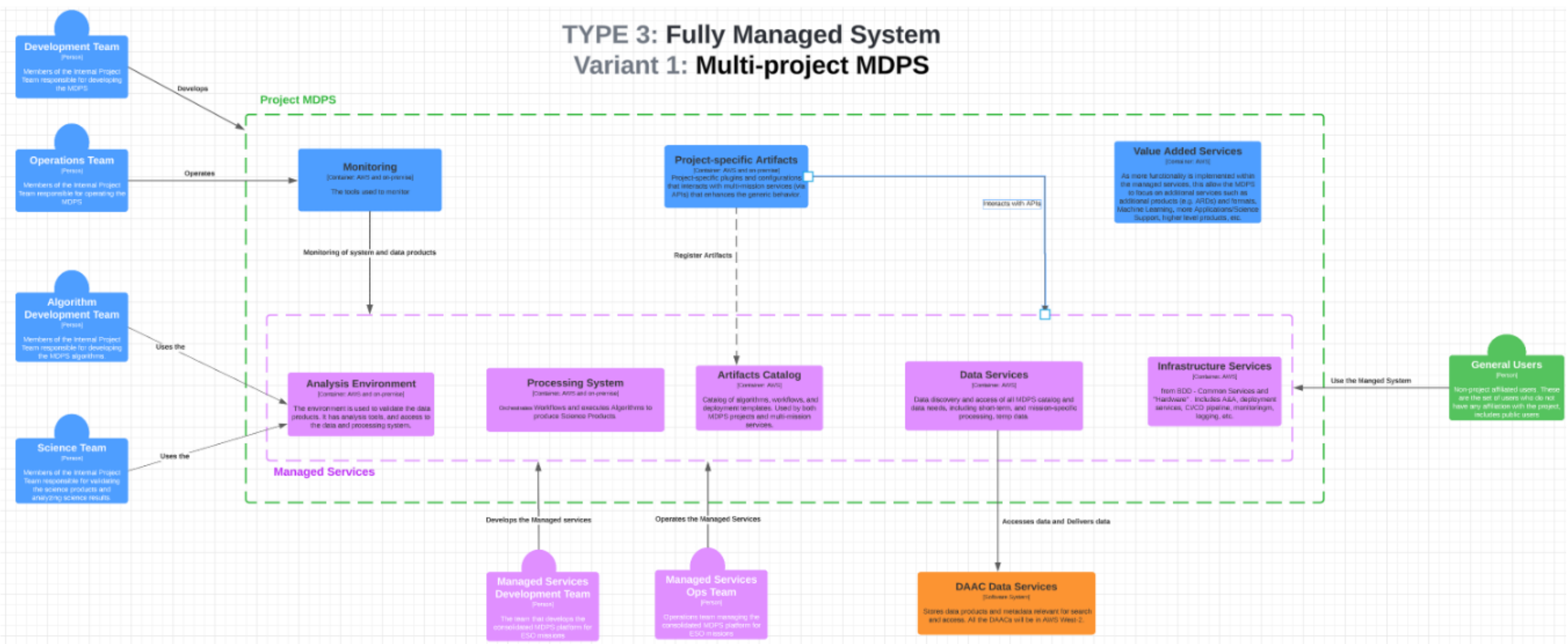
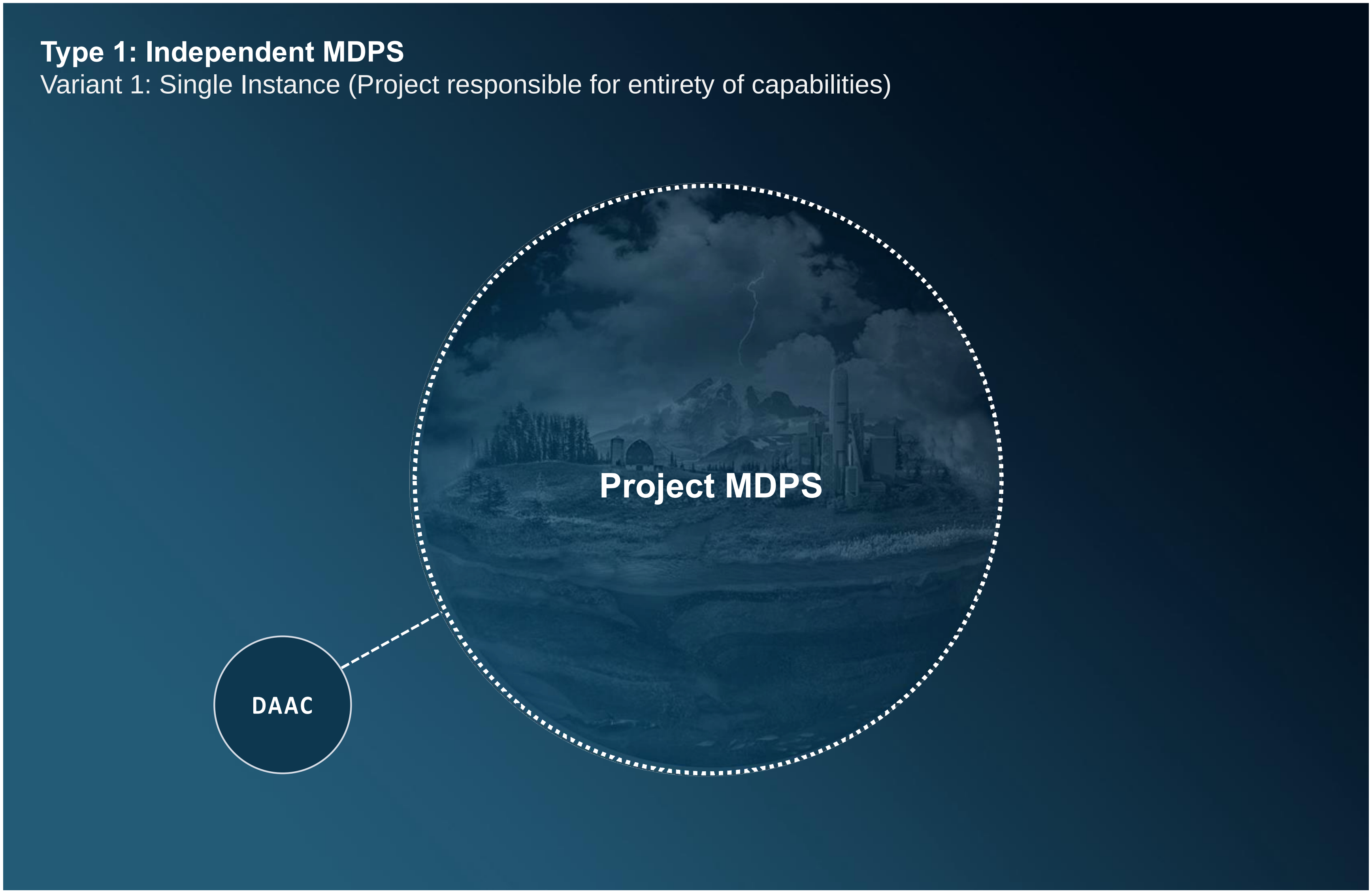
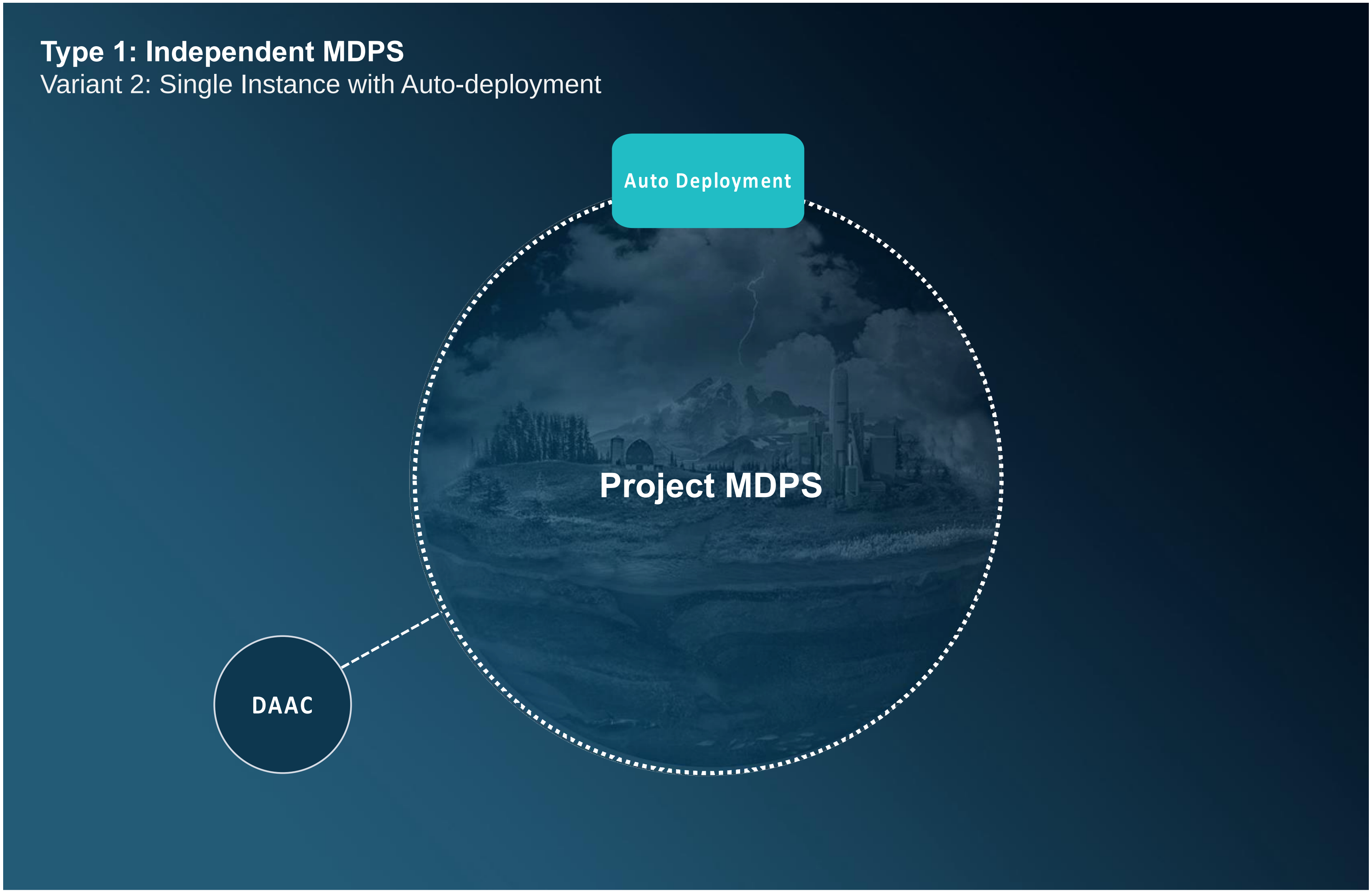
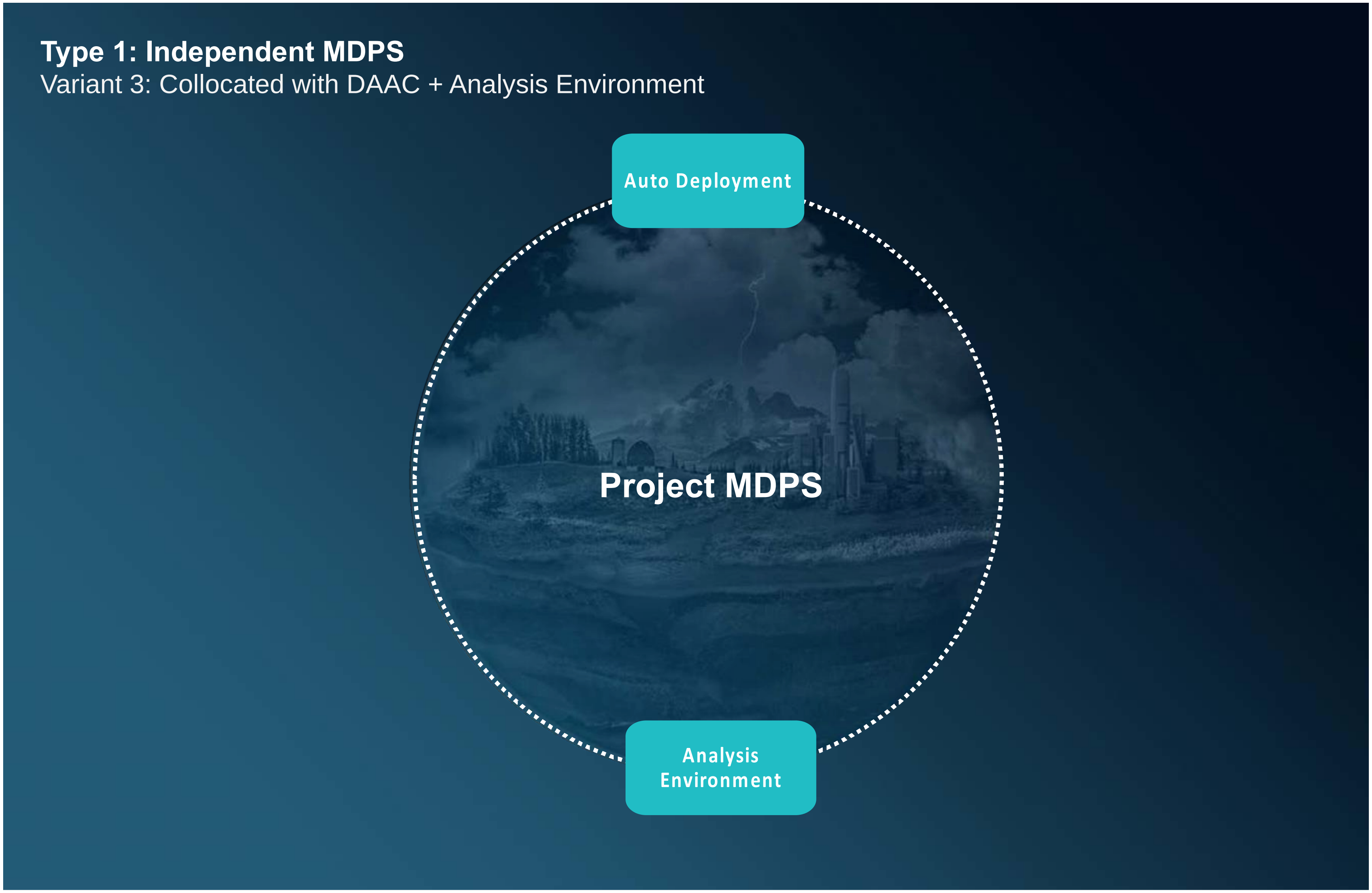


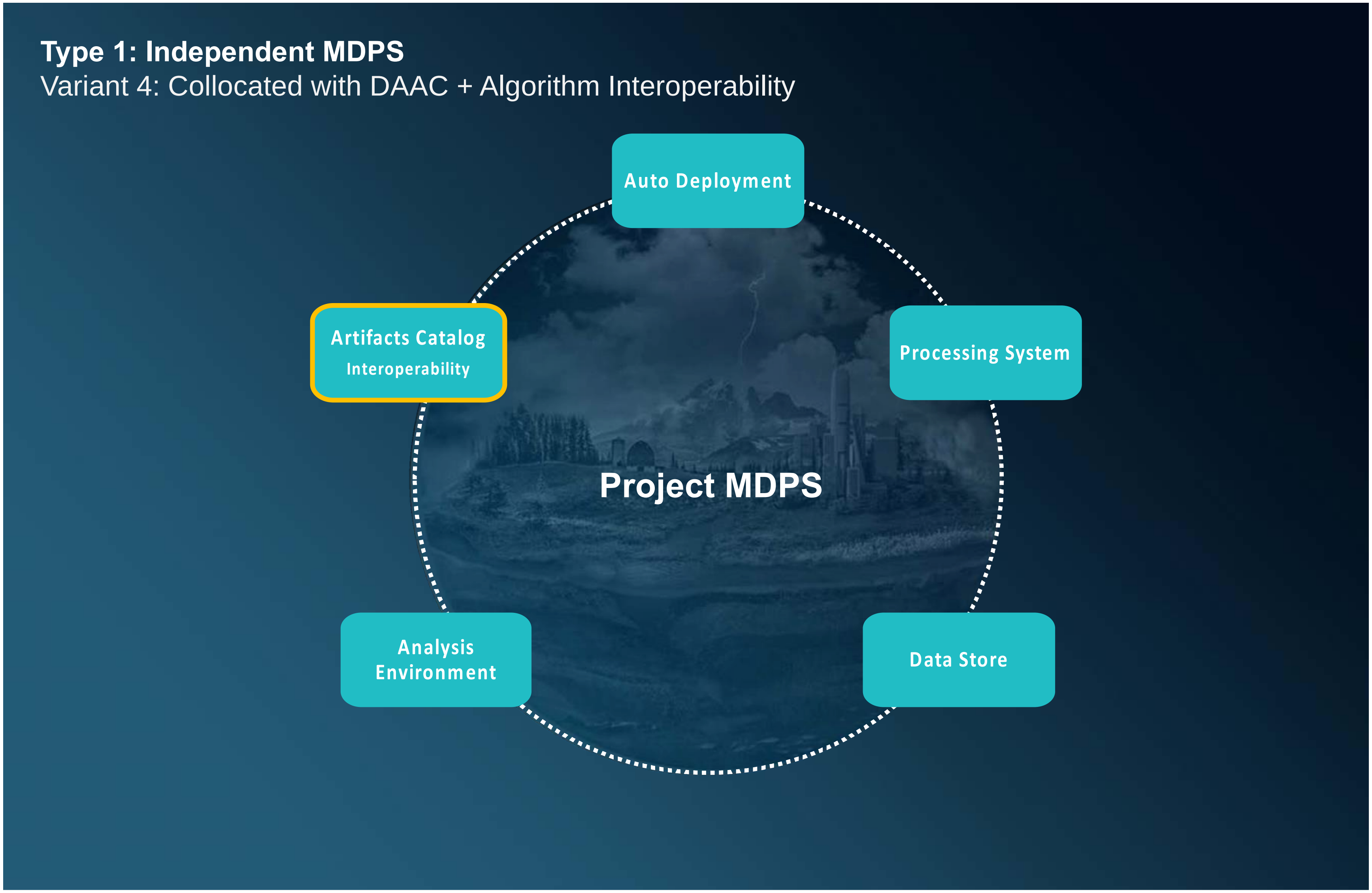
Figure A-5. T3V1: TYPE 3, Variant 1 - A Fully managed, multi-project MDPS system.

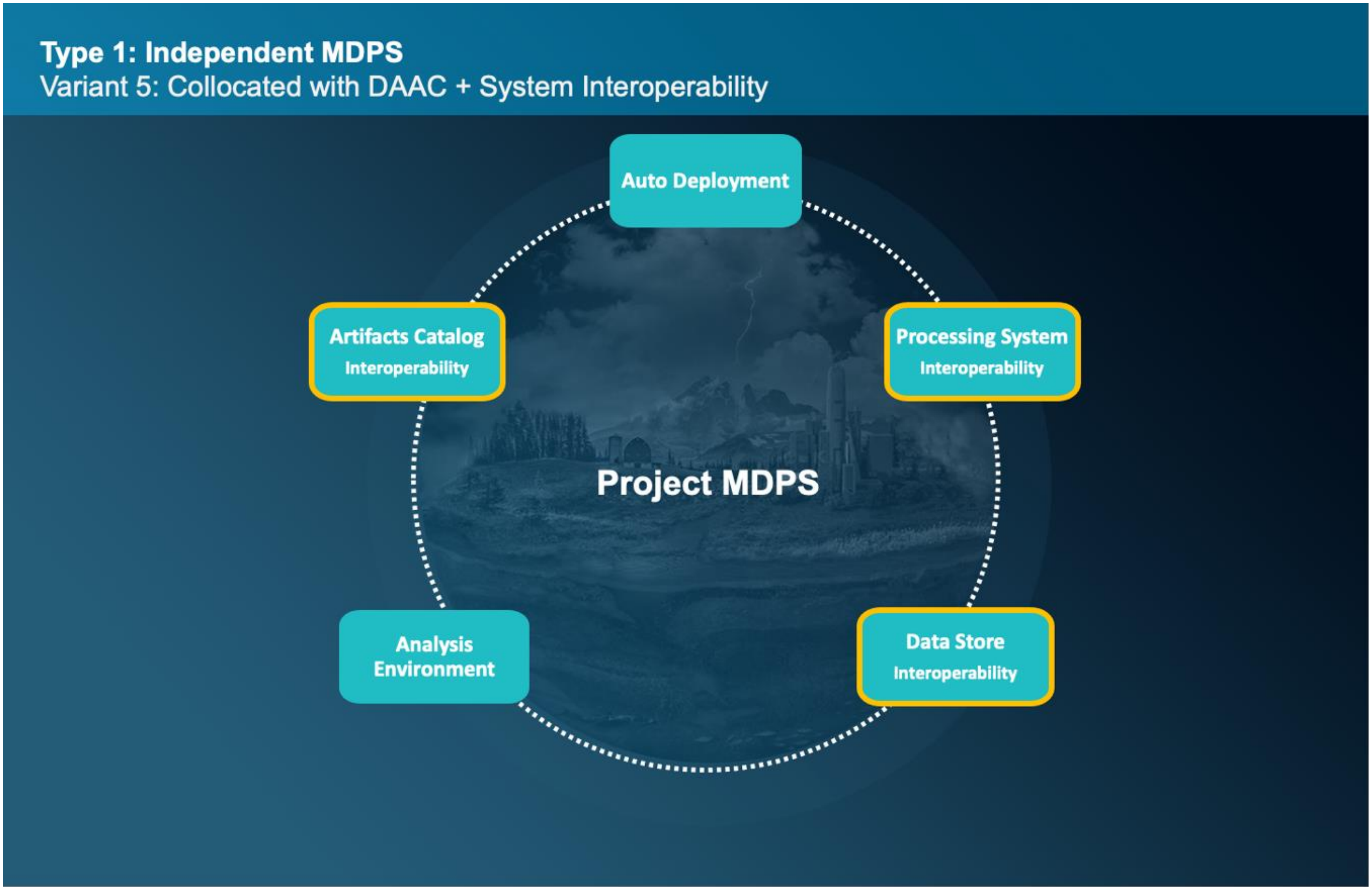
A.9 General Diagrams for the Collection of Architectures and Variants

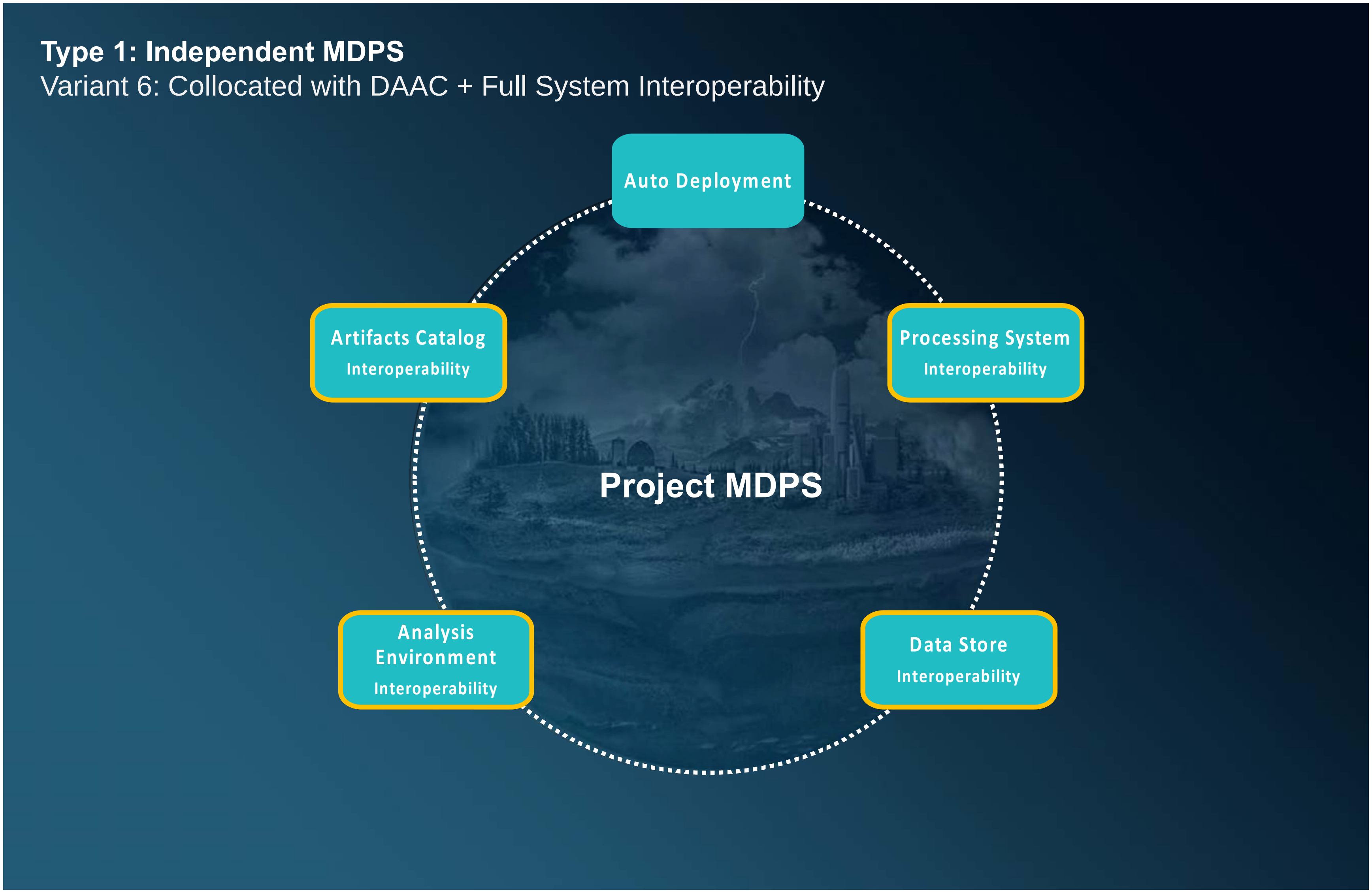


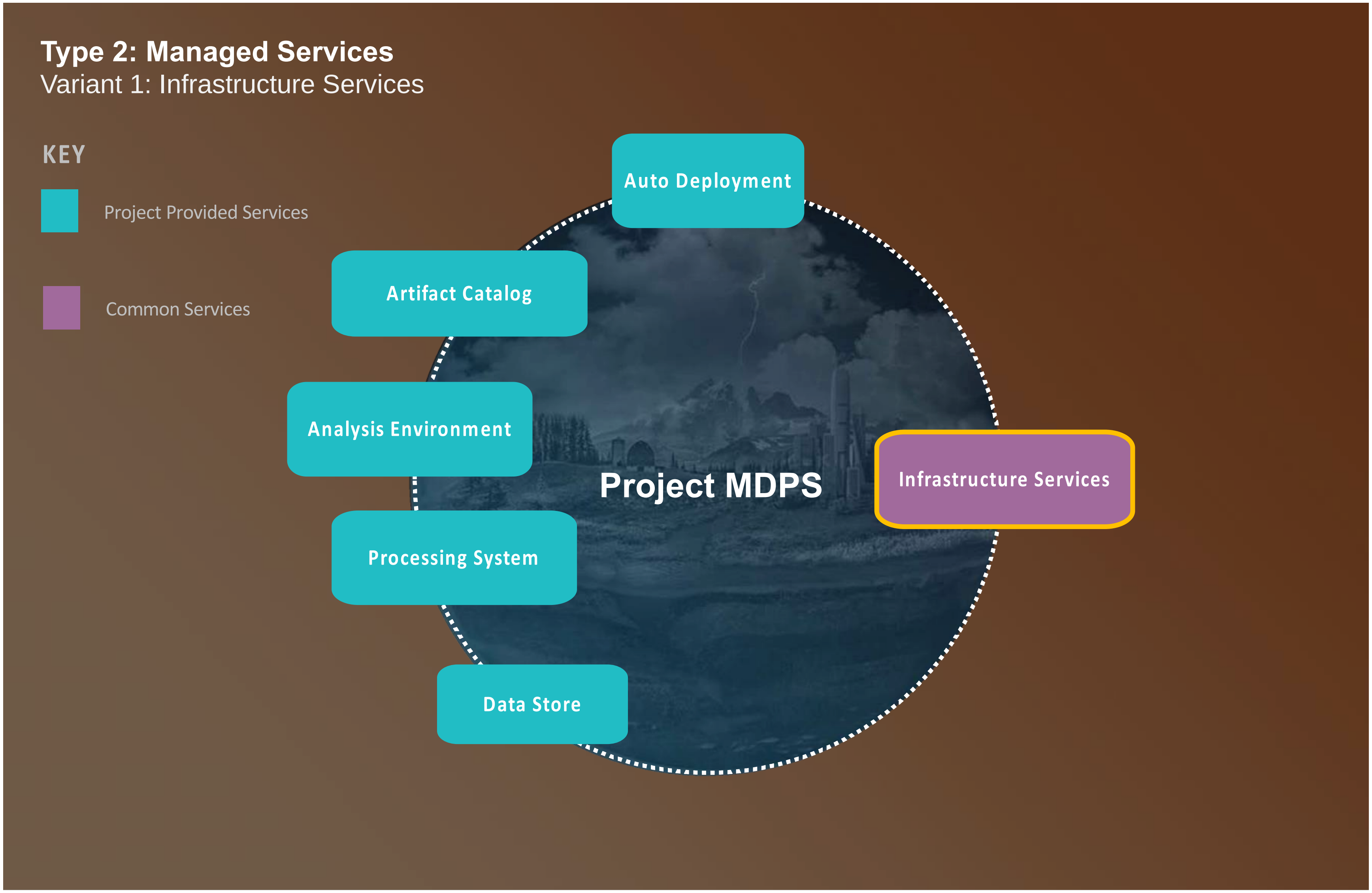

Type 1: Independent MDPS
Variant 2: Single Instance with Auto-deployment
Auto Deployment
Project MDPS
DAAC

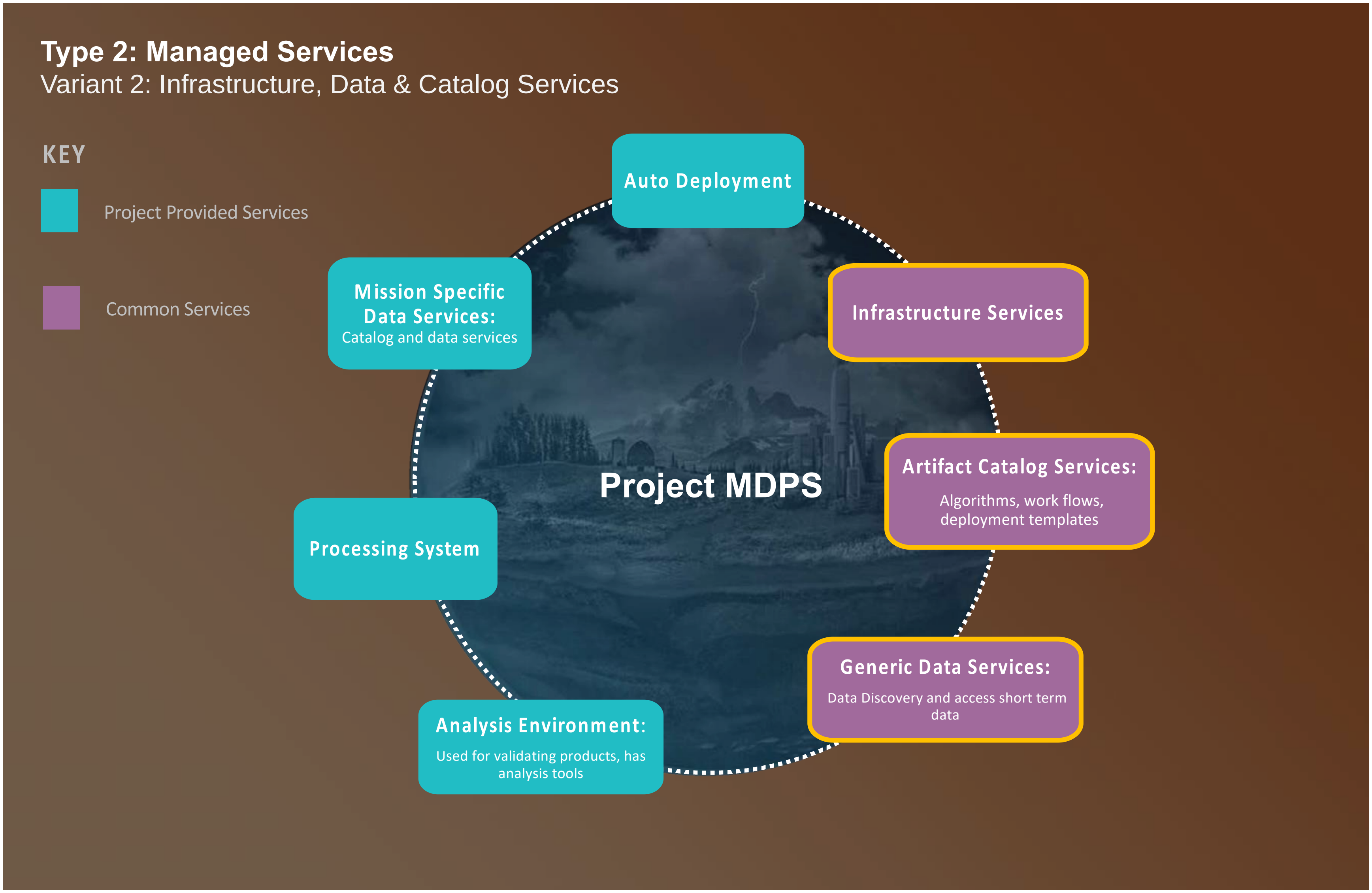

Type 1: Independent MDPS
Variant 3: Collocated with DAAC + Analysis Environment
Auto Deployment
Project MDPS
Analysis Environment


Type 1: Independent MDPS
Variant 4: Collocated with DAAC + Algorithm Interoperability
Auto Deployment
Artifacts Catalog
Interoperability
Processing System
Project MDPS
Analysis Environment
Data Store

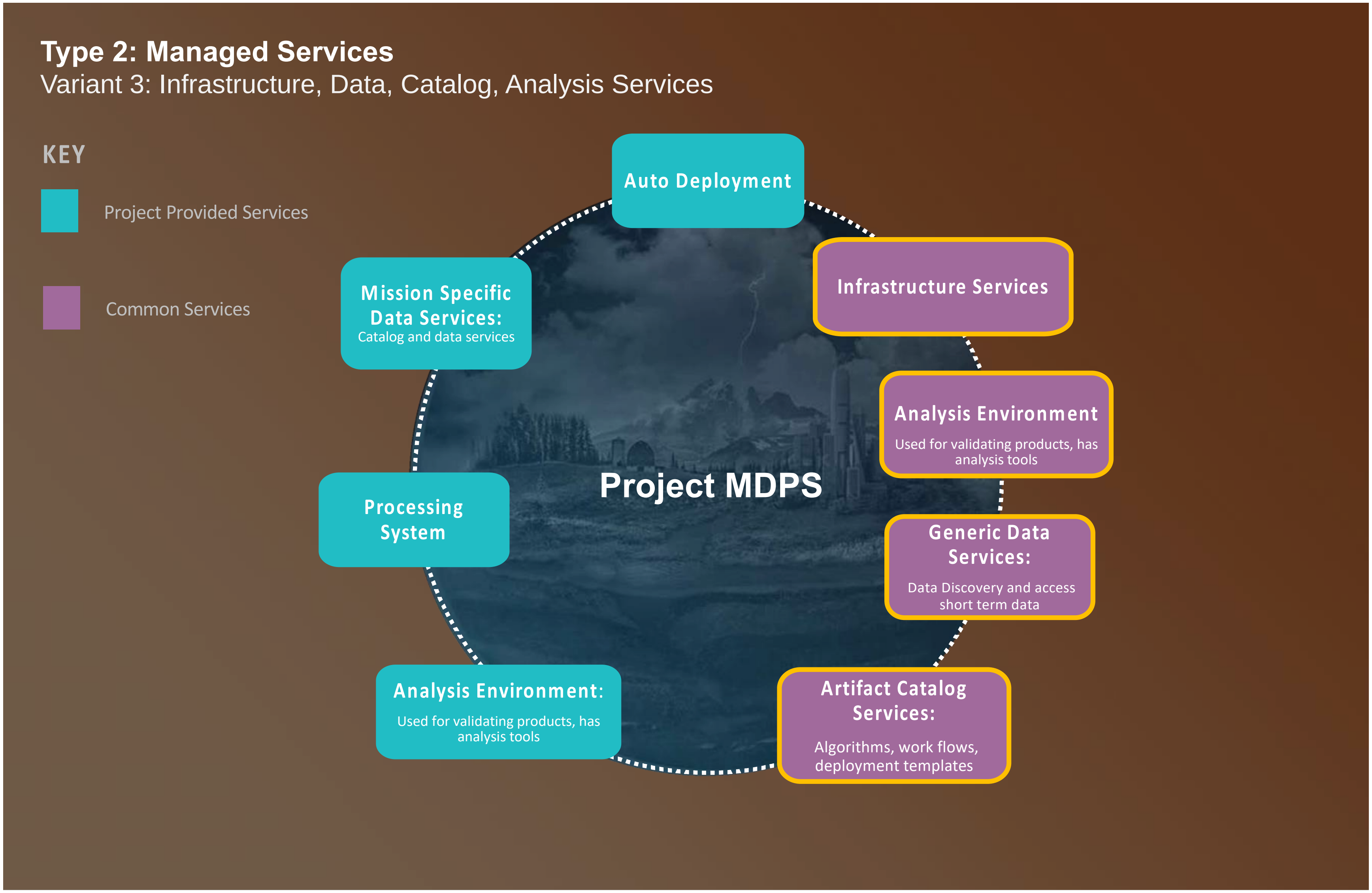

Type 1: Independent MDPS
Variant 5: Collocated with DAAC + System Interoperability
Auto Deployment
Artifacts Catalog
Interoperability
Processing System
Interoperability
Project MDPS
Analysis
Environment
Data Store
Interoperability

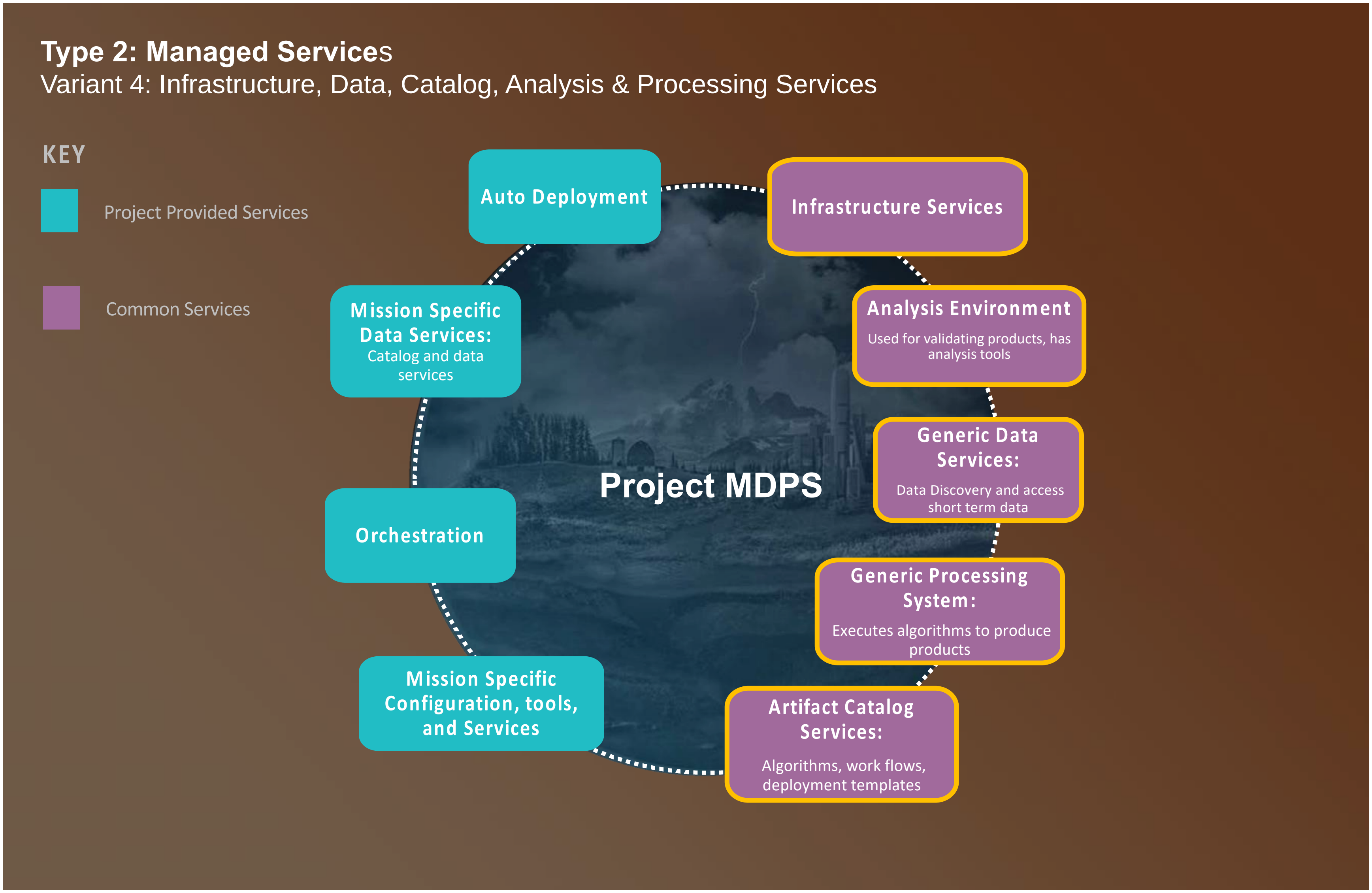

Type 1: Independent MDPS
Variant 6: Collocated with DAAC + Full System Interoperability
Auto Deployment
Artifacts Catalog
Interoperability
Processing System
Interoperability
Project MDPS
Analysis Environment
Interoperability
Data Store
Interoperability

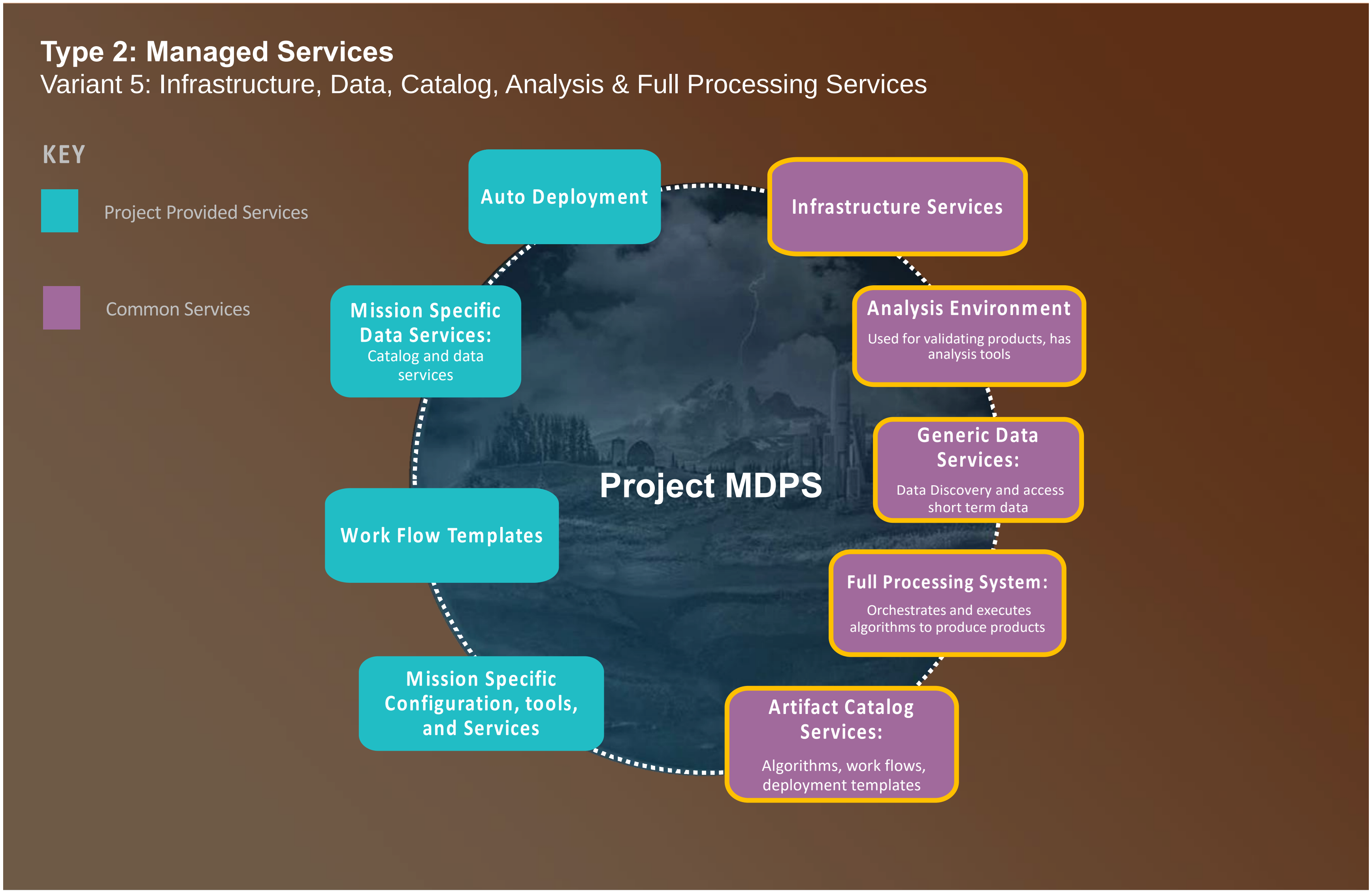

Type 2: Managed Services
Variant 1: Infrastructure Services
KEY
Project Provided Services
Common Services
Auto Deployment
Artifact Catalog
Analysis Environment
Processing System
Data Store
Project MDPS
Infrastructure Services

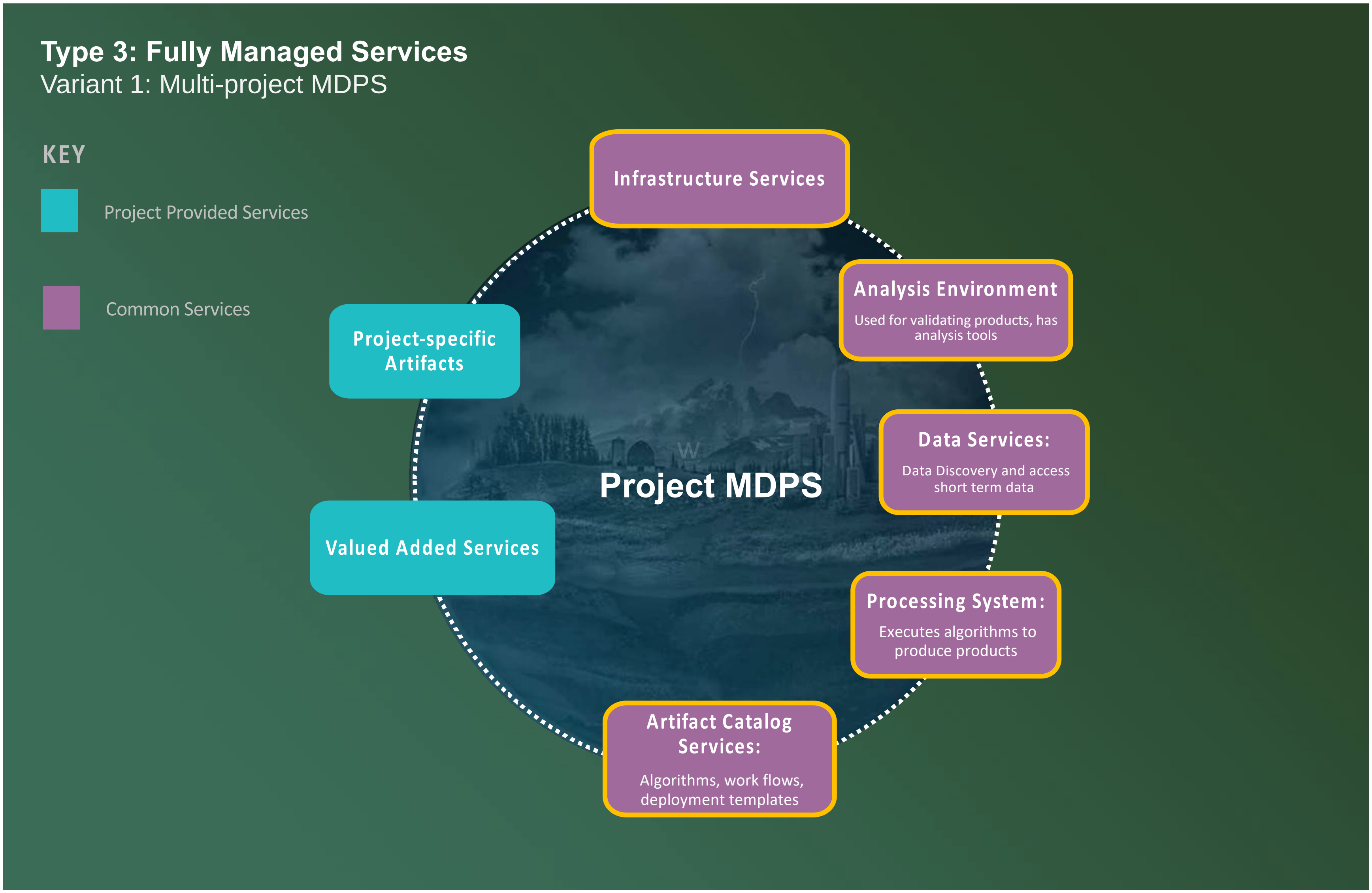


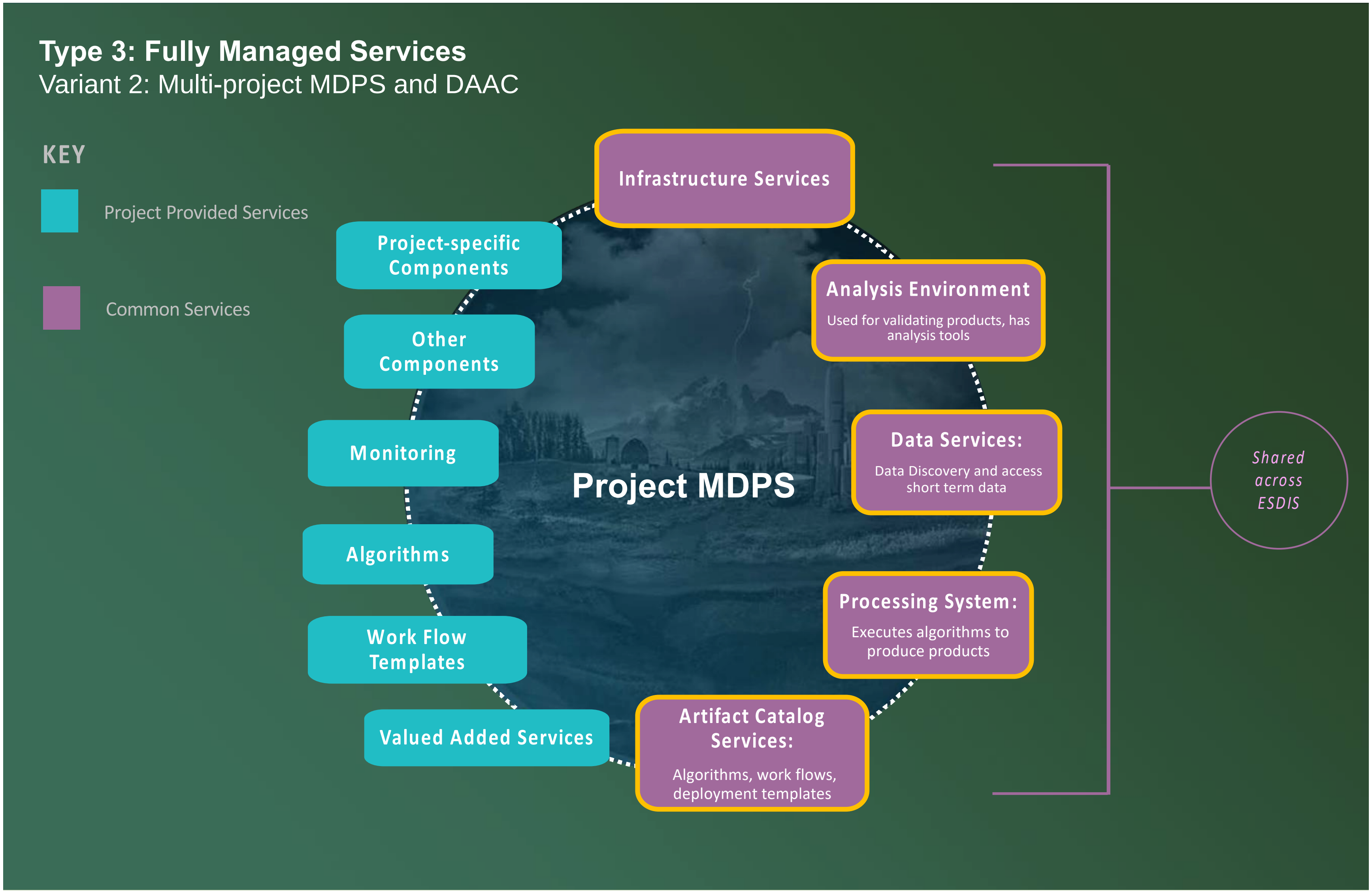
Type 2: Managed Services
Variant 2: Infrastructure, Data & Catalog Services
KEY
Project Provided Services
Common Services
Auto Deployment
Mission Specific Data Services:
Catalog and data services
Infrastructure Services
Project MDPS
Artifact Catalog Services:
Algorithms, work flows, deployment templates
Processing System
Generic Data Services:
Data Discovery and access short term data
Analysis Environment:
Used for validating products, has analysis tools


Type 2: Managed Services
Variant 3: Infrastructure, Data, Catalog, Analysis Services
KEY
Project Provided Services
Common Services
Auto Deployment
Infrastructure Services
Mission Specific Data Services:
Catalog and data services
Analysis Environment
Used for validating products, has analysis tools
Project MDPS
Processing System
Generic Data Services:
Data Discovery and access short term data
Analysis Environment:
Used for validating products, has analysis tools
Artifact Catalog Services:
Algorithms, work flows, deployment templates


Type 2: Managed Services
Variant 4: Infrastructure, Data, Catalog, Analysis & Processing Services
KEY
Project Provided Services
Common Services
Auto Deployment
Infrastructure Services
Mission Specific Data Services: Catalog and data services
Analysis Environment
Used for validating products, has analysis tools
Generic Data Services:
Data Discovery and access short term data
Project MDPS
Orchestration
Generic Processing System:
Executes algorithms to produce products
Mission Specific Configuration, tools, and Services
Artifact Catalog Services:
Algorithms, work flows, deployment templates


Type 2: Managed Services
Variant 5: Infrastructure, Data, Catalog, Analysis & Full Processing Services
KEY
Project Provided Services
Common Services
Auto Deployment
Infrastructure Services
Mission Specific Data Services: Catalog and data services
Analysis Environment
Used for validating products, has analysis tools
Generic Data Services:
Data Discovery and access short term data
Project MDPS
Work Flow Templates
Full Processing System:
Orchestrates and executes algorithms to produce products
Mission Specific Configuration, tools, and Services
Artifact Catalog Services:
Algorithms, work flows, deployment templates


Type 3: Fully Managed Services
Variant 1: Multi-project MDPS
KEY
Project Provided Services
Common Services
Infrastructure Services
Analysis Environment
Used for validating products, has analysis tools
Project-specific Artifacts
Data Services:
Data Discovery and access short term data
Project MDPS
Valued Added Services
Processing System:
Executes algorithms to produce products
Artifact Catalog Services:
Algorithms, work flows, deployment templates


Type 3: Fully Managed Services
Variant 2: Multi-project MDPS and DAAC
KEY
Project Provided Services
Common Services
Infrastructure Services
Project-specific Components
Other Components
Monitoring
Algorithms
Work Flow Templates
Valued Added Services
Project MDPS
Analysis Environment
Used for validating products, has analysis tools
Data Services:
Data Discovery and access short term data
Processing System:
Executes algorithms to produce products
Artifact Catalog Services:
Algorithms, work flows, deployment templates
Shared across ESDIS

A.10 Technical Scoring Tables

			Type 1 - Independent MDPS						Type 2 - Managed Services					Type 3 - Fully Managed System	
			Variant 1	Variant 2	Variant 3	Variant 4	Variant 5	Variant 6	Variant 1	Variant 2	Variant 3	Variant 4	Variant 5	Variant 1	Variant 2
		Weighting	Sub-score by criterion of architecture type and variant												
ESO Mission Science Processing Objectives ["Mission Dev"]															
The Data System should provide an accessible platform for the community to develop L3+ products that adhere to NASA metadata and provenance standards	MD1	2.38	0.00	0.00	4.75	7.13	7.13	7.13	2.38	4.75	7.13	7.13	7.13	7.13	7.13
The Data System should support on-demand product generation as soon as data is available from the ground data system.	MD2	2.38	0.00	2.38	4.75	4.75	4.75	4.75	2.38	4.75	4.75	4.75	7.13	7.13	7.13
The Data System should be portable to support deployment on-prem, in-cloud, multi-cloud, and hybrid infrastructure.	MD3	2.13	0.00	4.25	4.25	4.25	4.25	4.25	6.38	6.38	6.38	6.38	6.38	6.38	6.38
The Data system external interfaces should go through Authentication and Authorization.	MD4a	3.00	0.00	0.00	0.00	0.00	3.00	3.00	9.00	9.00	9.00	9.00	9.00	9.00	9.00
The Data system external interfaces should use standardized access protocols.	MD4b	2.88	0.00	0.00	2.88	5.75	5.75	5.75	8.63	8.63	8.63	8.63	8.63	8.63	8.63
The Data System should have the ability for forward-stream and bulk [re-]processing.	MD5a	2.75	2.75	2.75	2.75	5.50	5.50	5.50	8.25	8.25	8.25	8.25	8.25	8.25	8.25
The Data System should be compliant with DAAC archive retrieval.	MD5b	2.38	0.00	2.38	7.13	7.13	7.13	7.13	7.13	7.13	7.13	7.13	7.13	7.13	7.13
The Data System should be demonstrated (TRL6+) by the earliest ESO mission launch	MD6	2.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
The Data System is cost-constrained by ESO mission budget capacity	MD7	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Enable Data System Efficiencies															
The Data System should support a data lake.	SE1	2.38	0.00	2.38	4.75	4.75	4.75	4.75	2.38	4.75	4.75	4.75	4.75	4.75	7.13
The Data System should be flexible to efficiently (cost, bandwidth, processing capability) support large and small data volumes.	SE2	2.88	0.00	0.00	2.88	2.88	2.88	2.88	2.38	5.75	5.75	5.75	5.75	5.75	8.63
The Data system should accommodate variable compute needs over time is crucial to reducing costs.	SE3	2.88	0.00	0.00	2.88	2.88	2.88	2.88	2.38	2.88	2.88	5.75	5.75	5.75	8.63
The Data System should support services to create standard data formats (ESDIS Standards)	SE4	3.00	3.00	3.00	3.00	3.00	3.00	3.00	2.38	9.00	9.00	9.00	9.00	9.00	9.00
The Data System should keep up with forward-stream processing demand.	SE5	2.75	2.75	2.75	2.75	2.75	2.75	2.75	2.38	2.75	2.75	8.25	8.25	8.25	8.25
Support Earth System and Science Applications															
The Data System should enable ESO data sharing before data archive.	ESA1	1.75	0.00	0.00	0.00	0.00	1.75	3.50	0.00	5.25	5.25	5.25	5.25	5.25	5.25
The Data System should enable data access from non-ESO missions.	ESA2	2.00	0.00	0.00	2.00	2.00	2.00	4.00	0.00	4.00	4.00	4.00	4.00	4.00	6.00
The Data System should enable users to share algorithms.	ESA3	2.00	0.00	0.00	0.00	4.00	4.00	4.00	0.00	6.00	6.00	6.00	6.00	6.00	6.00
The Data System should enable development and sharing of data tools (e.g., software, code libraries, etc).	ESA4	2.75	0.00	0.00	0.00	2.75	5.50	8.25	0.00	5.50	5.50	5.50	5.50	8.25	8.25
The Data System should enable on-demand processing.	ESA5	2.38	0.00	0.00	2.38	2.38	2.38	2.38	0.00	2.38	2.38	7.13	7.13	7.13	7.13
The Data System should meet cross-ESO mission science goals.	ESA6	3.00	0.00	3.00	6.00	6.00	6.00	6.00	0.00	3.00	3.00	6.00	9.00	9.00	9.00

			Type 1 - Independent MDPS						Type 2 - Managed Services					Type 3 - Fully Managed System	
			Variant 1	Variant 2	Variant 3	Variant 4	Variant 5	Variant 6	Variant 1	Variant 2	Variant 3	Variant 4	Variant 5	Variant 1	Variant 2
		Weighting	Sub-score by criterion of architecture type and variant												
ESO Mission Science Processing Objectives ["Mission Dev"]															
Promote Open Science Principles															
The Data System should provide an analysis platform.	OSS1a	2.25	0.00	0.00	2.25	2.25	4.50	4.50	0.00	2.25	6.75	6.75	6.75	6.75	6.75
The Analysis Platform should be accessible by NASA and Non-NASA users while supporting algorithm sharing and on-demand batch processing.	OSS1b	1.75	0.00	0.00	1.75	3.50	3.50	3.50	1.75	1.75	5.25	5.25	5.25	5.25	5.25
The Analysis Platform should enable users to control user-generated resources: private and public code repositories, containers, binaries, data, etc.	OSS2	2.38	0.00	0.00	2.38	2.38	2.38	2.38	0.00	4.75	7.13	7.13	7.13	7.13	7.13
The Analysis Platform should enable public access.	OSS3	0.88	0.00	0.00	0.00	0.00	0.00	0.00	0.88	0.88	2.63	2.63	2.63	2.63	2.63
The Analysis Platform should enable user authentication and authorization for provenance of contributions.	OSS4	2.00	0.00	0.00	2.00	2.00	2.00	2.00	2.00	4.00	6.00	6.00	6.00	6.00	6.00
The Analysis Platform should automate standardized open source science guidelines (e.g., metadata standards, provenance, etc.).	OSS5	1.63	0.00	0.00	1.63	1.63	1.63	1.63	0.00	1.63	4.88	4.88	4.88	4.88	4.88
The Data System should allocate cost accounting by user activity.	OSS6	2.25	2.25	2.25	2.25	2.25	2.25	2.25	6.75	6.75	6.75	6.75	6.75	6.75	6.75
The Data System should be compliant with cybersecurity policies (e.g., authenticate and authorize, user management, etc.). NOTE: Intellectual Property compliance is covered by other evaluation criteria.	OSS7	2.88	0.00	0.00	0.00	0.00	0.00	0.00	8.63	8.63	8.63	8.63	8.63	8.63	8.63
TOTAL "Desirements" EVAL SCORE (Top 3 colored GREEN):			10.75	25.13	65.38	81.88	91.63	98.13	76.00	130.75	150.50	166.63	172.00	174.75	184.88
Feasibility Factors (FF)															
Cost of implementation to get to this Architecture from where we are now	D1	5	0	0	5	5	5	5	5	10	10	15	15	20	25
Technical Complexity	D2	6	0	6	12	18	18	18	6	12	12	18	24	24	30
Cost Tracking Complexity	D3	3	0	0	0	0	0	0	3	6	6	9	12	12	15
Requirements Complexity	D4	3	0	3	6	6	6	6	3	6	6	9	9	12	15
Team Complexity	D5	3	0	0	3	3	3	3	3	9	9	9	12	12	15
Schedule Complexity	D6	3	0	0	3	3	6	6	6	6	6	9	12	12	15
Cybersecurity Conformance	D7	2	0	2	2	2	2	2	2	6	6	6	6	8	10
	Total (FF)		0	11	31	37	40	40	28	55	55	75	90	100	125
Maturity Factors (MF)															
High TRL at ESO mission implement need	D8	5	0	0	0	5	10	20	5	5	5	10	20	20	25
Development process	D9	5	0	0	0	0	0	0	10	10	10	10	20	20	25
Operations model	D10	4	0	0	4	4	4	4	4	4	8	12	16	16	20
Maintenance model	D11	4	0	0	4	4	4	4	4	8	8	12	16	16	20
	Total (MF)		0	0	8	13	18	28	23	27	31	44	72	72	90
TOTAL (MF+FF) EVAL SCORE:			0	11	39	50	58	68	51	82	86	119	162	172	215

B. Glossary

Accessible - Data, tools, software, documentation, publications follow FAIR Data Principles.

Analysis-Ready Data (ARD)- are satellite data that have been processed to a minimum set of requirements and organized into a form that allows immediate analysis with a minimum of additional user effort and interoperability both through time and with other datasets.

Application - use of NASA data for decision support (policy, resources, etc.).

Analysis-Ready Cloud Optimized (ARCO) - ARD data stored in cloud-optimized data formats enabling rapid access to the ARDs.

Analysis Platform - used for scientific analysis of products generated and made available, not for generating products.

Application-Ready Data - GIS-ready data

Architecture - A MDPS architecture is a system as a collection of components and connectors. Architecture should not be considered merely a set of models or structures, but should include the decisions that lead to these particular structures, and the rationale behind them.

Architecture Type - the high-level grouping that the SAWG decided to divide architectures by. Architecture Type refers to the overall architecture’s topology, behavior, and deployment. Architecture Types are distinctly different from each other in the above characteristics.

Architecture Variant - the next level grouping of architectures after Architecture Type. Each Architecture Type may have multiple Variants. Variants of an Architecture Type retain the same topology, behavior, and deployment but are distinguished by differentiating factors. Differentiating factors are additional features and have a meaningful impact on the evaluation criteria.

Baseline Architecture - The best architecture that meets Workshop 1 evaluation criteria and is additive to a Threshold Architecture, such that should budget run over, components could be descoped without compromising our ability to meet the evaluation criteria.

Benchmark Architecture - a reference architecture for the current implementation.

Cloud - Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. (NIST SP 800-145, 2011)

Capability Need - functionalities of the system.

Common ESO data system - a standardized MDPS that services multiple ESO missions.

Common Service - a service that is needed by many projects but may be implemented, deployed independently by each project.

Core Services - collective term relating to NASA Science Mission Directorate (SMD) services provided across missions, science disciplines (e.g., Research and Analysis, Applied Sciences, etc.) and data archives; these are based on common services (e.g., cybersecurity or user registration) and will be standardized and provided across missions, as opposed to dictating how a mission serves its particular communities.

Distributed Active Archive Centers (DAACs) - are NASA data archives that serve different research communities but share common services to standardize NASA data management and archive through the NASA Earth Science Data and Information System (ESDIS).

Data Analysis - is the process of systematically applying statistical and/or logical techniques to describe and illustrate, condense and recap, and evaluate data.

Data Lake - The concept of centralized data storage in support of data-proximate processing where the data is stored and co-located with the processing. An example is the direction that NASA DAACs are going towards as far as moving all their data into Amazon Web Services.

Data Product Level - All definitions are assumed to be consistent with the NASA Data Processing Levels: https://Earthdata.nasa.gov/collaborate/open-data-services-and-software/data-information-policy/data-levels

Desirability - The degree to which the architecture responds to the study objectives.

Earth System Observatory (ESO) - a constellation of satellites will be launched by NASA in the 2020s to observe the Earth System as designated by the National Academies Decadal Survey (NRC, 2018) and classified as “designated observables”.

Evaluation Criteria - are design constraints by which to evaluate different architectures. This term is used in place of “requirements”, which are often traced for data systems from higher-level mission requirements; hence the avoidance of prescribing them for all ESO and future missions.

FAIR Data Principles - Data should be Findable, Accessible, Interoperable, and Reproducible by machines (Wilkinson et al., 2016).

Federated services - a service that is owned and operated by one organization, but is contributed to by many projects.

Full processing service - In addition to generic processing of batch execution of algorithms, full processing service adds workflow orchestration of the algorithms. Full processing service supports both workflow orchestration and execution of the algorithm steps to generate standard data products.

Generic processing service - Services for batch-oriented job execution of algorithms to create data products (i.e. bulk processing), but excludes workflow orchestration of the jobs (i.e. "full processing"). Only including job execution allows each mission to support its own mission-specific needs in the orchestration layer of the algorithms.

Ground Data System - The system responsible for receiving telemetry data from the observatory and providing it to the MDPS, which does the instrument specific processing.

Hybrid Cloud - Infrastructure that is a composition of two or more distinct cloud infrastructures (private, community, or public) that remain unique entities, but are bound together by standardized or proprietary technology that enables data and application portability. (NIST SP 800-145, 2011)

Inclusive - The process and participants welcome participation by and collaboration with diverse people and organizations.

Latency - defined as time between acquisition and data access by the users.

Long Term Data - Data that are curated and intended to outlast the duration of the mission. See also "Short Term Data."

Managed Service - Owned and operated by an organization team, includes a service level agreement, documentation, and is intended to be used by many projects. The intent of managed services is to consolidate and improve cost efficiencies across development, operations, and workforce for the service

Managed Service Teams - Managed service teams are expected to own, develop, operate, maintain, and evolve the managed service over time.

Mission Data Processing System (MDPS) - The set of algorithms, software, compute infrastructure, operational procedures, and documentation to automatically process raw instrument data through to science quality data products. This includes the software tools that support the development of the processing algorithms and validation and analysis of the processed data. MDPS process data for product generation with mechanisms in place to determine if products are ephemeral or worth long-term archive. It is worth noting that product creation could be by the public or ST, and this is not necessarily predefined. Products can include: 1) standard NASA project products that are within scope for a project to generate; 2) on-demand products from NASA-approved algorithms and workflows; 3) on-demand products of customized variants of NASA approved algorithms (e.g., locally calibrated); and 4) non-NASA products (e.g., state level - as opposed to global, applications focused, etc.).

On-prem Computing - Computing infrastructure that physically resides within an enterprise owned data center, server room, etc. On-prem may be referred to as “in-house”. Usually, an organization is fully responsible for procuring, deploying and managing on-prem computing.

Open Science - “a collaborative culture enabled by technology that empowers the open sharing of data, information, and knowledge within the scientific community and the wider public to accelerate scientific research and understanding” (Ramachandran et al., 2021).

Open Source Science - builds on concepts from Open Source Software revolution that expanded participation in developing code and applies it to the scientific process to accelerate discovery through open science from project initiation through implementation.

Open Source Software - The Open Source Initiative (OSI) defines Software to be Open Source if distributed under a license with a set of criteria: 1) license shall not restrict any party from selling or giving away the software, i.e. free redistribution, 2) source code is included with any program or set of programs, 3) license allows for derived works, 4) integrity of author’s source code, 5) license must not discriminate against a groups or persons, 6) license must not discrimination against fields of endeavor, 7) any rights must apply to all whom a program or source is redistributed to, 8) rights attached to the program must not depend on the program's being part of a particular software distribution, 9) license must not place restrictions on other software that is distributed along with, and 10) no provision of the license may be predicated on any individual technology or style of interface. (https://opensource.org/osd)

Permissive software - software that can be copied, modified, redistributed, etc.

Reproducible - The scientific process and results can be reproduced by members of the community.

Scientific Information - publications, data, and software

Services - Services provide a functional capability with well-defined interfaces.

Shared services - a service owned and managed by one organization that is used by many projects.

Short Term Data - Data that are produced incidentally during the mission, but are not necessarily curated or intended to be archived past the end of the mission. See also “Long Term Data.”

System Architecture Working Group (SAWG) - a team of system engineers, data system architects, software engineers, and ESO mission representatives tasked with conducting the ESO open source science data system architecture study. The SAWG is composed of science data system experts who represent the diversity of the data system community and are connected to the end-user science community and the ESO missions.

Steering Committee - the leadership team for the ESO open source science data system architecture study responsible for providing programmatic insights and steering the SAWG to conduct a programmatically relevant study.

Threshold Architecture - The bare minimum architecture needed to meet the evaluation criteria from Workshop 1.

Transparency - Both the scientific process and results are visible, accessible and understandable.

C. Acronyms

A&A	Authentication and Authorization
ACCESS	Advancing Collaborative Connections for Earth System Science
ACCP	Aerosol, Cloud, Convection, and Precipitation
ACF	Analytic Center Frameworks
ACF	Analytics Collaborative Framework
ADE	Application Development Environment
AGU	American Geophysical Union
AI	Artificial Intelligence
AIRS	Atmospheric Infrared Sounder
AIST	Advanced Information Systems Technology
ALIAS	Automated Labeling for Interactive Assisted Segmentation
AMMOS	Advanced Multimission Operations System
AMS	American Meteorological Society
AOS	Atmosphere Observing System
API	Application Programming Interface
ARCO	Analysis-Ready Cloud-Optimized data
ARD	Analysis-Ready Data
ARSET	Applied Remote Sensing Training
ASF	Alaska Satellite Facility
ASI	Agenzia Spaziale Italiana (Italian Space Agency)
ASP	Applied Sciences Program
ATBD	Algorithm Theoretical Basis Documents
ATLAS	Advanced Topographic Laser Altimeter System
AWS	Amazon Web Services
BDD	Block Definition Diagram
cal/val	Calibration and validation
CC	Cultural Complexity
CCAP	Containerized Cloud Algorithm Package
CCSDS	Consultative Community for Science Data Systems
CDR	Critical Design Review
CERES	Cloud and the Earth's Radiant Energy System

CHIME	Canadian Hydrogen Intensity Mapping Experiment
CI/CD	Continuous Integration/Continuous Delivery
CLARREO	CLimate Absolute Radiance REfractivity Observatory
CMR	Common Metadata Repository
CNES	Centre National d'Etudes Spatiales
CoC	Cost Complexity
COG	Cloud-Optimized GeoTIFF
CSA	Canadian Space Agency
CSDO	Chief Science Data Officer
CSP	Cloud Service Provider
D	Desirability
DAAC	Distributed Active Archive Center
DB	Data Base
DEVELOP	Digital Earth Virtual Environment and Learning Outreach Program
DLR	Deutsches Zentrum für Luft- und Raumfahrt e.V. (German Aerospace Center)
DMP	Data Management Plan
DMS	Data Management System
DOI	Digital Object Identifier
DPC	Data Processing Center
DPS	Data Processing Service
DS	Desirability Score
ECOSTRESS	ECOsystem Spaceborne Thermal Radiometer Experiment on Space Station
EDC	Earthdata cloud
EDOS	Earth Orbiting System (EOS) Data and Operations System
EIS	Earth Information System
EMIT	Earth Surface Mineral Dust Source Investigation
EMS	EOSDIS Metrics System
EO	Earth Observation
EOS	Earth Orbiting System
EOSDIS	Earth Observing System Data and Information System
ESA	European Space Agency
ESD	Earth Science Division
ESDIS	Earth Science Data and Information System

ESDIS	Earth System Data and Information System
ESE	Earth System Explorer
ESIP	Earth Science Information Partners
ESO	Earth System Observatory
ESTO	Earth Science Technology Office
ETL	Extraction, Transformation, and Loading
EVI	Earth Venture Instrument
EVM	Earth Venture Missions
FAIR	Findable, Accessible, Inter-operable, Reproducible
FF	Feasibility Factors
FORGE	Future Operationally Resilient Ground Evolution
GDS	Ground Data System
GEDI	Global Ecosystem Dynamics Investigation
GEE	Google Earth Engine
GES-DISC	Goddard Earth Sciences Data and Information Service Center
GFO	Gravity Recovery and Climate Experiment Follow-On
GFZ	Geoforschungszentrum
GIBS	Global Imagery Browse Services
GIS	Geographic Information System
GNSS	Global Navigation Satellite System
GNU	GNU's Not Unix
GPU	Graphics Processing Unit
GRACE	Gravity Recovery and Climate Experiment
GRACE-FO	Gravity Recovery and Climate Experiment Follow-On
GSFC	Goddard Space Flight Center
GUI	Graphical User Interface
HARP	Hyper Angular Rainbow Polarimeter
HCSA	Hybrid Computing and Storage Architectures
HEC	High-end Computing
HECC	High-end Computing Capability
HOSC	Huntsville Operations Support Center
HPC	High Performance Computing
HQ	Headquarters

HyP3	Hybrid Pluggable Processing Pipeline
IDE	Integrated Development Environment
IMGEOS	Integrated Multimission Ground Segment for Earth Observing Satellites
InSAR	Interferometric Synthetic Aperture Radar
IP	Internet Protocol
IRAD	Internal Research and Development
ISCE	InSAR Scientific Computing Environment
ISRO	Indian Space Research Organisation
ISS	International Space Station
ITAR	International Traffic in Arms Regulations
JAXA	Japan Aerospace Exploration Agency
JDK	Java Development Kit (Open JDK)
JEM-EF	Japanese External Module- Exposed Facility
JPL	Jet Propulsion Laboratory
JPSS	Joint Polar Satellite System
JSON	JavaScript Object Notation
KVM	Kernel based Virtual Machine
L#	Data Product Level # as defined by https://Earthdata.nasa.gov/collaborate/open-data-services-and-software/data-information-policy/data-levels
LAADS	Land And Atmosphere Distribution System
LANCE	Land Atmosphere Near real-time Capability for EOS
LC	LandSat Cloud
LEO	Low Earth Orbit
LIDAR	Light Detection and Ranging
LOE	Level of Effort
LRZ	Leibniz Supercomputing Centre
LSST	Legacy Survey of Space and Time
LSTM	Long Short Term Memory
MAAP	Multimission Algorithm and Analysis Platform
MADS	Mission Access Data System
MAIA	Multi-Angle Imager for Aerosols
MC	Mass Change

MCP	Microsoft Cloud Platform
MD	Mission Development
MDPAF	Mission Data Processing Application Framework
MDPS	Mission Data Processing System
MF	Maturity Factors
MMO	Multimission Organization
MMT	Metadata Management Tool
MOC	Mission Operations Center
MODAPS	MODIS Adaptive Processing System
MODIS	MODerate resolution Imaging Spectrometer
MODSIM	Modeling and Simulation
MSFC	Marshall Space Flight Center
NASA	National Aeronautics and Space Administration
NCAP	NESDIS Cloud Archive Program
NCCF	NESDIS Common Cloud Framework
NCEI	National Centers for Environmental Information
NCIS	Cloud-sandbox Infrastructure Services
NESDIS	National Environmental Satellite, Data, and Information Service
NEX	NASA Earth Exchange
NGAP	Next Generation Application Platform
NGE	NESDIS Ground Enterprise
NISAR	NASA-ISRO Synthetic Aperture Radar (SAR)
NIST	National Institute of Standards and Technology
NOAA	National Oceanic and Atmospheric Administration
NOS	New Observing Systems
NPR	NASA Procedural Requirements
NRT	Near Real Time
NSF	National Science Foundation
OBPG	Ocean Biology Processing Group
OCI	Ocean Color Instrument
OCO	Orbiting Carbon Observatory
OGC	Open GeoSpatial Consortium
ORNL	Oak Ridge National Laboratory

OSI	Open Source Initiative
OSS	Open Source Science
OVF	Open Virtualization Format
PACE	Plankton, Aerosol, ocean Ecosystem
PAL	Product Algorithm Laboratory
PB	Petabytes
PCF	Programmatic Complexity Factor
PDR	Preliminary Design Review
PEST	Policy, Economics, Sociocultural Factors, an Technologies/Tools
PGE	Program Generated Executables
PI	Principal Investigator
PiaB	Pipeline in a Box
PO.DAAC	Physical Oceanography DAAC
POR	Program of Record
PPM	Part Per Million
R&A	Research and Analysis
RC	Resource Complexity
RFI	Request For Information
RGT	Reference Ground Track
ROSES	Research Opportunities in Space and Earth Sciences
RTC	Radiometric-Terrain Correction (SAR Data product)
S3	Simple Storage Service (associated with AWS)
SAR	Synthetic Aperture Radar
SAT	Science Activity Timeline
SAWG	System Architecture Working Group
SBG	Surface Biology and Geology
SC	Steering Committee (Executive Summary)
SC	Schedule Complexity
SDAP	Science Data Analytics Platform
SDC	Surface Deformation and Change
SDS	Science Data System
SDST	Science Data Support Team
SIPS	Science Investigator-led Data System

SLC	Single Look Complex (SAR data product)
SMCE	Science Managed Cloud Environment
SMD	Science Mission Directorate
SME	Subject Matter Expert
SNPP	Suomi National Polar-orbiting Partnership
SOC	Science Operations Center
SPD	Science Mission Directorate Policy Document
SPS	Science Planning System
SQL	Structured Query Language
SQS	Simple Queue Service
ST	Science Teams
STAC	SpatioTemporal Asset Catalog
STSci	Space Telescope Science Institute
SWOT	Strength, Weakness, Opportunity, Threat
SWOT	Surface Water and Ocean Topography
TB	Terabyte
TC	technical complexity
TCO	Total Cost of Ownership
TESS	Transiting Exoplanet Survey Satellite
TIR	Thermal Infrared
TOPS	Transform to OPen Science
TPU	Tensor Processing Units
TRISHNA	Thermal infraRed Imaging Satellite for High-resolution Natural resource Assessment
TRL	Technology Readiness Level
TROPICS	Time-Resolved Observations of Precipitation structure and storm Intensity with a Constellation of Smallsats
UKSA	United Kingdom Space Agency
USGS	United States Geological Survey
USML	United States Munition List
V&V	Validation and Verification
VIIRS	Visible Infrared Imaging Radiometer Suite
VM	Virtual Machine

VNIR	Visible and Near-Infrared
VPC	Virtual Private Cloud
VSWIR	Visible to Short-Wave Infrared
W	Weights
WBS	Work Breakdown Structure
WPS	Web Processing Service
XML	eXtensible Markup Language
XSEDE	eXtreme Science and Engineering Discovery Environment
YARN	Yet Another Resource Negotiator
YOOS	Year of Open Science