36DOT/FAA/TC-693KA8-19-D-00003/TO 37

Federal Aviation Administration
William J. Hughes Technical Center
Aviation Research Division
Atlantic City International Airport
New Jersey 08405

Part 2 Cybersecurity Data Science Aviation (CSDS) Architecture Framework (AAF) - Technical Definition

August 1, 2025

Version 1.3

U.S. Department of Transportation
Federal Aviation Administration

NOTICE

This document is disseminated under the sponsorship of the U.S. Department of Transportation in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof. The U.S. Government does not endorse products or manufacturers. Trade or manufacturers' names appear herein solely because they are considered essential to the objective of this report. The findings and conclusions in this report are those of the author(s) and do not necessarily represent the views of the funding agency. This document does not constitute FAA policy. Consult the FAA sponsoring organization listed on the Technical Documentation page as to its use.

This report may be made available upon request to the FAA Aviation Research Division

Technical Report Documentation Page

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

1. Report No. DOT/FAA/TC-693KA8-23-D-00003/TO 37 – AAF 3 Part 2	2. Government Accession No.	3. Recipient's Catalog No.

4. Title and Subtitle Title of Report: Cyber Security Data Science Aviation Architecture Framework Subtitle of Report: Technical Definition	5. Report Date August 2025 6. Performing Organization Code
7. Author(s) Center for Aerospace Resilient Systems (CARS) Dan Diessner Dr. David Harvie Isidore Venetos	8. Performing Organization Report No.
9. Performing Organization Name and Address Embry-Riddle Aeronautical University Center for Aerospace Resilient Systems (CARS) 1 Aerospace Blvd. Daytona Beach, FL 32114-3900	10. Work Unit No. (TRAIS) 11. Contract or Grant No. DOT/FAA/TC-693KA8-19-D-0003/TO 37
12. Sponsoring Agency Name and Address The William J. Hughes Technical Center Aviation Research Division ANG-E271 Federal Aviation Administration - FAA. Atlantic City International Airport, Egg Harbor Township, NJ 08405	13. Type of Report and Period Covered 14. Sponsoring Agency Code ANG-E271

15. Supplementary Notes

16. Abstract

This is the second part of a series of four (4) core documents to provide an overview of the top-down output from the FAA Cybersecurity Data Science Aviation Architecture Framework (CSDS AAF) research program. The intent of this technical specification document is to provide an ontology for the CSDS AAF. It also provides a narrative to describe and explain all of the key AAF components and functions, coupled with diagrams to illustrate the overall AAF structure. The four (4) core CSDS AAF documents are:

- **Part 1 CSDS AAF – Overview & Value Proposition**: The primary purpose is to communicate aviation stakeholders the vision and potential value of the FAA CSDS research and generally how it could potentially be leveraged to address key aviation cybersecurity challenges.
- **Part 2 CSDS AAF – Technical Definition**: As an ontology for the CSDS AAF, this document provides a narrative to describe and explain all of the key AAF components and functions, coupled with diagrams to illustrate the overall AAF structure.
- **Part 3 CSDS AAF – Implementation Guidance**: This document provides guidance for the implementation of the CSDS AAF, which is defined in the AAF Technical Definition.
- **Part 4 CSDS AAF – Glossary & Acronyms:** This document provides the Glossary and Acronym material for all parts of the CSDS AAF documentation.

17. Key Words Cybersecurity Data Science (CSDS), Aviation Architecture Framework (AAF), Cyber Analytical Capability (CAC), Data Sphere	18. Distribution Statement This report may be made available upon request to the FAA Aviation Research Division.

19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages	22. Price

Contents

Figures

Tables

Acronyms

See Part 4: Cybersecurity Data Science (CSDS) Aviation Architecture Framework (AAF) – Glossary & Acronyms.

Executive summary

A critical challenge in cybersecurity is determining if a cyber incident has or is happening. Data science promises to more quickly and more effectively find anomalous data that could indicate a cyber incident. The Cybersecurity Data Science (CSDS) Aviation Architecture Framework (AAF) seeks to apply data science to the aviation ecosystem which involves both information technology (IT) and operation technology (OT) with various stakeholders such as airlines, airports, and original equipment manufacturers (OEMs). This document defines the conceptual elements and taxonomy of the CSDS AAF. The CSDS AAF Systems Architecture applies this framework in different Environments of Operation and involves Stakeholder Data-Stores, Cyber Analytical Capability (CAC), and Interconnected Individual Systems (IIS). The framework also introduces the CSDS AAF Data Life Cycle which consists of Acquire, Pre-Analyzed, Collect, Advanced Analytics, and Information Sharing. A critical component in this data perspective is collecting the appropriate data which is described using the Data Sphere concept. A chief goal in this document is to help inform future standards activities by providing and maturing the CSDS Aviation Architecture Framework.

1 Introduction

This is the second part of a series of four (4) documents to provide an overview of the top-down output from the FAA Cybersecurity Data Science Aviation Architecture Framework (CSDS AAF) research program. The intent of this technical definition is to provide an ontology for the CSDS AAF. It also provides a narrative to describe and explain all of the key AAF components and functions, coupled with diagrams to illustrate the overall AAF structure. The four (4) core CSDS AAF documents are:

- **CSDS AAF – Part 1: Overview & Value Proposition**: The primary purpose is to communicate aviation stakeholders the vision and potential value of the FAA CSDS research and generally how it could potentially be leveraged to address key aviation cybersecurity challenges (Embry-Riddle Aeronautical University Center for Aerospace Resilient Systems, 2025).
- **CSDS AAF – Part 2: Technical Definition**: As an ontology for the CSDS AAF, this document provides a narrative to describe and explain all the key AAF components and functions, coupled with diagrams to illustrate the overall AAF structure (Embry-Riddle Aeronautical University Center for Aerospace Resilient Systems, 2025).
- **CSDS AAF – Part 3: Implementation Guidance**: This document provides guidance for the implementation of the CSDS AAF, which is defined in the AAF Technical Definition (Embry-Riddle Aeronautical University Center for Aerospace Resilient Systems, 2025).
- **CSDS AAF – Part 4: Glossary & Acronyms:** This document provides the Glossary and Acronym material for all parts of the CSDS AAF documentation (Embry-Riddle Aeronautical University Center for Aerospace Resilient Systems, 2025).

1.1 Background of Data Science

Data is changing everything, as a result, there is an emergence of a new field – Data Science – that focuses on the processes and systems enabling the extraction of knowledge or insights from data in various forms, either structured or unstructured. In practice, Data Science has evolved as an interdisciplinary field that integrates approaches from data analysis fields such as Statistics, Data Mining, and Predictive Analytics, drawing on diverse observational domains (Rutenbar, 2016).

The term Data Science was coined in 2008 by D.J. Patil, and Jeff Hammerbacher, the respective leads of data and analytics efforts at LinkedIn and Facebook at the time (Cao, 2019) (Patil, 2012). As the world grows increasingly connected, so does the desire for new and even old technologies to be integrated into large operational networks. With so much data generated, it becomes increasingly difficult for data analysts to sift through that data to find useful, actionable information.

Data Science aims to solve this problem by leveraging considerable computing resources and automated data processing techniques to reduce the data into something more tangible and meaningful for data analysts to work on. Thus, CSDS is the application of data science to cybersecurity, i.e., the multi-disciplinary process to generate actionable cyber-analytical insights from large and ever-increasing volumes of cybersecurity data where machine learning (ML), a core part of AI, can play a vital role in discovering the insights from data (IBM Cloud Education, 15).

CSDS offers a scientific approach to identifying hostile attacks on digital infrastructures. It uses the data-focused approach that applies ML techniques to identify potential threats. Anomaly detection is a major feature that ML brings to cybersecurity. Attacks are often committed by malicious software that behaves differently from the norm when network traffic or software processes are monitored and compared. Creating a machine learning model to detect an anomaly is a great way to use data science techniques to support cybersecurity efforts (Torres, 2021).

Data Science, and therefore CSDS, is still in its infancy stages and proves to be a confusing industry landscape to navigate for organizations and academic institutions. There is not yet a consensus on what constitutes data science, while knowledge frameworks and standards bodies are still in the process of being formed (Hamutcu, 2020).

As discussed earlier, Data Science starts with defining a problem (see Figure 1 below). For this project the problem is simply how to detect and analyze cyber events in the aviation ecosystem. And now, via the AAF, there is a defined structured engineering approach for breaking down the aviation half of this problem statement into manageable chunks by decomposing aviation into specific use cases around aviation Environments of Operation. The next step is to apply the Data Science methodology to these use cases in the context of cybersecurity half of our problem statement (i.e., Cybersecurity Data Science).

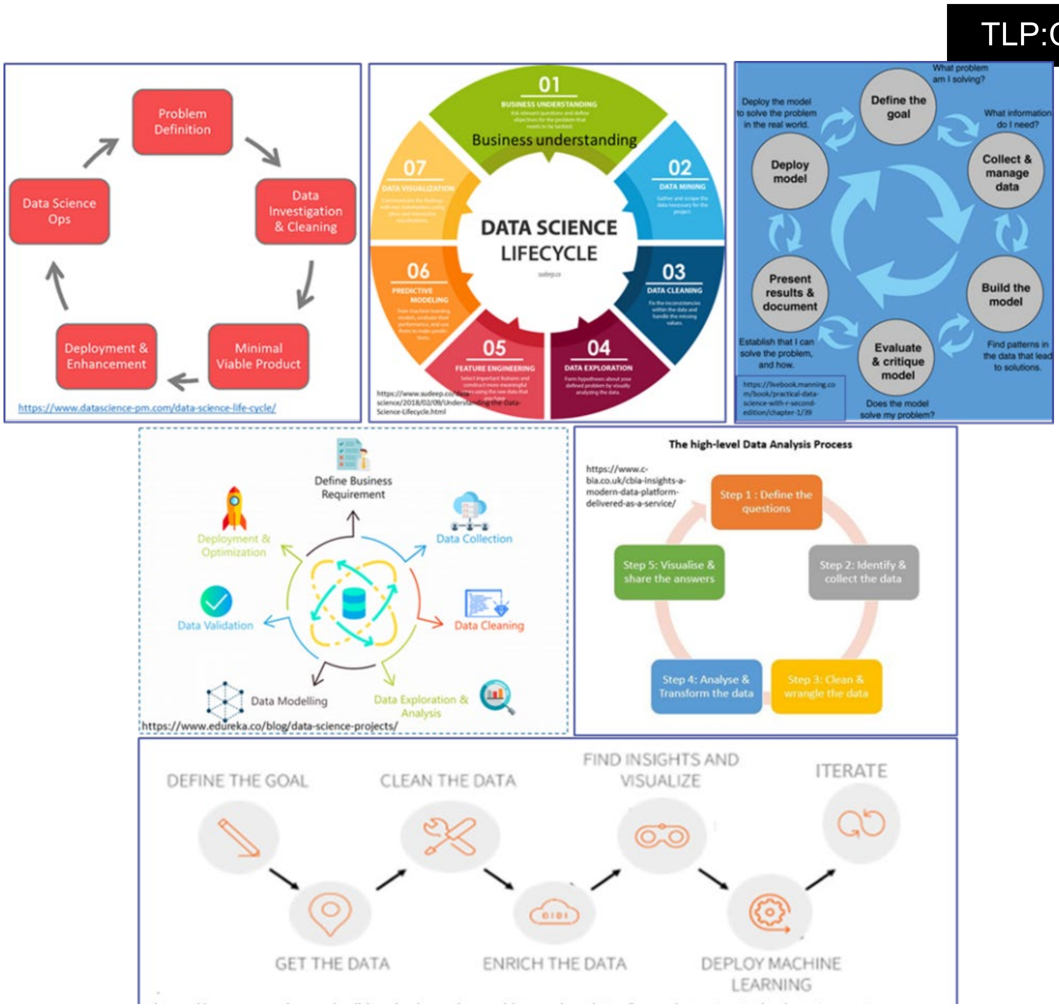

Problem Definition
Data Investigation & Cleaning
Minimal Viable Product
Deployment & Enhancement
Data Science Ops
https://www.datascience-pm.com/data-science-life-cycle/
01
BUSINESS UNDERSTANDING
Business understanding
02
DATA MINING
03
DATA CLEANING
04
DATA EXPLORATION
05
FEATURE ENGINEERING
06
PREDICTIVE MODELING
07
DATA VISUALIZATION
DATA SCIENCE LIFECYCLE
https://www.sudeep.co/data-science/2018/02/09/Understanding-the-Data-Science-Lifecycle.html
What problem am I solving?
Define the goal
What information do I need?
Collect & manage data
Build the model
Find patterns in the data that lead to solutions.
Evaluate & critique model
Does the model solve my problem?
Establish that I can solve the problem, and how.
Present results & document
Deploy model
Deploy the model to solve the problem in the real world.
https://livebook.manning.com/book/practical-data-science-with-r-second-edition/chapter-1/39
Define Business Requirement
Data Collection
Data Cleaning
Data Exploration & Analysis
Data Modelling
Data Validation
Deployment & Optimization
https://www.edureka.co/blog/data-science-projects/
The high-level Data Analysis Process
https://www.c-bia.co.uk/cbia-insights-a-modern-data-platform-delivered-as-a-service/
Step 1 : Define the questions
Step 2: Identify & collect the data
Step 3: Clean & wrangle the data
Step 4: Analyse & Transform the data
Step 5: Visualise & share the answers
DEFINE THE GOAL
GET THE DATA
CLEAN THE DATA
ENRICH THE DATA
FIND INSIGHTS AND VISUALIZE
DEPLOY MACHINE LEARNING
ITERATE

Figure 1. Various Data Science Lifecycle Representations

From the Data Science process model perspective, once the problem is defined, the next step (sometimes even identified as the initial step) is to “get the data”. Generally, Data Science starts with the assumption that the data exists somewhere out there, and as a Data Scientist, one simply needs to go and get it. By inserting the Data Science concept into the overall business and systems engineering processes, the AAF data lifecycle process can begin with more emphasis on the steps within the Environments of Operation (EOOs) where the systems acquire the data and decide what is to be collected for downstream use by the various functional elements like the Cyber Analytical Capability (CAC) within the Stakeholder’s business.

With this in mind, more practical Data Science applications for addressing systems engineering challenges within an operational system-of-systems, like the aviation ecosystem, can be considered. Data Science may support many areas within the business and engineering processes of our large and constantly evolving aviation ecosystem. For example, many different functional disciplines across the aviation ecosystem use advanced analytics and are pursuing the use of Data Science to improve their functional areas, such as reliability, safety, and cybersecurity. Most Stakeholders within the aviation ecosystem have dedicated functional organizations for each of these. For the purpose of establishing a generic AAF model and practical simplification, all of the core cybersecurity functional organizational capabilities of a given Stakeholder are placed into the CAC. This includes the cybersecurity data scientist, cybersecurity requirements engineering functions, as well as the day-to-day cyber analysts and incident response team functions. Just as they are focused on their activity of providing the cybersecurity functional capability across the entire Stakeholder's business, so are each of the other functional organizations (i.e., Reliability, Safety, …). Each of these functional organizations needs data from across the Stakeholder's same EOO.

Each functional organization defines requirements for what type of data they need from the various environments, and thus what data the specific systems need to acquire, pre-analyze, and collect for the Stakeholder's business. While each may be working to leverage Data Science methodologies across the entire business, the specific CSDS process efforts are associated with the CAC. Thus, in our model, the primary Data Science activities that map to the above data science models begin at the Curate Phase, where data is primarily extracted from the Stakeholder's Environments of Operation and stored within the CAC. The CAC members also continually define improvements that feed new CSDS engineer requirements back into the design and modification of the Stakeholder's Environments of Operation.

The CSDS AAF study effort leveraged the various existing data science knowledgebases to define a taxonomy and reference model that provides the best fit for aviation cybersecurity. The selection of terms and definitions was based on what is readily understood by the aviation community and the various rapidly evolving data science resources. As future phases of the project progress, processes will be developed based on what provides the most adequate and reasonable adoption given the unique and specialized nature of aviation-related Environments of Operation and associated systems. While it is possible to adopt processes from industry leaders in the field of Data Science, it is essential to understand that most of these industries are targeted towards a traditional IT environment, whereas the Aviation Industry has a mix of IT and various types of OT that need to be considered.

2 Cybersecurity Data Science Aviation Architecture Framework (CSDS AAF)

The aviation ecosystem comprises a multitude of stakeholders, each of which has unique business cases to support their slice of civil air transportation. To fulfill their interests and responsibilities, stakeholders' control and monitor merged enterprise information / operational technology (IT/OT) systems that support these cases during daily operations and act as backbone technologies throughout the aviation ecosystem. These merged IT/OT systems contain large amounts of raw data in the form of network traffic, system logs, and operational logs that the CSDS AAF seeks to harness, to provide next-generation cyber-analytical capabilities and cyber-threat intelligence for aviation.

2.1 CSDS AAF Conceptual Elements

"Cybersecurity Data Science (CSDS) encompasses the rapidly growing practice of applying data science to prevent, detect, and remediate cybersecurity threats. CSDS methods emerge from applying data analytics and machine learning to challenges associated with security assurance" (Mongeau, 2021).

This work has organized the CSDS AAF Conceptual Elements into three (3) categories: Data Acquire Elements, Data Categories, and Analytical Functional Elements.

Data Acquire Elements refer to both the Individual Interconnected System (IIS) components essential for acquiring data so that it can be used in the CSDS process and external sources of data that can feed into the CSDS process.

The data acquisition sensors monitor data generated on systems and networks and evaluate data relevancy. Once data is determined as relevant, it is encoded and often stored in Local Storage. When the relevant data is ready to be pulled, a Data Egress Point (DEP) is required to "offload" the data for analysis. This can be done through removable storage devices or over a networking interface.

In addition, there are a number of External Systems Information Feeds generated outside the EOO that be beneficial for use within the CSDS.

Data Categories include security data as the primary category as well as Network, Systems, and Application Data, as discussed earlier. Security data is typically a subset of network, systems, and application data that is cybersecurity relevant. For example, a login attempt into a system on the network is considered security-related, but it can be a network, system, or application-type log. Data from all external sources are collectively categorized as external data.

Information generated outside the EOO but used within the CSDS process is categorized as external information. Examples of external information include threat intelligence and other third-party data. Threat intelligence can include disclosed vulnerabilities and security bulletins from both private and public sources.

Sources of other third-party information can include, but are not limited to, suppliers, vendors, and sub-contractors. Some examples of this information can include quality assurance (QA) and maintenance information on third-party provided parts and components. Software bill of materials (SBOMs) for third-party provided software can also be classified as other third-party information.

Analytical Functional Elements interact with and act on the data. The first Analytical Functional Element is the collectors that store the data. Analyzers and Cyber Tool Sets are additional Analytical Functional Elements that analyze the data stored in the Collectors. The CAC is an organization that performs data analysis. Figure 2. CSDS Conceptual Elements shows the relationship between the three CSDS Conceptual Elements.

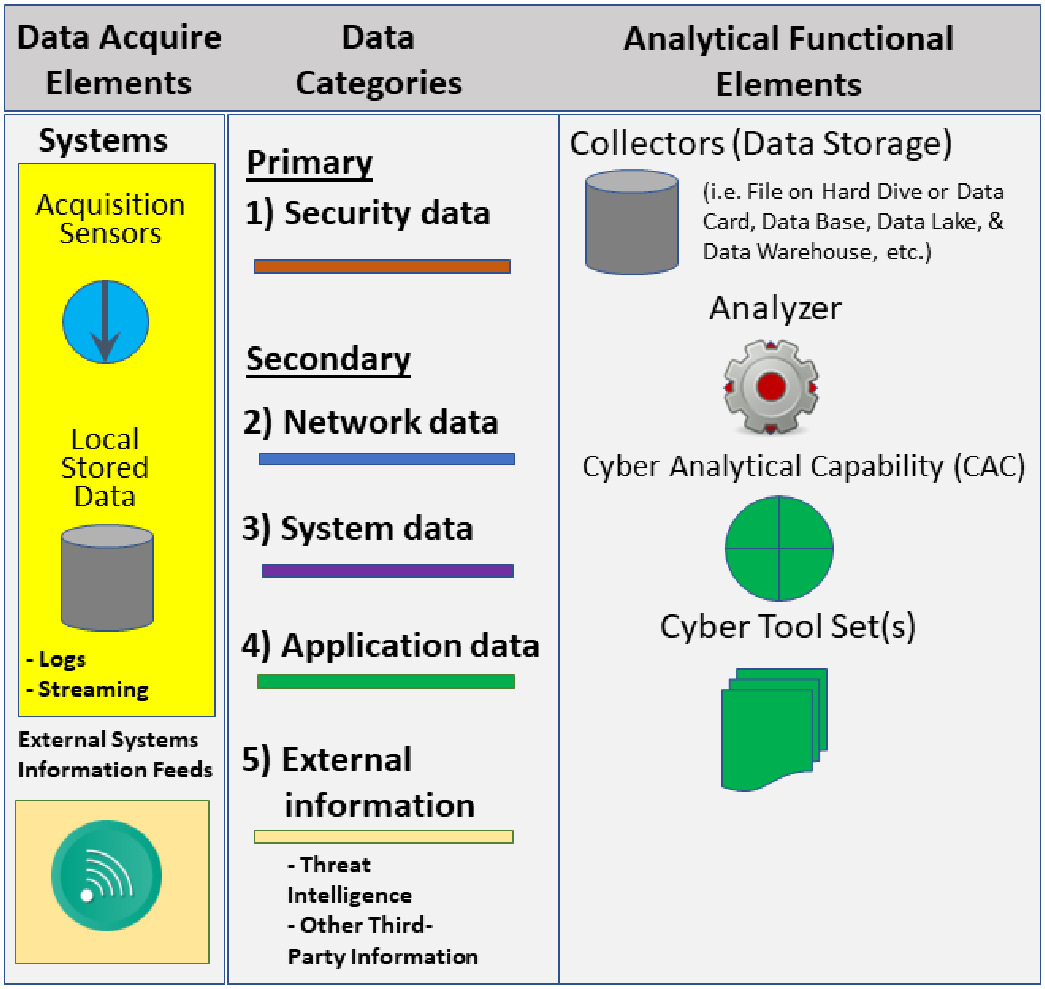


Figure 2. CSDS Conceptual Elements

2.1.1 CSDS AAF Data Life Cycle

The CSDS AAF Data Life Cycle consists of six (6) phases, as illustrated in Figure 3. Businesses will integrate/configure selected IIS with Data Acquisition Sensors to acquire and pre-analyze data. During the Acquire Phase, these sensors monitor and capture data from hardware components or software processes of the IIS across the EOO. The Pre-Analyze Phase occurs within the Data Acquisition Sensor, involving software-based logic that evaluates whether the data being acquired is relevant based on pre-defined data collection rules. The Pre-Analyze Phase includes automation to filter and perform data feature extraction, as well as tagging the data with additional meta data and properties to improve the later retrieval and analysis. The Collect Phase is concerned with storing relevant data on various non-volatile memory storage devices within a business' EOO. This can also include Cloud storage implementations. The collection of these storage devices makes up the Data-Store. A major objective of the Collection Phase is to ensure local storage devices are installed and configured correctly to handle data velocity requirements (i.e., maximum read/write speeds compared to the amount of data being generated by the Data Acquisition Sensors), and remote storage (e.g., Cloud storage services) is appropriately connected and secured. The Curate Phase aims to extract cyber-relevant data from the Data-Store and create data sets and models from it, depending on specific needs and interests.

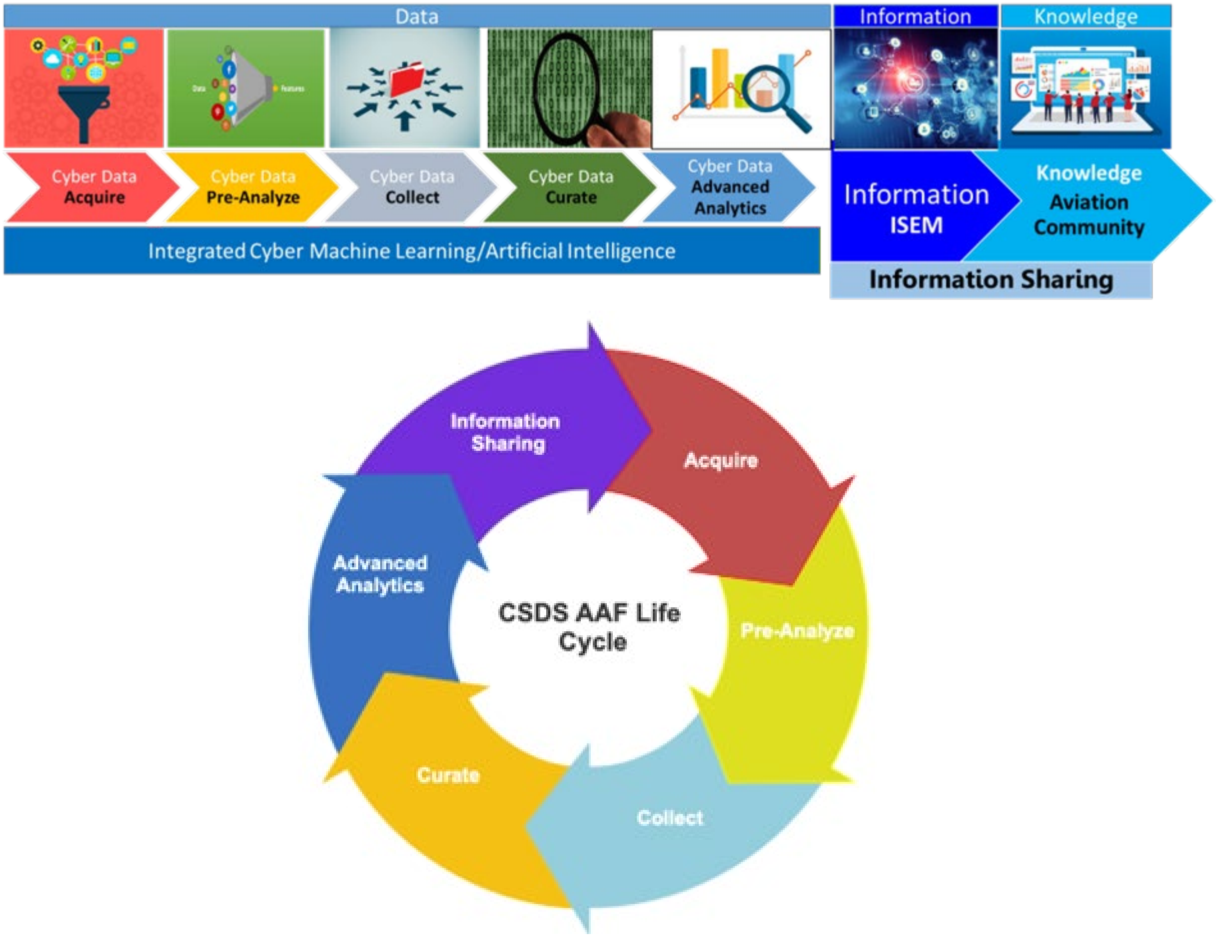


Figure 3. CSDS AAF Data Life Cycle

The Advanced Analytics Phase takes curated data and seeks to produce meaningful artifacts that include insights and actionable information. In terms of insights, analytical toolsets provide enhanced capabilities for human analysts to visualize and interpret data in various ways and may assist them in discovering valuable hidden information that can be provided to the business. These insights could include system user habits, network traffic patterns, data volume over time, etc. In terms of actionable information, AI/ML solutions act as an expert advisory system that provides suggestions to the business. An example of actionable information would be the recommendation of disabling specific network ports on an airline reservation system network as they may not be currently in use by a system. Disabling the port will reduce the risk of potential exploitation by an adversary. These artifacts are produced in various formats, including tabular information in the form of Comma-Separated Values (CSV) and Excel files, visual information in the form of PDF documents, or in the special case of AI automation, text-based alerts, and recommendation notifications to pre-authorized subscribers. These types of notifications are extremely useful in real-time analytics, where an action is required to be taken in a relatively short period of time. For example, if an AI/ML automation system detects a possible intrusion, an automated text message would be sent to a System Administrator to take proper action.

The Advanced Analytics Phase attempts to make sense of the data sets and data models to provide meaningful information and insights. From a cybersecurity perspective, this could be the detection of an ongoing cyberattack, or the presence of malicious software installed on a network, providing risk assessment and root cause analysis to determine proper mitigation strategies after a cyber incident has happened. The Information Sharing Phase takes results from the Advanced Analytics Phase and prepares them for internal Stakeholder teams notifications and disclosure to other stakeholders within the aviation community. This process ensures that confidential or sensitive domain stakeholder data such as Personally Identifiable Information (PII) of customers, employee data, International Traffic in Arms Regulations (ITAR)/ Export Administration Regulations (EAR), For Official Use Only (FOUO), or other information that may provide a competitive advantage to other stakeholders classified data is not inappropriately disclosed.

Acquire Phase Objectives

- Execute
    - Capture the correct data from the IIS that will be useful for CSDS efforts.
- Define and Design
    - Identify all IIS from which data needs to be acquired.
    - Define what data the IIS will acquire.
    - Integrate and configure Data Acquisition Sensors into the systems.
    - Ensure the Data Acquisition Sensors can successfully handle the volume of data being generated without negatively impacting the performance of the IIS.

Pre-Analyze Phase Objectives

- Execute
    - Evaluate the acquired data and select what will be collected to meet defined requirements and ignore the rest of the data.
- Define and Design

    - Develop the evaluation functionality that determines what should be collected and what should be ignored by the Data Acquisition Sensor (e.g., reducing terabytes of data to gigabytes of data).
    - Develop the algorithms to extract the features and data fields from the acquired data.
    - Provide mechanisms and processes for pre-analyzers to be easily re-configurable to support the evolution of CSDS needs and requirements as they change over time.

Collect Phase Objectives

- Execute
    - Gather and store all the pre-analyzed data across the Environment of Operation.
- Define and Design
    - Integrate and configure storage devices that may be distributed across the business' Environment of Operation or located remotely (e.g., Cloud storage) to collect data from the systems' Data Acquisition Sensors
        - Local Storage Devices – Storage within the IIS themselves (e.g., Local Hard Drive, Remove Storage Devices, or non-volatile memory).
        - On-Premise Storage Devices – Storage physically located in the same facilities as the IIS (e.g., Networked File Systems).
        - Remote Storage Devices – Storage located in an external geographical location (e.g., Cloud Storage).
    - Develop and implement data retention and redundancy mechanisms to ensure data is preserved (not lost or corrupted) for the data retention period, determined by technical and governance requirements of the business (i.e., data retention requirements).
    - Develop and implement mechanisms to allow business data users (e.g., CAC entities) to connect and extract data from the Data-Store when needed.
        - Manual Physical Extraction – Physically connect to and request data sporadically from the Data-Store.

        - Manual Remote Extraction – Remotely connect and request data sporadically from the Data-Store.
        - Automatic Remote Extraction (Bilateral) – Automatically connect and request data at scheduled intervals from the Data-Store.
        - Remote Near Real-Time Extraction (Unilateral) – Near Real-Time streaming of the collected data (directly from the IIS instead of the Data-Store).

Curate Phase Objectives

- Execute
    - Implement data extraction methods to get the available Desired Data from the Data-Store.
    - Perform Data Pre-Processing (Anunaya, 2021).
        - Data Cleaning
        - Data Transformation
        - Data Integration
        - Data Reduction / Dimension Reduction
    - Perform Data Maintenance – Ensure data is organized and preserved until it is no longer needed.
    - Perform Data Validation – Ensure data is correct.
    - Perform Data Verification – Ensure data is accurate.
    - Store newly curated data into the CAC Data Warehouse for future access.
- Define and Design
    - Identify cyber-relevant data based on specific CSDS Use Case efforts.
    - Develop requirements that define the Desired Data.
    - Identify Data-Store element locations that do/will have the Desired Data considered to be relevant.
    - Define data extraction methods to get the data from the Data-Store.

Advanced Analytics Phase Objectives

- Execute
    - Use various data analytical methods (including Data Science AI/ML algorithms) to perform advanced analytics on the available data pulled from the Data-Store and curated.
    - Generate analytical reports and visualizations by both AI Automation and Human Analysts.
    - Generate actionable insights, advisory, and recommendations by both AI Automation and Human Analysts.
- Define and Design
    - Apply various Data Science Algorithms (including AI/ML) to the curated data to produce data models that can be used for advanced analytics.

Information Sharing Phase Objectives

- Execute
    - Internal: CAC sharing information within its own Domain Stakeholder's business to take appropriate Incident Response Team actions, and make the necessary requirements, processes, and systems changes.
    - External: CAC sharing information with other Domain Stakeholders and Multi-Domain CACs for the benefit of the aviation community.
        - Redact the sensitive information or appropriately mark the artifacts.
        - Convert artifact format/structure to conform to that of the Shareable Artifact Template.
        - Approve Shareable Artifact for Distribution.
        - Publish Shareable Artifact.
- Define and Design
    - Identify types of analytical information and other CSDS bi-products appropriate to be shared under defined governance, i.e., considering both voluntary sharing vs

mandatory government reporting, and necessary sensitive data handling requirements & processes.

- Define the correct Shareable Artifact Templates to use.
- Integrate required Information Messaging Exchange System interfaces for publishing artifacts to the intended subscribers.

2.2 CSDS AAF Taxonomy & Reference Model

This section provides an overview of the aviation ecosystem today and seeks to define a taxonomy and reference model to be used throughout the rest of this report. It is important to start with understanding the highly interconnected nature of the CSDS AAF as applied to the aviation ecosystem. The high-level CSDS AAF Systems Architecture (see Figure 4) illustrates this interconnectedness across and within multiple domains. The CSDS AAF Systems Architecture will be more fully described in Section 2.3 of this report.

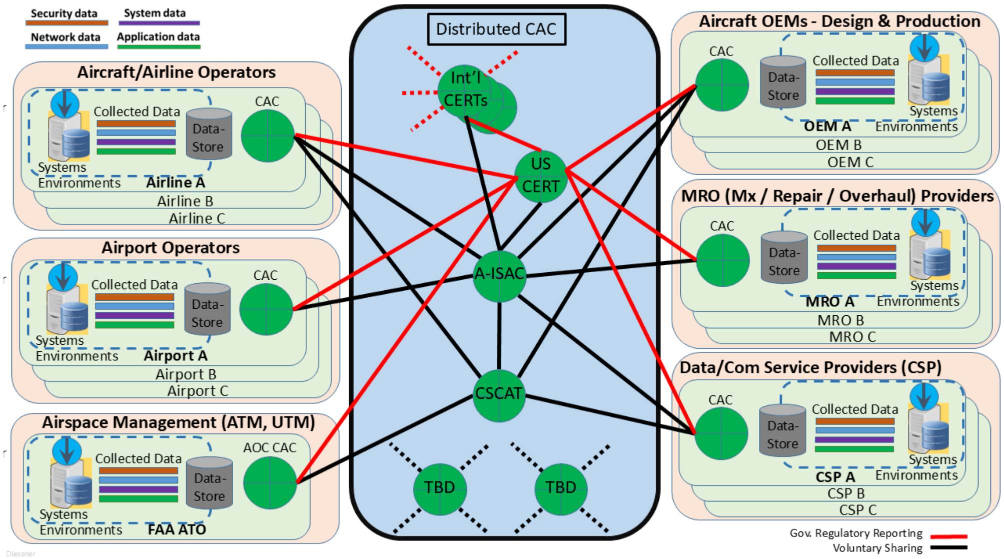


Figure 4. CSDS AAF Systems Architecture

The Aviation Ecosystem Reference Frame Model (see Figure 5) shows how data can be classified and organized for the purposes of CSDS. This classification and organization are important to define standards and guidance that the industry will use. In turn, this will help the industry to define best practices and design solutions for cost-effective and secure systems enabling easier sharing of cybersecurity data to protect the aviation ecosystem, as well as support each Stakeholder in better cyber-protecting the crown jewels of their businesses. The taxonomy and reference model seek to provide a common understanding of:

- Where does data originate/live?
- How is data organized?
- Who owns the data vs. who manages the data?
- Who are the key actors?
- What are the primary drivers for CSDS development?
- How does one determine what the common reusable components are?

2.2.1 CSDS AAF Reference Model

From a CSDS AAF perspective, the aviation ecosystem can be viewed using the reference architecture shown below (Figure 5). The reference architecture illustrates a layered hierarchy in which the aviation ecosystem is considered the top entity of the CSDS AAF (i.e., the aviation ecosystem encompasses the entire CSDS project scope). The aviation ecosystem is composed of multiple Aviation Domains, which can be considered the major branches that make up civil aviation infrastructure (i.e., Airlines; Airports; Maintenance, Repair and Overhaul Providers; etc.). Each Aviation Domain contains the Domain Stakeholders, the actual companies/organizations of the Aviation Domain. Each Domain Stakeholder owns and maintains multiple EOOs. Each EOO is composed of several IIS.

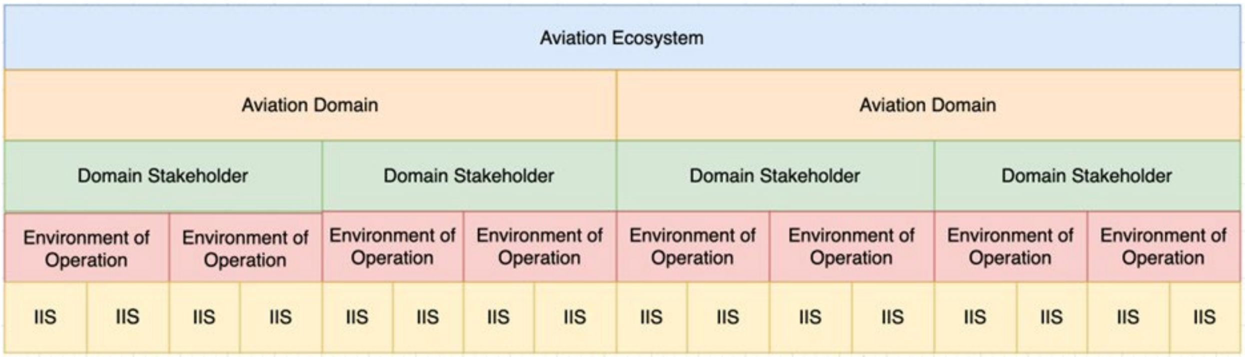


Figure 5. Top Level Aviation Ecosystem Reference Model from a CSDS Perspective for the AAF

Aviation Domains are the major, functional segments based on the typical aviation structure of core business operations and primary roles/responsibilities of stakeholders within the aviation ecosystem. Aviation Domains also typically share common regulatory compliance and oversight as established by the FAA and other international regulatory bodies. The CSDS AAF identifies the six Aviation Domains of interest as follows:

1. Aircraft OEMs / Supply Chains – Design & Production
2. Aircraft/Airline Operators
3. Maintenance, Repair, and Overhaul (MRO) Providers
4. Data / Communication Service Providers (CSPs)
5. Airspace Management / ANSP (ATM, UTM)
6. Airport Operators

Domain Stakeholder is a stakeholder of a specific Aviation Domain. A stakeholder is defined as an organization having a right, share, claim, or interest in the aviation system or in its possession of characteristics that meet its needs and expectations, as defined in DO-391 (ISO, 2015). These include but are not limited to the government, municipal and privately-owned organizations providing the primary business/service of a specific Domain in the aviation ecosystem. Examples of Domain Stakeholders include:

1. Airline Operator – Part 121
2. Aircraft OEM – Part 21
3. Airport Operator
4. MRO – Part 145, CAMP / CAMO

A generic reference model for a Domain Stakeholder from a CSDS perspective is shown below (Figure 6).

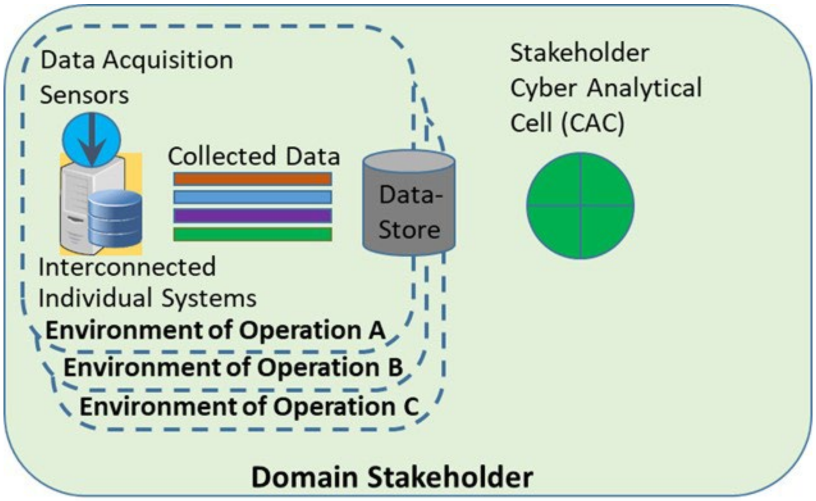


Figure 6. CSDS Domain Stakeholder Reference Model

It is important to note that a business like an airline may be a Domain Stakeholder in multiple different Aviation Domains. However, for the purpose of the CSDS AAF, they will be treated as separate Domain Stakeholders. Figure 7 below shows how an airline can be structured using the reference model and have businesses in more than one Aviation Domain.

In many instances, Domain Stakeholders may outsource product solutions to external vendors such as a Software as a Service (SaaS) or an Infrastructure as a Service (IaaS) provider. These are Cloud services considered to be Supporting Domain Stakeholders and can be a sub-tier contractor to the Primary Domain Stakeholders or another Supporting Domain Stakeholder.

Supporting Domain Stakeholders must be willing and able participants in the CSDS data sharing and collaboration process, and they must be enabled to do so by their higher tier stakeholders. Also, a sub-tier contractor may not share data directly with Multi-Domain CAC, unless given legal authority to do so on behalf of the higher-tier domain stakeholder. Thus, the appropriate domain stakeholder contracts, data privacy policies, and global requirements need to be established to support any viable cybersecurity data-sharing model, including CSDS. Whether a Stakeholder is labeled as Primary or Supporting, it is important to understand who both the Data Owners and the Data Managers are.

A **Data Owner** is accountable for who has access to information assets within their functional areas. A Data Owner may decide to review and authorize each access request individually or define a set of governing rules that determine who is eligible for access based on business function and support role (TCNJ Information Security Program).

Therefore, data ownership is the act of having legal rights and complete control over a single piece or set of data elements. It defines and provides information about the rightful owner of data assets and the acquisition, use, and distribution policy implemented by the data owner (Techopedia).

For example, one of the key challenges will be with respect to specific stakeholders (i.e., data owners) embracing the value of sharing elements of their data. The Aviation Information Sharing and Analysis Center (A-ISAC) and the Commercial Aviation Safety Team (CAST) represent just a few industry-driven aviation organizations. These have successfully embraced this value and are working to overcome the risks and trust barriers associated with sharing security data.

A **Data Manager** is responsible for enforcing policies and access to data as dictated by the Data Owners. Data Managers may not necessarily be Data Owners, particularly when business functions are outsourced to sub-tier contractors.

For example, a key airport stakeholder may outsource their ticketing kiosks to a 3rd party company specializing in flight check-in for passengers. While the airport stakeholder still owns the data being generated by these kiosks, the 3rd party company is considered the Data Manager, who is required to manage the data on the airport's behalf.

Alignment to DO-392

The six (6) Aviation Domains defined here align with the Aviation Stakeholder Framework of DO-392 that identifies the stakeholders as maintainers, manufacturers, operators, product suppliers, and service providers.

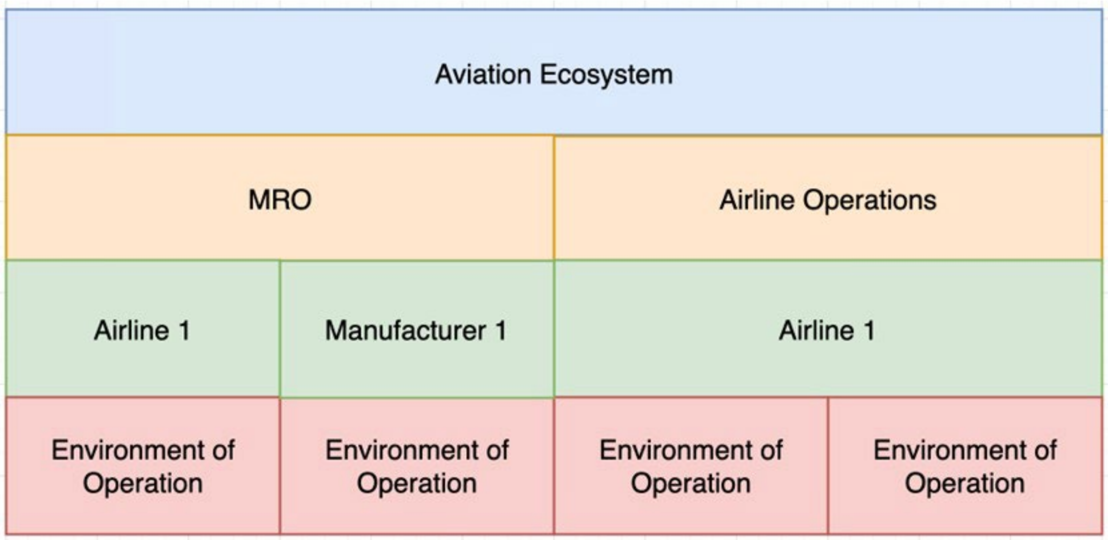


Figure 7. Application of CSDS AAF Reference Model to a Multi-Domain Airline

An **Environment of Operation (EOO)** is comprised of the systems and networks that provide an operational capability or specific mission or business function for aviation. For a generic definition of the EOO, refer to the National Institute of Standards and Technology (NIST) Computer Security Resource Center (CSRC) glossary (NIST Computer Security Resource Center). A more detailed discussion of the Aviation Environments of Operation is presented in Section 2.2.2.

Interconnected Individual Systems (IIS) contain the individual software/hardware components that provide the capabilities of the Environment of Operation. The IIS (i.e., systems) have several key characteristics, such as Interconnectivity Attributes, Information Systems Type, and the Acquired Data Categories of the system that are critical from a CSDS perspective. The interconnectivity attributes include several parameters such as the interconnectivity Continuous or Sporadic, does the system provide connectivity internal to the Environment of Operation or external (i.e., is it an edge node of an Environment of Operation network boundary, etc.) Each individual operational system can be classified as one of the following four types based on the characteristics and constraints of its design and sustainment within an Environment of Operation of the domain stakeholders: Aviation IT Commercial off the Shelf (COTS) Fully Maintained, Aviation Maintenance Constrained COTS, Aviation Customized COTS, and Aviation Industry Custom / Unique. These systems' characteristics are defined in more detail in Section 2.2.3.3.

2.2.2 Primary Actors and Drivers for CSDS AAF

Understanding the primary drivers and responsible actors at each layer of the Aviation Ecosystem Reference Model will help to identify where to target the implementation of CSDS engagement (see Figure 8). This diagram is meant to clarify CSDS AAF parties' primary role concerning CSDS implementation. The primary drivers represent the key roles/responsibilities that each layer plays in supporting the CSDS AAF. The primary actors are the entities responsible for developing and/or implementing the policies, procedures, and technologies at each layer in supporting the CSDS AAF. This diagram also shows where various Human Analysts sit within the reference model. For example, a Stakeholder Analyst sits within the Domain Stakeholder boundary and has no analysis function outside a given Domain Stakeholder. Multi-Domain Analysts typically sit at the cross-aviation domain or aviation ecosystem level (e.g., Aviation ISAC) and may aggregate and analyze data coming in from multiple cross Domain Stakeholders.

An example of an Industry-led multi-Domain Analyst organization is the A-ISAC.

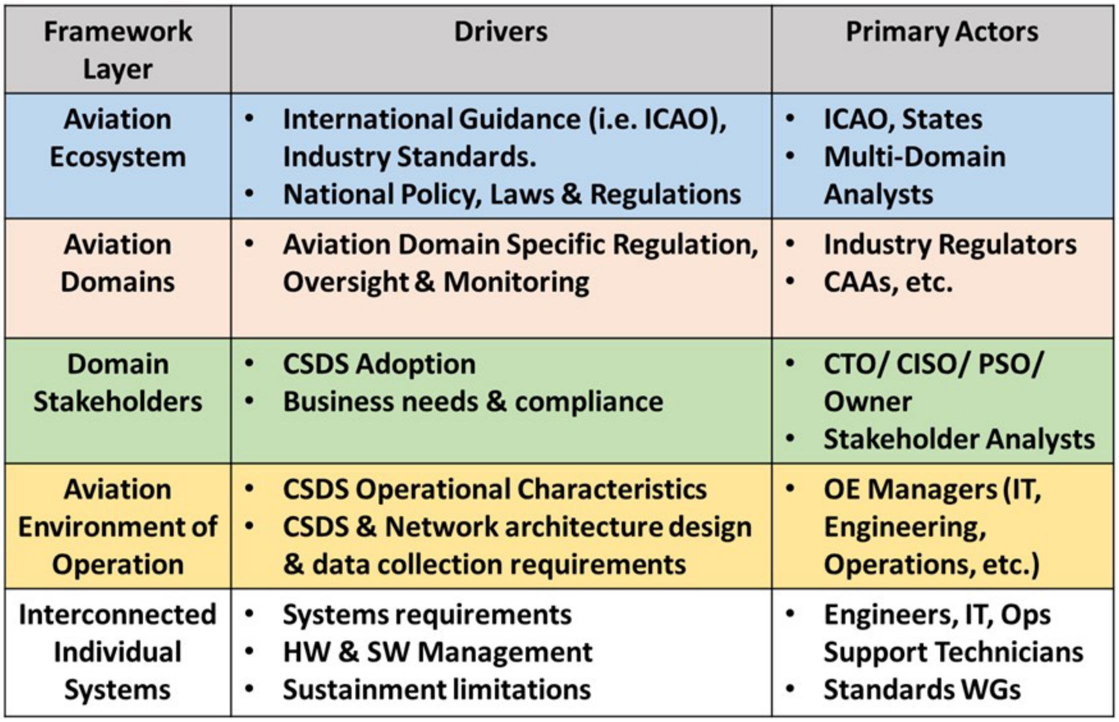
Aviation Ecosystem – International and National Policies: The aviation ecosystem layer of the reference model is primarily driven by both International and National Policies set forth by governing bodies to ensure a safe and secure aviation environment. Examples of these include ICAO guidance, as well as the National Strategy for Aviation Security.

Aviation Domains – Specific Regulations, Oversight, and Monitoring: At the Aviation Domain level, CSDS-specific guidance must be in place so that stakeholders of that domain can implement CSDS correctly within their organizations. Oversight will also be required to ensure stakeholders are correctly implementing CSDS in a timely manner. Primary actors at this layer are the Civil Aviation Authorities and National or Regional Airspace Operators.

Domain Stakeholders – CSDS Adoption: The primary driver at this layer is "getting the ball rolling" and approving security business changes with respect to CSDS. The Chief Technology Officer (CTO), the Chief Information Security Officer (CISO), the Product Security Officer (PSO), similar C-Suite leaders, and the subservient organization owners must be able to approve and facilitate the development of CSDS programs/activities. They also allow for additional budgeting and staffing required to support CSDS.

Environments of Operation – CSDS Design, Integration, and Operations: Each Environment of Operation will be responsible for the Design and Operations of CSDS activities. Given that each Environment of Operation's structure is unique regarding the Information System Types that exist, as well as the architecture, IT and Operational Managers must design, implement, and operate CSDS in a way that is consistent with the rest of the aviation ecosystem.

Interconnected Individual Systems – Hardware/Software Management for CSDS Operations and CSDS Data Collection: Within Information System Types, Engineers must identify the Networked Operational Elements and Data-Rich networks with respect to CSDS activities and implement methods for raw data collection. Technicians are responsible for the configuration/re-configuration of networked operational elements to support CSDS raw data acquisition activities and verify that the appropriate network traffic is being transmitted and the correct log files are being produced.



Framework Layer	Drivers	Primary Actors
Aviation Ecosystem	• International Guidance (i.e. ICAO), Industry Standards. • National Policy, Laws & Regulations	• ICAO, States • Multi-Domain Analysts
Aviation Domains	• Aviation Domain Specific Regulation, Oversight & Monitoring	• Industry Regulators • CAAs, etc.
Domain Stakeholders	• CSDS Adoption • Business needs & compliance	• CTO/ CISO/ PSO/ Owner • Stakeholder Analysts
Aviation Environment of Operation	• CSDS Operational Characteristics • CSDS & Network architecture design & data collection requirements	• OE Managers (IT, Engineering, Operations, etc.)
Interconnected Individual Systems	• Systems requirements • HW & SW Management • Sustainment limitations	• Engineers, IT, Ops Support Technicians • Standards WGs

Figure 8. Drivers and Primary Actors with Respect to the AAF Reference Model

2.2.3 Environment of Operation

An Environment of Operation (EOO) consists of many IIS that are typically networked together to fulfill a business objective/service for a Domain Stakeholder. These networks of IIS do not have to be contained in a single physical location and can span multiple geographic regions.

Environments of Operation (EOOs) may also consist of IIS of different Information System Types. An example would be in an aircraft manufacturing facility where there may be a mix of OT and IT equipment that are networked together.

A Domain Stakeholder may own and operate multiple EOOs of the same type. For example, an airline Tech Operations Center in San Francisco can be viewed as a separate EOO from an airline Tech Operations Center in Tokyo. Note that in this case the airline is a Domain Stakeholder within the Maintenance, Repair, and Overhaul (MRO) provider domain.

A Domain Stakeholder may also have one or more EOOs of different types depending on the various day-to-day operations and services offered to customers. For example, an airline as an airline/aircraft operator Domain Stakeholder operates a fleet of aircraft EOO, as well as an airline operation center, passenger reservation systems, and crew resource management systems EOOs. It also may operate one or more MROs as discussed above.

The regulations, guidance, and business needs/requirements for a given EOO will drive the detailed design and technical implementation. The EOO, cybersecurity considerations, and the information system types implemented will drive the requirements for network segmentation and separation. From a CSDS perspective, EOOs should provide logical partitioning requirements of information systems such that all CSDS components operate correctly. Also, a given Stakeholder's physical facility will likely contain multiple EOOs (e.g., Office & Factory, Airport Passenger Check-in & Aircraft Flight Line, etc.). These should provide multiple logically and/or physically isolated information systems to support appropriate security and resilience.

2.2.4 Interconnected Individual Systems

The IIS within the EOO includes hardware and software that provide its aviation operational capabilities. These systems acquire various types of raw data, often performing pre-analyzing before storing data determined to be relevant per the programmed system data collection instructions. Each IIS includes the Data Acquisition Sensors that acquire data and processors that may pre-process the data and store/forward the raw data. An IIS mostly provides local storage for log data. Examples include Aircraft Line Replacement Units, Factory ICSs, Airline reservation systems, etc. For instance, modern aircraft have Aircraft Condition Monitoring Systems (ACMS) to provide a predictive maintenance tool consisting of a high-capacity flight data acquisition unit and the associated sensors. These systems sample, monitor, and record information and flight parameters from significant aircraft systems and components, which can be an important source of OT data for determining if an event is occurring. Considerations for incorporating analogous onboard cybersecurity analytics capabilities are an ongoing industry discussion.

2.2.4.1 Interconnected Individual System Characteristic: Interconnectivity Attributes

It is also important to understand the IIS's interconnectivity attributes from an operational perspective. The following questions can be used as guidelines:

- Is system connectivity Continuous or Sporadic?

- Does the system act as a network edge node? (e.g., is the connectivity internal or external to the OE.)
- Is the system hard wired, wireless, sneaker net, or does it have human interfaces like keyboards or readers (e.g., card, biometric, etc.)?
- Are there IIS access points beyond classic network connections? (i.e., open USB ports, human interface devices, etc.), and how are these being secured and monitored?

2.2.4.2 Interconnected Individual System Characteristic: Acquired Data Categories

Prior to identifying the (four) 4 types of acquired CSDS relevant data, a set definition for a log record will be defined. According to the NIST Special Publication 800-92:

> *A log is a record of the events occurring within an organization's systems and networks. Logs are composed of log entries; each entry contains information related to a specific event that has occurred within a system or network. Originally, logs were used primarily for troubleshooting problems, but logs now serve many functions within most organizations, such as optimizing system and network performance, recording the actions of users, and providing data useful for investigating malicious activity.*

Within the aviation ecosystem, the CSDS AAF identifies four types of acquired data as relevant for the purposes of CSDS.

Security Data – Security related data collected as logs or streamed from the systems (e.g., antimalware software, intrusion detection systems, intrusion prevention systems, remote access software, web proxies, vulnerability management software, authentication servers, and other potentially security event relevant potential data). Typically, the security data is pulled from and is a subset of other categories of data listed below.

Network Data – Network data collection by IIS. Typically, this includes network activity logs from routers, switches, and DNS servers.

Operational Systems Data – This is the OT systems' data and typically includes the collection of data such as system operational parameters such as flight parameters for an aircraft and out-of-tolerance conditions, system fault reports, and other systems health management data (e.g., AHM data). Erroneous Systems Data is often the first indication of a security event.

Application Data – "Some applications generate their own log files, while others use the logging capabilities of the OS on which they are installed. Applications vary significantly in the types of information that they log" (Souppaya, 2006). Nevertheless, the following information is commonly logged as application data: Client requests, server responses, authentication and account information, application usage and operation actions like startup and shutdown.

2.2.4.3 Interconnected Individual System Characteristic: Information System Types

From a CSDS perspective and from an aviation industry product and systems infrastructure design and support perspective, it is helpful to define four (4) different Information System Types (IST) based on the requirements that drive how systems are defined, developed, and maintained over their useable lifetime. Thus, each IIS can be classified as one of these four types based on, or sometimes broken into functional elements that can be classified by these four types. These ISTs are defined around the characteristics and constraints of their design and sustainment within an EOO of the domain stakeholders. The aviation EOOs are a varied blend from pure IT to deep OT.

- **Type A: Aviation IT COTS Fully Maintained Information System Type**
- **Type B: Aviation Maintenance Constrained COTS Information System Type**
- **Type C: Aviation Customized COTS (Hardware -HW-/Software -SW-/Config.) Information System Type**
- **Type D: Aviation Industry Custom / Unique Information System Type**

These definitions incorporate considerations for the definition and design of the protocols, the hardware, the software, as well as operational factors that drive the upgradeability & patchability of systems within each of the four different ISTs. Any specific EOO within a specific Aviation Stakeholder's business will typically include more than one of these different Information System Types. Understanding different aviation ISTs and how they are implemented for a given EOO will significantly assist in understanding how to apply CSDS and the extent of its applicability for the Environment of Operation under evaluation.

In addition, from a CSDS perspective, the focus should be on understanding the cyber risks in each type of aviation information system and how CSDS can provide additional or unique analysis capabilities to identify potential cyber intrusions to help mitigate the cyber risks.

See Table 1 below for a summary of the characteristics associated with each of the four ISTs. Figure 9 (see below) shows the relationship between the Information System Architecture Types and Environments of Operation.

Table 1. Information System Types (ISTs)

	COTS Technologies		**Customized/Custom Technologies**	
	Type A: Aviation IT COTS Fully Maintained Information System Type	**Type B: Aviation Maintenance Constrained COTS Information System Type**	**Type C: Aviation Customized COTS (HW/SW/Config.) Information System Type**	**Type D: Aviation Industry Custom / Unique Information System Type**
Technology	Uses contemporary IT COTS industry technologies.	Uses COTS technologies but not the latest due to business constraints.	Uses customized COTS technologies due to specific industry product, business, or regulatory requirements.	Aviation industry uniquely defined and developed protocols, technologies, and products.
Upgrade Status: HW/SW Upgrades and Patches (Security, etc.)	Maintained regularly to standard IT industry best practices.	Upgrades need to be carefully scheduled due to operational limitations.	Upgrades require the redesign of customized HW/SW and potential recertification.	Upgrades require the redesign of customized HW/SW, recertification, and may require changes to industry specific standards.
Changeability: Impacts to cycle time & cost due to operational limitations & certification requirements	Minimal (Typically in Hours / Days)	Some (Typically in Months / Years)	Moderate (Typically in Years)	Extensive (Typically in Years / Decades)
Aviation Example: Airline / Aircraft	Airline Back Office Systems where aircraft log data is collected & stored	IFE Networks on some airplanes.	Some AISD networks on some aircraft	Most ACD networks on aircraft
Aviation Example: OEM Aircraft Factory	Aircraft Factory IT networks.	Aircraft Factory OT networks.	Handheld aircraft data loader.	Likely N/A
CSDS Perspective	Most commonly addressed across industries. Most data available & easiest to change data collection.	CSDS Perspective	Most commonly addressed across industries. Most data available & easiest to change data collection.	CSDS Perspective

2.2.4.3.1 Type A: Aviation IT COTS Fully Maintained Information System Types

These types are typically the globally ubiquitous and standardized IT type systems (protocols, hardware, software, cloud services, etc.). They are usually found in or associated with almost every EOO across all aviation stakeholder domains to varying degrees. For example, these include the typically COTS IT network systems that are most readily maintained with up-to-date software and security patches. These comprise the front and back office of most aviation industry stakeholders, as well as much of the interconnectivity between and within their industrial facilities and are typically managed at least in part by the IT departments. For example, this category of information system is fully implemented within the factory and MRO environments alongside and often connected (to lesser or greater extents) to the OT Industrial Control System (ICS) Network Systems.

2.2.4.3.2 Type B: Aviation Maintenance Constrained COTS Information System Types

These ISTs tend to be found in different Aviation OT environments like the commercial ICS used in factories, MROs, and airports. They incorporate both classic computer IT-driven as well as controller-driven OT equipment. What makes this category of network system architecture unique is not that it always differs significantly from IT systems design, but that these networks and their connected equipment are not easily modifiable for maintaining current HW or SW (OS & Config Files, etc.) and patch levels. ISTs integrated with operational equipment (e.g., in a factory setting) may not be easily taken offline without causing significant operational disruptions. Specific industry compliance certifications, particularly for safety-critical systems, may make it prohibitive, in terms of time and cost, to keep these ISTs fully updated.

2.2.4.3.3 Type C: Aviation Customized COTS (HW/SW/Config.) Information System Types

The aviation industry has sought to leverage and customize COTS network systems capabilities to save cost and development time. However, to preserve the resilience and safety capabilities of the aviation products for specific EOOs, this has required the implementation of highly constrained configurations and customized hardware & software designs that are different or more restrictive than IT industry solutions. This makes these networks and connected equipment even more challenging to modify, maintain, and update with current HW or SW (OS & Config Files, LRU OPS & OPC, etc.) and patch levels. For example, these system designs often include highly constrained configurations and configuration files, removal of all unused code in COTS software, re-writing the software to the Radio Technical Commission for Aeronautics (RTCA) DO-178 practices and controls, and in some cases, even limiting protocol options that require changes to core hardware, firmware, and software components may also be needed.

2.2.4.3.4 Type D: Aviation Industry Custom / Unique Information System Types

These are custom-defined aviation network systems unique to aviation, and typically encapsulated in the ICAO, RTCA/EUROCAE, ARINC, and other aviation-specific standards. Aircraft-specific networks include ARINC 429, ARINC 629, ARINC 664, etc. Multi-domain custom aviation networks that cut across aircraft, airspace management, Air Navigation Service Providers (ANSPs), and Com Service Providers (CSPs), include examples like Internet Protocol

Suite (IPS), Controller Pilot Data Link Communications, etc. The development of these networks has historically been driven by a safety design philosophy and methodologies, which have many parallels with the Zero Trust Architectural approach that is becoming popular across the IT industry.

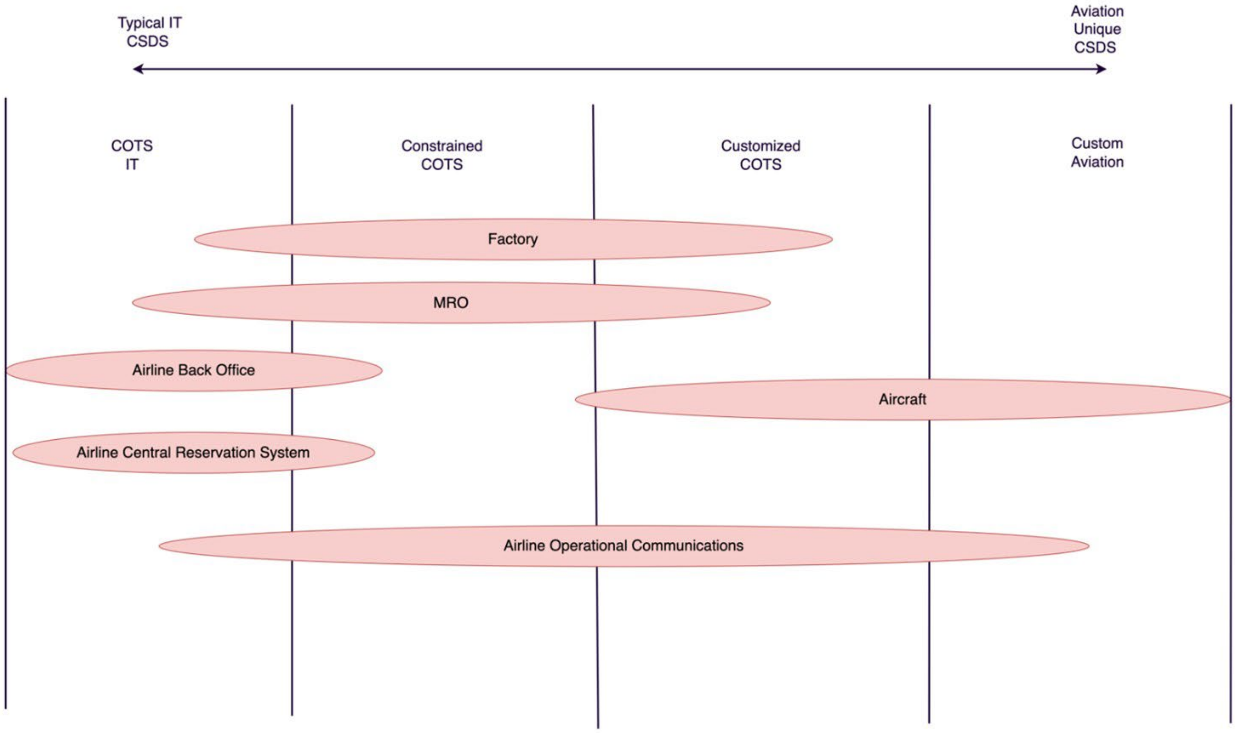


Figure 9. Relationship between Information System Architecture Types and Environments of Operation

2.3 CSDS AAF Systems Architecture: The Systems Perspective

This section lays out the Aviation Architectural Framework from the systems architectural perspective. Figure 10 illustrates the high-level systems architecture diagram for the AAF.

The architecture includes the Distributed CACs with vested interests in the cyber analytical information generated and shared by Domain Stakeholders. The architecture is segmented into six (6) Aviation Domains, each containing numerous Domain Stakeholders that own, maintain, and manage various Systems and EOOs. From this diagram, it can be observed the multiple connections between CACs, both from a Domain Stakeholder CAC perspective and Distributed CAC perspective. A key goal in designing the CSDS AAF has been to provide fast and efficient sharing of information among all key players while ensuring the privacy and security of confidential information. This is accomplished using the proposed Information Exchange Messaging System (IEMS) discussed in more detail in Section 2.3.6.

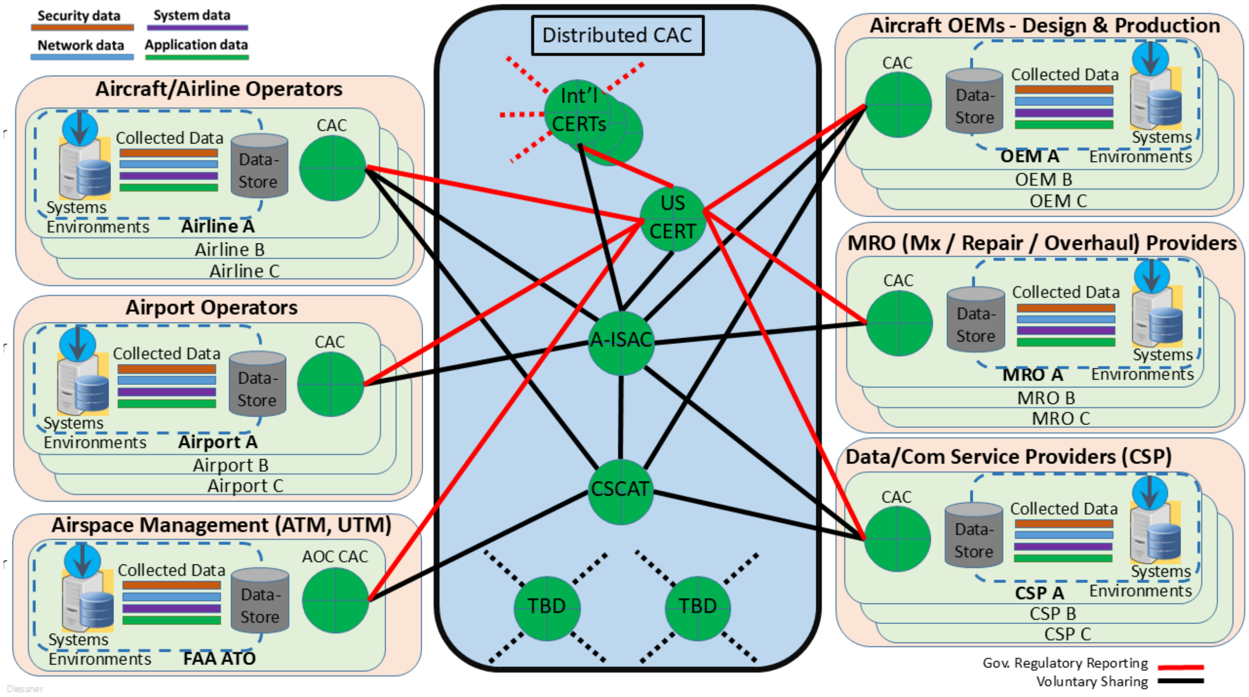


Figure 10. CSDS AAF System Architecture Perspective

2.3.1 AAF Operational Concept (System Architecture)

The AAF Operational Concept Diagram (see Figure 11) shows how various system components are interconnected to share cyber analytical information among Domain Stakeholders and Multi-Domain CAC through an IEMS.

EOOs contain the IIS that acts as a source of available data (Collected in the Data-Store) that may be valuable for CSDS-related activities. CACs must work closely with the Primary Actors identified at the EOOs layer in Figure 8 to identify all Data-Store elements that make up this Data-Store and implement strategies to extract this data when needed for a given CSDS Use Case. From a systems perspective, this would involve configuring/re-configuring the Data Acquisition Sensors within the IIS to allow CACs to acquire this data using various acquisition methods discussed in future sections.

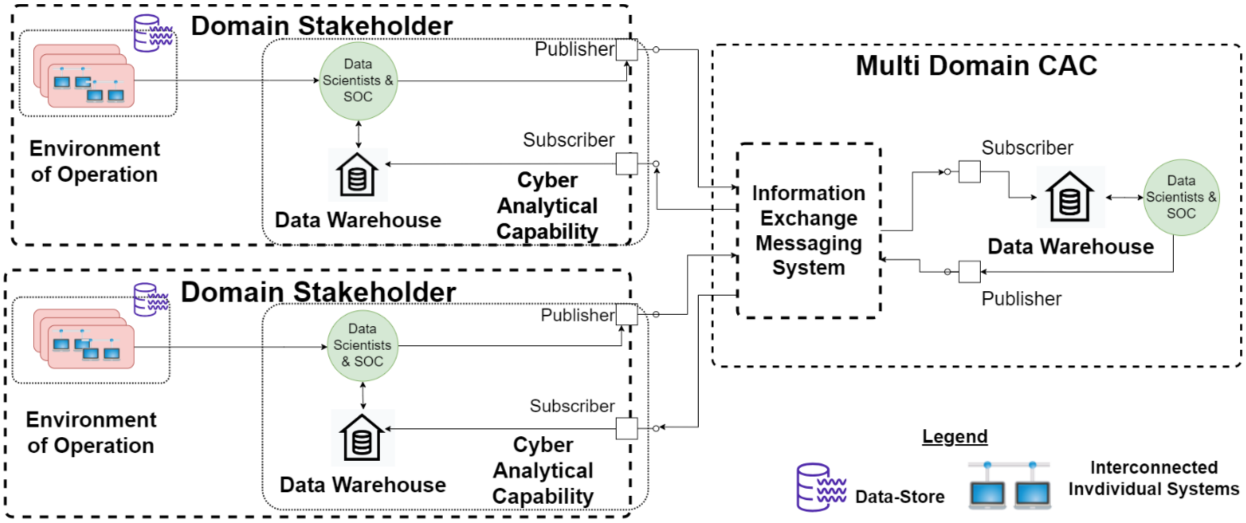


Figure 11. AAF Operational Concept Diagram

The Data-Store is not a single system, but a collection of storage elements scattered throughout an Environment of Operation that each contributes to making up the Data-Store. This is somewhat synonymous with how the “Cloud” refers to a collection of servers. The Data-Store will usually store sensitive and confidential data. This includes network data such as IP Addresses, device hardware identifiers, customer data, diagnostic data, and other uniquely identifiable data that should be kept undisclosed within the Domain Stakeholder itself. For this reason, the AAF requires that no data should be shared with external entities directly from the Data-Store but must instead always be shared through the CAC. The CAC’s responsibility is to ensure that all confidential and sensitive information is redacted or removed before publication (see Information Sharing Phase of the CSDS Data Life Cycle).

The Data Warehouse represents the central data/information storage used by the CAC to store various results (the artifacts) that are created throughout the CSDS Life Cycle process (specifically the Curation, Advanced Analytics, and Information Sharing phases). These artifacts may be owned by the Domain Stakeholders themselves, or from other Domain Stakeholders and Multi-Domain CAC that they are subscribed to through the IEMS. From a system requirements perspective, Data Warehouses shall be designed to store data in a flexible, organized fashion that allows for easy lookup and retrieval of data/information. For example, if a CAC needs to conduct an analysis of an EOO over a large period, they must be able to easily query for data/information to perform analysis. Data Warehouses shall also be designed with resiliency and fault tolerance in mind to ensure data is not corrupted or lost.

A good example of how this external Domain Stakeholder information may be used is to perform a comparative analysis of specific EOOs and to detect operational anomalies. Traditionally a Domain Stakeholder will only have information about their own EOOs, and it may be challenging to identify what is a "normal" operating state. If a Domain Stakeholder can compare their EOO operational metrics and telemetry to the rest of the aviation community, it may help to detect when something abnormal is occurring within their networks. To complete the objective stated above, CAC will utilize the Domain Stakeholder's Data-Store in conjunction with data already in the Data Warehouse to apply CSDS and detect these anomalies. The IEMS acts as the backbone for information sharing and collaboration. Multi-Domain CACs will be in control of the IEMS and determine the access rights of all Domain Stakeholders using message Topics. Domain Stakeholders may publish/subscribe to the various topics they have access to. The IEMS is discussed in more detail in Section 2.3.6.

CACs represent the primary facilities for CSDS by Human Analysts and Data Scientists. Data Scientists extract data from the Data-Store to conduct preliminary analysis on the data, develop various analytical toolsets that Human Analysts downstream will use, as well as AI/ML automation software that can perform real-time analytics. Human Analysts will acquire data from the Data-Store and any data in the Data Warehouse using toolsets that will perform advanced analytics to produce insightful cyber analytical information that can be shared with the rest of the aviation community.

2.3.2 System Architecture Aspects of Environments of Operations

A given Environment of Operation (EOO) of a specific stakeholder will be made up of many IIS, typically of multiple Information System Types, which may or may not be appropriately segmented. Operations include things like flying aircraft, building aircraft, managing passenger reservations, managing airport gate assignments/routing, managing baggage & handling at an airport, managing airspace, etc.

The compositional makeup of an EOO will typically differ between Domain Stakeholders, though they likely share similar key characteristics. It is expected that EOOs are not architected in the same way, nor should the CSDS AAF Framework impose strict implementation requirements. Recognizing this system's architectural diversity, the CSDS AAF provides a general set of guidelines to appropriately identify existing EOOs and architect new ones to leverage CSDS.

EOOs are composed of a set of IIS, which can be categorized into the previously described four (4) Aviation Information System Types, COTS IT, COTS ICS, Constrained IT, and Custom

Aviation depending on the criteria in Table 1. A Domain Stakeholder may have an IIS that supports multiple EOOs, as opposed to each IIS being located within only one EOO. If there are systems that support multiple EOOs, it could be critical to review security and network segmentation implementations.

From a CSDS perspective, key characteristics of the EOOs include: the types and topologies of networks, understanding where all network edge node systems that form the network boundary of the EOOs are located, understanding all IIS access points within the EOO beyond classic network connections (i.e., open USB ports, human interface devices, etc.) [note how they are being secured and monitored is addressed at the System level below], EOOs network segmentation, what CSDS Capabilities exist or need to exist within the EOOs, etc.

For the purpose of the CSDS AAF, EOOs are categorized into four (4) Class Designations shown below in Figure 12, depending on the CSDS Capabilities that exist.

It is important to note that ISTs are naturally segmented (topologically) and require some routing mechanism to send/receive data/information between them. Furthermore, Interconnected Individual Systems that wish to communicate between EOOs must also be done using a routing mechanism (which is also an IIS).

Domain Stakeholders should identify and classify their EOOs accordingly and should strive towards a class 4 as much as possible (see Table 2 below)

Table 2. Operational Class Designations

Operational Class Designations	**CSDS Capabilities**
1	No CSDS Capabilities. Requires upgrade to a Class 2 or 3 or connection to a Class 2,3 or 4 EOO. Example: manufacturing systems that store limited data.
2	Has the ability to store Available Data for the Data-Store. Requires manual extraction or connection to a Class 3 or 4 EOO. Example: an aircraft in which data relevant to CSDS may be available.
3	Has the ability to connect and exfiltrate data out to other EOO. Example: factory environment that connects and transmits data back to an IT Back Office.
4	Has the ability to extract, analyze and store cyber analytical information to a CAC capable of conducting CSDS. Also has the authorization to share information using the IEMS.

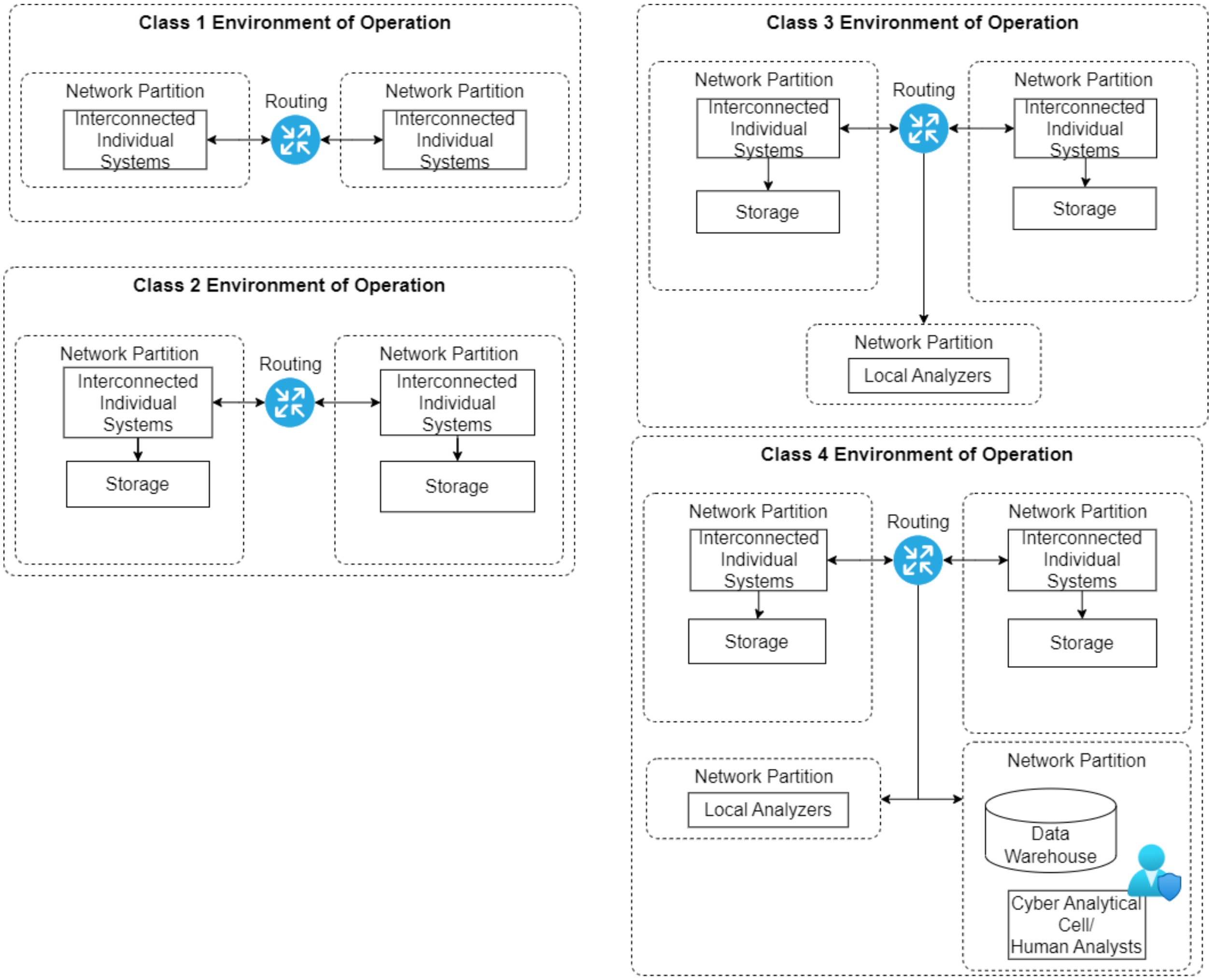

Class 1 Environment of Operation
Network Partition
Interconnected Individual Systems
Routing
Network Partition
Interconnected Individual Systems
Class 2 Environment of Operation
Network Partition
Interconnected Individual Systems
Storage
Routing
Network Partition
Interconnected Individual Systems
Storage
Class 3 Environment of Operation
Network Partition
Interconnected Individual Systems
Storage
Routing
Network Partition
Interconnected Individual Systems
Storage
Network Partition
Local Analyzers
Class 4 Environment of Operation
Network Partition
Interconnected Individual Systems
Storage
Routing
Network Partition
Interconnected Individual Systems
Storage
Network Partition
Local Analyzers
Network Partition
Data Warehouse
Cyber Analytical Cell/ Human Analysts

Figure 12. CSDS Class Designations for Environments of Operation

2.3.2.1 Interconnected Individual Systems

An IIS includes the physical nodes on a network that functions to accomplish some objective of the Environment of Operation. An IIS is any device that can be detached relatively easily. Examples include a Line Replacement Unit of an aircraft or a ticketing Kiosk in an airport. It may also consist of supporting equipment, such as routers, switches, and other server equipment. Figure 13 shows the block diagram of a generic IIS.

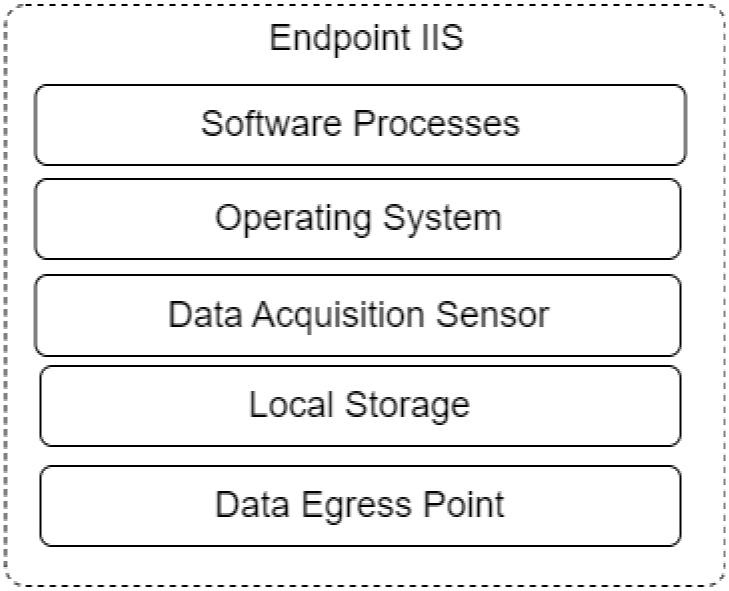


Figure 13. Data Acquisition Placement within an Interconnected Individual System

2.3.2.2 Data Acquisition Sensors

Figure 14 below shows the block diagram of a Data Acquisition Sensor. Data Acquisition Sensors can either be a software or hardware component of an IIS that provides four (4) core functions:

- **Acquire**: Monitor Software Processes running on the IIS and detect application/system level logs/events.
- **Pre-Analyze**: From the logs/events that are detected, evaluate them, and determine if they are useful for CSDS purposes.
- **Encode**: Encode the logs/events into a format that can be stored locally or transmitted over the Data-Interconnect Interface. This also involves feature extraction, where only a subset of all data properties is actually collected.
- **Local Storage/Data Egress Point**: Store the encoded log/event into the Local Storage of the IIS or to a Data Egress Point, which forwards the data up to a Data-Store.

It is important for the Primary Actors of IIS layer of the Aviation Ecosystem Reference Model to understand their Data Acquisition Sensors' placement and capabilities and put processes in place to successfully acquire the correct Cyber-Relevant Data. The Data Acquisition Sensor may be comprised of multiple components or a single component. From the CSDS AAF perspective, the Data Acquisition Sensor is logically viewed as a single entity.

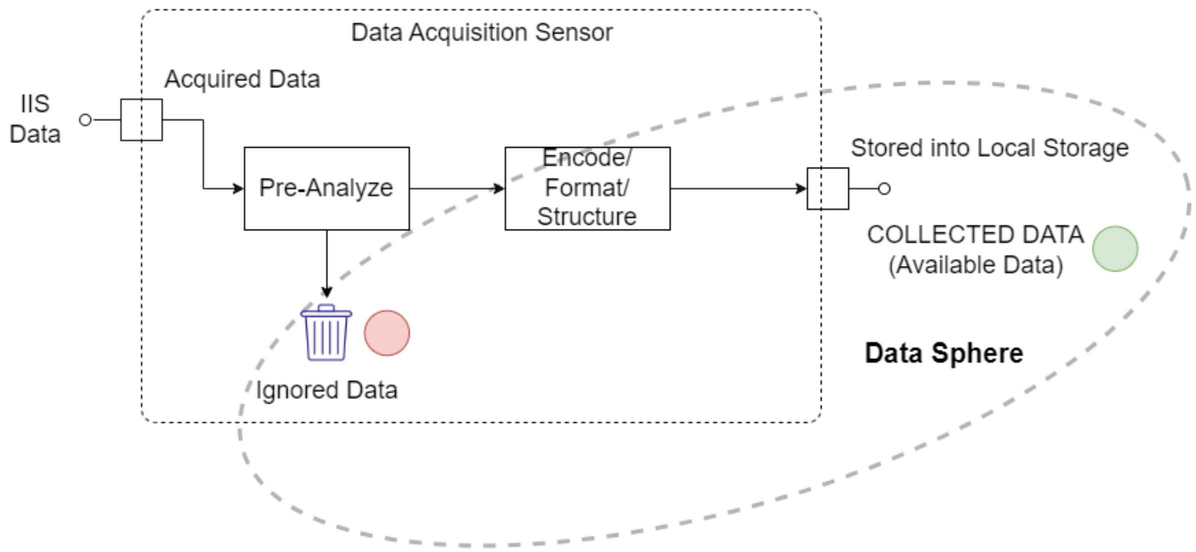

Data Acquisition Sensor
Acquired Data
IIS Data
Pre-Analyze
Encode/ Format/ Structure
Stored into Local Storage
COLLECTED DATA (Available Data)
Ignored Data
Data Sphere

Figure 14. Data Acquisition Sensor

Figure 14 shows the relationship between the Data Acquisition Sensor and the Data Sphere concept. The figure illustrates that the Data Sphere is composed of the sum of the Ignored and Collected data. For the Data Sphere to increase in size, more data needs to be acquired.

2.3.2.3 Local Storage

CSDS Local Storage are primitive storage locations that can be located throughout an Environment of Operation. Local Storage can be embedded into an Interconnected Individual System or can be an Interconnected Individual System itself that specializes in data storage such as a Network Attached Storage server. There is nothing special about CSDS Local Storage, and it just provides basic data storage and retrieval capabilities. CSDS Local Storage does not contain any data analyzers or pre-processors.

Primary Actors of the Individual Interconnected Systems layer of the Aviation Ecosystem Reference Model must be aware of the placement and capabilities of CSDS Local Storage so that Cyber-Relevant Data can be successfully extracted and stored in the Data-Store.

2.3.2.4 Data Egress Points

Data Egress Points are specialized Interconnected Individual System devices within an Environment of Operation whose job is to receive and transmit cyber-relevant data to Data-Stores. Data Egress Points have special access rights to connect to the Data-Stores and should be treated as a CSDS critical resource that requires additional security. Compromise of Data Egress Points may cause incorrect or incomplete data to be sent to Data-Stores or may cause complete disruption of the system overall. EOOs may contain zero or more Data Egress Points and depends on the EOO class designation (see Section 2.3.2).

Primary Actors of the IIS layer of the Aviation Ecosystem Reference Model must be aware of the placement and capabilities of Data Egress Points and ensure a continuous connection exists with the Domain Stakeholder's Data-Store. They must also manage the access controls of Data Egress Points and ensure access is terminated and any confidential access keys removed when a Data Egress Point is decommissioned.

2.3.2.5 Acquisition Modes

Data Acquisition is how CSDS cyber-relevant data within Interconnected Individual System Local Storage is retrieved/acquired by CACs to be analyzed. The data acquisition process may vary depending on the Environment of Operation as well as the capabilities of the Interconnected Individual System where the available relevant data exists. Data Acquisition methods are listed below and discussed in more detail in Section 2.3.3.1.

1. Physical Manual Extraction
2. Remote Manual Extraction
3. Remote Automatic Extraction
4. Remote Near Real-Time Extraction

The Data Acquisition modes chosen depends on the capabilities of the Data Acquisition Sensors located within the Operational Environments.

2.3.2.6 Data Acquisition Placement

Data Acquisition Sensors can be completely software-based components installed into the operating system or runtime environments of IIS. Examples include the Windows Event Viewer on Windows Server, which generates event logs from system and application events. Another example is Cisco Traffic Analysis, which captures CSDS cyber-relevant network traffic data.

2.3.3 Stakeholder Data-Store Concept

2.3.3.1 System Architecture for a Data-Store

The Data-Store is the cumulative data storage of potential CSDS Cyber-Relevant data captured from its Environment of Operation. The data stored within the Data-Store is said to be available if CACs readily have access to it. Local Storage within various IIS contributes to the Data-Store. While they do not exist on their own, they must instead be incorporated into an Interconnected Independent System and can always be represented in a form as illustrated in Figure 15 below.

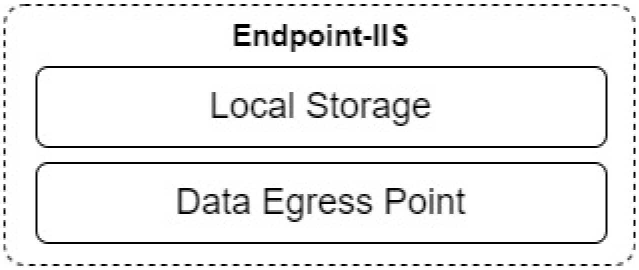


Figure 15. Basic Interconnected Individual System (IIS) Block Diagram

Local Storage must be peered at with the appropriate DEP to allow for the successful extraction of the data by CAC entities. Within a stakeholder's Environment of Operation, Local Storage can be found in many locations as well as in different system configurations. The CSDS AAF identifies three Local Storage configurations.

Configuration #1

In the first configuration (Figure 16), the Local Storage is located within an Endpoint IIS device. These Endpoint devices include standard IT equipment such as Laptops, Desktops, Servers, Routers as well as OT equipment such as Line Replace Units (LRU) and equipment controllers.

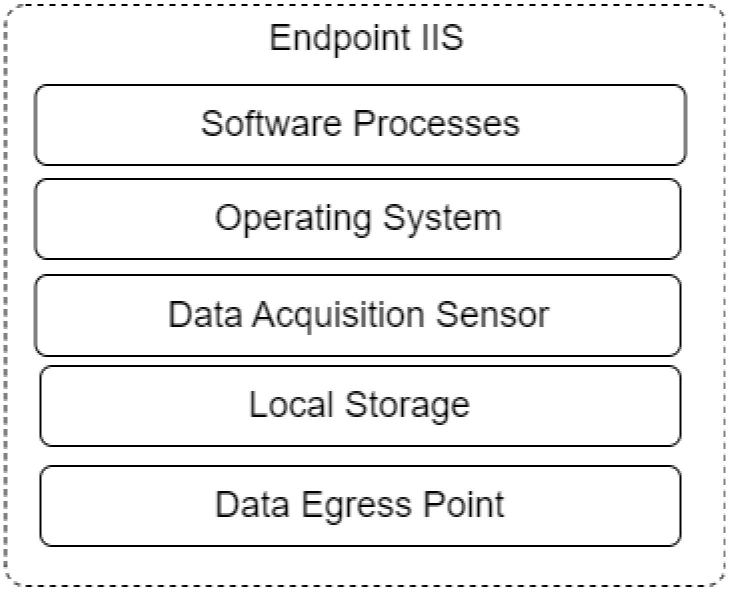


Figure 16. Local Storage Interconnected Individual System (IIS) Configuration

Configuration #2

In the second configuration (Figure 17), the local storage is in a Central Storage IIS that is dedicated to storing/collecting data for multiple Endpoint IIS devices. This option is typically seen in OT technologies in which the Endpoint IIS devices may not have the necessary storage capabilities to host a local storage itself. This option may also be used with systems that are not expected to "retain" their data for long periods and in which a more long-term storage solution is needed. Examples of local storage in a Central Storage IIS include a Network Attached Storage device in the case of IT networks.

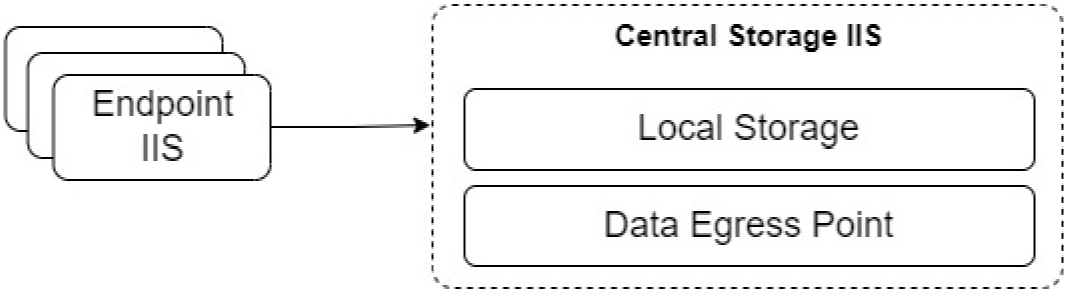


Figure 17. Centralized Local Storage Configuration for IIS

Configuration #3

The third configuration (Figure 18) builds on Configuration #2, in which a multi-storage IIS configuration is employed. This configuration may be required in EOOs that have different security requirements or are regulated such that data compartmentalization is required. For example, on an aircraft, network data for the Passenger Information and Entertainment Systems Domain (PIESD) networks are typically not stored in the same location as the Hardware Security Module (HSM) logs.

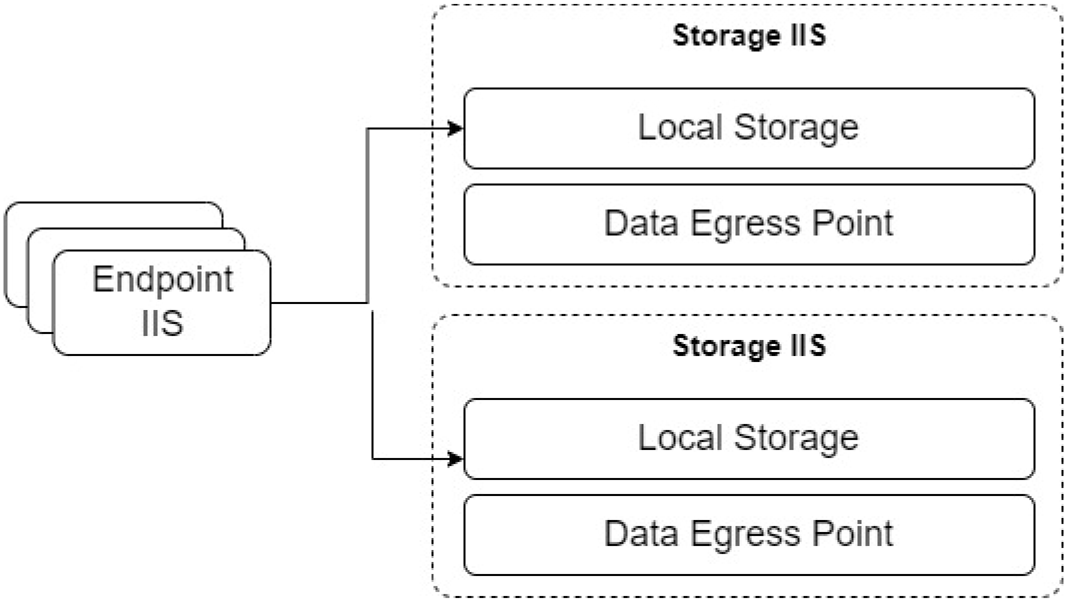


Figure 18. Distributed Local Storage Configuration

Local Storage Data Extraction Interface

Any IIS that contains local storage must be peered with an appropriate DEP. This is required so that data can be successfully acquired from it by CACs. The DEP does not need to be an independent subsystem but can be existing interfaces re-purposed for CSDS.

Section 2.3.2.5 defines four (4) methods for Data Acquisition

- Manual Physical – Physical ports on the IIS that allows for Flash Drives, SD Cards, and other Removable Storage devices to be connected to it so that data can be extracted from it.
- Manual Remote – A networked interface that allows remote connection to the local storage as well as a defined protocol such as FTP, SSH, HTTP etc. to extract the data. Manual Remote extractions methods are used by Human Analysts and Data Scientists that explicitly connects to the local storage to pull data.
- Automatic Remote – The same as Manual Remote, with the only difference being that this method is used by automated software/toolsets that connect to the IIS over a remote connection on a pre-defined schedule such as once a day.
- Automatic Continuous – This mode requires a perpetual connection between the IIS and the CAC. No Human Analyst nor automated software is needed to make initial contact to the IIS, instead data from the IIS is constantly being streamed into the CAC (near real-time Use Case).

Local Storage Security

Local Storage for a Domain Stakeholder must enforce additional access control policies for the purpose of data extraction by Human Analysts, Data Scientists, and automated toolsets. For example, in the case of Manual Physical data extraction methods, additional physical security access policies are required for Human Analysts to physically access the local storage. In the case of other remote acquisition methods, additional network security controls are required for software toolsets to connect, such as creating unique FTP/SSH user accounts etc.

Local Storage Data Retention

CACs must be mindful of how long each local store retains data. This will vary greatly depending on hardware, operational and regulatory constraints. For example, an IIS may only have enough storage capacity to store 1 weeks' worth of data, before the data gets overwritten. Another example may be a regulation that requires data to be retained for 30 days which, if data is stored locally, will drive hardware and operational requirements on the local storage. CACs should maintain an inventory of all the available local storage and develop routine data extraction reschedules to acquire the data before it gets lost.

Local Storage Downtime During Extraction

During data extraction of the local storage, some system downtime may be necessary. This is prevalent mostly in the manual physical data extraction mode, where data extraction requires the IIS to not be in service. Other IIS such as a Network Intrusion Detection System, may need to be in maintenance mode to extract data remotely and will incur some downtime. CACs must factor in downtime when determining appropriate times for extracting data from the local storage to not impair operation during business hours.

2.3.4 Cyber Analytical Capability (CAC)

CACs are represented by a collection of Human Analysts using software-based toolsets to perform analytics on data to produce cyber-analytical information within a highly secured networking environment. CACs are usually enclosed within a secure facility such as the Security Operation Centers (SOCs) that currently exist within many IT Organizations. However, it could be an independent facility as well.

Figure 19 below illustrates the reference model for a CAC. The CAC perspective does not care about Environments of Operation or how the data gets stored in the Data-Store, it only cares about the cyber-relevant data available in the Data-Store and the various methods to acquire the data.

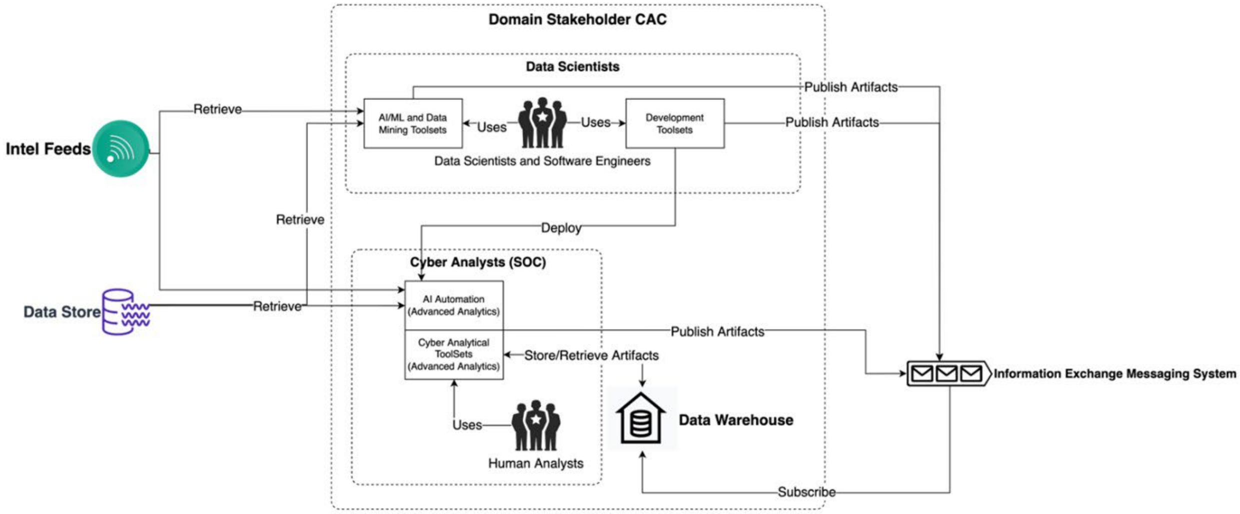


Figure 19. Cyber Analytical Capability Reference Model

The CAC Reference Model identifies two (2) distinct groups that perform unique functions within the CAC, the Data Scientists and Software Engineers group and the Cyber Analysts group. The Data Scientists and Software Engineers group uses AI/ML and Data Mining toolsets to retrieve data from the Domain Stakeholder Data-Store and combines it with Threat Intel Feeds to perform exploratory analytics on the data and generate AI/ML Models, visualization software, and other applications using development toolsets to create Cyber Analytical Toolsets and AI Automation for Human Analysts to use.

AI/ML Automation differs from toolsets. Whereas Human Analysts typically use toolsets, AI/ML Automation automatically checks new analytical results and provides recommendations that can be taken by Human Analysts continuously in the background. Toolsets provide insights into large collections of data and give meaning to these insights for Human Analysts to understand. AI/ML Automation Engines takes insights and provides advisory based on desired end goals. For example, if a toolset creates risk profiles for various systems on the network, the AI/ML Advisory Engine will determine ways for the risk profile to be lowered.

2.3.4.1 CAC Required Capabilities

1. Must be able to extract available desired data from Data-Stores and threat intelligence feeds.

2. Must contain a collection of AI/ML, Data Mining, Development, and Cyber Analytical Toolsets that Data Scientists and Cyber Analysts use.

3. Toolsets must provide a User Interface for Human Analysts to run and generate analyses and create reports.
4. Toolsets must also store analytical results in a CAC Data Warehouse for access in the future.
5. CAC may also contain AI/ML Advisory Engines that interpret analytical results and provide meaningful recommendations to Human Analysts.
6. CAC must have an Information Exchange Messaging System (IEMS) that can be used to publish information out of the CAC to other stakeholders.
7. The IEMS must provide some form of approval mechanism (i.e., one or more designated personnel with authority that gives final approval to anything leaving a CAC)
8. The IEMS must allow Human Analysts to redact/anonymize any message leaving the CAC.
9. The IEMS must allow other external CAC to send messages to it and allow an inbox system for Human Analysts to process.
10. Human Analysts must be able to collect incoming messages from the IEMS and store them in the Data Warehouse for use with analytical toolsets. AI/ML models may aggregate external and internal information and perform correlations.
11. CAC may contain R&D departments that develop internal toolsets as well. R&D Departments may extract data from CAC Preprocessor Collectors to create Training Sets for AI/ML development. R&D departments will then provide these tools to Human Analysts to use or publish them through the Information Message System for other CAC to utilize.

2.3.4.2 Human Analysts in a Multi-Work Environment

While it is ideal for CAC to be fully contained in one physical facility with Human Analysts working from dedicated workstations connected to the local LAN, it may not be possible or feasible. Human Analysts may be located anywhere in the world and may even work remotely. Human Analysts may work from multiple devices as well (e.g., one at home and one in the office), yet may need access to various toolsets on the go. While there are many NIST Security Guidelines for Enterprise Telework, there are some CSDS-specific considerations with respect to Human Analysts.

1. Human Analysts may need access to data stored in the Cyber Analytical Preprocessor Collector while working remotely.
2. Human Analysts may need to cache copies of data stored in the Data Warehouses; in the event the specific toolsets being used require access to the local file system. How can Human Analysts ensure they are using the correct data sets?
3. The packaging and deployment of various toolsets may differ depending on the software design. Some toolsets may be a binary installed on the given laptop or workstation, web-based, or may require Human Analysts to remotely access the IT network using a VPN. In the event toolsets are installed on each device, Human Analysts will need to ensure the software is up to date.

2.3.4.3 Toolset Types

Within the CAC, four (4) toolset types are identified, AI/ML and Data Mining toolsets, development toolsets, Cyber Analytical Toolsets, and AI automation.

- **AI/ML Data Mining Toolsets-** Data Scientists, use these in conjunction with the data in the Central Data-Store to successfully fulfill the Maintain and Process phases of the CSDS Life Cycle. Artifacts generated from these toolsets can be published.
- **Development Toolsets** – Data Scientists use these to create and update Cyber Analytical Toolsets and AI Automation systems that successfully fulfill the Analyze phase of the CSDS Life Cycle. Artifacts generated from these toolsets can be published.
- **Cyber Analytical Toolsets** – Human Analysts use these to generate Cyber Analytical Information and generate publishable artifacts for the IEMS. Artifacts generated from these toolsets can be published.
- **AI Automation** – These are backend systems running “behind-the-scenes”. They generate artifacts without Human Analyst intervention. Artifacts generated from these toolsets can be published.

2.3.4.4 Distributed Collector Data (Configuration Management i.e., Synchronicity)

In a distributed environment, an analysis may not be performed within the CAC facility itself. For example, a Human Analyst working from home may need access to collector data to perform analysis. Data may be retrieved and cached locally on the device for analysis (in the event the Toolset requires a local file for analysis) or may retrieve data in real-time as it is processing (in the event of high I/O intensive analysis operations). In the event of live analysis, a VPN should be suitable enough to maintain a secure connection to the main CAC facility, at least for the duration of the analysis.

2.3.4.5 Access Control Mechanisms

In larger CACs that employ more than ten (10) human analysts, it may be important to establish various levels of access to toolsets and data within the CAC. For example, newly employed Human Analysts may only be given permissions to certain toolsets, perform only specific analysis with those toolsets or access a specific subset of data to analyze.

2.3.4.6 Cloud-Implemented CACs

Cloud-Implemented CACs typically consist of virtual servers and networking equipment, although some may include co-located physical hardware. Cloud-Implemented CACs may also include the use of virtualized workstations for analysts to run toolsets, which would eliminate the need for any local caching of files that toolsets may need, thus allowing all data to remain in a secure cloud-based environment. Virtualized environments are also easy to scale when additional processing power or speed is needed for complex workflows. Because of the nature of a cloud-based environment, all analysts are effectively remote and working from a commonly maintained set of toolsets that can be centrally version-controlled and updated with the latest settings and algorithms.

There are generally two (2) types of clouds, Private Clouds and Public Clouds. Private Clouds utilize an organization’s own infrastructure and are maintained and secured either by their own or contracted employees. Public Clouds are offered by large third-party vendors and can take advantage of economies of scale while offering both shared and dedicated infrastructure depending on the needs of the tenant. Many providers offer a hardened GovCloud option. This option is certified to meet the needs of government customers that comply with the FedRAMP High baseline; the DOJ’s Criminal Justice Information Systems Security Policy; U.S. ITAR; EAR; Department of Defense Cloud Computing Security Requirements Guide (SRG) for Impact Levels 2, 4 and 5; FIPS 140-2; IRS-1075; and other compliance regimes. Some organizations may use a Hybrid concept based on a combination of Private and Public clouds for redundancy, backup, or to reduce costs. There have been rare instances in the past where a Public Cloud provider has experienced temporary interruptions of service, either due to internal maintenance or external threats such as Distributed Denial of Service (DDOS) attacks. However, Public Cloud providers have the advantage of economy of scale to offer multiple redundant data centers and network providers to detect threats such as DDOS attacks early on and quickly take action to block such traffic. An organization with a Private Cloud may only have a single network provider to rely on to block such attacks from saturating their network with traffic.

Additionally, Public Cloud providers can effectively act as a large honeypot, performing their own analytics to detect intrusion attempts across all their tenants.

2.3.5 Data Warehouse

Data Warehouses provide localized long-term storage for Curated Data and Shareable Artifacts for CACs of both individual stakeholders and multi-domain users, to enable long-term analysis of patterns for multiple CSDS Use Cases. Data Warehouses typically handle thousands or millions of queries a day. Data inside a Data Warehouse has its own schema that determines how the structured data is organized for optimal query performance. Individual stakeholders make the warehoused Cyber Information available to other authorized stakeholders using an Information Messaging System. Stakeholders may in turn cache interesting, published data within their local Data Warehouse for more efficient searches.

2.3.6 Information Exchange Messaging System

The CSDS AAF may use an IEMS for the exchange of Cyber Information between the CACs. One of the well-known examples of such systems is the Publisher/Subscriber model.

The Publisher/Subscriber model generally has the main elements shown in Figure 20 below. The key components include a Publisher, Message Broker, Topic, and Subscriber.

- **Publisher**: A software-based messaging client that encodes a message and sends it to a Message Broker with a specific topic.
- **Message Broker**: A software-based service that retrieves encoded messages from authorized Publishers and stores them in the correct Topic queue.
- **Topic**: A logic queue-like partition within the Message Broker to segment and separate various messages that are published.
- **Subscriber**: A software-based messaging client that retrieves messages from authorized Topics and decodes the message to be used in the future.
- **Configuration**: A set of rules or policies that keep track of various Publishers and Subscribers and the Topics they have access to.

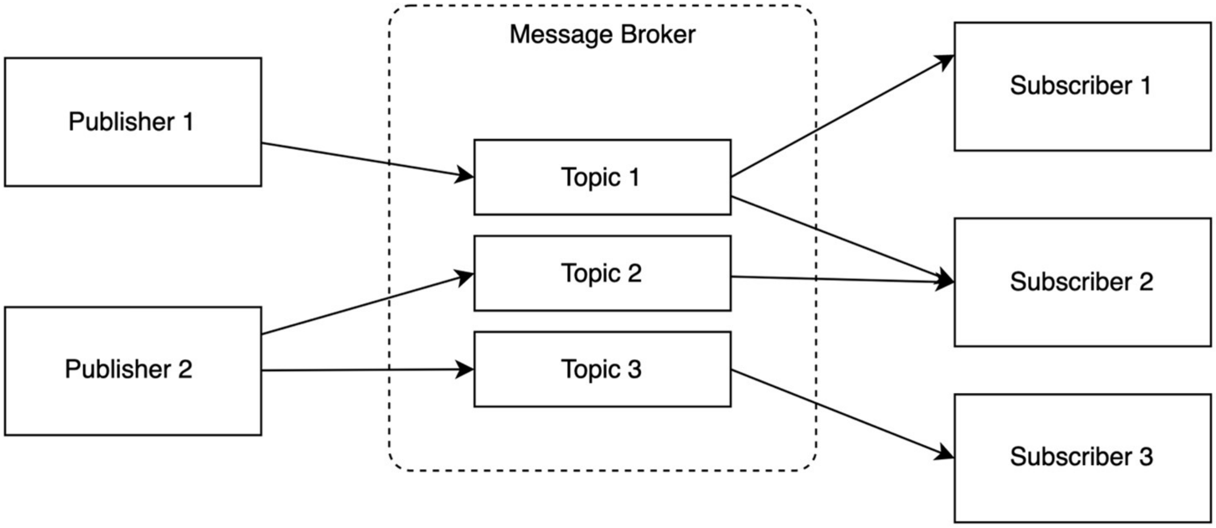


Figure 20. Publisher/Subscriber Model

2.3.6.1 Inter-Domain Interfaces

This research recommends a Centralized Messaging Broker Service, like the System Wide Information Management System (that is currently in place). This approach greatly reduces the amount of development overhead required by each CAC to (1) implement their own Messaging Broker Service and (2) integrate with the Messaging Broker Services of other CAC.

An independent 3rd party will be required to operate and maintain the Messaging Broker Service as well as put business processes in place to onboard/offboard new CAC and Security Operations Centers as they begin to adopt the CSDS AAF. The 3rd party will be responsible for creating and destroying topics used by CACs to publish and subscribe to messages.

The 3rd party will also be responsible for generating credentials that provide access to the messaging system, as well as revoking credentials to remove access.

The Centralized Messaging Broker Service will operate as a traditional IP Based system to appeal to the larger ecosystem that is primary IP Based. In the event CACs are located in ICS typed environments, the appropriate software adapter (translator) must be used to convert non-standard messages to IP.

2.3.6.2 Multi-Domain Communication Architecture

In a Multi-Domain CAC environment, information must be shared with two key goals in mind.

1. Share Information without disclosing confidential/proprietary/sensitive information to other CAC.
2. Share information without disclosing the CAC that originated the message.

Goal 1 can be solved using policies and procedures to ensure Human Analysts are reviewing and redacting confidential information before publishing messages.

Goal 2 is more difficult to accomplish but is doable with the implementation of a Multi-Domain CAC as a proxy.

Figure 21 shows the proposed setup of a Multi-Domain CAC and the message flows using 2 Domain Stakeholder CAC and 1 Multi-Domain CAC. A Message Broker is set up and under the control of the Multi-Domain CAC SOC that has full authority over the Publishers and Subscribers. CAC 1 will take Full Analysis Reports and redact any confidential/sensitive information from them to product Redacted Reports. The Redacted Reports are then published to the CAC 1 Topic provided by some Message Broker. CAC 2 does the same thing. The Multi-Domain CAC SOC will subscribe to both the CAC 1 and CAC 2 Topics and retrieve the Redacted Reports that were published. The Multi-Domain CAC SOC will combine both reports and perform high-level analysis (the analysis requires information for multiple CAC) and produce a Consolidated, Anonymized, Redacted Report that can then be published to all CAC that are subscribed to the Multi-Domain SOC Topic.

- **Artifact**: Results of an analysis produced by a toolset. The report will contain information that may contain sensitive/proprietary information that is used internally within the CAC only.
- **Shareable Artifact**: An Artifact with sensitive/proprietary information removed. During this process, sensitive information such as System ID’s, company names, user information is removed or replaced by random characters.

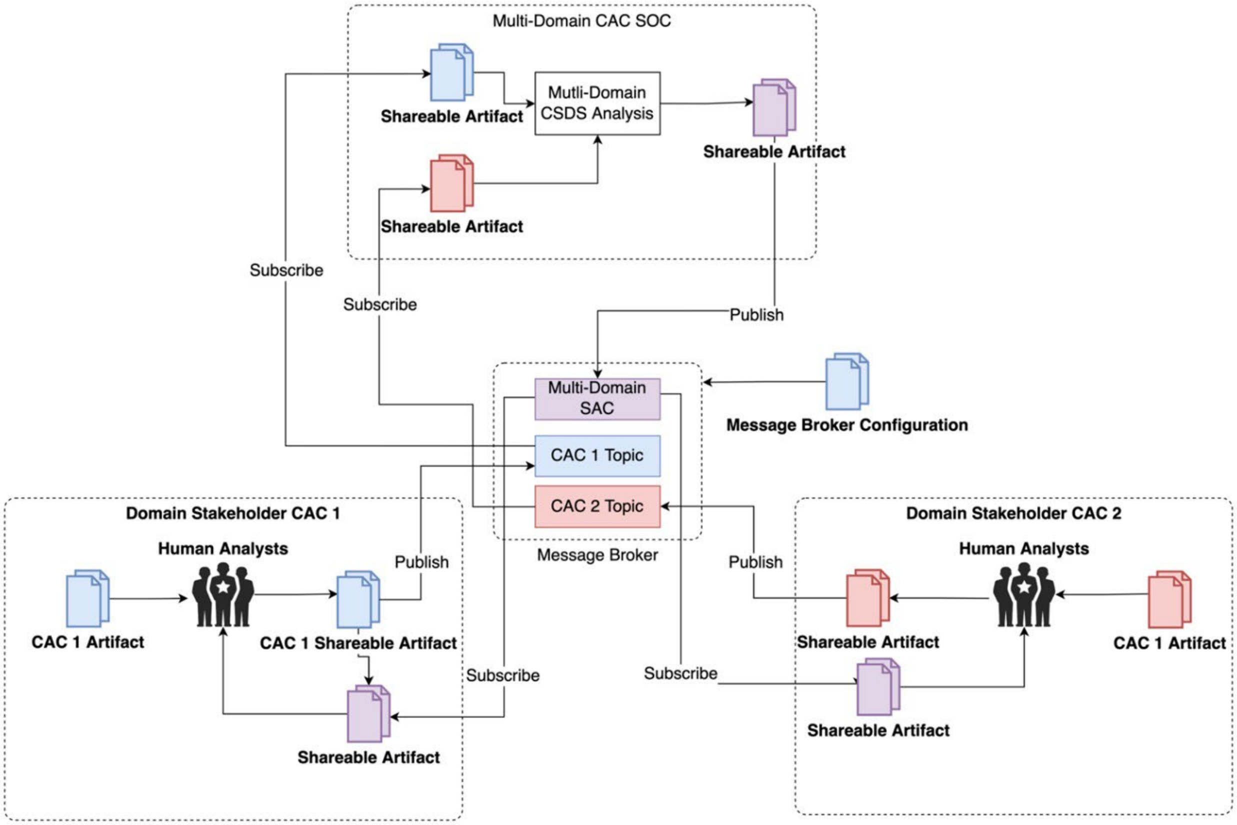


Figure 21. Multi-Domain CAC Architecture

2.4 CSDS AAF Data Architecture: The Data Perspective

This section builds on the concepts of Section 2.3, which discusses the AAF from a systems architecture perspective and how various system components are connected and interact more at a physical, network level. With the system architecture in place, it is now possible to show how the CSDS AAF Data Life Cycle (Figure 22) process can be implemented, and how the key CSDS AAF Data Life Cycle phases (Acquire, Pre-Analyze, Collect, Curate, Advanced Analytics, Information Sharing) are supported from a Data Perspective. This section will also discuss potential AI/ML integration opportunities to improve overall quality and efficiency of CSDS.

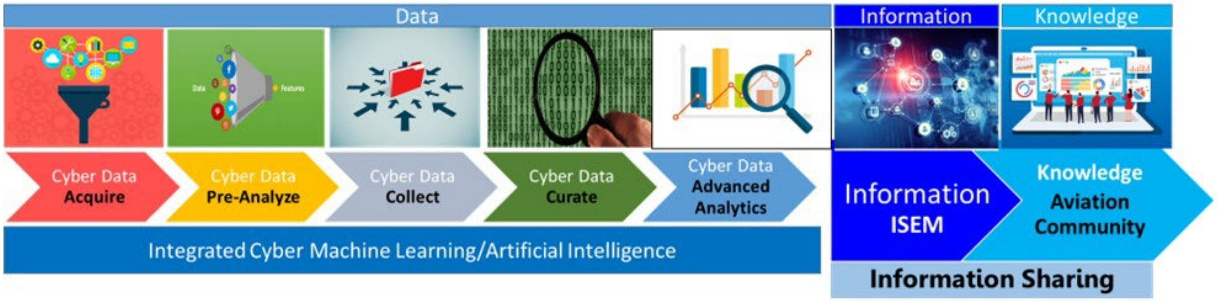


Figure 22. CSDS AAF Data Life Cycle

This section also introduces the concepts of Data Governance and the Common Operating Picture (as illustrated in Figure 23) and shows how these can be provided using careful construction of CSDS Use Cases and Shareable Artifact data specifications. With Data Governance and Common Operating Pictures in place, Human Analysts at a Domain Stakeholder or Multi-Domain CAC level may communicate and collaborate consistently, even facilitating the development of analytical toolsets and visualization technologies to support real-time cyber threats intelligence capabilities.

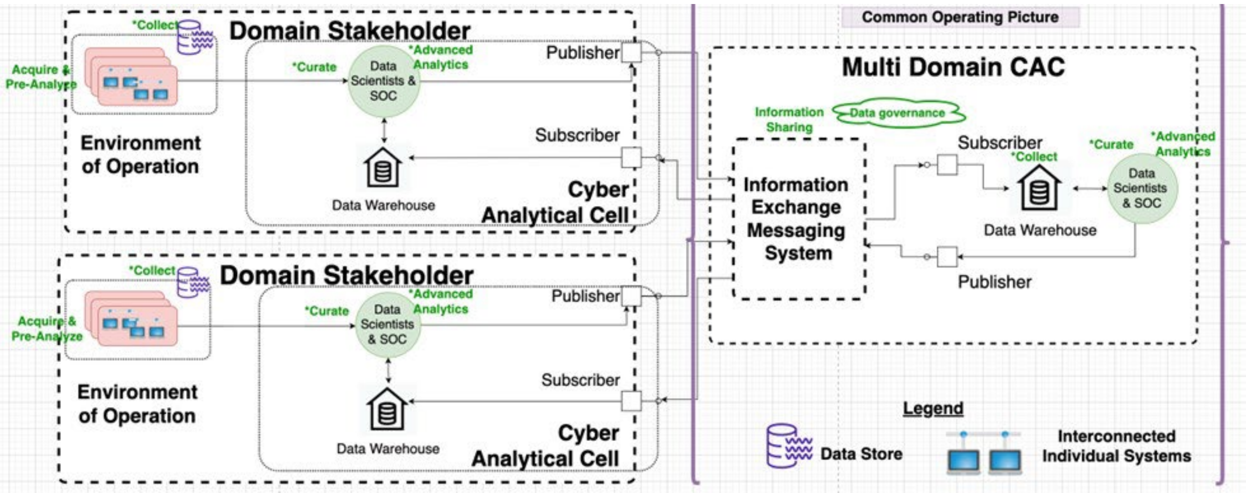


Figure 23. CSDS AAF Operational Concept Diagram including Data Governance & Common Operating Picture

Figure 23 also illustrates (in green text) where various phases of the CSDS AAF Data Life Cycle take place with respect to the proposed CSDS AAF System Architecture. The Acquire Phase occurs between the CAC and the Data-Stores. The Curate and Advanced-Analytics phase takes place within the CAC. The results that have been generated throughout the CSDS AAF Data Life cycle will be sent to the Data Warehouse for storage. Information Sharing occurs with the IEMS that brokers communication between the Domain Stakeholders and Multi-Domain CAC and is used to communicate Shareable Artifacts among authorized subscribers.

Multi-Domain CAC's will be responsible for defining Data Governance requirements for the IEMS, where they will enforce what types of information are permitted. Data Governance is discussed in more detail in Section 2.4, and Shareable Artifacts are in Section 2.4.4.

2.4.1 Data Relevancy and the Data Sphere

From the context of CSDS, it is imperative that the relevant data and their associated data types are acquired from systems and networks to ensure success in accomplishing three (3) primary CSDS objectives:

1. Is there a cyber-event pending?
2. Is there an attack occurring now?
3. Was an incident/event caused by cyber activity?

A Data Sphere (Figure 24) can be defined as the set of all data that is acquired from IIS, of which typically only a small portion of that acquired data is collected into the Data-Store. If the Data Sphere contains incorrect or insufficient data as an input into the CSDS process, it may lead to various issues downstream, such as:

1. Slow performance as analytical toolsets process large amounts of data unnecessarily.
2. Skewed results as analytical toolsets and AI/ML systems create biases and faulty conclusions.
3. Inability to produce meaningful results due to high levels of noise and unrelated data.
4. Excessive/unnecessary configuration and re-configuration of network monitoring devices and IIS that acquire data which may put unnecessary strain on networks and systems engineers.

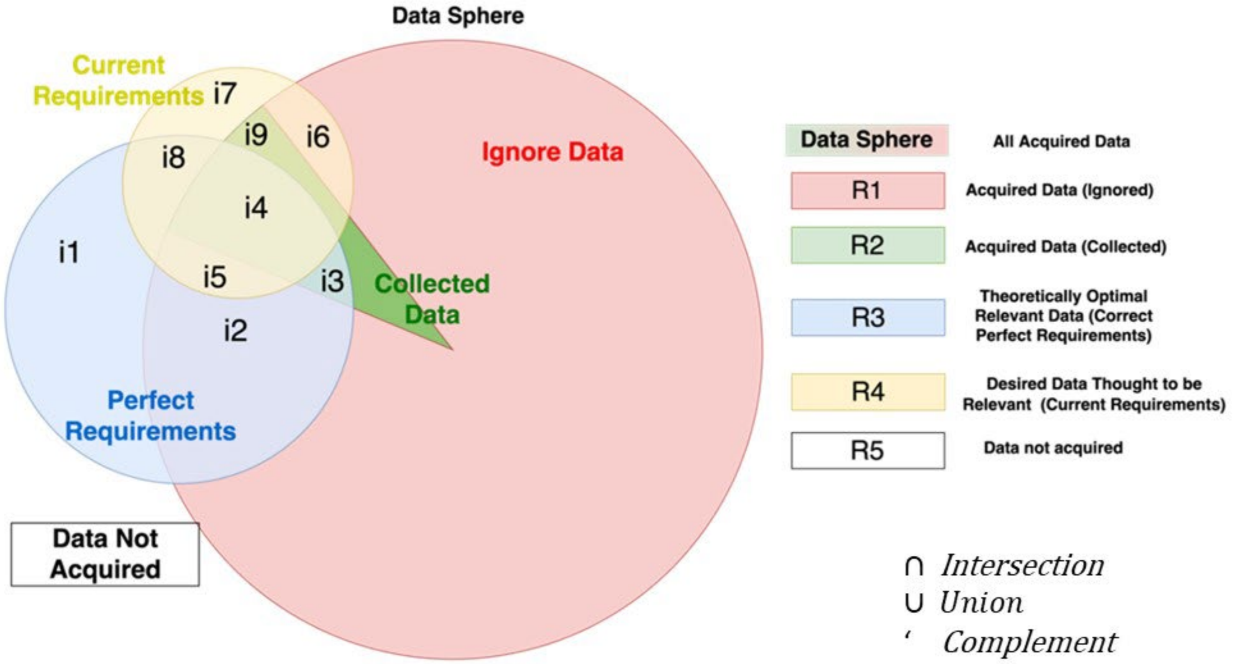


Figure 24. The Data Sphere, Regions, and Intersections

The determination of the correct data requirements (relevant data) for CSDS is not trivial. Figure 24 shows the diagram of an Operational Data Sphere depicting five (5) regions and nine (9) intersections. The bullets below provide a detailed explanation:

1. The Data Sphere $[R1 \ \cup R2]$ represents all acquired data within a given Stakeholder's Environment of Operation from IIS. For data to be acquired from these systems, a Data Acquisition Sensor must be integrated into the IIS. As more Data Acquisition Sensors are integrated into a Domain Stakeholder's Environment of Operation, the Data Sphere will expand accordingly.

2. $R1$ represents all data that has been Acquired by Data Acquisition Sensors but ignored due to certain Pre-Analysis logic. $R1$ will grow or shrink in size as re-configuration to pre-analyzers occur.

3. $R2$ represents all data that has been Acquired and Collected due to certain pre-analyzer logic that deems it "potentially relevant data". This data is said to be "Available" for CSDS and will require the CAC to correctly extract the available data from the Data-Store based on the 4 Extraction Modes discussed. $R2$ will grow or shrink in size as re-configuration to pre-analyzers occurs.

4. $R3$ is a theoretical region representing the “Optimal data” for a specific CSDS Use Case. $R3$ indicates the correct CSDS data requirements for a particular Use Case (i.e., what data must be collected). It is important to understand that the Optimal data for CSDS is not always obvious, known at the start of a CSDS effort, or may not be possible to discover. Via adjustments to the yellow region, it typically takes multiple iterations of trial and error by Data Scientists and Human Analysts to determine the Optimal Data.

5. $R4$ represents data that the CAC desires to extract and curate for a specific CSDS Use Case (e.g., Malware Detection, Intrusion Detection, LMD, Spam Filtering, DDoS Detection, etc.). This represents the data that is part of the current data requirement for doing CSDS, but these requirements have not yet been validated as being part of the theoretically optimal $R3$. Note that just because a CAC considers the data requirements to be optimal and therefore desires this data, it does not mean the data is actually optimal (i.e., the requirements have not yet been proven to be valid), see Figure 33.

6. $i1\ [R3\ \cap\ \text{Data Sphere}']$ represents an intersecting region in which the Optimal data has not yet been acquired. To get this data, changes to the IIS configuration are required to acquire, pre-analyze, and collect this. This can often be an expensive and time-consuming process to accomplish.

7. $i2\ [R1\ \cap R3]$ represents an intersecting region in which the Optimal data is being Acquired but has been ignored due to faulty/incorrect pre-analyze logic. To address this problem, the Pre-Analyzer in the Data Acquisition Sensors must be reconfigured to account for this new data.

8. $i3\ [R2\ \cap R3]$ represents an intersecting region in which the Optimal data is being collected, but the CAC has not yet recognized/identified the data as relevant to the specific CSDS Use Case. To fix this, requires a process-driven scientific approach by Data Scientists to assist them in recognizing that available data is missing from the analysis, see Figure 33. For near-term CSDS Use Cases, this represents the Optimal Data Set that is most useful to current efforts.

9. $i4\ [R2\ \cap R3\ \cap R4]$ represents an intersecting region in which the theoretically optimal data has been identified for a specific CSDS Use Case and is actively being collected/extracted by the CAC for conducting advanced analytics. This is the best outcome.

10. $i5\ [R3\ \cap R4\ \cap\ \text{Data Sphere}\ \cap R2']$ represents an intersecting region in which the CORRECT data has been identified to be relevant for a specific CSDS Use Case but is currently not being collected due to faulty/incorrect pre-analyze logic.
11. $i6\ [R3' \cap R4\ \cap\ \text{Data Sphere}\ \cap R2']$ represents an intersecting region in which data that has been recognized/identified as relevant for a specific CSDS Use Case is not the CORRECT data to be used. To fix this, requires a process-driven scientific approach by Data Scientists to assist them in recognizing that some data being analyzed is not needed and should be removed, see Figure 33.
12. $i7\ [R4\ \cap\ \text{Data Sphere}' \cap R3']$ represents an intersecting region in which data that has been recognized/identified as relevant for a specific CSDS Use case is not being acquired. $i7$ is detrimental to CSDS efforts and the business as expensive and time-consuming changes to systems will be made to acquire NEW data that is not optimal for the specific CSDS Use Case. (i.e., wrong requirements)
13. $i8\ [R4\ \cap\ \text{Data Sphere}' \cap R3]$ represents data that is not being acquired and has correctly been identified as a CSDS requirement.
14. $i9\ [R2\ \cap R3' \cap R4]$ represents data that is being collected and has incorrectly been identified as a CSDS requirement.

This view of the Data Sphere and relevant data poses two major challenges for the Aviation Architectural Framework:

- How can the Aviation Architectural Framework be used to guide Human Analysts and Data Scientists in generating the correct set of requirements for CSDS?
- How can the Aviation Architectural Framework be used to assist network and systems engineers in implementing/re-configuring Data Acquisition Sensors to increase the amount of Available Relevant Data as much as possible?

Theoretically Optimal Relevant Data and Relationship to Use Cases & Threat Scenarios

The Theoretically Optimal Relevant Data represents the theoretically correct data requirements which provides the CSDS Use Cases the best chance of success when analyzed and provides the most accurate results when answering the 3 key CSDS questions 1) Is there a cyber-event pending? 2) Is there an attack occurring now? 3) Was an incident/event caused by cybersecurity? Note that it is possible, and even likely, that a current Data Sphere does not contain the full set of

Theoretically Optimal Relevant Data, and this data gap may be an area to consider for new Environment of Operation design requirements.

The Theoretically Optimal Relevant Data region is an important concept to understand because its location with respect to the Data Sphere is highly dependent on the CSDS Use-Cases and Threat Scenarios in scope of ongoing CSDS efforts for a Domain Stakeholder. For example, if the CSDS Use Case was to provide Intrusion Detection of an Aircraft Wi-Fi network, only a small subset of the Airlines' total available data would be relevant.

Figure 25 below illustrates how multiple Theoretically Optimal Relevant data regions may exist within a given Data Sphere and is based on the given Threat Scenario and CSDS Use Case definitions. It is possible for these regions to intersect, which simply means that there are common relevant data among the Threat Scenario/Use-Cases. Defining CSDS Use-Cases and Threat Scenarios that create these intersections are desirable, as it means that the same data can be re-used multiple times, increasing the efficiency and effectiveness of CSDS efforts.

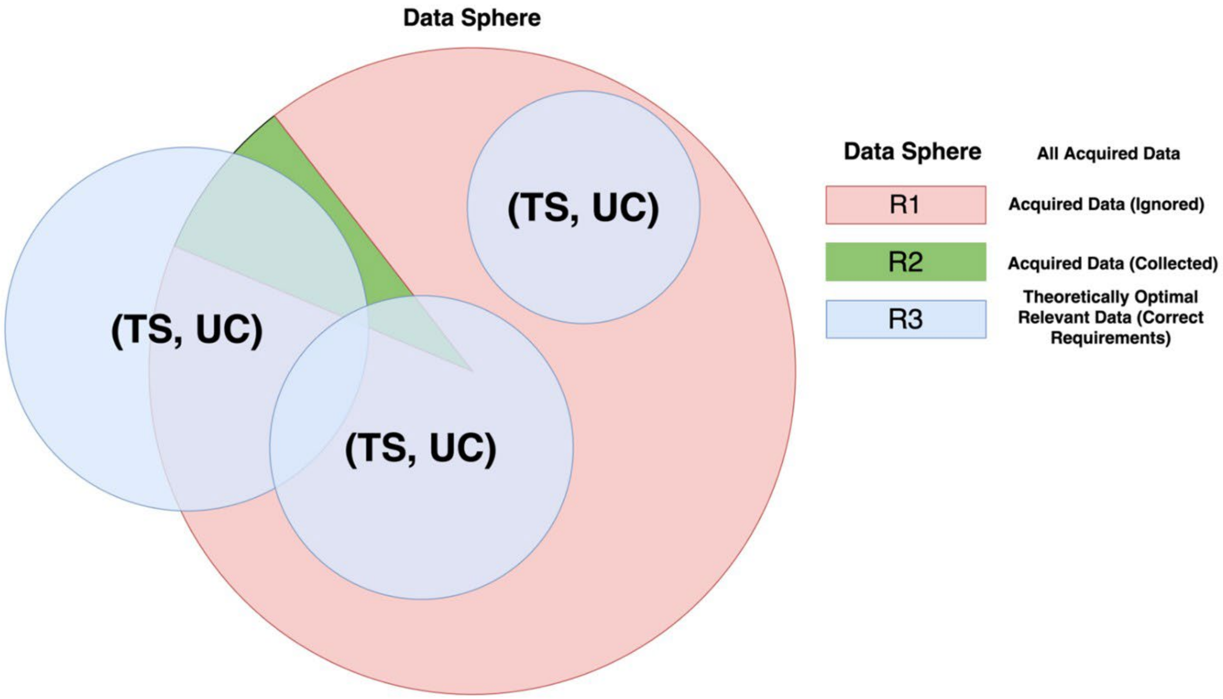


Figure 25. Relationships between Data Sphere and Theoretically Optimal Relevant Data

2.4.2 CSDS AAF Data Life Cycle Data Flows

The elements of the CSDS AAF Data Cycle (see Figure 26) are explained in more detail below.

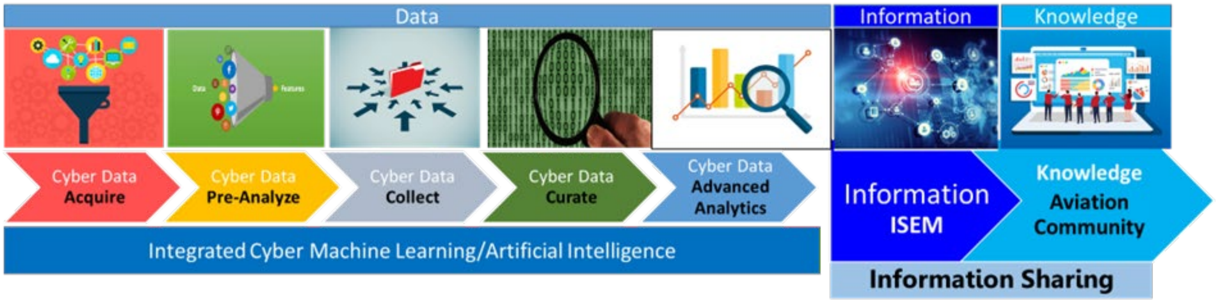


Figure 26. CSDS AAF Data Cycle

2.4.2.1 Acquire Phase

The primary goal of the Acquire Phase is to integrate/configure Data Acquisition Sensors into IIS so that processes running on those systems, and therefore the data they produce, can be monitored and pre-analyzed. Table 3 below illustrates an example of the Data Requirement Specification format.

Table 3. Example of Data Requirement Specification format

IIS and Category Types	**Processes**	**Data to Acquire**

AI/ML Considerations

The researchers have not identified opportunities for AI/ML during the Acquire Phase, given its reasonably straightforward, rudimentary nature.

2.4.2.2 Pre-Analyze Phase

The Pre-Analyze Phase seeks to take the data acquired in the previous phase and determines if it is appropriate to be collected based on a set of pre-defined rules.

The Pre-Analysis process is illustrated in Figure 27 below. Acquired Data first goes through a Data Type Detection process where Human Analysts or automated software tries to determine the type of the data. For example, it could be a network log or an application log. If it is a network log, the network type could be identified.

Once the data has been detected, a set of pre-analysis tasks are executed. This can be done manually by Human Analysts using specific toolsets or through automation scripts developed by Data Scientists.

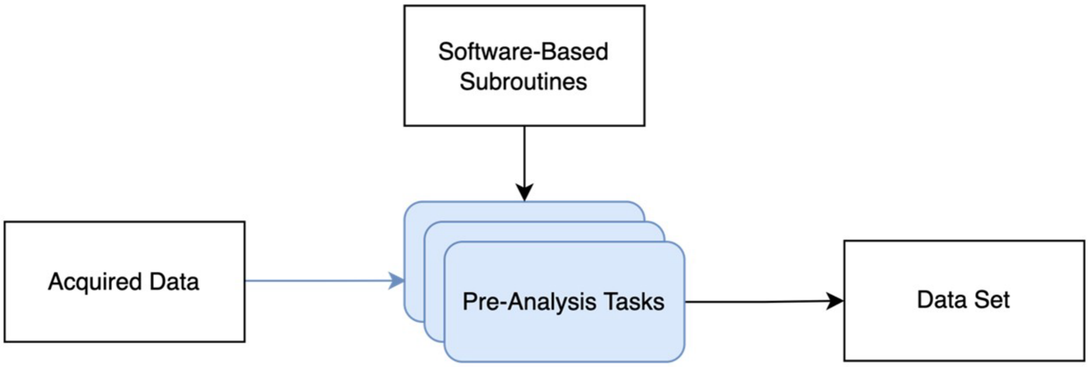


Figure 27. The Pre-Analysis Process

AI/ML Considerations

Opportunities for AI/ML application are plentiful in the Pre-Analysis phase. Most of this report addresses AI/ML at the CAC layer. It is recognized that in time, it would make sense to incorporate AI/ML capabilities at the edge. The pre-analysis phase will be a good first target for this implementation. Today this is not a capability that is used in general and is not a technique that would currently be certified on an aircraft. The following are areas which AI/ML can assist in the Pre-Analysis phase.

- Data Type Classification- AI/ML can be developed to classify certain data or detect unique signatures automatically. This may assist Data Scientists in developing pre-analysis tasks to determine if there is enough coverage based on the types of data acquired.
- Meta Data Tagging – Descriptive terms that can be added in the form of meta data tags to provide context and additional meaning to that data. This meta data can be useful in data findability for future analysis. AI/ML can be utilized to automate the process of adding helpful meta data tags.
- Smart Filtering – Not all data is useful in analysis, and unnecessary data can hinder analysis. AI/ML can help to identify unnecessary data and filter out the data which prevents unnecessary data collection.

2.4.2.3 Collect Phase

The collect phase aims at storing Data Sets in various Data-Store elements where they can be accessed by Human Analysts, Data Scientists, and automated toolsets that need them. Similar to the Acquire Phase, a data management strategy is required to ensure data is organized and searchable. An example of data organization is shown in Table 4.

Data requirements include proper versioning of the data sets so that they can be tracked over time, a log of pre-analysis tasks performed on the data as well as a reference to the original, acquired data that was used to generate that data set. Note that the latter can only be done if acquired data is being stored somewhere, although this may have a great impact on data storage requirements.

Table 4. Log Formatting example

Data Set ID	**Version ID**	**Creation Date**	**Reference to Original Raw Data**	**Meta Data/Tags**	**Data Set**

AI/ML Considerations

The use of AI/ML in managing the Collect Phase associated with data across all the Data-Store elements of a business is a complex potential opportunity. AI/ML application to Data-Store data collection, management, and retention (e.g., data velocity and volume) strategies could be worth further investigation.

2.4.2.4 Curate Phase

The Curate phase is the beginning of the CAC’s CSDS processes. Since Data-Store elements are decentralized, Human Analysts and Data Scientists must be aware of all the “pockets” of data that are accessible to them. The CACs implement strategies to extract the Desired Relevant Data from Data-Store elements (pulling them into the CAC for analysis) either manually or via software automation.

The CSDS Use Case provides context to the CSDS Data Set required to Curate the data.

When data is extracted from the Data-Store, a data management strategy is required to organize the data appropriately. This ensures that each piece of data is traceable back to a single origination point and ensures that important network and system characteristics can be measured accurately and monitored over time. This becomes relevant during a cyber event where Human Analysts need to determine the source of the attack.

Data that is curated is expected to be stored in the CAC's Data Warehouse. Table 5 below shows example key metadata to be included with the curated data stored in the Data Warehouse for successful CSDS application of the upcoming Advanced Analytics phase.

Table 5. Example of Metadata

Environment of Operation	**IIS ID**	**Acquisition Timestamp**	**Curated Data**

Figure 28 illustrates the Curation process in more detail. Data Sets are used as an input along with a Particular Use Case. For example, Lateral Movement Defense in a manufacturing environment may be a particular CSDS Use Case threat scenario. The Curation process generates Data Models that will be used as an input for the Advanced Analytics phase.

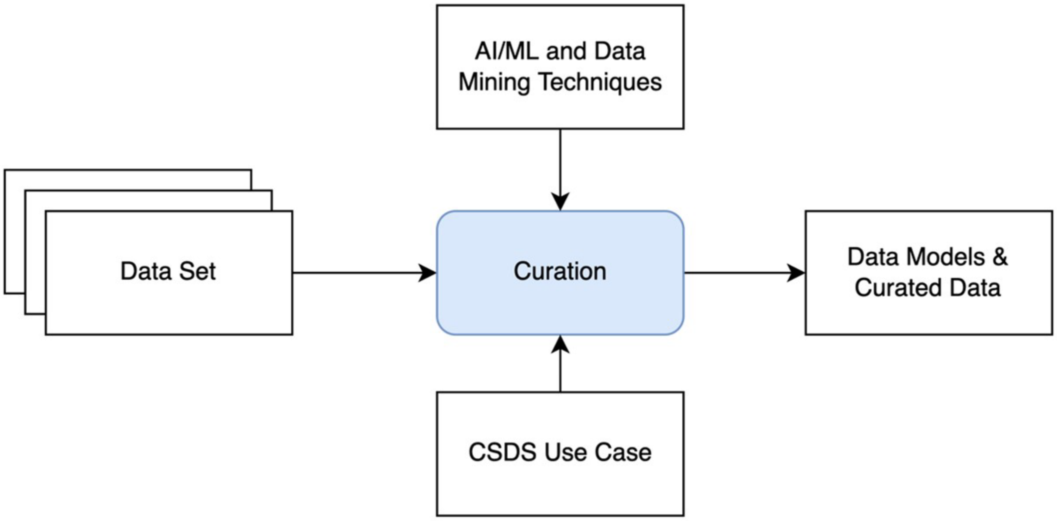


Figure 28. The Data Curation Process

AI/ML Considerations

There are many opportunities for applying AI/ML to the Curation Phase. In particular, the Curation Phase is most appropriate for Data Mining techniques (https://bootcamp.pe.gatech.edu/blog/10-key-data-mining-techniques-and-how-businesses-use-them/)

- Clustering
  - Partitioning
  - Hierarchical Method
  - Density-Based method
  - Grid-Based method
  - Model-Based Method
- Association
  - Single Dimensional Association
  - Multi-Dimensional Association
- Classification
  - Logistic Regression
  - Decision Trees
  - K-Nearest Neighbors
  - Naïve Bayes
  - Support Vector Machines
- Outlier Detection
  - Numerical Outlier
  - Z-Score
  - DBSCAN
  - Isolation Forest
- Prediction
  - Forecast Modeling
  - Classification modeling
  - Clustering Modeling
  - Time Series Modeling

2.4.2.5 Advanced Analytics Phase

The Advanced Analytics phase uses one or more Data Models generated from the curation process and generates cyber analytical information to accomplish CSDS Use Case Objectives, as illustrated in Figure 29. This phase also generates visualization and other graphics that are more appropriate for consumption by business executives and other functional roles within the Stakeholder environment.

Data Visualization and Graphics may include:

- Comparison Charts
- Maps
- Heat Maps
- Density Plots
- Histograms
- Network Diagrams
- Scatter Plots

Advanced Analytics may also generate security recommendations and alerts that Human Analysts should execute actions. These forms of analytics are more technical in nature and usually involve a deep understanding of the information system and the Use Case in question.

Advanced Analytics is accomplished by using specialized Cyber Analytical Toolsets designed for specific purposes.

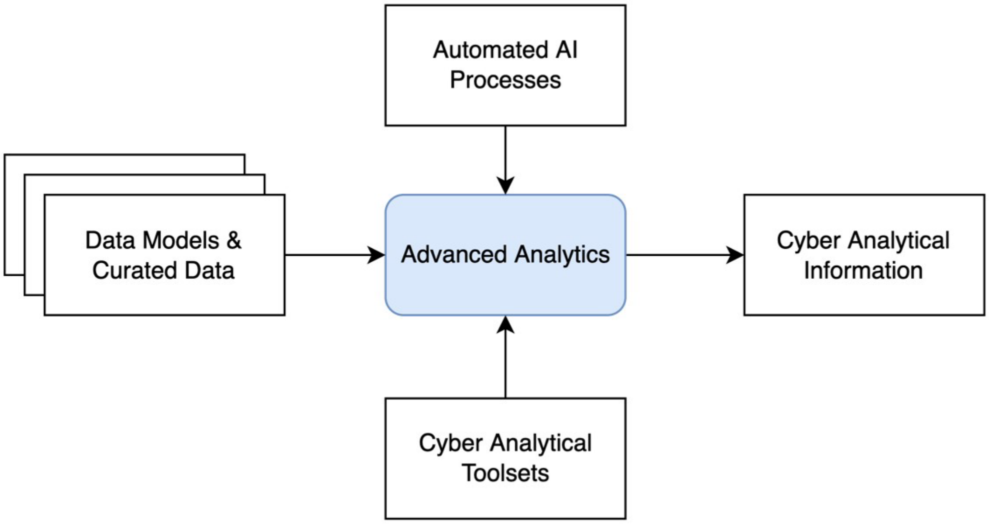

Automated AI Processes
Data Models & Curated Data
Advanced Analytics
Cyber Analytical Information
Cyber Analytical Toolsets

Figure 29. The Advanced Analytics Process

2.4.2.6 Information Sharing Phase

The Information Sharing Phase uses one or more pieces of Cyber Analytical Information to create Shareable Artifacts, as illustrated in Figure 30. This is governed primarily based on information-sharing policies and regulations put in place by National regulators or the Domain Stakeholders themselves. Shareable Artifacts are also constrained by a specific specification that information must conform to to be valid.

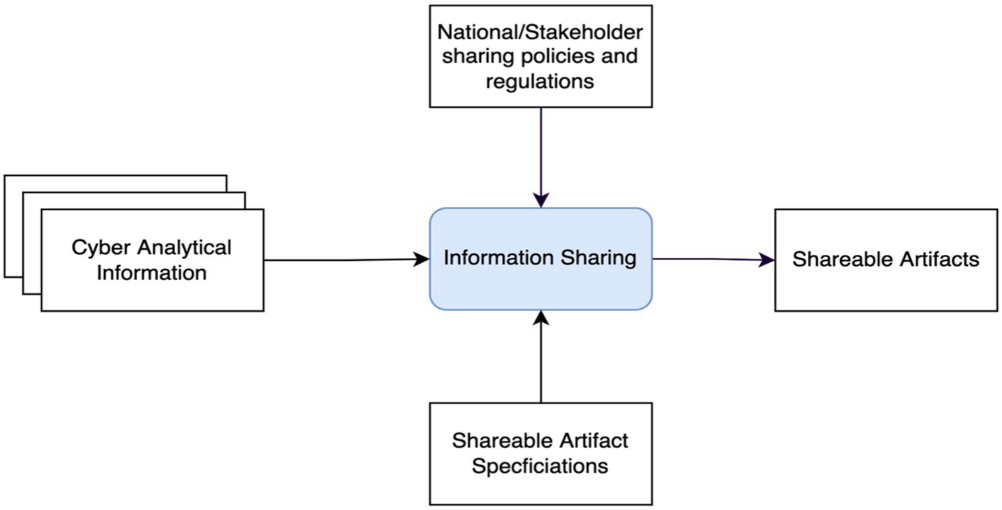

National/Stakeholder sharing policies and regulations
Cyber Analytical Information
Information Sharing
Shareable Artifacts
Shareable Artifact Specficiations

Figure 30. Information Sharing Process

2.4.3 The CSDS AAF Data and Information Flows Summary

Figure 31 below shows the entire Data Flow from Acquired Data to Shareable Artifacts which encompass all phases of the CSDS Data Life Cycle. With the complete picture, it can now be shown what phases will be under the jurisdiction of a business's engineers and the CAC team.

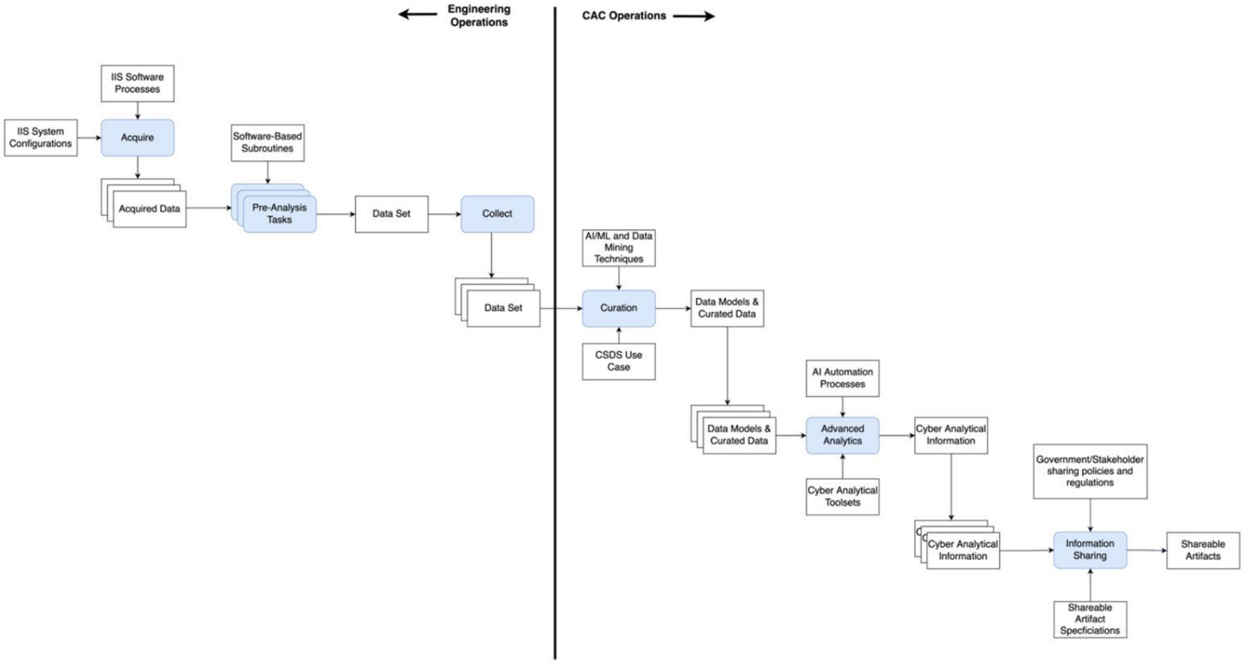

Engineering Operations
CAC Operations
IIS Software Processes
IIS System Configurations
Acquire
Acquired Data
Software-Based Subroutines
Pre-Analysis Tasks
Data Set
Collect
Data Set
AI/ML and Data Mining Techniques
Curation
CSDS Use Case
Data Models & Curated Data
Data Models & Curated Data
AI Automation Processes
Advanced Analytics
Cyber Analytical Toolsets
Cyber Analytical Information
Cyber Analytical Information
Government/Stakeholder sharing policies and regulations
Information Sharing
Shareable Artifact Specficiations
Shareable Artifacts

Figure 31. Entire Data Flow from Acquired Data to Shareable Artifacts

2.4.4 Shareable Artifacts

Shareable Artifacts may come in a wide variety of formats and data types and may have different purposes based on where in the CSDS life cycle the artifact was generated or where it will be used in the next step of the life cycle.

Why are Shareable Artifacts Important?

Shareable Artifacts allow for faster collaboration and progress within the aviation community to accomplish mutual CSDS Objectives. Traditionally, Domain Stakeholders would only have access to their data when doing data analytics. Shareable Artifacts allow Domain Stakeholders to work together and build on each other's progress to form a common body of knowledge. The CSDS AAF acknowledges the fact that Domain Stakeholders may not have the cyber capability to establish a full-fledged CSDS program. Shareable Artifacts considers this fact and allows stakeholders to share "what they can".

The goal of CSDS AAF is to apply CSDS tools and techniques to big data to gain Cyber-Analytical Information. The journey from raw data to valuable information, however, is complex and necessitates many intermediary stages reflected in the Pre-Analyze, Curate, and Advanced Analytics phases of the CSDS AAF Data Life Cycle. Shareable Artifacts represent the outputs of these intermediary stages with an added constraint that the outputs are deemed "shareable" with the rest of the Aviation cyber-community. Examples of Shareable Artifacts include:

1. Conformed Data
2. Modeled Data (generated from AI/ML techniques)
3. Data Sets (generated from AI/ML techniques)
4. Statistical Reports
5. Cyber-Analytical Reports
6. Machine Learning Models/Algorithms
7. Custom Software Toolsets
8. Standardized Messaging and Notifications
9. Reportable Events (i.e., per RTCA/EUROCAE cybersecurity means of compliance documentation)

Shareable Artifacts Considerations

1. Shareable Artifacts must be well-formed and comply with Data Governance policies set by Information Exchange governing body. This may require mapping data fields/properties to ensure alignment with what is expected from a Multi-Domain CAC point of view.
2. Shareable Artifacts must be sanitized and redacted from all confidential and sensitive information.
3. Shareable Artifacts must be approved for distribution by a designated Data Custodian.

A Data Custodian is an employee that belongs to a Domain Stakeholder CAC or Multi-Domain CAC. Their job is to ensure Shareable Artifacts are formatted to meet acceptable data governance guidelines.

A Shareable Artifact is the product of the CSDS AAF Data Life Cycle process, with the added constraint that the artifact is cleaned and redacted all company-sensitive information. With this pre-requisite, the artifact is fit for distribution throughout the aviation community. The competitive advantage of this collaboration can now be realized as stakeholders have more information about what's happening to advance their own cyber-analytical capabilities.

2.5 CSDS AAF Use Cases

2.5.1 CSD AAF Use Case Development

Use Cases can be mapped to the CSDS AAF for different Environments of Operations in the aviation ecosystem. Generically, a Use Case is a list of actions or event steps typically defining the interactions between an actor and a system to achieve a goal. The actor can be a human or another external system. The development process for a CSDS Use Case involves selecting and defining the Environment of Operation intended for application, identifying Threat Scenarios, and documenting the Use Case in the context of CSDS AAF.

2.5.2 CSDS AAF Use Case Lifecycle

A Use Case follows the lifecycle of the Environment(s) of Operation it describes, which includes the associated systems and/or product. This Use Case is the primary systems engineering tool for defining, verifying and re-defining the CSDS requirements that need to be implemented within the target Environment(s) of Operation. Figure *32* describes a 5-stage Use Case lifecycle process which starts with Use Case Development where the "CSDS Desired Relevant Data Requirements" are understood and defined.

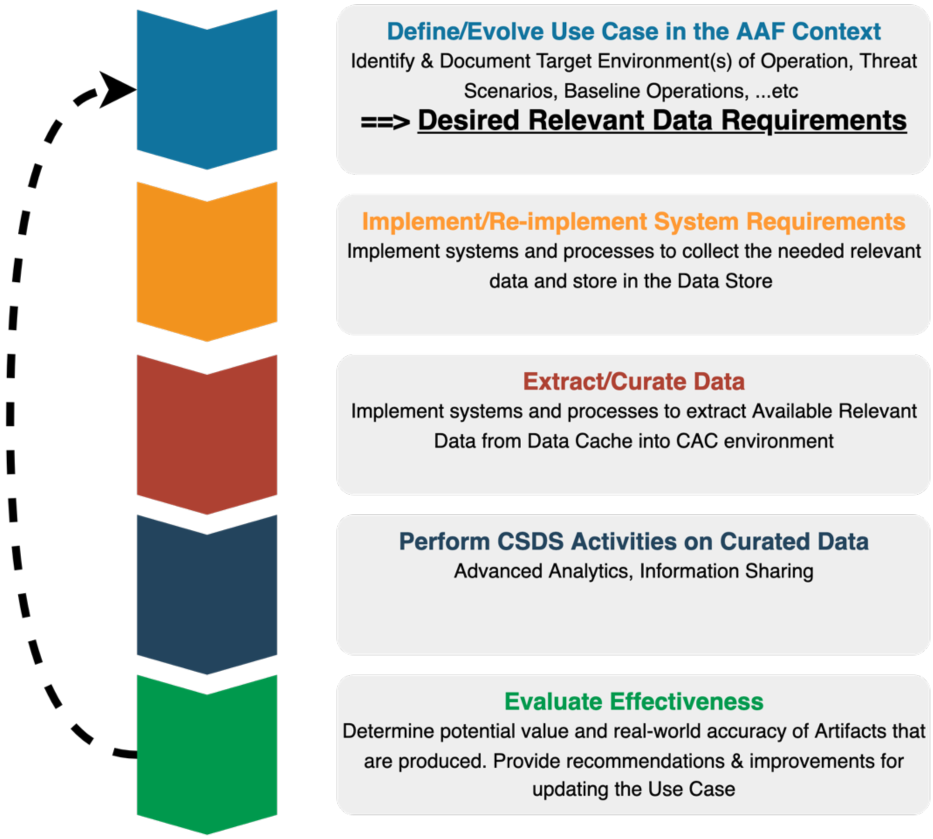

Define/Evolve Use Case in the AAF Context
Identify & Document Target Environment(s) of Operation, Threat Scenarios, Baseline Operations, ...etc
==> Desired Relevant Data Requirements
Implement/Re-implement System Requirements
Implement systems and processes to collect the needed relevant data and store in the Data Store
Extract/Curate Data
Implement systems and processes to extract Available Relevant Data from Data Cache into CAC environment
Perform CSDS Activities on Curated Data
Advanced Analytics, Information Sharing
Evaluate Effectiveness
Determine potential value and real-world accuracy of Artifacts that are produced. Provide recommendations & improvements for updating the Use Case

Figure 32. CSDS AAF Use Case Lifecycle

2.5.2.1 Stage 1: Define / Evolve Use Case in the AAF Context

Stage 1 is the primary systems engineering process for defining the CSDS "Desired Relevant Data Requirements". Stakeholders define the CSDS Use Case, which includes possible Threat Scenarios within the Environment(s) of Operation. These scenarios are the basis for understanding the CSDS relevant data requirements. The involved systems and their operation must be defined in order to identify what Desired Relevant Data is already being collected within the target Environment(s) of Operation. From this, the CAC cyber systems engineers can define what CSDS requirement changes need to be committed within the business to meet their needs for the acquisition and collection of data from the Environment(s) of Operation.

2.5.2.2 Stage 2: Implement / Re-Implement System Requirements

This is the primary challenge for many cyber teams, as it necessitates the demonstration of the cyber systems engineers to show a viable Business Case to leadership. By demonstrating this, leadership could have a better view of the situation and spend the resources to change the processes and systems across the business so that the CAC can eventually get the Desired Relevant Data to meet their requirements. As is often the case with the cyber landscape of evolving threats and new technologies, the cyber requirements are difficult to define with accuracy, certainty, and longevity.

As illustrated in Figure 33, some requirement changes lead to process changes within the CAC or changes to which data is extracted from the existing Available Data Collected by the Business. These are typically the least expensive changes to implement. Other requirement changes require making systems changes to what the Pre-Analyzer chooses to Collect vs. discard. These requirement changes often only require the need to make configuration file changes on the appropriate individual systems within the Environment(s) of Operation, and are typically at a reasonable level of cost to implement, but the specific costs and timing for implementation are driven by Aviation Information Systems Types involved. Lastly, if the requirement changes necessitate the Acquisition of new data that is outside of the current Data Sphere. In that case, this will require systems changes that can be very significant, costly, and can take a long time to implement in an ICS/OT environment.

Once the requirements of the new system have been implemented, we are ready to start Acquiring/Pre-Analyzing/Collecting the new Desired Relevant Data of the CAC systems engineers/data scientists. It is worth noting that people, even engineers and data scientists, are not all knowing, so rarely is the Desired Data captured in the Use Case a perfect match to the Theoretically Optimal Relevant Data of a Perfect Requirements world.

Note that sometimes a requirement change may also require a major or minor systems change, and at other times a requirement change prevents a business from investing in implementing bad requirements, as illustrated in the $i7$ region of Figure 33 by removing bad requirements.

2.5.2.3 Stage 3: Extract / Curate Data

Now that the processes and systems have implemented the new CSDS AAF Use Case Desired Relevant Data requirements, the CAC Extracts and Curates new Collected data for CAC CSDS analysis.

2.5.2.4 Stage 4: Perform CSDS Activities on Collected Data

Domain Stakeholder CACs or Multi-Domain CADs then perform CSDS (Maintain, Process, and Analyze) activities to produce analysis Artifacts supported by the CSDS Use Case. The AAF assumes that most of the time will be spent at this stage of the CSDS execution efforts, and it is anticipated that it will require the most human involvement (data analyst, data scientist, systems engineer, etc.).

2.5.2.5 Stage 5: Evaluate Effectiveness

The final stage in the Use Case lifecycle requires the appropriate Stakeholders and Multi-Domain Stakeholders to determine potential value and real-world accuracy of artifacts that are produced and provide recommendations and improvements for updating the Use Case.

2.6 CSDS Data Sphere Engineering Requirements

Figure 33 incorporates both the CSDS AAF Data Life Cycle (Figure 3) and the Data Sphere (Figure 24). The phases of the Data Life Cycle are displayed across the top of the figure.

2.6.1 Data Sphere Regions

Underneath these phases are relative, though not necessarily to-scale, areas of the Data Sphere (indicated by the colored circles) that generally align to those phases. The Data Sphere represents all the acquired data (regions $R1$ and $R2$). During the Pre-analyze phase, $R1$ data is ignored, and $R2$ data is collected and available in the Data-Store. The $R4$ data represents the desired data described in the CSDS requirements. However, some of that desired data may not be available for various reasons. The final circle, colored with the gradient of $R2$ and $R4$ colors, represents the available desired data ($R2 \cap R4$). The amount of data decreases through the progression of the Data Life Cyle phases of Acquire, Pre-Analysis, Collect, and Curate.

2.6.2 Advance Analytics

The Advanced Analytics phase incorporates the Cyber Analytical Capability Reference Model (Figure 19) to include both the inputs and outputs. The inputs include the internal available desired data and the external threat intelligence feeds and shared artifacts. Process changes, such as changes to the techniques, tools, and models used by the CAC, is another category of inputs to the Advanced Analytics phase. The outputs of this phase are artifacts that are used internally evaluate CSDS requirements, toolsets, and analytics. A subset of these artifacts is sanitized to become community shareable artifacts.

2.6.3 Changing CSDS Requirements

The evolution of CSDS requirements over time, as documented and maintained in the Use Case, will follow a multi-feedback loop design as shown in Figure 33. Feedback is created because of the Artifact generation process in conjunction with CSDS retrospective exercises, toolset/AI evaluations and analytics evaluations that will trigger modifications in four feedback loops. These feedback loops include process changes, requirements changes, minor system changes, and major system changes, which increase in time and cost, respectively. For example, a change to the size of the Data Sphere (i.e., the data being acquired) requires major system changes like installing new sensors and associated infrastructure. This change falls on the most expensive end of the cost spectrum as shown in Figure 33. In contrast, modifications to the format of the extracted data are less expensive in terms of resources and time and does not require as much system change.

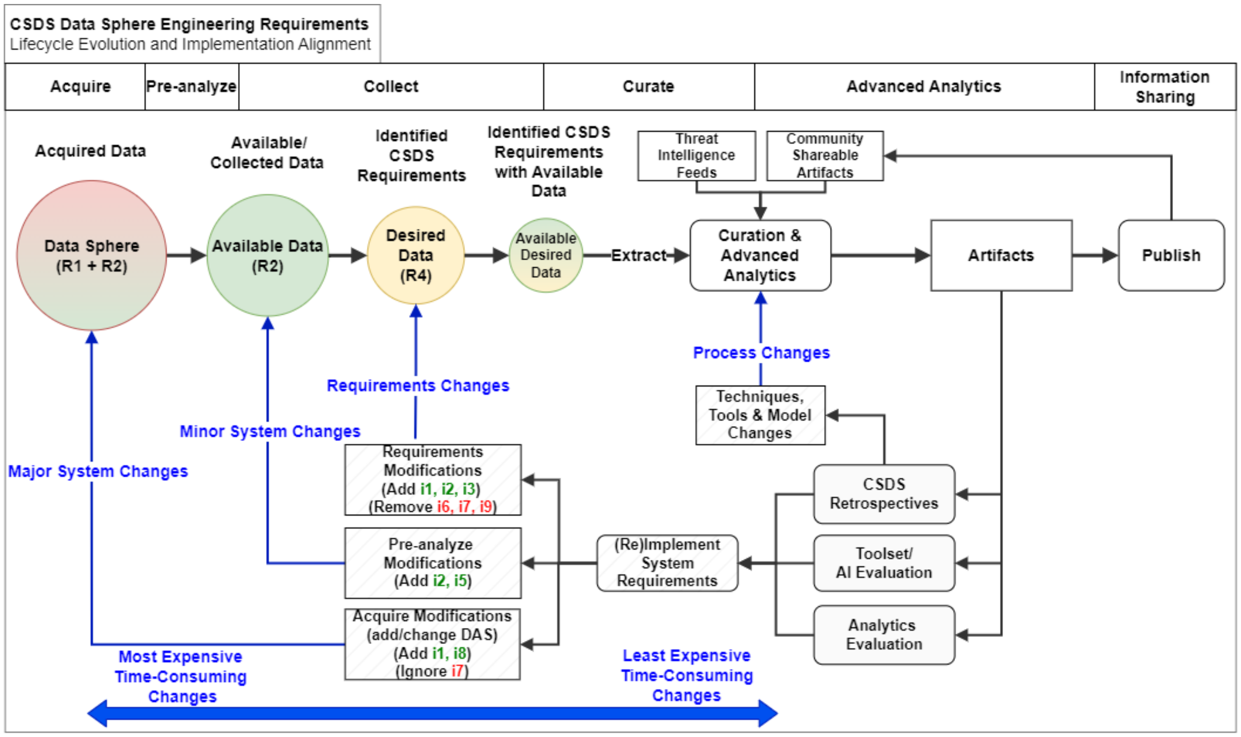

CSDS Data Sphere Engineering Requirements
Lifecycle Evolution and Implementation Alignment
Acquire
Pre-analyze
Collect
Curate
Advanced Analytics
Information Sharing
Acquired Data
Available/ Collected Data
Identified CSDS Requirements
Identified CSDS Requirements with Available Data
Threat Intelligence Feeds
Community Shareable Artifacts
Data Sphere (R1 + R2)
Available Data (R2)
Desired Data (R4)
Available Desired Data
Extract
Curation & Advanced Analytics
Artifacts
Publish
Process Changes
Requirements Changes
Minor System Changes
Major System Changes
Techniques, Tools & Model Changes
Requirements Modifications (Add i1, i2, i3) (Remove i6, i7, i9)
CSDS Retrospectives
Pre-analyze Modifications (Add i2, i5)
(Re)Implement System Requirements
Toolset/ AI Evaluation
Acquire Modifications (add/change DAS) (Add i1, i8) (Ignore i7)
Analytics Evaluation
Most Expensive Time-Consuming Changes
Least Expensive Time-Consuming Changes

Figure 33. CSDS Data Sphere Engineering Requirements Lifecycle Evolution and Implementation Alignment

3 References

Anunaya, S. (2021, August 10). *Data Preprocessing in Data Mining -A Hands On Guide.* Retrieved January 2022, from analyticsvidhya.com: https://www.analyticsvidhya.com/blog/2021/08/data-preprocessing-in-data-mining-a-hands-on-guide/

Aviation Research Advisory Committee (ARAC) Aircraft Systems Information Security/Protection (ASISP) Working Group. (2016, October 3). Retrieved from Final Report # B-H020-REG16-TLM-66: https://www.faa.gov/regulations_policies/rulemaking/committees/documents/media/ARACasisp-T1-20150203R.pdf

Cao, S. (2019, November 09). *What on Earth Is a Data Scientist? The Buzzword's Inventor DJ Patil Spills All.* Retrieved 2021, from Observer.com: https://observer.com/2019/11/data-scientist-inventor-dj-patil-interview-linkedin-job-market-trend/

Embry-Riddle Aeronautical University Center for Aerospace Resilient Systems. (2025). *Part 1: Cybersecurity Data Science (CSDS) Aviation Architecture Framework (AAF) - Overview & Value Proposition.* Atlantic City, NJ: FAA NextGen Cybersecurity Data Science Project.

Embry-Riddle Aeronautical University Center for Aerospace Resilient Systems. (2025). *Part 2: Cybersecurity Data Science (CSDS) Aviation Architecture Framework (AAF) - Technical Definition.* Atlantic City, NJ: FAA NextGen Cybersecurity Data Science Project.

Embry-Riddle Aeronautical University Center for Aerospace Resilient Systems. (2025). *Part 3: Cybersecurity Data Science (CSDS) Aviation Architecture Framework (AAF) - Implementation Guidance.* Atlantic City, NJ: FAA NextGen Cybersecurity Data Science Project.

Embry-Riddle Aeronautical University Center for Aerospace Resilient Systems. (2025). *Part 4: Cybersecurity Data Science (CSDS) Aviation Architecture Framework (AAF) - Glossary & Acronyms.* Atlantic City, NJ: FAA NextGen Cybersecurity Data Science Project.

FAA AFS-300. (2015, September 30). *AC 119-1 - Airworthiness and Operational Approval of Aircraft Network Security Program (ANSP).* Retrieved from https://www.faa.gov/regulations_policies/advisory_circulars/index.cfm/go/document.information/documentid/1028288

Google Cloud. (n.d.). *What is Data Governance.* Retrieved 2022, from cloud.google.com: https://cloud.google.com/learn/what-is-data-governance

Hamutcu, U. F. (2020, June 30). *Toward Foundations for Data Science and Analytics: A Knowledge Framework for Professional Standards.* Retrieved March 2021, from hdrs.mitpress.edu: https://hdsr.mitpress.mit.edu/pub/6wx0qmkl/release/3

IBM Cloud Education. (15, May 2020). *Data Science.* Retrieved March 2021, from ibm.com: https://www.ibm.com/cloud/learn/data-science-introduction

ICAO. (n.d.). *About ICAO.* Retrieved January 2022, from icao.int: https://www.icao.int/about-icao/Pages/default.aspx

ICAO. (n.d.). *Assembly.* Retrieved January 2022, from icao.int: https://www.icao.int/about-icao/assembly/Pages/default.aspx

ICAO. (n.d.). *ICAO Secretariet.* Retrieved January 2022, from icao.int: https://www.icao.int/secretariat/Pages/default.aspx

ICAO. (n.d.). *The History of ICAO and the Chicago Convention.* Retrieved January 2022, from icao.int: https://www.icao.int/about-icao/History/Pages/default.aspx

ISO. (2015, May). *Systems and software engineering — System life cycle processes.* Retrieved March 2022, from iso.org: https://www.iso.org/standard/63711.html#:~:text=ISO%2FIEC%2FIEEE%2015288%3A2015%20establishes%20a%20common%20framework,hierarchy%20of%20a%20system's%20structure

Mongeau, S. (2021, September). *Cybersecurity Data Science (CSDS): Emerging Trends.* Retrieved January 2022, from Resources.sei.cmu.edu: https://resources.sei.cmu.edu/library/asset-view.cfm?assetid=739704

NIST Computer Security Resource Center. (n.d.). *CSRC: Environment of Operation.* Retrieved January 2021, from csrc.nist.gov: https://csrc.nist.gov/glossary/term/environment_of_operation

Patil, T. H. (2012, October). *Data Scientist: The Sexiest Job of the 21st Century* . Retrieved January 2021, from hbr.org: https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century

RTCA. (2014, 8 6). *DO-326A Airworthiness Security Process Specification.* Retrieved from https://my.rtca.org/NC__Product?id=a1B36000001IcfuEAC

RTCA. (2021). *About Us.* Retrieved January 2022, from rtca.org: https://www.rtca.org/about/

RTCA. (2021). *RTCA Standards Documents.* Retrieved January 2022, from rtca.org: https://www.rtca.org/standards/

Rutenbar, F. B. (2016, December). *Realizing the Potential of Data Science.* Retrieved 2021, from https://www.nsf.gov/cise/ac-data-science-report/CISEACDataScienceReport1.19.17.pdf

Souppaya, K. K. (2006, September). *Guide to Computer Security Log Management.* Retrieved December 2021, from csrc.nist.rip: https://csrc.nist.rip/library/NIST%20SP%20800-092%20Guide%20to%20Computer%20Security%20Log%20Management,%202006-09.pdf

TCNJ Information Security Program. (n.d.). *Information Security Roles & Responsibilities.* Retrieved 2022, from security.tcnj.edu: https://security.tcnj.edu/program/security-responsibilities/third-party-system-administrator-guidelines/

Techopedia. (n.d.). *Data Ownership.* Retrieved March 2021, from techopedia.com: https://www.techopedia.com/definition/29059/data-ownership

Torres, P. (2021, January 16). *Data Science for Cyber Security.* Retrieved March 2021, from medium.com: https://medium.com/codex/data-science-for-cyber-security-32e2f81e15d3