# The Role of Cybersecurity Data Science in Aviation Cybersecurity: Applicability for Enhancements to RTCA/EUROCAE Standards

## AIA / US ACCESS Joint Paper
## April 2023

## Contributors


| | |
|---|---|
| David Harvie | Embry-Riddle Aeronautical University |
| Stefan Schwindt | GE Aerospace (AIA) |
| Siobvan Nyikos | Boeing (US ACCESS/AIA) |
| Bill Trussell | FAA Cybersecurity Data Science |
| Isidore Venetos | FAA Cybersecurity Data Science |
| Dan Diessner | Embry-Riddle Aeronautical University |
| M. Ilhan Akbas | Embry-Riddle Aeronautical University |
| Bobby Anderson | Shift 5 (US ACCESS) |
| Jessica Carroll | Shift 5 (US ACCESS) |
| Joshua Cruse | Shift 5 (US ACCESS) |



Civil Aviation Cybersecurity Subcommittee
Aerospace Industries Association of America, Inc.
1000 Wilson Boulevard, Suite 1700 | Arlington, VA 22209-3928 | 703.358.1000 | www.aia-aerospace.org

# Contents



# Figures





Civil Aviation Cybersecurity Subcommittee
Aerospace Industries Association of America, Inc.
1000 Wilson Boulevard, Suite 1700 | Arlington, VA 22209-3928 | 703.358.1000 | www.aia-aerospace.org

# 1 Introduction and Motivation

The aviation ecosystem generates a tremendous amount of data on large informational and operational networks. The scale of this data generation can be in the range of terabytes a day for an aviation original equipment manufacturer (OEM). Airlines, airports, and other entities within the aviation ecosystem produce data on a similar scale. That data also presents a lucrative target for cyber attackers to steal, manipulate, or make unavailable. The indicators that such an attack has happened or is happening exists somewhere in that data. The challenge is how to find that those indicators amid a sea of data in a timeframe that enables cybersecurity personnel to take action.

The Cybersecurity Data Science (CSDS) Aviation Architecture Framework (AAF) seeks to apply data science to aviation ecosystem which consists of both Information Technology (IT) and Operational Technology (OT). CSDS AAF can be applied to different Environments of Operation such as airlines, airports, and Original Equipment Manufacturers (OEMs). The CSDS AAF identifies and clearly defines system level architecture such as Stakeholder Data-Stores, Cyber Analytical Cells (CACs), and Interconnected Individual Systems (IIS). The CSDS AAF also introduces the Data Life Cycle consisting of Acquire, Pre-Analyze, Collect, Curate, Advanced Analytics, and Information Sharing (Embry-Riddle Aeronautical University Center for Aerospace Resilience, 2023).

Ongoing research efforts, via the FAA CSDS Research Project, to evaluate the applicability to specific aviation Environments of Operation include working with AIA over the past nine months on a generic aircraft OEM factory CSDS use case. This first phase of this effort culminated in an Analytical Exercise/Table Top Exercise which was conducted at FAA NextGen Research Test Bed in Daytona Beach in February 2023. This use case effort and others will be continued in the future to include incorporating laboratory prototyping demonstrations. The lessons learned from this research project are being incorporated into a CSDS AAF Definition Documentation currently consisting of the following three parts:

1. Part 1: Cybersecurity Data Science (CSDS) Aviation Architecture Framework (AAF) - Utilization Strategy
2. Part 2: Cybersecurity Data Science (CSDS) Aviation Architecture Framework (AAF) - Technical Specification Document
3. Part 3: Cybersecurity Data Science (CSDS) Aviation Architecture Framework (AAF) - Systems Guidance Document

The chief contribution of the CSDS AAF is to provide guidance and recommendations to aviation standards organizations to assist their members meet current and future requirements. The CSDS AAF most closely aligns with current ongoing ISEM EUROCAE WG-72 / RTCA SC-216 efforts to update ED-206 / DO-392, respectively (RTCA, 2022).

### 1.1 Definition of Cybersecurity Data Science

Cybersecurity Data Science (CSDS) is the use of data analytics of security event data to detect security incidents or vulnerabilities. The raw security data, which can come from various sources, can be converted into useful information for decision making (Thanh, 2021).

### 1.2 Use of Data Science in Cybersecurity

Data Science is an interdisciplinary field that integrates various data analysis disciplines such as Statistics, Data Mining, and Predictive Analytics (Rutenbar, 2016). The goal of Data Science is to sift through large data sets to discover useful, actionable information. This goal is achieved by applying considerable computing resources and automated data processing techniques to reduce the raw data into more tangible and meaningful artifacts for data analysts to work with. Artificial intelligence (AI), to include machine learning (ML), are some of the automated data processes that can provide insight (IBM Cloud Education, 2023). Another benefit of data analysis is that it can transform noisy and incomplete data from diverse sources into cleaner and more useful information for human analysis (Rawat, Doku, & Garuba, 2021). Data Science could be applied to cybersecurity to assist cyber analysts to find actional information such as the indicators of a cyber incident.

.

## 2 CSDS Aviation Architectural Framework (AAF) Overview

The CSDS AAF is an aviation cybersecurity framework which incorporates Data Science. This framework can assist in the development of standards, such as the collection and analysis of data to meet current and future requirements. The framework can also inform and influence the development and adoption of tools, technologies, and processes to meet those requirements. The CSDS AAF, however, is not a new requirement.

The Aviation Ecosystem consists of numerous stakeholders, such as airlines, airports, and OEMS. Each stakeholder is unique in its operation and structure, and some stakeholders operate in multiple environments. A common characteristic of these stakeholders is the reliance of merged information / operational technology (IT/OT) systems to support their operations individually and the Aviation Ecosystem as a whole. These merged IT/OT systems generate tremendous amounts of raw data in form of network traffic, system logs, and operational logs. Within this vast amount of data, there may be indicators of a future or current cyber event.

### 2.1 CSDS AAF Conceptual Elements

The Conceptual Elements of the CSDS AAF can be divided into three categories: Data Acquire Elements, Data Categories, and Analytical Functional Elements. The CSDS AAF Conceptual Elements are visually depicted in Figure 1.

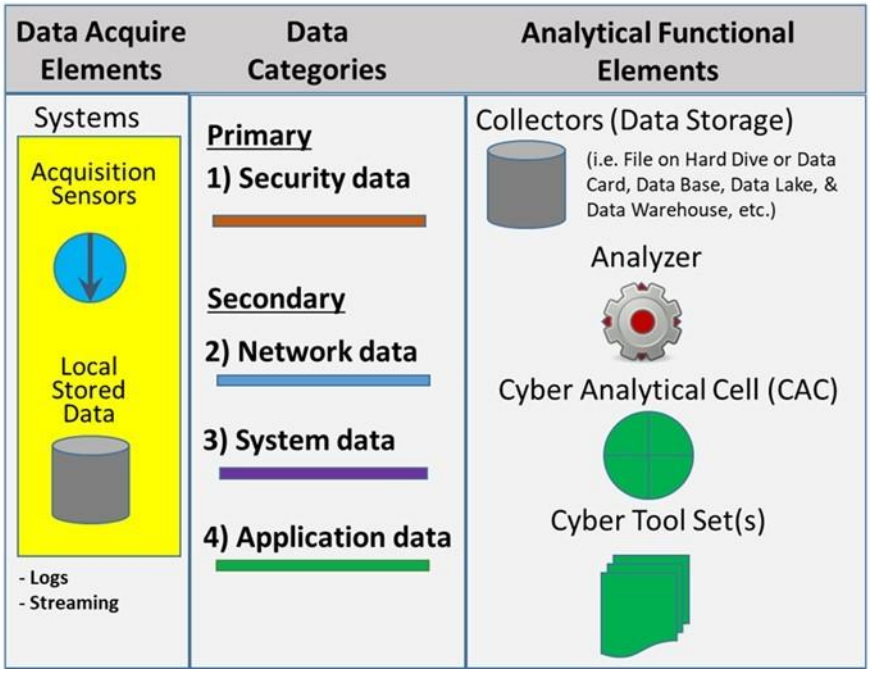


*Figure 1. CSDS AAF Conceptual Elements (Embry-Riddle Aeronautical University Center for Aerospace Resilience, 2023)*

The Data Acquire Elements category refers to the Individual Interconnected System (IIS) components that are essential for acquiring data. Within this category, Acquisition Sensors monitor data generated on the systems and networks. They then evaluate the data for relevancy. Relevant data is encoded and likely stored in Local Storage. When the relevant data is ready to be extracted, a Data Exchange Interface is required to pull the data for analysis. This can be accomplished using a removable storage device or utilizing a network interface.

The Data Categories include Security Data as the primary category as well as Network, System, and Application Data as secondary categories. Security Data is usually a subset of Network, Systems, and Application data that is considered cybersecurity relevant. Examples of Security Data include logs or data generated from antimalware software, intrusion detection systems and authentication servers. Network Data can include network activity logs from routers and switches. Operational Systems Data can include the collection of system operation parameters, fault conditions, and system health status. Application Data can include logs generated by applications to record events such as authentication, client requests, server responses, and other application usage events.

An important element of System Data is the data produced by the onboard OT components which allow the aircraft to operate in nominal fashion. These are non-traditional OT devices, avionics, instrumentation, and other hardware devices which typically run either a lightweight, real-time operating system or firmware and intercommunicate via serial bus(es) such as ARINC, ASCB and the like. These devices are in constant communication in a manner that is highly ephemeral, with only a small fraction of this communication ever being retained or logged. It is this data, when captured and retained in total, that provides the forensic record and insight into real-time behavior which all for problems to be rapidly identified and sourced via inspection, retention and analysis. This data, when combined with other traditional data, paints a complete picture.

The Analytical Functional Elements interact with the data. Collectors are data storage devices that could be found in various locations throughout an environment. Analyzers and Cyber Tools Sets analyze the data stored in the

Collectors. The Cyber Analytical Cell (CAC) is the organization that employs the Analyzers and Cyber Tool Sets to analyze the data found on the Collectors. CACs are often located within a secure facility such as a Security Operations Center (SOC) that are prevalent within many IT organizations. Figure 2 illustrates the reference model for the CAC.

The CAC Reference Model recognizes two distinct groups with the CAC. The Data Scientists and Software Engineers group utilize Data Analytical Tools to extract data from the Data Store which represents all the data that is collected. This data combined with threat intelligence feeds are input into data toolsets to generate cyber toolsets for use by the Human Analysts. The Human Analysts group then utilize the cyber toolsets to provide insights into the large data and generate reports.

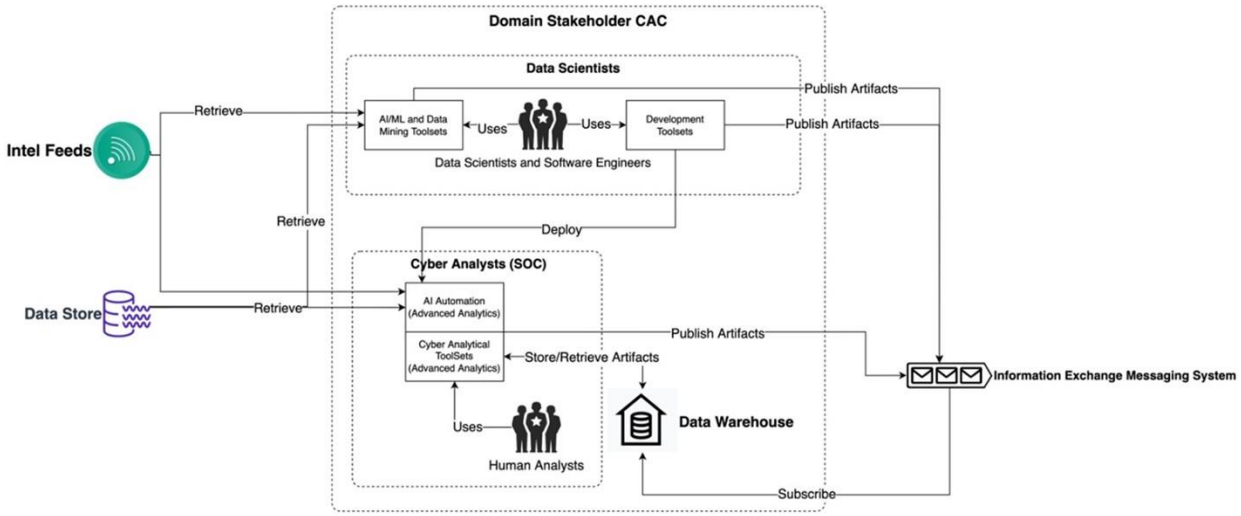


*Figure 2. Cyber Analytical Cell (CAC) Refence Model (Embry-Riddle Aeronautical University Center for Aerospace Resilience, 2023)*

## 2.2 Application to ISEM Documents

The concept of the CAC can augment Section 3.1 Organization & Key People Identification of DO-392. DO-392 divides Information Security Event Management (ISEM) roles into ISEM Core Roles and Related Supporting Roles. The Human Analysts group in the CAC Reference Cell could incorporate multiple ISEM Core Roles such as Vulnerability Management team, Security Incident Response Team (SIRT), and Operation team. The Data Scientists and Engineering group roles can be added to the list of ISEM Core Roles.

# 3 Identifying the Right Data to Collect/Analyze

The three primary objectives that CSDS seeks to answer are:

1. Is there a cyber-event pending?
2. Is there an ongoing cyber-event?
3. What caused a cyber-event to happen?

Collecting and analyzing the right data that can provide insight to those questions is critical to CSDS. It is critical to have a comprehensive understanding of how data is collected, transmitted, stored, and analyzed. The CSDS AAF Data Life Cycle provides such a comprehensive model. Merely collecting all data available is a naïve approach to data science because it makes two assumptions. First, it assumes that all data being collected is useful or at least

its collection is not detrimental to data analysis. Collecting and storing unnecessary data can impede timely analysis of relevant data. Also, there are significant costs in resources in storing unnecessary data. Second, this approach assumes that all necessary data is already being generated and collected. The Data Sphere Model illustrates this challenge of collecting the correct data.

## 3.1 CSDS AAF Data Life Cycle

The CSDS AAF Data Life Cycle consists of six phases: Acquire, Pre-Analyze, Collect, Curate, Advanced Analytics, and Information Sharing. Figure 3 illustrates the flow of these six phases. In the first phase, the Acquire Phase, sensors acquire data from hardware components and software processes in the various IIS in the Environment of Operations. In the Pre-Analyze phase, automated processes filter and extract features from data. Meta data may be added to assist in future retrieval and processing. The Pre-Analyze phase occurs within the sensor. The pre-analyzed data is then stored using non-volatile memory storage devices. These devices could be local storage or remote storage. The collection of these various non-volatile memory storage devices is referred to as the Data Store. It is critical that local storage devices are configured correctly to handle the velocity of data created by the sensors and that remote storage devices are appropriately connected and secured.

During the Curate Phase, cyber-relevant data is extracted from the Data Store. This data will be used to create data sets and models for further analysis. The Advanced Analytics Phase transforms the curated data into useful and meaningful artifacts. Advanced analytics enable human analysts to visualize and interpret previously hidden data. Such insights could include network traffic patterns, user habits, and data volume over time. Meaningful artifacts could include recommendations to close an unused logical port or flagging suspicious user activity. This analysis could assist in the detection of an ongoing cyber attack or providing risk assessment and root cause analysis after a cyber incident. The final phase, the Information Sharing phase, extracts meaningful and useful information to share with internal stakeholders and other stakeholders within the aviation ecosystem while protecting confidential and sensitive stakeholder data.


*Figure 3. CSDS AAF Data Life Cycle (Embry-Riddle Aeronautical University Center for Aerospace Resilience, 2023)*

## 3.2 Data Life Cycle Objectives

The following are the Execution Objectives and Define and Design Objectives for each phase during the Data Life Cycle.

**Acquire Phase Objectives**

- Execute
  - Acquire 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., Cyber Analytical Cell (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
    - 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 (to include 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 the 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 CAC's 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.

## 3.3 CSDS AAF Data Life Cycle Application to ISEM Documents

### 3.3.1 Security Risk Management

Section 3.3 Security Risk Management Contribution for ISEM in DO-392 recommends that an "organization should define and deploy event collection (e.g., logging policies) and detection capabilities based on the results of security risk assessments" (RTCA, 2022). The event collection process is a practical instance of the CSDS AAF Data Life Cycle. This life cycle illustrates how the data containing evidence of a security evidence is acquired, collected, and analyzed. This illustration provides understanding to the data collection process without specifying how to implement.

### 3.3.2 Information Sharing

Section 3.5 Information Sharing in DO-392 and the Information Sharing Phase in the CSDS AAF Data Life Cycle describe the same subject. The Information Messaging System of the CSDS AAF (Embry-Riddle Aeronautical University Center for Aerospace Resilience, 2023) may using a Publisher/Subscriber model. Some of the key components include:

- 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 can access.

This information sharing would require a Centralized Messaging Broker Service, such as the System Wide Information Management (SWIM) system.

### 3.3.3 Event Detection

Section 4.2 Detection Strategy in DO-392 describes the two general approaches to event detection via use cases. The first use case involves an Original Equipment Manufacturer (OEM) providing guidance on what to monitor during operation of the equipment. The second use case involves an organization conducting risk analysis on what elements to monitor. Both use cases can benefit from the CAC model that is collecting and analyzing security events. There are inferences that a CAC-like organization must be monitoring element to detect cyber events, but this is not clearly delineated.

Section 4.3 Security Event Information Sources to Monitor in DO-392 identifies multiple sources to monitor for detecting security events. Some of those sources include notifications from intrusion detection and prevention systems, system logs, vulnerability scan reports, component failures, and unexplained system failures. The CAC model (Figure 2) and CDSD AAF Data Model (Figure 3) provide a holistic view of how data that could indicate the presence of a security event is acquired, transmitted, collected, and then analyzed by security personnel in a CAC-like organization. The CSDS AAF Data Model and CAC add details to the process of event detection in a non-prescriptive manner.

The Data Sphere model could be referenced in Section 4.2 Detection Strategy of DO-392 as a design principle of what data is acquired. The model succinctly communicates that only a small subset of available data is useful and

desirable. The full description of the Data Sphere would best be communicated in a separate appendix of DO-392 so that technical definitions would not slow down the communication flow of the main document.

# 4 References

Embry-Riddle Aeronautical University Center for Aerospace Resilience. (2023). *Part 1: Cybersecurity Data Science (CSDS) Aviation Architecture Framework (AAF) - Utilization Strategy.* Atlantic City, NJ: FAA NextGen Cybersecurity Data Science Project.

Embry-Riddle Aeronautical University Center for Aerospace Resilience. (2023). *Part 2: Cybersecurity Data Science (CSDS) Aviation Architecture Framework (AAF) - Technical Specification Document.* Atlantic City, NJ: FAA NextGen Cybersecurity Data Science Project.

Embry-Riddle Aeronautical University Center for Aerospace Resilience. (2023). *Part 3: Cybersecurity Data Science (CSDS) Aviation Architecture Framework (AAF) - Systems Guidance Document.* Atlantic City, NJ: FAA NextGen Cybersecurity Data Science Project.

IBM Cloud Education. (2023, January 19). *Data Science.* Retrieved March 2021, from ibm.com: https://www.ibm.com/cloud/learn/data-science-introduction

Rawat, D. B., Doku, R., & Garuba, M. (2021). Cybersecurity in Big Data Era: From Securing Big Data to Data-Driven Security. *IEEE Transactions of Services Computing*, 2055-2072.

RTCA. (2022, June 23). *DO-392 Information Security Event Management.* Retrieved from https://my.rtca.org/productdetails?id=a1B1R00000zbOOMUA2

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

Thanh, C. T. (2021). A Study of Machine Learning Techniques for Cybersecurity. *2021 15th International Conference on Advanced Computing and Applications (ACOMP)*, 54-61.

# Appendix A: Written Recommendations for DO-392

The following are recommended changes and additions to DO-392. They are organized by section in DO-392. Recommended additions are highlighted in green, changes in yellow, and deletions in red.

## A.1 Section 3.1 Organization & Key People Identification

The roles and responsibilities associated with the core team can include:

- Management Team: is responsible for coordinating among various stakeholders, establishing ISEM budget, and staffing. Management Team is also responsible for strategic and business-related decisions. The accountable manager or suitably authorized delegate is included in the management team.
- Security Incident Response Team (SIRT) Leader: is responsible for operating the overall ISEM process. The SIRT leader establishes procedures, policies and develops the communication strategy to enable seamless communication that is validated by the management team. During incidents or vulnerabilities, the SIRT leader oversees activities and keeps contact with main stakeholders.
- Monitoring team: is responsible for security events monitoring activities, identification of relevant events, recording of security events and tracking of vulnerabilities and incidents.
- Operations teams: are responsible for operating the various IT systems in the ISEM scope. They implement and control technical security measures, approve changes to the IT infrastructure and support the analysis and response to incidents and vulnerabilities.

- Vulnerability Management team: is responsible for the analysis and mitigation or correction of vulnerabilities.
- Security Incident Response Team (SIRT): is responsible for handling incidents when they occur, and for performing in-depth analysis and mitigating the damage of incidents. The SIRT may require several roles to ensure that incidents are managed and coordinated effectively.
- Safety manager: is responsible for managing security incidents that may have potential safety impacts, and for ensuring an effective safety event management process.
- Accountable manager: is responsible for coordinating with authorities and ensuring that the whole ISEM process is effectively managed (i.e., resourced and skilled appropriately).
- Product engineering teams: are responsible for the design, development, verification, and maintenance of assets in the ISEM scope. They implement technical security measures in products, assess the impact of changes, and implement remediation in response to incidents and vulnerabilities.
- Communication team: may be involved to define and establish an information sharing program between the different stakeholders.
- Cyber Analytical Cell (CAC): collection of Human Analysts using software-based toolsets to perform analytics on data to produce useful cyber-analytical information. The Human Analysts can include other core roles such as the Vulnerability Management team, Security Incident Response Team (SIRT), and Operation team. CACs are often located within a secure facility such as a Security Operations Center (SOC).

The CAC organization, inputs, outputs, and activities is illustrated below in Figure 3-1:

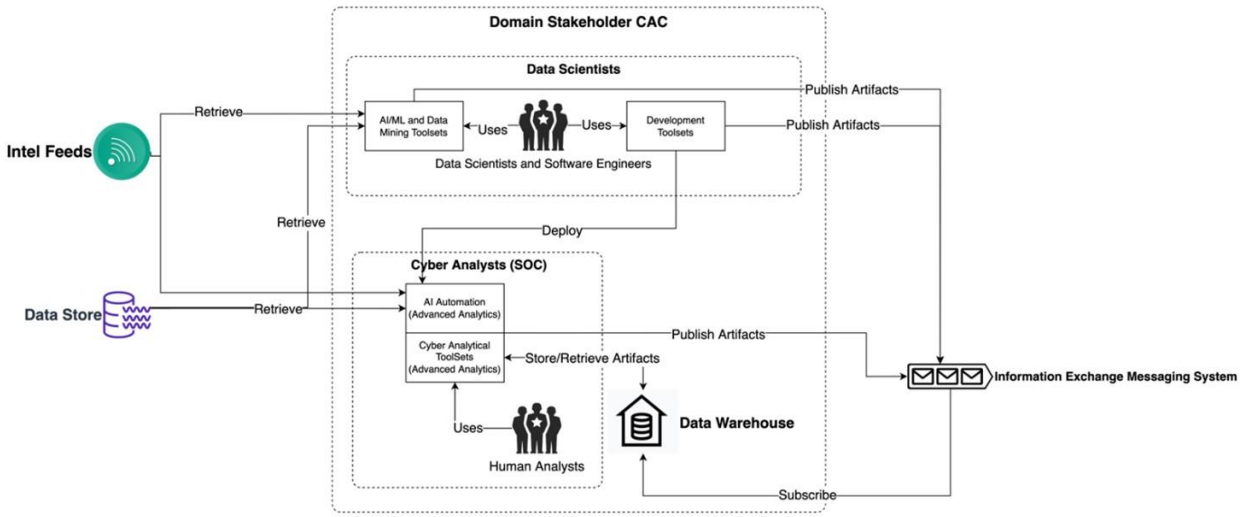


*Figure 3-1. Cyber Analytical Cell (CAC) Model*

The roles and responsibilities associated with the supporting team (optional) may include:

- Audits and risk management team: may be involved in identifying the assets to be monitored within the ISEM scope.
- Legal/Regulator expert team: may be involved to ensure that an organization's legal and regulatory compliance is well protected even in the case of a security incident.

- Crisis management team, business continuity team: these teams may be involved when there is a crisis situation due to a security incident. A crisis being a situation in which Hazardous or Catastrophic conditions exist with the service or product.
- Data Scientists and Software Engineers: this group uses Data Science techniques to perform analytics on data. They also develop visualization software, cyber analytical toolsets, and automation for Human Analysts to use. Human Analysts would include key roles such as the Monitoring Team and SIRT.

## A.2 Section 3.5.2.3 Communication

Communication within the information sharing community, or using information sharing platform, can be time-intensive. Communications activities are drafting and distribution of bulletins and reports across organizations, exchanging with public media, providing further technical analysis or denying fake news. The burden can be reduced by using some good practices: using pre-built templates, gathering different information in the same communication and create awareness of time needed to perform activities. Requirements may be cascaded to external organizations such as suppliers, service providers and vendors to facilitate: the detection of security events; the analysis of vulnerabilities; and to improve the efficiency of incident handling.

A Publisher/Subscriber model is a well-known example of an Information Exchange Messaging System. 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.

## A.3 Section 4.2 Detection Strategy

O4.1: Security events are collected and screened that indicate deviations from predetermined functional performance baselines.

Section 3.3 discusses security risk management as a contribution to the ISEM organize and prepare process. The following text expands on the concepts in greater detail for clarity regarding the ISEM detection process.

An organization should define and document their event detection strategy. It should be structured into clearly defined goals or objectives.

Developing an efficient detection strategy for security events is about establishing the monitoring of variable/evolving elements associated to security risks. Security risk level is determined by likelihood (Level of threat condition) and impact. The enrollment dossier should be reviewed for relevant information concerning vulnerabilities and impact. Reference section 3.4.2.3 Enrollment Dossier.

One method is to consider all risks identified by a security risk assessment and according to the enrollment dossier. Each risk is defined within the context of use cases or scenarios. That is, each risk associated with an asset should be analyzed and described against each of the three attributes of security risk, Confidentiality, Integrity, and Availability (CIA). For each of these attributes, a use case can be described that demonstrates how the attribute was identified by the risk assessment. For example, for each asset defined in the risk assessment:

- Vulnerability - Consider the likelihood that the vulnerability would be exploited, that a vulnerability-threat pairing would be realized. Consider which security attributes need to be protected and what are the potential consequences of each compromised asset.
- Security Events – What are the possible scenarios that could describe the details of events and method of detecting each possible security event?
- Security Events – Implement event triage and consider the possible scenarios that are consistent with the indications of the event. Determine if the event is a security incident.

Another method to consider applying is the Cybersecurity Data Science (CSDS) Aviation Architecture Framework (AAF). The CSDS AAF seeks to answer the following three key questions using data analytics:

1. Is there a cyber-event pending?
2. Is there an ongoing cyber-event?
3. What caused a cyber-event to happen?

There are five stages to the implement of CSDS AAF to support event detection. Stage 1 is defining the CSDS Desired Relevant Data Requirements. The specific use cases where the CSDS AAF is implemented will be the basis for defining the data requirements. The Data Sphere Model, further described in Appendix H, illustrates the challenge of identifying relevant data in the midst of both collected and uncollected data.

Stage 2 is implementing or re-implementing system requirements to support the acquisition of the relevant data defined in Stage 1. These system requirements may be as minimal as changing which data is collected to a more extensive requirements as emplacement of sensors and logs to acquire data that previously was not collected. Stage 2 implementation can be significant in terms of both time and resources to implement.

Stage 3 is the extraction and curation of the relevant data. The curated data is collected by appropriate Cyber Analytical Cells (CACs) for further investigation.

Stage 4 is the application of CSDS activities on the data to produce useful and actionable artifacts for human analysts to help answer the three key questions. This stage requires the most human involvement involving analysts, data scientists, and other core team personnel.

Stage 5 is the evaluation of the effectiveness of the CSDS AAF for the specific use case. Feedback loops are critical to the refinement of the CSDS relevant data, the effectiveness and efficiency of acquiring the relevant data, and the effectiveness of the CSDS AAF in assisting to answer the three key questions.

## A.4 Section 4.3 Security Event Information Sources to Monitor

Evidence of information security events can be observed and noted from several different sources. The Monitoring Team should establish and monitor all relevant sources for security events identified by the risk assessment. Some sources may be defined without a risk assessment using the enrollment dossier (see event sources list below in this

section). However, a minimum set of internal sources are determined by risk assessments. Additional sources such as vulnerable data feeds can be monitored also.

Many sources are identified by the risk assessment as described in sections 3.3.3 and 4.2.2. However, some sources may be defined without a risk assessment. See sections 3.3.2, 4.2.1, and event sources list below within this section.

A model that may be helpful in identifying the relevant sources to monitor is the Data Sphere model (see Appendix H). The Data Sphere model illustrates the non-trivial challenge of determining the correct data requirements.

Additionally, organizations should foster a security culture with governance established by organizational management. The SIRT leader should establish guidance of what people can look for that may indicate suspicion of security risks.

The following are typical sources to monitor for security events. Most items in the list point to a section where the security risk source is described in detail:

- Event detection notifications from security event detection software tools. Some systems, especially ground systems, have intrusion detection and prevention systems (IDPSs) that alert security personnel if there is an anomaly indicating the presence of a security event
- Onboard OT bus data that contains the communications between components, avionics, controllers, and other devices that use these buses to interact during normal operations.
- Embedded information system security Log File Data. Refer to section 4.3.1
- External Notification from system manufacturers, DAHs, or other users regarding events and/or incidents specific to systems, system components, and software. Refer to section 4.3.2
- Unexplained system failure if root cause is not found. Refer to section 4.3.3
- Physical evidence of possible tampering, hacking, intrusion. Refer to section 4.3.4
- Other sources (Media report, Organizational feedback, etc.). Refer to section 4.3.5
- Vulnerability databases and other vulnerability information sources, including Common Vulnerabilities and Exposures (CVE) programs, that provide collections of vulnerabilities that have been identified and publicly disclosed. Many sources also provide Common Platform Enumerations (CPEs) that associate CVEs to products. These databases typically provide live data feeds that can be used to automate some aspects of vulnerability monitoring. Refer to section 4.3.6.
- Vendor bulletins, which may indicate discovered vulnerabilities and if they have been remediated as part of a new release. The bulletins may also describe the characteristic or abnormality associated with the vulnerability.
- Vulnerability scan reports, which may be provided by the OEM and others during development. These can also be provided by internal or external security audit centers. These can provide information regarding discovery of new vulnerabilities.
- Test and audit findings performed by the audit team generate reports that could indicate vulnerability to attacks.
- Operational and incident reports generated by organizations such as operators, MROs and OEMs, provide current and historical information that can be used to indicate possible sources.

# Appendix B: Overview of Cybersecurity Data Science in Context of Aviation (Proposed Appendix H in DO-392)

## B.1 Data Sharing in the Aviation Ecosystem

The Aviation Ecosystem is highly interconnected both within and across the six Aviation Domains which align with the Aviation Stakeholder Framework described in Section 2.3. Those six Aviation Domains are:

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

Within each Stakeholder, there is a Cyber Analytical Cell (CAC), or similar type organization, that hosts a collection of human analysts that perform analytics on cybersecurity related data. The results of these analyses are Artifacts used internally within the CAC. However, these Artifacts may contain sensitive/proprietary information. Any sensitive/proprietary information must be removed prior to sharing the information to any external agency. These Shareable Artifacts are shared on a voluntary or mandatory basis to both non-governmental and governmental organizations and agencies. Figure 4, shown below, illustrates the interconnectedness and sharing of artifacts from individual Stakeholder CACs to governmental and non-governmental Distributed CACs which have vested interests in the cyber analytical information generated and shared by the Domain Stakeholders.

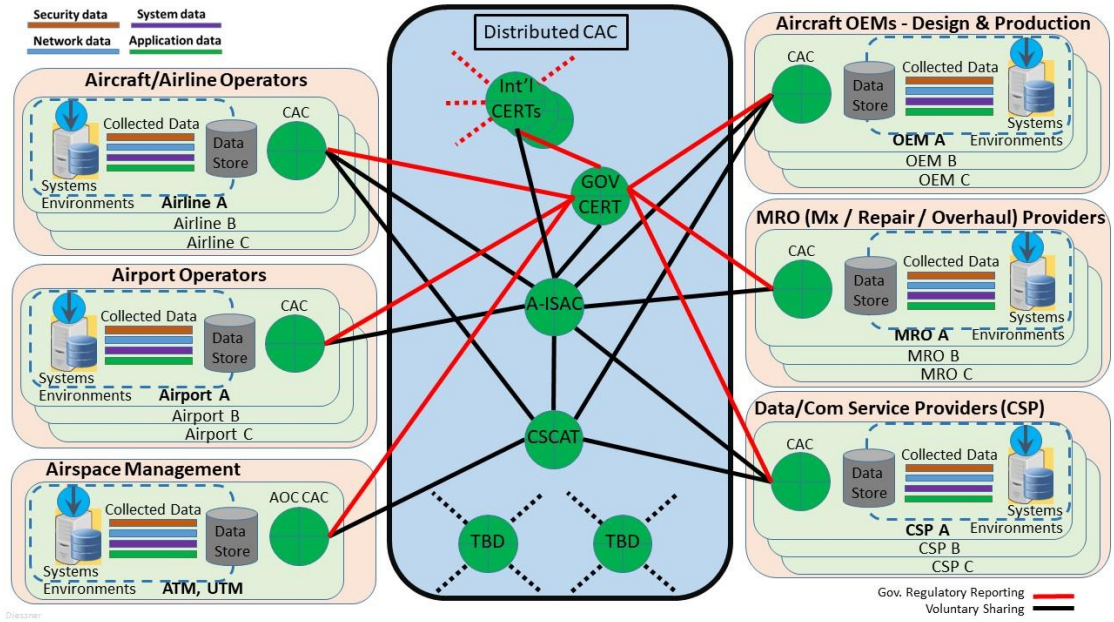


*Figure 4. System View of Aviation Ecosystem Interconnectedness (Embry-Riddle Aeronautical University Center for Aerospace Resilience, 2023)*

## B.2 Data Life Cycle

The Stakeholders in the Aviation Ecosystem control and monitor merged enterprise information / operational technology (IT/OT) systems to support daily operations and act as backbone technologies throughout the ecosystem. These merged IT/OT systems contain enormous amounts of raw data in the form of network traffic, system logs, and operational logs. The sheer volume of data generated on a daily basis surpasses human ability to adequately analyze in a timely manner to provide actionable information in regard to cybersecurity. CSDS applies data science and analytics to provide human analysts with increased capabilities to provide actionable information in a more timely manner.

Data Acquisition Sensors integrated with Individual Interconnected Systems (IIS) acquire and pre-analyze the raw data. During the Acquire Phase, the sensors monitor and acquire the raw data either hardware components or software processes. The Pre-Analyze Phase also occurs within the Data Acquisition Sensor. This pre-analysis uses software-based logic to determine if the acquired data is relevant. Relevant data undergoes feature extraction and tagging with meta data that will improve later retrieval and analysis. During the Collect Phase, the relevant data is stored on various non-volatile memory storage devices which can include Cloud storage implementations. These storage devices are collectively referred to as the Data-Store. During the Curate Phase, cyber-relevant data is extracted from the Data-Store for the purpose of creating data sets and models, depending on specified needs and interests. The Advanced Analytics Phase takes the curated data and seeks to produce meaningful Artifacts that include insights and actionable information for the internal Stakeholder CAC. These Artifacts may include detection of an ongoing cyber attack, the presence of installed malicious software, or risk assessment and root cause analysis following a cyber attack. During the Information Sharing Phase, the sensitive/proprietary data is removed from the Artifact to make it a Shareable Artifact with external agencies as described above. Figure 5, shown below, illustrates this CSDS Data Life Cycle.

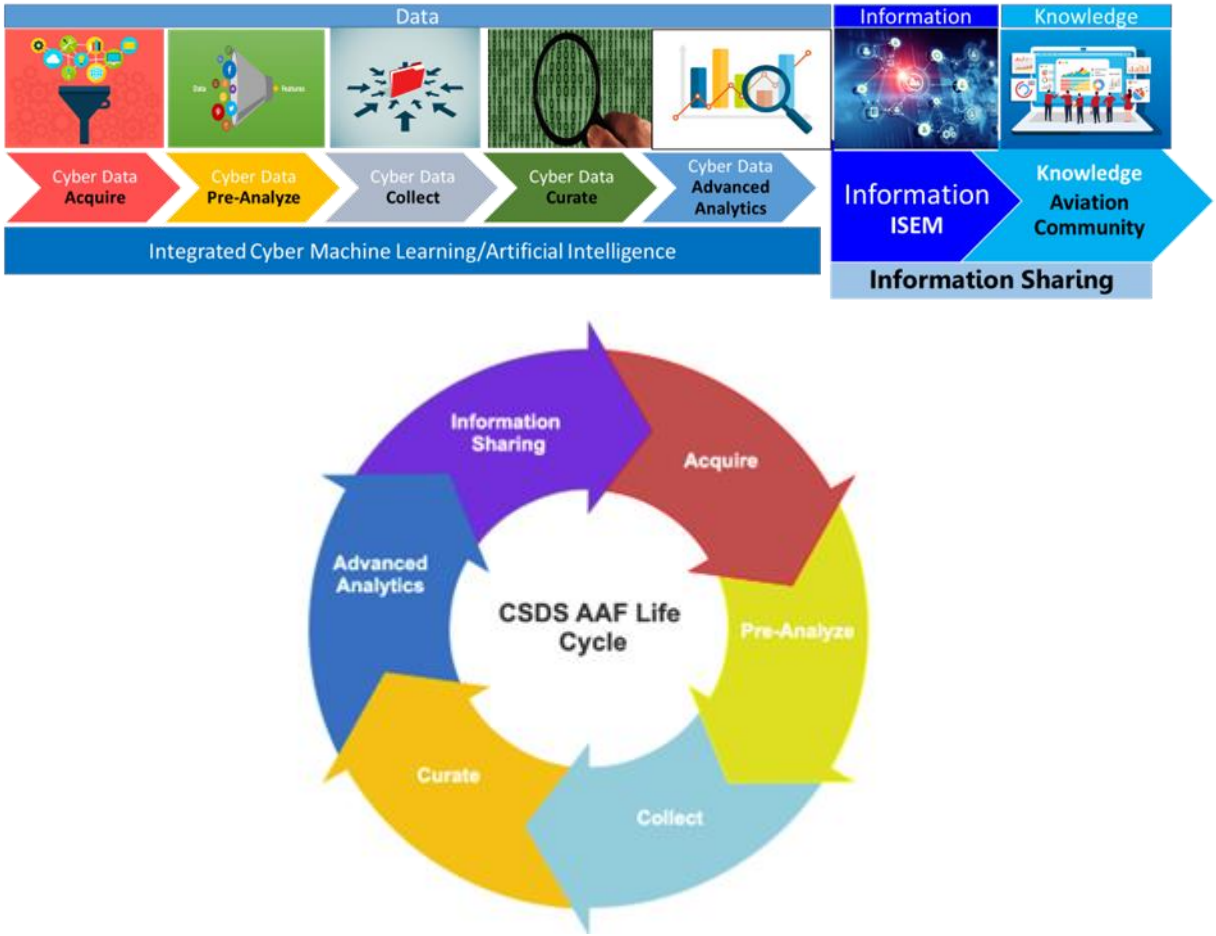


*Figure 5. CSDS AAF Data Life Cyle (Embry-Riddle Aeronautical University Center for Aerospace Resilience, 2023)*

## B.3 Data Sphere

Data relevancy is a critical concept in CSDS because not all data is useful or helpful. Also, there may be gaps between the data currently being collected and what should be collected. From a CSDS perspective, data is relevant if it can help answer one of the three primary questions:

1. Is there a cyber-event pending?
2. Is there an ongoing cyber-event?
3. What caused a cyber-event to happen?

The Data Sphere (Figure 6) is defined as the set of all data that is acquired from all IIS. Only a small portion of the originally acquired data is collected into the Data Store. If the portion of the Data Sphere that serves as an input to the CSDS process is incorrect or insufficient, then the following issues may happen:

- Analytical toolsets are slowed down processing unnecessary data.
- Analytical toolset results are skewed leading to biased or faulty conclusions.

- Unnecessary noise and data prevent the production of meaningful results.
- Unnecessary or excessive configuration of network monitoring devices and IIS may stress the networks and system.

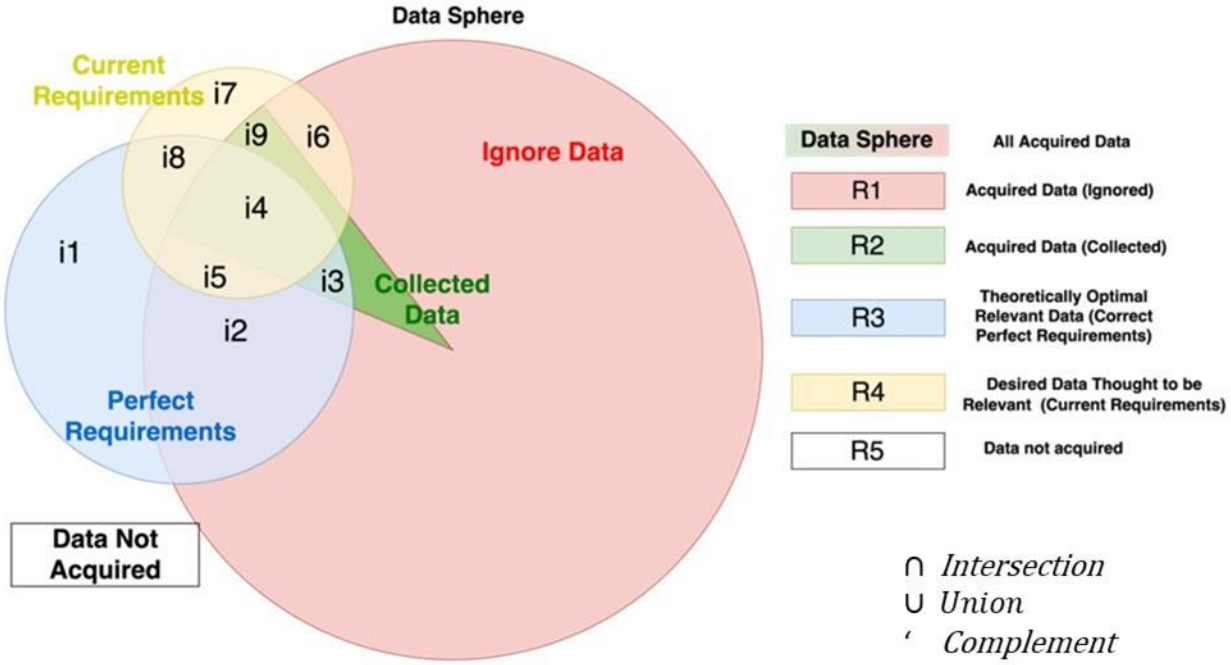


*Figure 6. The Data Sphere, Regions, and Intersections (Embry-Riddle Aeronautical University Center for Aerospace Resilience, 2023)*

Determining what is relevant data for CSDS is difficult. Figure 6 shows the Data Sphere with five regions and nine intersections which are described below:

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).

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. 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.
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. $i9$ [$R2 \cap R3^{\wedge\prime} \cap R4$] represents data that is being collected and has incorrectly been identified as a CSDS requirement.

### B.3.1 Theoretically Optimal Relevant Data

With the Data Sphere model, there is Theoretically Optimal Relevant Data that represents the data requirements that will provide the CSDS AAF the theoretical best chance of success. It is quite possible that the current Data Sphere does not cover the full set of Theoretically Optimal Relevant Data. This data gap may necessitate adjusting the Environment of Operations design requirements.

The Theoretically Optimal Relevant Data region of the Data Sphere is highly dependent on the CSDS Use-Cases and Threat Scenarios for a specific Domain Stakeholder. For example, if a Use-Case/Threat Scenario involved Intrusion Detection of an aircraft Wi-Fi network, then just a small subset of the airline's total available data would be relevant.

Figure 7 illustrates how multiple Theoretically Optimal Relevant Data regions may exist within a Data Sphere based on the Threat Scenario and CSDS Use-Case. It is possible for these regions to intersect which means that there are common relevant data for multiple Threat Scenarios/Use-Cases. These data intersections are desirable as the same data can be re-used, increasing the efficiency and effectiveness of the CSDS AAF.

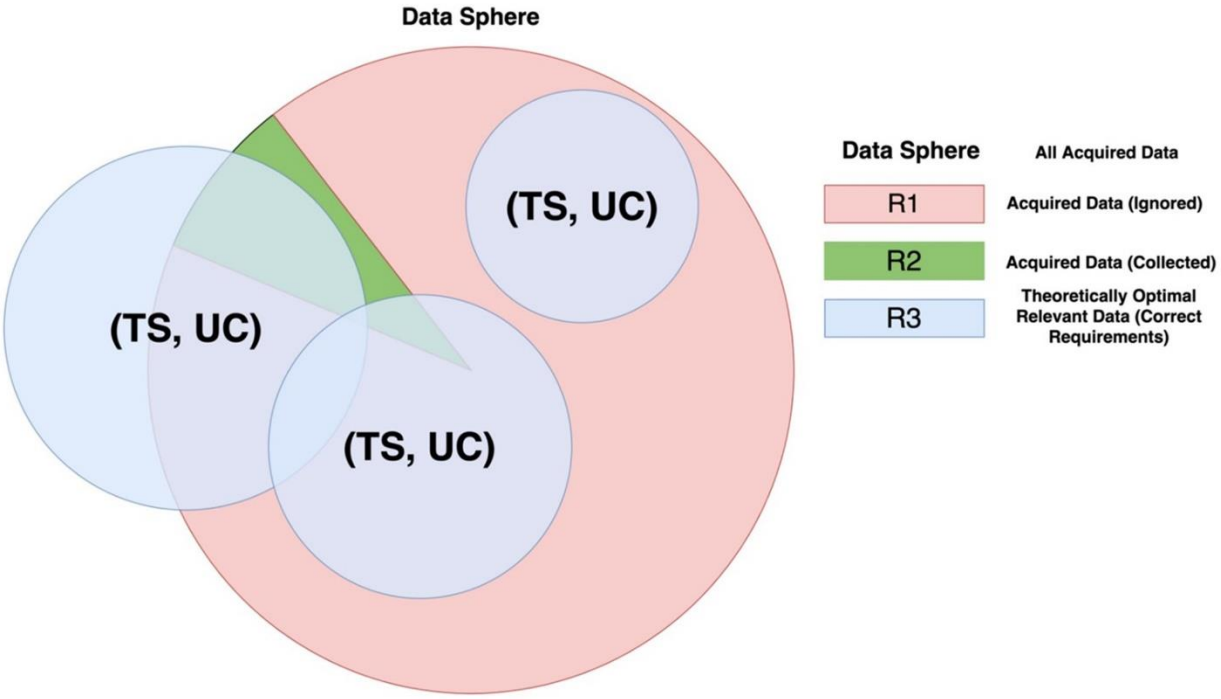


*Figure 7. Relationship between Data Sphere and Theoretically Optimal Relevant Data (Embry-Riddle Aeronautical University Center for Aerospace Resilience, 2023)*