ISAO
Standards Organization

ISAO 300-2: Automating Cyber Threat Intelligence Sharing

v1.0

April 5, 2019

ISAO 300-2

Automating Cyber Threat Intelligence Sharing

Version 1
ISAO Standards Organization
April 5, 2019

Acknowledgements

This publication was developed by the Information Sharing and Analysis Organization Standards Organization (ISAO SO) with representatives from the private, professional, and government communities in an ongoing effort to produce a unified voluntary set of guidelines for information sharing. The ISAO SO and the Working Group leadership are listed below.

ISAO Standards Organization

Gregory B. White, Ph.D.

ISAO SO - Executive Director

Director, UTSA Center for Infrastructure Assurance and Security

Allen D. Shreffler

ISAO SO - Deputy Director

LMI

Working Group Three—ISAO Information Sharing

Kent Landfield	Roger Callahan
Chief Standards and Technology Policy Strategist	*Consultant*
McAfee LLC	*FS-ISAC*

The ISAO SO leadership and authors of this document would also like to acknowledge those individuals who contributed significantly to the development of this publication, including:

Tyler Bent of IT-ISAC, Adam Buteux of PWC, Cory Casanave of Model Driven Solutions, Michael Darling of NTT Security, Jerry Eastman State of Colorado, Steve Hitch of NTT Security, Bill Kloster of SEH Inc., Jim Lippard of American Express, Tom Litchford of the National Retail Federation, Betsi McGrath of MITRE, Chris Needs of NC4, Rajat Ravinder Varuni of Amazon Web Services, Michael A. Vermilye of Johns Hopkins University Applied Physics Laboratory and Ole Villadsen of IBM.

Special thanks from the authors go to the ISAO SO advisors and staff who provided amazing support and guidance in the development of this document: Marlis Cook, James Navarro, and Jeremy West.

Table of Contents

Figures:

Revision Updates

Item	Version	Description	Date
	V1	Final Published Version 1	05 April 2019 13:00

1. EXECUTIVE SUMMARY

The purpose of this document is to provide a description and implementation guideline for automating key elements of the cyber threat intelligence (CTI) lifecycle process of collection, identification, ingesting, processing and correlation to establish derived actions. As envisioned, the document is targeted at organizations wanting to automate and use cyber threat intelligence processes for defending their enterprise. This document is equally useful to Information Sharing and Analysis Organization (ISAO) members and the ISAOs that are participating or considering participation in automated sharing efforts.

This document comprises a technical discussion and guidelines to assist organizations implementing automated cyber threat intelligence information sharing and its utilization in mitigating cybersecurity risks. Intelligence efforts have been generally characterized as strategic, operational or tactical.[1] This guide is focused on the area of tactical intelligence utilization that can benefit an enterprise and is dependent on an information sharing ecosystem that can support automated sharing of cyber threat intelligence.

Throughout the document, the terms cybersecurity information sharing, and information sharing are used synonymously. Additionally, cyber threat intelligence and cyber threat information are also used synonymously.

2. INTRODUCTION

The "ISAO 300-1 Introduction to Information Sharing"[2] document published by the Information Sharing and Analysis Organization Standards Organization (ISAO SO) in September 2016 provided an overall context for the critical importance of information sharing among those addressing and engaged in the management of cybersecurity risks.

An essential element within the context of those dealing with their organizational cyber risks is the availability of cyber threat intelligence. This intelligence provides the information and analysis needed to better understand the situational awareness of the environment in which they are operating. This knowledge supports the decision-making and actions taken to justify and manage risks to organizations. Figure 1, from the referenced document, depicts the overall context for information sharing discussed in the ISAO 300-1 document.

[1] See the Intelligence and National Security Alliance (INSA) resources discussing this breakout at https://www.insaonline.org

[2] See ISAO 300-1 at https://www.isao.org/products/isao-300-1-introduction-to-information-sharing/

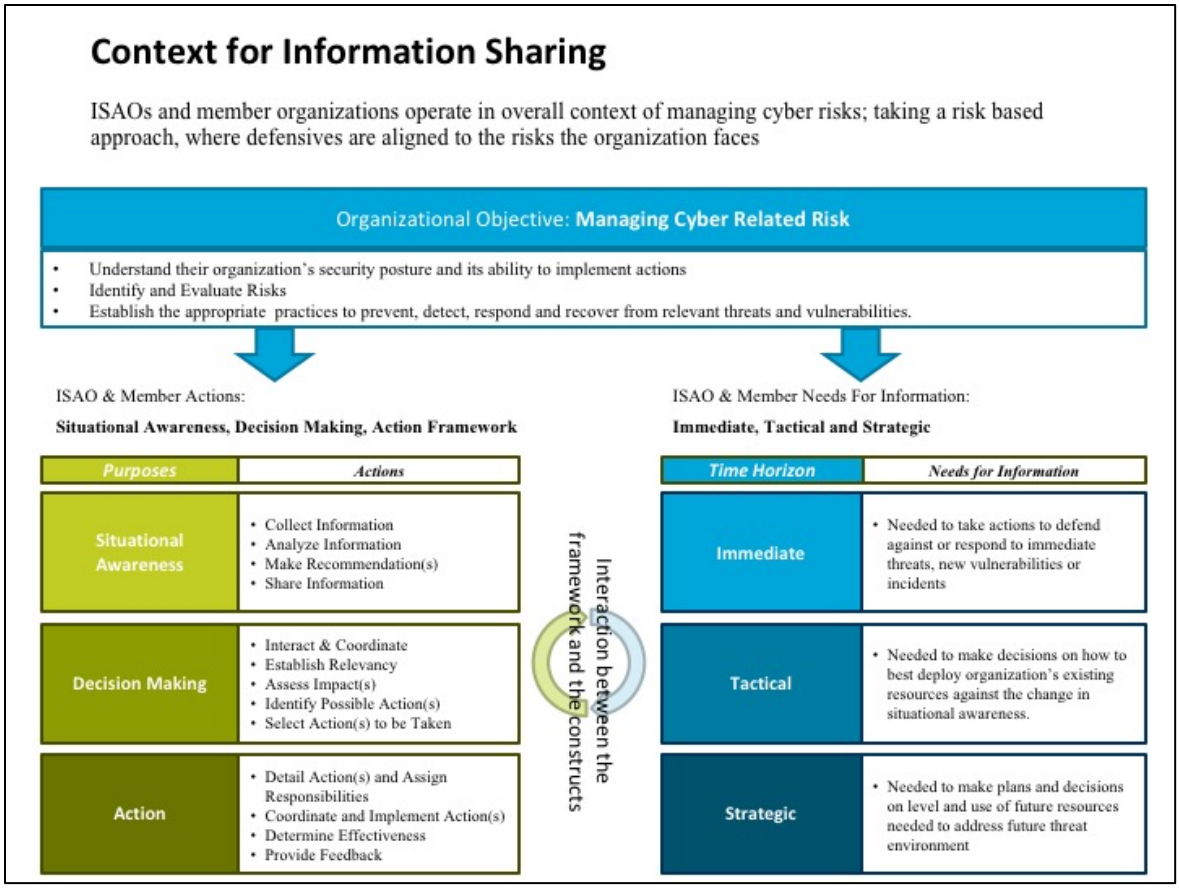

Figure 1: Context for Information Sharing

Further, 300-1 noted[3], "Threat intelligence reports are a broad category of cyber threat information ranging from high-level trending reports to detailed analysis of specific campaigns. Vendors, governments, and independent organizations produce various types of reports, including open source intelligence reports. Some are targeted at specific incidents; some are predictive, while others describe the current state of the cyber threat landscape. These reports can include the full range of cyber threat intelligence providing strategic, tactical, and immediate response value. The report can include campaign, threat actor, tactics, techniques and procedures, and other threat indicator information. Some reports are the result of several years of analysis and tracking of cyber threats."

This guide focuses on tactical considerations organizations should be addressing when automating cyber threat intelligence information for their internal consumption and use.

2.1. FRAMING CONCEPTS

To support the understanding of what automation is, where it can be applied, and how it can be applied to threat intelligence sharing, it is important to understand the following three concepts:

1. How threat intelligence is used: This is described in *An Information Life Cycle Model, Section 2.1.1.*

[3] See ISAO 300-1, Section 10 at https://www.isao.org/wp-content/uploads/2016/10/ISAO-300-1-Introduction-to-Information-Sharing-v1-01_Final.pdf

2. The notion of structured and unstructured data and how that impacts the ability to automate processes associated with it: This is described in the *Structured and Unstructured Data, Section 2.1.2*.
3. What is meant by automation: This is described in the *Different Types of Automation, Section 2.1.3.*

2.1.1. INFORMATION LIFE CYCLE MODEL

The first framing concept relates to activities that are basic elements of a threat intelligence process and its use. By understanding how threat intelligence is used, it helps identify where automation can best be applied.

Cyber threat information consists of threat indicators, tactics, techniques, procedures, behaviors, motives, specific adversaries and their targets, vulnerabilities exploited, courses of action that should be taken, or other warnings regarding an adversary and their intentions or actions against operational systems.

One common example of useful threat indicators is "Indicators of Compromise (IOCs)", which generally are pieces of information that if observed on a network or operating system will indicate with high confidence a computer intrusion. Some examples of indicators of compromise can include unusual outbound traffic, anomalies in privileged user account activity, geographic irregularities, etc.[4] To use shared information, you first must collect it and provide it to systems which can process it. An example is the collection and use of IOCs as part of an intrusion detection capability.

The use of shared information can be described by using the information life cycle. For an enterprise, the "information life cycle" relates to the application of cyber threat information sharing" designed to improve the detection and mitigation of cyber threats and consists of six basic activities[5] as shown in Figure 2.

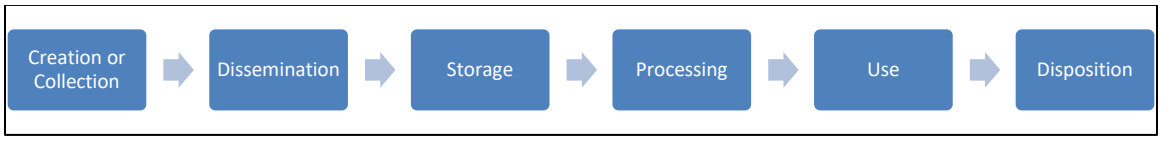

Figure 2: Information Life Cycle

1. **Creation or Collection**: generating or acquiring cyber threat information
2. **Dissemination**: distributing information to those elements and systems that will use, process, and analyze the information
3. **Storage**: short and long-term retention of information for use in analytical processing, alerting and forensic analysis or hunting efforts using databases, or other searchable repositories

[4] Ericka Chickowski, "Top 15 Indicators of Compromise" October 9, 2013. https://www.darkreading.com/attacks-breaches/top-15-indicators-of-compromise/d/d-id/1140647

[5] The information life cycle is taken from "OMB Circular A-130, Transmittal Memorandum #4"

4. **Processing**: aggregating, transforming, correlating, and analyzing stored information to identify applicability of the information or derived information to the operational security of the enterprise or its information.
5. **Use**: automating the application of measures to counter identified threats to the enterprise or applying the threat information to support operational actions to detect or minimize the impact of threats of primary importance and for use in any organizational decision-making
6. **Disposition**: implementing and enforcing policies for the retention and disposal of information to retain the effectiveness of automation efforts.

2.1.2. STRUCTURED AND UNSTRUCTURED DATA

The second framing concept is on the nature of the information being shared. Automation lends itself well to structured data, especially that which is machine readable, whereas humans are often better at working with some forms of unstructured data, such as verbally communicated information. Structured data is associated with a predefined data model, whereas unstructured data may consist of a narrative.

By using or selecting a more structured form of data, an organization can increase the options for automation. Some examples of structured formats are those employing Structured Threat Information Expression (commonly referred to as STIX), Common Vulnerability Reporting Framework (CVRF), other Extensible Markup Language (XML) approaches, or some product specific, potentially proprietary format.

Technologies do exist for supporting the transformation of unstructured data into more structured and machine-readable information. For example, the technology that underpins the ability of various home assistants (Amazon Alexa or Google Home) to turn voice commands into actions. For some forms of unstructured data, especially large datasets, artificial intelligence, augmented intelligence or machine learning, data analytics, and other specific technologies can provide levels of analysis that would not otherwise be available through other means. Incorporated into threat analytics processes, these technologies enable threat analysts to identify patterns of potential compromises within the data that human analysis alone may miss.

2.1.3. DIFFERENT TYPES OF AUTOMATION

The third framing concept is defining automation in the context of threat intelligence sharing. To help organizations think about automation and assess where automation can be used, we define four tiers of automation for information sharing. These can be used to categorize existing systems as well as used to define target states for future information sharing systems. These categories (as described in Table 1) are:

- No automation
- Manual processes supported by automation
- Automation with human oversight
- Full automation

The individual tier of automation that a system falls into is not necessarily a comment on the level of maturity of that process. While it is encouraged that organizations move away from manual processes, the decision as to the degree to which a process should be automated benefits from:

- An assessment of the process itself
- The resources available to support that processes
- The capabilities of available automation technology

If there is a need to combine an assessment of the level of automation for a process with a process maturity assessment, the tiers can be combined with an assessment of process effectiveness, such as the four implementation tiers from NIST CSF 1.1[6]. This can provide each system within an organization with both a description of the level of automation and the effectiveness of the process.

No automation	• Communication, processing, decision making, and actions all require human involvement. • Tools such as email, telephone, voice over internet protocol (VoIP), chat tools would be used but their use is initiated by humans, and the consumption, processing, and action are all initiated by humans. • Example: Threat intelligence is shared via a phone call between two or more individuals who make the decision on how to act on that information and manually make changes to their firewall rules based on the information shared.
Manual processes supported by automation	• Communication, processing, decision making, and actions contain substantial manual elements and the automated processes partially support or make suggestions that end users can act on. • This can be thought of as the traditional use of tools to support a process carried out by people. • Example: A user initiates a process to update company firewalls. Based on parameters entered by the user, the automation technology suggests changes to firewall rules, which the human reviews and instructs the technology to make.
Automation with human oversight	• Communication, processing, decision making, and action are automated, but there remains active human oversight. • This can be thought of as the machine doing most of the work, but a human needing to be present to make sure that the machine does not make mistakes. • Example: The technology automates changes to firewall rules based on provided threat intelligence. Humans actively review alerts and change logs at regular intervals, which provide details of what has changed and the information that led to the automated decision to make a change.
Full automation	• Communication, processing, decision making, and action are automated and human oversight is minimal or non-existent.

[6] See https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.04162018.pdf

	• Example: Malware is detected on a device. A calculated hash of the malware is automatically sent to a centralized internal threat repository supporting a publish-subscribe capability. The subscribed firewalls, intrusion prevention and mail gateways can now recognize the malware at the perimeter. Internal devices are then alerted to search for the specific instance of the malware. No human is needed to be involved.

Figure 3: Description of Tiers of Automation

3. PART 1: PLANNING

This section contains information that organizations can use to help plan introducing automation into an existing information sharing process or introduce a new automated information sharing process.

3.1. ESSENTIAL CONSIDERATIONS FOR AUTOMATING CYBER THREAT INTELLIGENCE SHARING

The following need to be considered and discussed when planning for the automation of threat intelligence:

- What is the ecosystem where information sharing is taking place, and what level of comprehension of this ecosystem do we have?
- Who are the stakeholders providing, consuming, and processing shared information?
- What are our goals and objectives in sharing information?
- What are the goals and objectives of the other identified stakeholders?
- What information is meaningful to share?
- Where do we have agreement on the meaning of information?
- Where do we have agreement on standards?
- Where do we have agreement on protocols for exchange?
- How will information be shared and used?

This information should be recorded and reviewed as part of the planning for information sharing.

3.2. CYBER THREAT INTELLIGENCE ECOSYSTEM

The cyber threat intelligence ecosystem is formed by companies, governmental entities (such as the Automated Indicator Sharing (AIS) system), groups, and individuals whose interactions may be formal or informal. Those interactions result in the sharing of various types of cyber threat-related information to help others know, understand, analyze, and react to threats to information and information system components. Some elements of this "community" or ecosystem are sources of indicators of newly identified cyber threats and others serve as aggregators and may provide searchable data bases of historical and new threat information. Some may provide analysis of the threats and procedures or capabilities to prevent or mitigate the effectiveness of threats. A number of service providers offer an array of digital products to automate the receipt of threat data of interest. Often interactions among members of this "community" can further broaden the knowledge of threats and collective methods of deterring, reducing the effectiveness, or negating specific threats or categories of threats.

Organizations wanting to capitalize on the vast array of cyber threat intelligence must fully understand what produces value for their efforts; as well as how they can become more effective users of cyber threat information by capitalizing on the use of appropriate automation capabilities.

3.3. STAKEHOLDER ENGAGEMENT

An organization pursuing information sharing to capitalize on available threat intelligence will require they engage multiple stakeholders both within their organization and others external to it. Some or all stakeholders may need to be engaged when automating information sharing processes. An organization must determine which stakeholders will be key to providing, processing, analyzing and consuming threat intelligence.

External stakeholders can include:

- Governmental entities, such as the Department of Homeland Security (DHS), law enforcement entities and/or regulatory agencies.
- Some organizations may be within sectors that have established formal information sharing organizations and in which they can participate.
- Commercial companies that provide unique or aggregated threat intelligence.
- Open source threat intelligence that may be freely available from organizations or individuals.

Internal Stakeholders can include:

- The organization's sponsor for directing and funding automation efforts.
- Departments responsible for engineering and integrating automation efforts into the organization operating environment.
- The group or groups responsible for operating, using and analyzing the threat intelligence.
- Those within management and operations who will consume and act upon the threat intelligence or authorize systems that will take autonomous action for specific threat information.

It is critical that the organization's leadership recognize the broad and varied engagement of stakeholders the automation of threat intelligence will entail and leads the establishment of both the strategic and tactical visions and plans the organization will require.

3.4. A GREEMENT ON THE ORGANIZATIONAL GOALS AND PURPOSES FOR INFORMATION SHARING

Agreement on the organizational goals and purpose for information sharing within an organization, and with other members of the information sharing ecosystem that the organization belongs, is essential. It is helpful to define success criteria for programs to automate information sharing processes so that all parties are aligned or understand the needs of others. For example, focusing resources on automating processes that add most value to the organization.

Communication and agreement on goals becomes more important for peer-to-peer sharing[7], especially where any programs to automate the sharing have substantial cost implications for the parties involved.

Information sharing can be a human to human, machine to machine, or machine to human process. For both humans and machines there must be some agreement as to what exchanged data means, how it is to be communicated, with whom, and how it is protected. For machine-based communications those agreements must be in a structured and standards-based form that enables such communications to be effective, accurate and secure. Humans are more able to handle "unstructured" information.

The layers of agreement must ultimately include:

- What information is meaningful to exchange within a community
  - Based on business needs, use cases, and processes
- The meaning of information to be exchanged
  - Based on vocabularies, conceptual models, and semantics
- Patterns and protocols for exchange
  - Based on kinds of interactions and protocols
- The terms, codes, and syntax used to exchange the information
  - Based on natural languages, data formats, and schema
- How information is to be exchanged
  - Utilizing voice, paper, networks, communications links or information repositories
- The parties or roles of parties that have the need and authority to exchange specific information
  - Based on the access rights to specific information, sharing agreements, identity, and authorization

The above must be agreed upon because ultimately all parties in a communication must agree on these things or act through some mediator that participates in such an agreement. Without all these agreements in place, the usefulness and security of information sharing is severely diminished, regardless of how it is realized. With those agreements in place resources can be allocated by each party to enable communications based on those agreements and leverage the resulting information sharing in support of their internal processes and objectives. Note that sometimes multiple layers of agreement are compressed into a single artifact.

For machines to be able to share information these agreements must be in some machine processable and formalized form – preferably based on recognized standards. Standards reduce the time, cost, and risk of sharing information and provide for leveraging information sources, technologies, products, and services built around those standards. For human to human communications natural languages are often used, however in many cases human centric information may be structured as forms, spreadsheets or reports.

[7] See appendix for definition.

Fortunately, many of these agreements come "pre-packaged" in industry standards, open source, and commercial products. Users and communities can leverage these packaged capabilities. While standards have advantages, it should be recognized that there will be no one technology, data format or schema that will be used for all information sharing relevant to cyber security – agility and flexibility in being able to communicate with many diverse parties and technologies, and understanding their information, is key to being a successful collaborator in any community.

3.5. DETERMINATION OF WHAT INFORMATION IS MEANINGFUL TO SHARE

The scope and detail of information sharing is based on the common needs, evolving knowledge of the threat environment, use cases, and processes within a community. These drive the requirements for the other layers of agreement that are the foundations for any successful sharing initiative. Meaningfulness within a community is derived from the needs and capabilities of the participants and a negotiation of what is to be shared.

An information sharing community is important as it identifies the current and potential parties that may want to share information for specific purposes. It provides scope and context for sharing agreements at all levels. Cyber threat intelligence is such a community that may also have more specific communities within that scope (like malware reporting) and may interact with other communities, such as law enforcement. That communities interact suggests the need for communities to have agreements and common standards that include but go beyond cyber threats.

The smallest information sharing community is two specific parties who have agreed to share some specific information in a specific way, this is referred to as point-to-point sharing. This point-to-point sharing is typical of many legacy systems and processes. The issues with point-to-point sharing is that it is very costly and anti-agile. Every point-to-point interaction must be agreed upon, designed and implemented. As organizations participate in many (sometimes hundreds or thousands) of such point-to-point agreements it becomes almost impossible to change their processes, systems or internal databases.

At the community level, flexibility and inclusiveness is key. The ability to share information within a community should not be confused with the rights or agreement for a specific entity to share specific information with another entity. In identifying scope, anything that may be of interest within the community for any process or specific set of actors should be considered. Rights, agreements and privacy are then managed after the community level needs are established.
The ISAO 100-2, *Guidelines for Establishing an ISAO*, provides a set of guiding questions around what information to share and how to share that information.[8]

3.6. AGREEMENT ON MEANING OF INFORMATION

For any set of parties to communicate, they must have a shared understanding of the meaning of the information – there must be agreement as to what the data is about and what the data represents. For informal human to human communications subject matter expertise and a shared vocabulary may be sufficient. For automated information sharing the meaning, or semantics, must be explicit

[8] ISAO 100-2: https://www.isao.org/wp-content/uploads/2016/10/ISAO-100-2-Guidelines-for-Establishing-an-ISAO-v1-01_Final.pdf

to guard against risky misinterpretation and costly redundant implementations. The degree to which semantics is explicit and independent of the data formats and technologies will, to a large degree, determine how flexible and safe information sharing will be. This will be discussed below. Explicit semantics may come in many forms, at various levels of formality and generality. At one end of the spectrum are vocabularies and definitions. Good terms and definitions are essential but may suffer from being a "human only" artifact that machines can't understand. Vocabularies also tend to be human language specific (e.g., written in French) such that communications across different countries remain difficult and error prone.

At the other end of the spectrum are conceptual models and ontologies that are intended to capture semantics represented in formalized languages such as Simple Knowledge Organization System, Unified Modeling Language, Web Ontology Language or Common Logic. These models may be used as "reference models" to mediate between different data formats and technologies and may also be leveraged to automate application needs like reasoning, correlation, simulation or pattern matching.

Even information sharing communities with no explicit formalized semantics must have some implicit semantics behind the information they share, otherwise data would be meaningless. However, failure to specify explicit semantics in some way risks dangerous misunderstandings or failure to enable meaningful communications among all parties.

3.7. AGREEMENT ON STANDARDS

Any information exchange will have a syntax and some form of structure or set of terms used within that syntax to identify data elements representing the semantics of meaningful information. Humans use natural language syntax, while machines typically use some form of data structure or schema. Common examples include XML Schema, Entity–relationship Model E/R Models, Resource Description Framework Schema and Integration Definition and Function Modeling (IDEF-0).

Data schema specify a specific way to efficiently "package" data representing meaningful semantics, using a specific technology, for some specific purpose, exchange or process. Internal applications and database management systems (DBMS) also have schema, frequently representing the same semantics as what is shared; however, it is not required and generally not effective to require internal application schema to have to match external information sharing schema, even when they share the same semantics. It is best to "decouple" internal systems and databases from external information sharing to allow each to evolve and be managed independently. Also, most organizations will have multiple sharing partners that use different schema.

The same or related information semantics may be packaged in different schema for different purposes, applications or different exchange partners. In some legacy systems semantics are only specified in terms of data schema definition text, which makes it difficult to share and correlate information across different schema. It is best practice to define semantics independently based on stakeholder-relevant concepts and then map technology focused data schema to the semantic definitions. Requiring this separation of concerns makes it less risky and costly to manage change and support multiple applications and exchange partners.

3.8. AGREEMENT ON MECHANISMS FOR EXCHANGE

There are a limited number of patterns for information exchange implemented by many technology protocols. The basic exchange patterns are:

1. **Query of information repositories:** This is a "client driven" model where some data store, service, repository or "data lake" is "queried" for information the client requires. There must be some prior agreement or specification of the information in the repository or how to determine that information. Think of this like a trip to the library or a "data call". Typical technologies include Structured Query Language (SQL), Hyper Text Markup Language (HTML), and REpresentational State Transfer (REST- Query).
2. **Broadcast:** The broadcast pattern is provider driven. The provider "broadcasts" information they determine is relevant to some group or community able and authorized to receive the broadcast. The syntax and semantics of the broadcast must be mutually understood. Think of this like an email to a group. Typical technologies include message queuing mechanisms like Java Message Service (JMS) and Data Distribution Service (DDS).
3. **Directed:** In a directed exchange information is sent to one recipient or a set of specific recipients based on some pre-determined exchange agreement. Think of this like an email to an individual or a person to person conversion. Typical technologies include Electronic Data Interchange (EDI), email, and Simple Object Access Protocol (SOAP).
4. **Negotiated:** A negotiated exchange may be client or provider driven and requires negotiation and agreement on a per-message or per-process basis. This exchange pattern is typically used for very sensitive information that may require approval on a per-partner basis. The "directed" technologies may be used for negotiated exchanges, typically with a specific exchange agreement.

Based on the basic exchange pattern a technology specific protocol specification and a data schema is used to implement the exchange for a specific purpose or process. There are multiple technical standards for each pattern.

3.9. HOW INFORMATION IS TO BE EXCHANGED

The actual sending and receiving of information, and even the same exchange patterns, may be implemented over a variety of technical media. TCP/IP is by far the most common, but other technologies are used in specific communities. The low-level exchange mechanisms are almost always pre-packaged and based on industry standards. Refer to *ISAO 300-1 Introduction to Information Sharing*, Section 7.3 Sharing Mechanisms[9] for a listing of means to consider.

4. PART 2: DESIGN

This section contains information organizations can use to help design automated processes for capitalizing on information sharing. The assumption is that an organization would have gone through a planning process that can be used as an input into the design phase.

[9] ISAO 300-1 Section 7.3: https://www.isao.org/wp-content/uploads/2016/10/ISAO-300-1-Introduction-to-Information-Sharing-v1-01_Final.pdf

4.1. ESTABLISHING AN ENTERPRISE REQUIREMENT: REFERENCE TO THE MISSION AND GOALS OF THE ORGANIZATION

As with any initiative, the processes in this section should reference the mission and goals of the organization. For example, if the mission of the organization includes providing support and services during a crisis, then the threat data feeds and surrounding processes need to be sufficiently resilient that they continue to operate during crisis situations.

4.1.1. DATA CONSUMER REQUIREMENTS

Start by understanding who and what within the organization requires cyber threat data – these are referred to as data consumers. Data consumers may be the end user of the data, or automation software that uses or enriches the data prior to it being sent to another data consumer.

It is important to identify all of the intended data consumers. It may be possible to consolidate the list to avoid duplicates, e.g., a process and the team who performs that processing could be consolidated into a single data consumer.

The names and descriptions of the data consumers should be recorded in a data requirements document. The following are examples of who and which teams may require data:

Organizational:

- Security Operations team
- Threat Intelligence team
- Network security / Change management
- Cyber Risk Management staff
- Internal cyber risk status reporting
- C-Suite and the Corporate Board of Directors
- Organizational automation and processes that utilize the information for detection, prevention, alerting and reporting

Sharing partnerships:

- Partnership Members

Product capabilities:

- Customer reporting
- Product intelligence and capabilities

The list above is representative of the types of consumers who should be considered when trying to identify the consumers of the data relevant to the organization. This is by no means a complete list.

4.1.2. CONSUMER NEEDS

For each data consumer recorded in a data requirements document, the data requirements should be recorded. Data requirements can include:

- Type of data

- Level of detail
- Format of the data
- Frequency of data
- Whether data is pulled on demand or pushed
- Quality of data
- Amount of data
- Trustworthiness
- Potential value of the data
- Applicability of the data
- Cost to acquire data
- Ease of filtering/searching
- Whether relationships to other data elements are already established (e.g., is the data in a graph database?)
- Need for associated meta data (e.g., audit trails and information supporting traceability)

4.1.3. CREATE DATA REQUIREMENTS DOCUMENT

The information collected should be recorded in a data requirements document. If the organization has many data consumers it may prove useful to create a consolidated set of data requirements. This is so a simplified set of requirements can be presented to vendors and/or used for implementation activities.

The expectation is that this document can act as a guide to the rest of the data ingestion activities. As such, version and other good document management practices are recommended.

A single document outlining the data needs of the stakeholders within the organization can prove useful if the organization is made up of stakeholders who have differing needs or preferences for one data provider or technology over another. This is because it allows the organization to conduct vendor selection based on agreed upon requirements.

4.1.4. IDENTIFY DATA SOURCES

Identification of data sources is the next logical step in the process after establishment of requirements. This section covers identification and assessment of potential data sources, review of gaps and data transformation requirements, source selection, architecture, integration, data quality, and data model integration.

4.1.4.1. DETERMINE POTENTIAL SOURCES OF DATA THAT CAN MEET THE DATA REQUIREMENTS

Based on a data requirements documentation, a long list of vendors and other data sources can be generated. Where there are multiple stakeholders within the organization, canvasing these stakeholders to understand if there are any data sources or vendors they would like added to the long list can be beneficial.

Data can be obtained from many sources such as:

- Public and commercially available intelligence feeds
- U.S. Government agencies

- Governmental sources in foreign countries
- Members
- ISAOs
- Formal and informal affinity groups of subject matter experts and researchers

4.1.4.2. ASSESS EACH DATA SOURCE AGAINST REQUIREMENTS AND CREATE DATA SOURCE SHORT LIST

Using a data requirements document each data provider should be assessed to understand to what degree they can meet the requirements of the organization.

4.1.4.3. ASSESSING AND ESTABLISHING TRUST IN A DATA SOURCE

Trust in a data source can be viewed as the data source provider's ability to meet a set of expectations about the type, frequency, quality, etc. of the data it provides.

Trust can be assessed and established with a data provider in the following ways:

- Reputation – the data provider is used by many other organizations, who can attest to their level of trust
- Controls and processes that the data provider has in place – verify the data provider has processes ensuring the ongoing quality of their data
- Contractual agreements and service level agreements (SLAs) with the organization – a contractual agreement that defines the level of service to be provided
- Communication to set clear expectation – trust can be established by each party communicating their needs, ability to provide services, and indicating when there is a change in either
- Track record – working with a data provider over time builds trust as each party understands the needs and abilities of the other

4.1.4.4. UNDERSTAND ANY CONSTRAINTS OR REQUIREMENTS ASSOCIATED WITH THE USE OF DATA FROM LISTED SOURCES

While some information is open and freely available, other critical information can only be shared with specific parties for specific purposes. One simple model used in some information sharing environments to identify a sharing policy is the Traffic Light Protocol (TLP)[10]. Safety, security and privacy must be designed into the foundation of information sharing environments and specifications.

Some data providers may place conditions on the organization if they are to send/share their data. These may include ensuring that the organization or the data consumers have a sufficient clearance level. Data providers may also want to understand and/or ensure that the consuming organization has sufficient controls in place or adheres to necessary standards for handling the data provided.

[10] See the Forum of Incident Response and Security Teams (FIRST) discussion of TLP at https://www.first.org/tlp/

Producers and consumers must have a clear understanding of how shared information can and cannot be used. Creating clear policies and agreements will minimize misinterpretation of requirements. An Information Exchange Policy (IEP)[11] framework, as an example, identifies areas that should be addressed in such policies.

4.1.4.5. DETERMINE GAPS OR DATA TRANSFORMATION REQUIRED FOR SHORT LISTED DATA SOURCES TO MEET REQUIREMENTS

It is possible that no single data provider will be able to supply data that meets the data consumer's requirements. Where this is the case the organization will need to determine what processes will need to be put in place to transform the data into something that meets the needs of the data consumers.

Examples of transformation processes are:

- Data cleansing
- Combining data from multiple sources
- Data enrichments
- Filtering

Any data transformation or data processing requirements will need to be factored into the selection of data sources.

4.1.4.6. DATA SOURCE SELECTION

The selection of data source providers should be based on a clear understanding of:

- Who or what are the data consumers
- The needs of the data consumers
- A long list of potential data providers, where stakeholders have been given opportunity to suggest vendors and data providers for consideration
- An assessment of how well the data providers meet the needs of the data consumers
- An understanding of the requirements that potential data providers would have on the organization ingesting the data
- An assessment of where any gaps between the requirements and what can be provided are, and any data transformations that are needed

The selection of a data provider should be fact based and auditable. This is especially true where an organization has accountability to a board or other stakeholders who may favor certain vendors or stakeholder's requirements over others. To support the auditability of data source selection, the selection process should be defined, and the results documented.

4.1.4.7. DEFINE DATA INGESTION ARCHITECTURE

A data ingestion architecture consists of:

- The data model
- The data policies, procedures, and controls

[11] See the Forum of Incident Response and Security Teams (FIRST) discussion of IEP at https://www.first.org/iep/

- Processes, technologies, etc. that support data quality and the needs of the data consumers and the organization

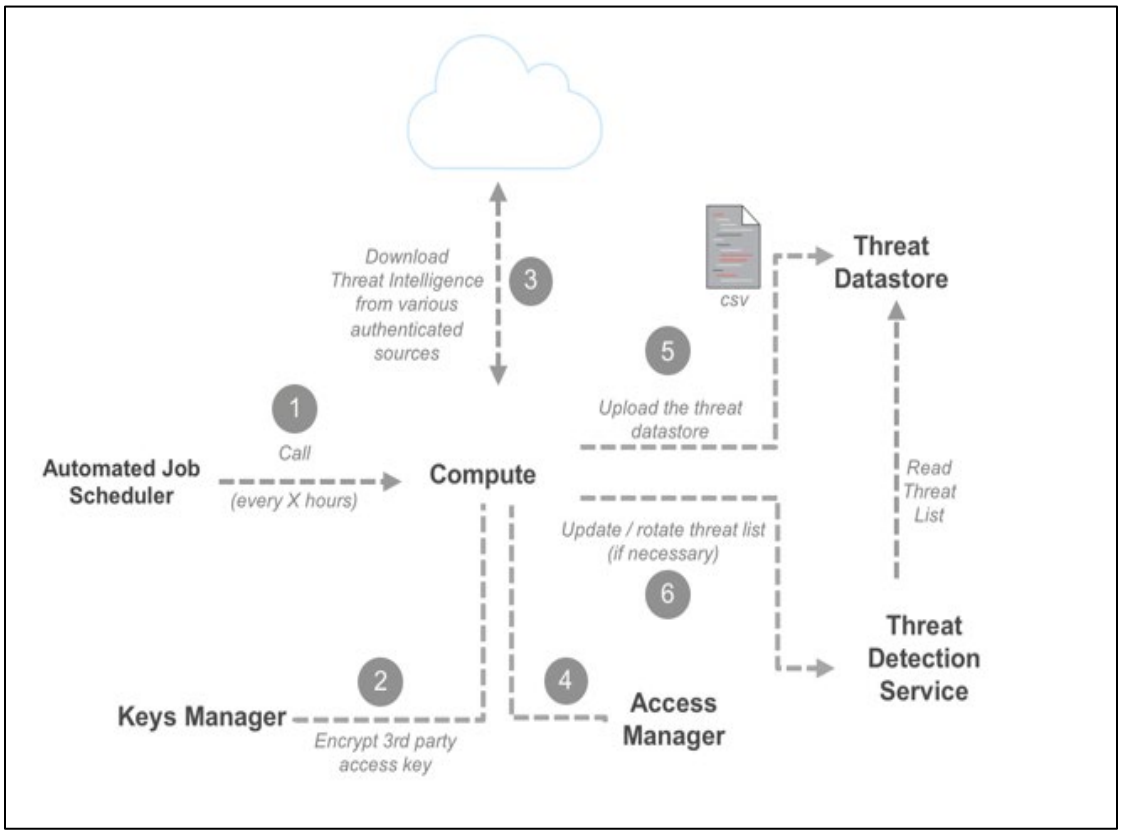

Figure 4: Sample Reference Architecture Diagram

The following resources are used in this solution:

1. An Automated Job Scheduler periodically invokes a Compute function.
2. The Keys Manager securely stores the public and private keys that you provide. These keys are required to download the threat feeds.
3. The compute function that consists of a script that programmatically imports a licensed Threat Intelligence feed into Threat Detection Service.
4. Access Manager that gives the Compute function access to the following:
    a. Threat Detection Service, to list, create, obtain, and update threat lists.
    b. Logs Store Service, to monitor, store, and access log files generated by Compute Service.
    c. Threat Datastore, to upload threat lists and ingest them into the Threat Detection Service.

The Compute function that updates your threat lists is invoked right after you provision the solution. It's also set to run periodically to keep your environment updated. However, in scenarios that require faster updates to your threat intelligence lists, such as the discovery of a new Zero Day vulnerability, you can manually run the Compute function to avoid waiting until the scheduled update event.

This solution provides a large amount of individual threat intelligence data to process and report findings on. Furthermore, as newer threat feeds are published by the threat intelligence feed provider of your choice, they will be automatically ingested into the Threat Detection Service.

4.1.4.8. INTEGRATION WITH EXISTING DATA MANAGEMENT PRACTICES WITHIN THE ORGANIZATIONS

It is important that the ingestion of the data works with the existing data management practices of the organizations. If the organization is forming, it may be necessary for the organizations to define data management practices.

4.1.4.9. DATA QUALITY

Ingestion should work with the organization's policies, procedures, processes, controls, and technologies that support data quality.

There are multiple definitions of data quality. What is important is selecting a definition that is meaningful to the organization and the needs of the data consumers (located in a data requirements document). Example characteristics of data quality are:

- Timeliness
- Existence
- Completeness
- Integrity
- Consistency
- Accuracy
- Interpretability
- Uniqueness
- Availability

The organization will need policies, procedures, processes, controls, and technologies that support the needed level of data quality.

4.1.4.10. DATA MODEL INTEGRATION

If the organization has an existing data model, it will be important that the ingested data be integrated into this model, If the use of data is relatively simple, then the data model should be relatively simple. A simple data model would contain the sources of data and how they relate to each other.

4.1.4.11. DATA CONTROLS AND COMPLIANCE

The ingested data should comply with existing policies, procedures, and control for data within the organization.

4.1.5. DEFINING A MASTER LIST OF DATA SOURCES

Defining a master list of all data sources is an important process. Understanding what data the organization is consuming, the source of that data, information about the source (e.g., company

name, SLAs, contact details), and relevant information about the data better enables the organization to manage its data.

The master list should be maintained, and processes put in place to ensure that it remains accurate and up to date. This could include processes followed when adding, changing, or removing a data source.

4.1.6. DEFINING END POINTS FOR DATA AND THE FLOW OF DATA TO THESE SOURCES

Mapping the flow of data through the organization from source to final consumers will enable the organization to understand how the data is used within the organization. The goals of mapping the data flow are:

- Understanding what processes use the data
- Understanding what technologies/systems use the data
- Understanding who should have access to the data
- Knowing where the data is being stored
- Knowing where and how the data is being transformed/manipulated
- Understanding the impact that a loss of a data source would have
- Determining if there are any bottlenecks in the process that uses the data

The mapping of data does not need to be a complex process, and while there are techniques like data flow diagrams available, the organization should focus on performing the mapping in a way that meets the above goals. The mapping of the data flow is used as input into the data ingestion process described in Section 4.5.

4.2. TECHNOLOGY STACKS AND AUTOMATING INFORMATION SHARING

Fortunately, a lot of agreement has already been found for the "lower levels" of information sharing. This agreement appears in technology stacks – web servers, enterprise service busses, and messaging systems available from open source and multiple commercial vendors. Most of these technology stacks leverage industry standards such that they are interoperable at the technology level – that is they provide the technology infrastructure to implement some or all of the exchange patterns using compatible schema languages, protocols, identity management and authorization. By using one of these pre-built stacks, or multiple stacks that implement the same standards, users and communities do not have to worry as much about the mechanics of exchange, they can concentrate on what is to be exchanged and with whom. Identification and selection of a technology stack for automated information sharing necessitates thoughtful consideration of requirements and evaluation of available offerings. Users can begin researching open and/or commercial cyber threat intelligence and information sharing and analysis systems through investigation on the internet or other sources such as an existing ISAO.

As there will be multiple internal and external schema representing the same or related data about the same things, it is necessary to map data formats and to combine multiple data sources into a common form for advanced analytics – to "connect the dots". The semantic model as defined above, when combined with a suitable infrastructure, facilitates the automation of these mappings

and data federations. Once a suitable schema is defined, a semantic model federation and mapping can be automated to every other information source in the same way. This kind of "semantic mediation" can dramatically lower the time, cost and risk of information sharing.

4.3. OPERATIONAL CONSIDERATIONS

An organization's operational considerations are directly affected by the business strategy the organization employs for its information technology (IT), networking and information services it uses, along with those of critical partners interconnected with its IT environment. This overall "enterprise architecture" will dictate the essential types of threat intelligence the organization should be receiving.

As an example, if the enterprise is only employing endpoint devices to access services and data contained in a service provider's environment, an essential operational consideration is assuring receipt of threat intelligence germane to managing the risks of its endpoint systems and network connections. Conversely, this organization can be a valuable source of threat intelligence related to endpoint systems.

More complex operational considerations apply to those enterprises where it is operating and managing the security of their own IT infrastructure and applications and a large array of customer or business services, especially those with Internet facing operations. Throughout this spectrum of operational considerations, the timeliness of the threat intelligence can materially affect its effective use. The application of automation to the processes of receipt (ingesting), correlating applicability, and incorporating it to mitigate risks are becoming more broadly recognized as the best practice needed to deal with the expanding threat environment in which organizations operate.

If your enterprise is receiving cyber threat intelligence but it takes an unacceptable amount of time before new or an adjustment to defensive measures are implemented or appropriate remediation is acted upon, you have not effectively operationalized the use of threat intelligence.

The continuous efforts of attackers working to exploit cybersecurity and those defending their enterprise are well illustrated through the importance of a long-standing military concept known as the "OODA loop", which refers to the decision cycle elements of observing, orienting, deciding and acting. Critical to effective use of this concept is determining where to direct one's energies to defeat or minimize the impact of an adversary's efforts and to act quickly.

This guideline is focused on the operational considerations an organization must address as a recipient of threat intelligence information and identifying where and how automation can improve an enterprise's risk management efforts and decision cycle. Further, operational and other considerations are discussed that can permit an enterprise to be more effective in developing and sharing threat intelligence, it may create.

There are some basic operational considerations an enterprise must consider as it assesses the "what" and "how" of automating cyber threat intelligence use. The very first step in automating cyber threat intelligence for your organization must be an examination of the nature of your organization's operations. What are its business applications and supporting IT infrastructure assets; along with its approach to cybersecurity risk management. This inventory will begin to guide the

type, where and how threat intelligence could be applied. This guideline can also help to identify shortcomings in operations where more effective defensive and remediation security processes can be employed driven by cyber threat intelligence.

For discussion purposes, enterprises are grouped into just three categories to consider the operational use of cyber threat intelligence. The three categories are shown in Figure 4.

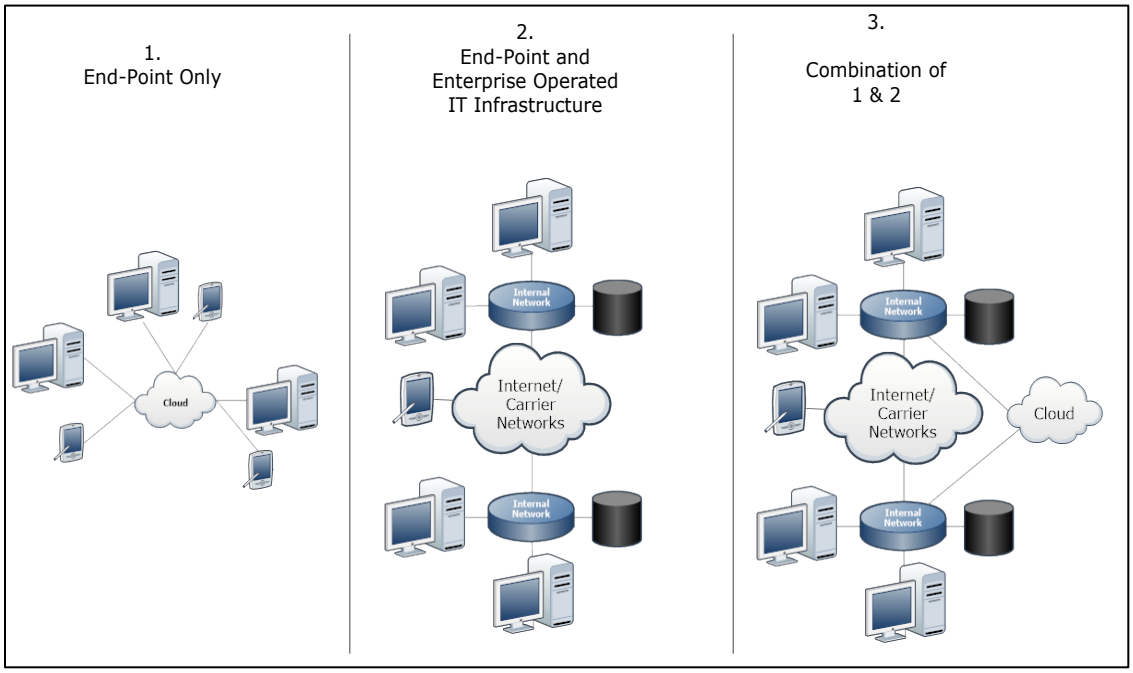

Figure 5: Enterprise Categories

- In category one, the enterprise has employees who access IT services and applications using end-point devices and utilize communications services, software applications, storage and other IT services provided by a third party.
- In category two, the enterprise operates and manages its own IT infrastructure to include end-point devices, communication services, software applications, storage, and other services likely with the assistance of some contracted services and third-party devices; especially for network connectivity if the organization has a geographically dispersed operations structure.
- Today, and more so in the future, most organizations will have operations that fall into category three; e.g., a range of cloud-based applications/services provided by a third party, its end-point devices, and some of its own IT infrastructure.

Given that context, the operational considerations addressed in this guideline are directly applicable to those operating and managing significant IT systems and infrastructure themselves, even if third party "cloud-based" services are involved. The guidance provided in this document can then be decomposed to specifically address categories 1 and 2, which are subsets of 3.

Another useful set of information is the identification of the key business objectives of the organization and a current risk management assessment of cybersecurity practices; perhaps, using the National Institute of Standards and Technology Cybersecurity Framework (NIST CSF) approach,

to identify the most operationally critical systems and organizational business processes, and details of how cybersecurity risks are being managed.

For those organizations with an established IT and security enterprise architecture and standards, that information provides a leg up for establishing the specific technology-based threat intelligence of most importance to the operations. Otherwise, the inventory of the systems used within the enterprise must be catalogued by the organization.

The next step is to determine how the current security and operational processes use threat intelligence today. Throughout this spectrum of operational considerations, the timeliness of the threat intelligence can materially affect its effective use. The application of automation to the processes of receipt (ingesting), correlating applicability, and incorporating it to mitigate risks are becoming more broadly recognized as the best practice needed to deal with the expanding threat environment in which organizations operate.

4.4. ARCHITECTURAL HIGH-LEVEL MODEL FOR ENTERPRISE AUTOMATION OF CYBER THREAT INTELLIGENCE

An organization's ability to detect and respond quickly, if not immediately, against cyber-attacks is critical to a successful defense against a developing threat campaign. To accomplish this, many organizations are looking to enhance their ability to automate responses to these threats. According to a 2018 survey conducted by the SANS Institute, "39% of respondents cite the lack of interoperability and automation as a key inhibitor to fully implementing and using" cyber threat intelligence.[12]

There are many ways an organization can automate the use of machine-readable threat intelligence within their network. The way this is accomplished will depend on a variety of factors, such as the organization's security budget, personnel training and experience, risk of experiencing an advanced or sophisticated cyber-attack, and existing network defense infrastructure, such as Security Incident and Event Management (SIEM) systems, Next Generation Firewalls (NGFW), Intrusion Detection/Prevention Systems (IDS, IPS), Threat Intelligence Platforms (TIP), and Endpoint Detection and Response solutions (EDR).[13] Organizations with smaller budgets, less risk, and fewer personnel may rely on threat intelligence feeds provided through existing vendors and integrated with existing network defenses (e.g., SIEMs, IDS, IPS, NGFWs, EDRs). In many cases, these feeds can be "turned on" by the vendors as part of existing packages or for an additional fee. There are also many open source solutions to meet this capability in which openly available intelligence feeds can be integrated into existing network devices using APIs, depending on the device and source of the feeds.[14] Organizations at a higher risk of cyber-attacks, e.g., the financial or manufacturing industries, and with larger budgets and more personnel, may be more likely to implement

[12] SANS Institute, CTI in Security Operations: SANS 2018 Cyber Threat Intelligence Survey, February 2018.

[13] Gartner, Market Guide for Security Threat Intelligence Products and Services, 20 July 2017

[14] Ibid

processes consistent with the concept of Security Orchestration, Automation, and Response (SOAR). According to Gartner, SOAR references "technologies that enable organizations to collect security threats data and alerts from different sources, where incident analysis and triage can be performed leveraging a combination of human and machine power to help define, prioritize and drive standardized incident response activities according to a standard workflow."[15]

In these cases, organizations may use a product designed to manage threat intelligence, including a TIP that would allow personnel to analyze external threat information, correlate this activity with internal network activity, and respond to threats through automating incident response "playbooks." Benefits of this approach to automation include the ability to better analyze existing and emerging threats, identify their presence in the network, and mitigate these threats quickly through automated and semi-automated responses that benefit from the direct integration with network defenses.[16]

The high-level model depicted in Figure 5 identify six key elements of any cyber threat automation effort with an enterprise.

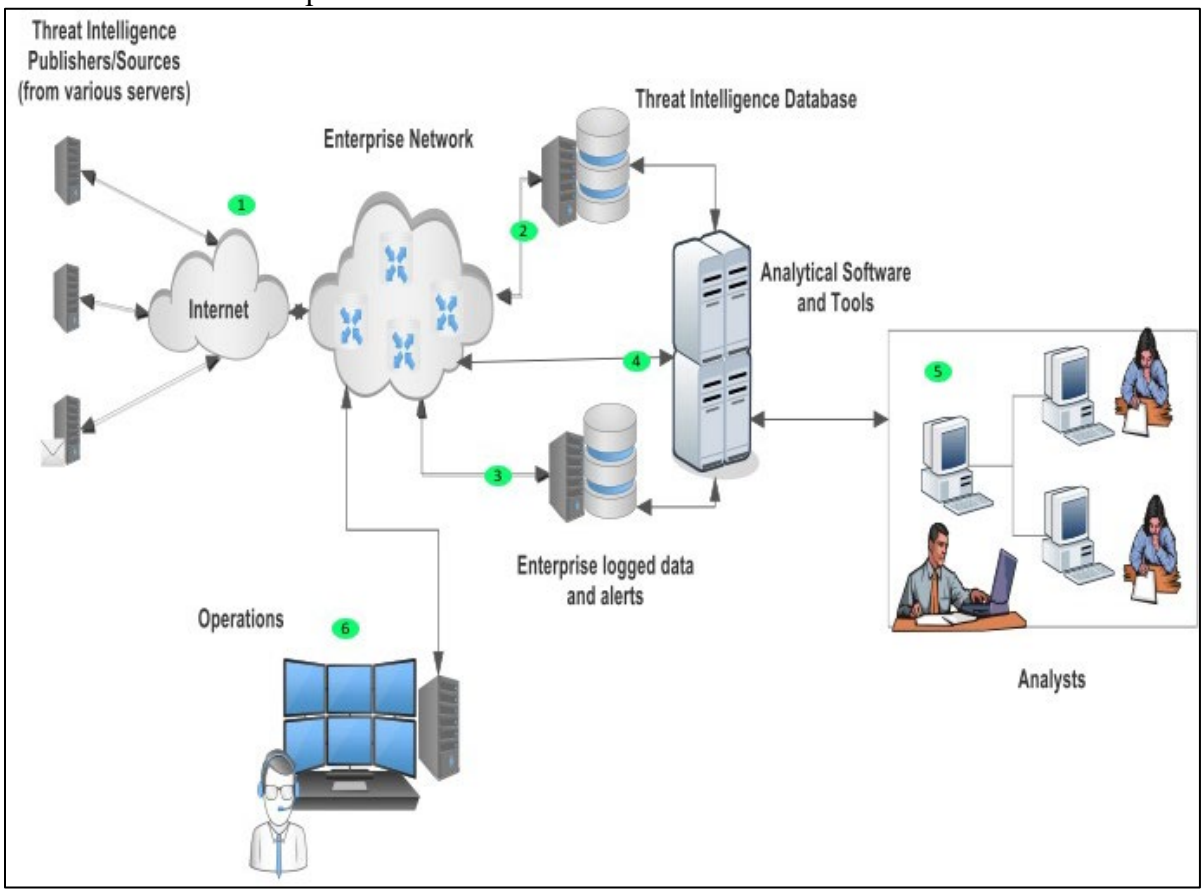

Figure 6: Aspects of Automating Cyber Threat Intelligence

Element 1 – represents sources of digital cyber threat intelligence. E.g., open source or commercially available threat feeds, indicators received through Trusted Automated eXchange of Indicator

[15] Gartner, Innovation Insight for Security Orchestration, Automation and Response, 30 November 2017

[16] Ibid

Information (commonly referred to as TAXII) servers or other automated means of information sharing.

Element 2 – represents the translation and storage of that information to be used in various analytical processes. This function could be met through a TIP or similar product that would also combine some or all of the capabilities in Item 4.

Element 3 – represents internal enterprise data used for analysis or additional internally generated alerts or logged information from the enterprise. This function can often be met through a SIEM.

Element 4 – represents any analytical, forensic or cyber hunting software or tool used by analysts or automated instructions sent to operations or systems designed to defend or mitigate threat to the enterprise's systems (e.g., SIEM, NGFW, IDS/IPS, EDR). The capabilities in Item 2 and Item 4 are often found to varying degrees in TIPs.

Element 5 – operational analysis capabilities taking advantage of the analytical processes in examining any combination of threats, vulnerabilities, incidents, or practices that results in methods to protect specific data, infrastructure, or functions.

Element 6 – operations can implement or undertake actions based on the analytics and/or analysis results. This can be done using various approaches as discussed in Section 2.1.3, Different Types of Automation, of this document.

4.5. DATA INGESTION PROCESSES

The data ingestion process defines the processes that are needed to effectively ingest the data and ensure it reaches the data consumers in a format that meets their requirements. The data ingestion process can span multiple systems, which are used to transport data from source to destination.

The following considerations are needed for the data ingestion process:

1. The technology solutions available to ingest data.
2. The formats, standards, and protocols that data to be ingested adheres to.
3. The required capacity, availability, security, and resilience of the data ingestion process.
4. How the ingestion process will be monitored and managed.

4.5.1. THE TECHNOLOGY SOLUTIONS AVAILABLE TO INGEST DATA

An understanding of the systems and technologies used is required to transport the data from source to where it is ultimately used.

The Information Life Cycle process steps discussed in section 2.1.1 earlier can be used to assist in collecting this information.

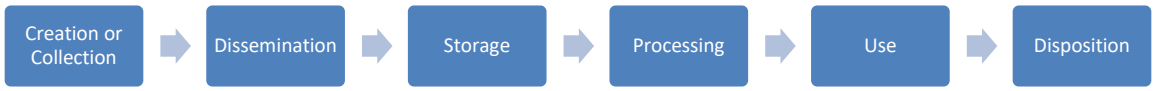

Figure 7: Information Life Cycle

At each step in the process, an understanding of the technologies used, how the technologies communicate with each other, and data formats that each technology can ingest and disseminate.

For effective automation, the technologies involved in transporting of data should be able to communicate with each other.

A formal process for data ingestion, with appropriate controls, in needed. This would include transformation, cleansing, and enrichment of data, along with appropriate security controls. As such a level of understanding needs to be reached regarding automation and human involvement at each step.

4.5.2. THE FORMATS, STANDARDS, AND PROTOCOLS FOR DATA TO BE INGESTED

Consideration should also be made for how information can pass from unsecure or public networks to secure areas within an organization's network, as well as the controls that will need to be in place to enable the data to cross from an untrusted domain to a secure domain. If areas of the network are "air gapped" workable solutions for getting information to its intended end user will need to take this into account.

The data ingested should adhere to the technical standards and protocols discussed in Section 3.8. Consideration should also be given to how the information falls into the standards and classification of the consuming organization.

It is also important to determine how information being ingested falls into an organization's data classification model. Ideally this information should be available from the provider of the data. However, organizations may want to consider tools to monitor incoming information to detect potentially sensitive or classified data. The Traffic Light Protocol[9] describes rules on how shared information can be handled. Data can also come from sources that require levels of official governmental clearance to access.

4.5.3. THE REQUIRED CAPACITY, AVAILABILITY, SECURITY, AND RESILIENCE OF THE DATA INGESTION PROCESS

The following questions can help determine the required levels of capacity, availability, security, and resilience in the data ingestion process. What would happen if:

- The process that ingests data stops working?
- The provider of the data increases the volume of data one hundred-fold?

- A threat actor attempted to use the data ingestion process as a point of ingress into your system?
- The provider accidently sent a different format of data?

Proper consideration of the availability, capacity, security, and resilience of the data ingestion process should be made in its design. The level of availability, capacity, security, and resilience of the data ingestion process should be in proportion to the criticality of the ingestion process to goals of the organization, and weigh the costs required to achieve it.

4.5.4. HOW THE INGESTION PROCESS WILL BE MONITORED AND MANAGED

The level of monitoring of the ingestion process should be in proportion to the criticality of the ingestion process to goals of the organization and weigh the costs required to achieve it. While options from constant real-time monitoring to reviewing of log files are possibilities, it is suggested that any errors in an automated ingestion process integrate with an organization's event management tools and processes. So that if errors occur in the data ingestion process, events are triggered and sent to the organization's event management toolset, where appropriate rules to address any errors can be defined.

4.6. DEFINING DATA TRANSFORMATIONS

At any stage in the Information Life Cycle data may need to be transformed. Data transformation can include one or more of the followings:

- Data cleansing
- Data enrichment
- Conversion of format

Any transformation process should be well defined and documented. The level of detail needed in the documentation will vary based on the needs of data consumers. For example, if the data consumer is a software tool that has very precise requirements about the data fields, then the documentation will need to take this into account.

4.6.1. DATA CLEANSING

Data cleansing aims to increase the quality of the data. Data quality, as discussed in Section 4.1.4.9, can be defined as having the following dimensions[17]:

- Completeness – the degree to which the data represents 100 percent of the data that is available.
- Uniqueness – that each piece of data is recorded only once and there are no duplicate records.
- Timeliness – the degree to which the data represents reality at a point in time.
- Validity – data are valid if it conforms to the syntax (format, type, range) of its definition.

[17] Adapted from "THE SIX PRIMARY DIMENSIONS FOR DATA QUALITY ASSESSMENT", DAMA UK. See https://www.whitepapers.em360tech.com/wp-content/files_mf/1407250286DAMAUKDQDimensionsWhitePaperR37.pdf

- Accuracy – The degree to which data correctly describes the "real world" object or event being described.
- Consistency – The absence of difference when comparing two or more representations of a thing against a definition.

As the quality of data increases its value to the organization, processes to improve data quality are desirable. Data cleansing performed manually can be time consuming and therefore costly, and as such any automation of data cleansing processes is recommended, where possible.

Tools are available for both assessing levels of data quality and supporting cleanse activities.

4.6.2. DATA ENRICHMENT

Data enrichment seeks to add value to data by enhancing, refining, and otherwise improving raw data. This can include:

- Combining data from multiple sources.
- Making the information more specific for the organization using it.
- Making the data easier to read by (human) end users.

There are tools available to support data enrichment.

The specific scenario that is most relevant for an organization regarding information sharing is making generic information provided to it specific for that organization's environment. At a high-level filtering shared information for those data points that only affected the systems and technologies deployed within an organization should be achievable. Two key considerations for achieving this are 1) ensuring that an accurate and up-to-date list of deployed technologies is maintained, and 2) ensuring that ingested data is tagged with the technologies it is applicable to.

4.6.3. CONVERSION OF FORMAT

Data format conversion may be required to ensure different systems can read or process the data. It may also be necessary to transform unstructured or semi-structured data into structured data to allow another system to process it.

While tools are available to support the conversation of data from one format to another, these may require customization for the systems that the organizations are using to process the data. If these systems require a non-standard format of data, custom scripts or other methods will need to be deployed to convert data in a useable format.

Several off-the-shelf services are available, including those from leading cloud providers, to support the conversation of unstructured data—in the form of speech or written text—into commands or data formats that can be processed by other systems.

4.7. DATA DISPOSITION

In the final step of the Information Life Cycle, data disposition also needs careful consideration. Processes and solutions need to be designed to delete data once it is no longer needed. This becomes especially relevant if the information shared with an organization contains any Personally

Identifiable Information (or other information that may be subject to regulatory scrutiny). In addition, storage and backing up of information that is no longer needed has costs associated with it.

Assuming all data has been digested and categorized following an organization's information classification policies and procedures, the disposition of this data should be in alignment with these policies and procedures.

4.8. DATA SUPPLIER MANAGEMENT

Data suppliers should be managed in accordance with the organization's supplier management processes.

Data mapping should be used to determine which data suppliers are critical for the operations of the organization. These suppliers need to be managed, such as with an SLA, in accordance with how critical they are to the organization's operations.

4.9. DEFINING ROLES AND RESPONSIBILITIES

Roles and responsibilities for the management of data should be clearly defined. The data flow mapping process should be able to provide information to support this process.

4.10. DEFINING DATA ASSESSMENTS AND FEEDBACK PROCESSES

Feedback processes can be useful for both the consumers and producers of data. Consumers should provide feedback about how the data was used and its effectiveness. Producers then may tailor or improve the quality of data they produce. This feedback also aids other consuming organizations in determining whether a piece of data is valuable.

Data consumers can provide feedback in qualitative and quantitative means. These means can be automated, or they can be manual in nature. The simplest form of feedback is one person providing written or verbal feedback on the services provided to them by the data provider. More complex feedback processes could be automated to provide feedback when the data is used.

There are a number of areas where a consumer of data can provide feedback, these include:

- Whether the data was used
- Where the data was applicable to the organization
- Whether the data enabled the organization to detect a threat
- The cost to the organization to use the data (in terms of CPU, network, and memory usage)
- How easy the data was to use
- Where the data had any data quality issues

As there will be costs or defining and implementing feedback processes, if and where feedback processes are used, they should be designed in such a way that they produce useful information for either the data provider or other consumers of the data. It may also be useful to determine whether the data provider is able to act based on the feedback provided.

5. PART 3: IMPLEMENTATION

Implementation should follow the planning and design phases, and the outputs from these phases should be used during implementation.

5.1. AN IMPLEMENTATION GAME PLAN

Improving an organization's effectiveness in the use of cyber threat intelligence will require a broad commitment across the organization. As such, the decision to support the effort must have broad executive level support and buy-in from the involved organizations. Development of the plan may be led by the operational and security organizations. The plan should address the "as is" state and how increased automation will be phased into the organization. Some practical considerations are described in Appendix A and efforts done by the Department of Homeland Security, National Security Agency (NSA) and John Hopkin's University Applied Physics Laboratory on an Integrated Adaptive Cyber Defense (IACD) framework[18] initiative, offers additional considerations.

Implementing the plan may be a multi-year effort depending on the size of the organization and should be considered a major initiative in any organization. Since the effort will engage and require cross-organizational support and commitments, all single points of failure should be well documented, and the roles and responsibilities of others fully addressed. For major initiatives, sponsorship by the Chief Operating Officer may be most appropriate and necessitate regular reviews of the plan.

5.2. IMPLEMENTATION OF THE TECHNOLOGY INFRASTRUCTURE TO CONSUME AND MANAGE DATA

Usage of technology in automated information sharing requires careful planning and consideration. These planning factors include selection of vendors, scalability of applications and infrastructure, integration with existing data sources, capabilities of analysts, evaluation of reports and metrics, and communicating lessons learned.

5.2.1. VENDOR SELECTION

In addition to the selection of data source vendors (if used), it may be necessary to select a vendor which can automate the collection of data. As with the data source selection, the data collection vendor selection should be based on the needs of the organization. They will also need to be compatible with any requirement that data providers place upon the organization (e.g., encryption of data).

With a clear understanding of the organization's needs, the data sources available, the processes associated with the data sources, and any requirements placed on the organization by data providers, it should be possible to conduct a fact-based approach to selecting vendors.

[18] See https://www.iacdautomate.org/

5.2.2. SCALABILITY, ELASTICITY, AND CAPACITY OF APPLICATIONS AND INFRASTRUCTURE

Organizations should anticipate their future needs and those of its data consumers. Organizations can grow, shrink, and change in unexpected ways. To take this into account, the organization should select applications and hosting infrastructure enabling them to scale to meet changes in future needs. The areas where the organization may need flexibility include:

- Number of users of an application
- Processing power
- Storage space
- Throughput capacity
- Ability to add or remove services or product features

5.2.3. INTEGRATION AND CORRELATION

It is essential to have access to analysts capable of interpreting and understanding the outputs from various systems including logged information, alerts of anomalous activity, and suspicious events and/or behavior indicating potential intrusions. However, a reliance on manual correlations for a significant volume of threat intelligence data is highly impractical for most enterprises.

Besides selecting threat intelligence sources permitted to be automatically ingested into your analytical system, you must also have the capability to identify and automatically ingest the various log and sensor data being created by your enterprise that are needed by analysis software/systems and analysts. Various vendors provide applications with appropriate standard or custom application program interfaces for ingesting this data into your storage database. Understanding the data models being used and what the various data elements represent is critical for accurate correlation and analysis.

Use of automated, machine-based analytical applications, machine learning and artificial intelligence capabilities to support as near as real time the flagging of suspicious or known exploitation within an enterprise is an engineering challenge.

Evaluating vendor products to meet the business and enterprise needs in pilot initiatives can confirm the threat intelligence and automation can be operationalized. Correlating supplied threat intelligence has the effect of amplifying the value of detected internal indicators by connecting internally suspicious activity or indicators with externally shared threat information.

5.2.4. MAKING RESULTS RELEVANT

Practitioners and analysts performing “cyber hunting” efforts are often challenged with processing a significant number of false positives from analytical systems being identified as suspicious activity, which require analysts to investigate forensically to resolve their relevance.

Often ramping up the number of the required analysts may not be possible. Therefore, the analytical systems must provide superior forensic tools and capabilities to efficiently support analysts. The integration of a variety of internal and external forensic tools and information coupled with

the ability of the analyst support systems to drill down on enterprise information without moving from one support system to another can materially affect timely analysis.

5.2.5. DERIVED ACTIONS

Some cyber threats happen at machine speed and efforts to interrupt activity early, in what has been referred to as the "Cyber Kill Chain®"[19], can be critical. This requires the enterprise to define "derived actions" to be taken when the analytical automation systems detect activity prominent in attack and exploitation efforts.

Will some control be instituted automatically to throttle the potential effects of the detected suspicious activity? For example, interrupt communication to specific domains or prevent certain protocols from executing that might be responsible for exfiltration of data, while further investigations are undertaken.

Are the defensive products and services employed by the enterprise themselves taking advantage of threat intelligence and automated responses within their capabilities? The dynamic and changing nature of cybersecurity issues requires that strategies for the needed services be employed that are adaptable. If cloud-based capabilities offer the performance and security, acceptable to an enterprise, then this approach should be evaluated. With any vendor dependence the due diligence investigation must be thorough and consider backup solutions if issues arise.

This capability can also provide other potential benefits by providing indicators of unauthorized activity by employees, authorized vendors, or potential fraudulent or illegal activity. Processes to involve the human resources and legal counsel departments when employee issues are a focus must be engaged early.

5.2.6. MANAGEMENT REPORTING AND PERFORMANCE METRICS

As discussed in Section 3.4 stakeholders must have agreement on what are success factors to be achieved through automation efforts. This involves the often-difficult task of creating objective and measurable metrics for these factors.

Translating and depicting these metrics within dashboards for management will be essential to demonstrating the value of the large investments that will be authorized to implement automation of threat intelligence, analytical capabilities, and acquisition of the required human resources. The efforts that were undertaken in the planning and design phase; where goals and objectives for information sharing were established, the agreement on what information was shared and the meaning of the shared information and determining where information is used are all vital foundational activities for producing meaningful reports.

Reporting capabilities must be very responsive to management inquiries that will arise from operational problems/incidents, prominent news reporting or impacts suffered by others, especially those with an organization same business sector. Having a clear understanding of what information

[19] See https://www.lockheedmartin.com/en-us/capabilities/cyber/cyber-kill-chain.html

is shared, how it is used, and a well-managed repository for this information where queries can be easily run from is vital to being able to respond to ad hoc requests. Be prepared to support a variety of ad hoc requests for information from management, more so during any incident affecting the organization to answer the often-expected question: “Was this caused by a cyber-attack?”

5.2.7. LESSONS LEARNED AND PARTNER COMMUNICATION

Over time, accumulation of valuable information about the operations of the enterprise information technology environment can inform risk management practices and areas of valuable investment strategies. This offers opportunities to identify potential operational problems and where improved efficiencies may be warranted. Specific processes should implemented to communicate relevant findings to partners across the enterprise while also incorporating threat data into the enterprise risk management process.

APPENDIX A – PRACTICAL ACTIONS

Using information in this guide, the following is offered as a practical set of actions to consider for improving or beginning efforts focused on automating intelligence sharing processes. The important action is to start and establish some common objectives for your organization. Organizations wishing to automate their threat intelligence need to answer three basic questions:

1. Where and what can we automate?
2. What key benefits of automation are to be achieved?
3. How can we implement automation?

This guide covers all three of these questions.

PROCESS FOR DETERMINING WHAT TO AUTOMATE

Organizations are assumed to have an existing operating environment and, as such, most likely need to apply a "crawl, walk, and run" approach to automation. The following process is suggested to determine where automation can be used, the benefits of applying automation, and possible ways that automation can be applied.

1. List sources of threat intelligence that you are either currently using or plan to use.
2. For each information source determine and record the following to arrive at a list of potential options for automation and process improvement across information sources:
    a. Using the information life cycle, describe how each source of threat intelligence is or will be used.
    b. Determine who the stakeholders are at each stage of the life cycle. This can be organizations, departments, and individuals.
    c. Determine the technologies used at each stage of the life cycles. At this stage this can be at a fairly high level; listing the systems involved is sufficient for now.
    d. Determine the level of automation currently used, or available for use, at each stage of the life cycle.
    e. Identify any constraints, "pinch points", "pain points" in the stages of the information life cycle that are limiting your ability to make effective use of the information source.
    f. For each of the identified constraints, identify possible solutions. These solutions do not specifically have to involve automation, as automation in another part of the information life cycle may require non-automation-based solutions elsewhere for the benefits of the automation to be fully realized.
    g. For each life cycle stage, assess options for automation. Record the possible sources of automation in the life cycles for the information source, add any information on costs, implementation, and operation available currently.
    h. Describe the future state of the information life cycle for the information source when both remediation to constraints and automation have been applied.
        i. Assess the benefits of the future state. Use the list of stakeholders generated earlier to help determine benefits to all parties (as you may need to persuade these stakeholders of the merits of ideas).

ii. Assess and estimate the costs of achieving the future state. Use the list of stakeholders to help determine the costs for all relevant stakeholders.

3. When all information sources have been assessed, look for commonalities (such as the ability to use the same software platform to automated process for multiple information sources) across potential solutions and information sources.
4. Determine the levels of automation that can be potentially used.
5. Create a short list of potential options that offer the most benefit.
6. Review these options with stakeholders to help determine which ones you will choose to investigate in greater depth.

APPENDIX B – NIST CSF ALIGNMENT

This section outlines how information sharing practices support the National Institute of Standards and Technology's "Framework for Improving Critical Infrastructure Cybersecurity" (commonly referred to as NIST CSF) version 1.1.[20], and where information sharing impacts activities aligned with the NIST CSF.

Organizations are increasingly using the NIST CSF as a way to gauge their level of cybersecurity maturity and as a way to structure their information security and risk programs. This section provides guidance on which areas of the framework are supported by information sharing or where information sharing should be considered.

Within the NIST CSF there are subcategories that specifically mention information sharing, and there are subcategories where information sharing can play a supporting role, and where considerations need to be made for organizations sharing cybersecurity information.

B.1 SUBCATEGORIES MENTIONING INFORMATION SHARING

Information sharing is core to two NIST CSF subcategories, specifically:

- ID.RA-2: Cyber threat intelligence is received from information sharing forums and sources
- RS.CO-5: Voluntary information sharing occurs with external stakeholders to achieve broader cybersecurity situational awareness

For these two subcategories the guidance provided by ISAO.org and NIST 800-150 provides foundational and supporting information to help organizations align with the NIST CSF. While membership in an ISAO is not the sole way to achieve alignment with these NIST CSF subcategories, ISAO membership or other information sharing approaches provide a way of improving NIST CSF maturity with respect to the above subcategories.

B 2. WHERE INFORMATION SHARING SHOULD BE CONSIDERED

There are a number of NIST CSF subcategories where organizations should consider how information supports or affects these subcategories. Below are the areas where information sharing supports/affects the NIST CSF. The areas are:

- Awareness
- Communications and coordination
- Detecting, responding, and recovering from events and incidents
- Knowledge management

Awareness

[20] Accessible at https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.04162018.pdf

Shared information can be a valuable source of data for organizations. This is especially true when sharing information with other entities that the organization is dependent on, or where other organizations are subject to similar risks. For example, other organizations who are part of the same supply chain, where there is a dependency on critical infrastructure, or where organizations operate in the same industry.

Where information sharing is taking place, formalizing the inclusion of this information into the risk management practices of the organization is suggested. This can include updating policies and procedures, roles and responsibilities, and technologies to support the availability and effective use of shared information.

The following are NIST CSF subcategories where information sharing can play a supporting role or should be considered.

- ID.BE-1: The organization's role in the supply chain is identified and communicated
- ID.BE-2: The organization's place in critical infrastructure and its industry sector is identified and communicated
- ID.BE-4: Dependencies and critical functions for delivery of critical services are established
- ID.BE-5: Resilience requirements to support delivery of critical services are established for all operating states (e.g. under duress/attack, during recovery, normal operations)
- ID.RA-4: Potential business impacts and likelihoods are identified
- ID.RM-3: The organization's determination of risk tolerance is informed by its role in critical infrastructure and sector specific risk analysis
- ID.SC-1: Cyber supply chain risk management processes are identified, established, assessed, managed, and agreed to by organizational stakeholders
- ID.SC-2: Suppliers and third-party partners of information systems, components, and services are identified, prioritized, and assessed using a cyber supply chain risk assessment process

Communication and Coordination

Formalized information sharing agreements or membership in an ISAO can support communication and coordination between organizations.

The following are NIST CSF subcategories where information sharing can play a supporting role or should be considered.

- ID.GV-2: Cybersecurity roles and responsibilities are coordinated and aligned with internal roles and external partners
- ID.SC-3: Contracts with suppliers and third-party partners are used to implement appropriate measures designed to meet the objectives of an organization's cybersecurity program and Cyber Supply Chain Risk Management Plan.
- ID.SC-5: Response and recovery planning and testing are conducted with suppliers and third-party providers

Detecting, Responding To, and Recovering from Events and Incidents

Where an organization is part of an ISAO or other information sharing agreement, the role of information sharing in detecting, responding to and recovering from an event should be formalized.

Information sharing can play a supporting role in detecting, responding to, and recovering from events. This is especially true where more than one organization is impacted by an event. As such, the role of shared information should be determined, and formalized through the use of policies and procedures, roles and responsibilities, and technology.

It is also important to formalize when information about an event that the organization has experienced is shared with external parties. As events can often be sensitive in nature, having policies, processes, controls, guidelines, technology, etc., in place to control what can be shared, with who, and under what circumstances is important.

The following are NIST CSF subcategories where information sharing can play a supporting role or should be considered.

- DE.AE-4: Impact of events is determined
- DE.DP-4: Event detection information is communicated
- RS.RP-1: Response plan is executed during or after an incident
- RS.CO-3: Information is shared consistent with response plans
- RS.CO-4: Coordination with stakeholders occurs consistent with response plans
- RS.AN-2: The impact of the incident is understood
- RC.RP-1: Recovery plan is executed during or after a cybersecurity incident
- RC.CO-3: Recovery activities are communicated to internal and external stakeholders as well as executive and management teams

Knowledge Management

Information sharing can provide a valuable source of knowledge. This can be used to improve processes.

Involving external parties in lessons learned can be facilitated through ISAOs or other information sharing agreements. This can be effective for widespread or new types of events. The formalized use of information sharing organization and practices can be built into improvement processes.

The following are NIST CSF subcategories where information sharing can play a supporting role or should be considered.

- PR.IP-8: Effectiveness of protection technologies is shared
- DE.DP-5: Detection processes are continuously improved
- RS.AN-5: Processes are established to receive, analyze and respond to vulnerabilities disclosed to the organization from internal and external sources (e.g., internal testing, security bulletins, or security researchers)
- RS.IM-1: Response plans incorporate lessons learned
- RC.IM-1: Recovery plans incorporate lessons learned

APPENDIX C – GLOSSARY

Selected terms used in the publication are defined below.

Alert: Timely information about current security issues, vulnerabilities, and exploits.

Analysis: A detailed examination of data to identify malicious activity and an assessment of the identified malicious activity to existing threat information to say something greater about the data at hand.

Automated Cybersecurity Information Sharing: The exchange of data-related risks and practices relevant to increasing the security of an information system utilizing primarily machine programmed methods for receipt, analysis, dissemination, and integration.

Campaigns: In the context of cybersecurity, a campaign or attack via cyberspace that targets an enterprise's use of cyberspace for disrupting, disabling, destroying, or maliciously controlling a computing environment/infrastructure, destroying the integrity of the data, or stealing controlled information.

Computer Security Incident: See "Incident."

Cyber Threat Information: Information (such as indications, tactics, techniques, procedures, behaviors, motives, adversaries, targets, vulnerabilities, courses of action, or warnings) regarding an adversary, its intentions, or actions against information technology or operational technology systems.

Cybersecurity Information: Data-related risks and practices relevant to increasing the security of an information system. "Examples include hardware and software vulnerabilities, courses of action, and warnings".

Cybersecurity Information Sharing: The exchange of data-related risks and practices.

Cybersecurity Threat: An action on or through an information system that may result in an unauthorized effort to adversely impact the security, availability, confidentiality, or integrity of an information system or information that is stored on, processed by, or transiting an information system. The term does not include any action that solely involves a violation of a consumer term of service or a consumer licensing agreement.

Cyber Threat Indicator: Information that is necessary to describe or identify—

- malicious reconnaissance, including anomalous patterns of communications that appear to be transmitted for gathering technical information related to a cybersecurity threat or security vulnerability;
- a method of defeating a security control or exploitation of a security vulnerability;
- a security vulnerability, including anomalous activity that appears to indicate the existence of a security vulnerability;

- a method of causing a user with legitimate access to an information system or information that is stored on, processed by, or transiting an information system to unwittingly enable the defeat of a security control or exploitation of a security vulnerability;
- malicious cyber command and control;
- the actual or potential harm caused by an incident, including a description of the information exfiltrated because of a cybersecurity threat; or
- any combination thereof.

C-Suite: C-Suite gets its name from the titles of top senior executives which tend to start with the letter C, for chief, as in chief executive officer (CEO), chief financial officer (CFO), chief operating officer (COO), and chief information officer (CIO). Also called "C-level executives."

Data Consumers: Those within an organization who use data. I.e., data in an input for their role. Systems, individuals, teams, and processes can all be described as data consumers.

Data Provider: Organizations, teams, departments, systems, or individuals who provide data.

Defensive Measure: An action, device, procedure, signature, technique, or other measure applied to an information system or information that is stored on, processed by, or transiting an information system that detects, prevents, or mitigates a known or suspected cybersecurity threat or security vulnerability.

Event: Any observable occurrence in a network or system.

False Positive: An instance in which a security tool incorrectly classifies benign content as malicious.

Incident: A violation or imminent threat of violation of computer security policies, acceptable use policies, or standard security practices.

Incident Handling: The mitigation of violations of security policies and recommended practices.

Incident Response: See “Incident Handling.”

Indicator: An artifact or observable evidence that suggests that an adversary is preparing to attack, that an attack is currently underway, or that a compromise may have already occurred.

Information Life Cycle: A model that describes the six basic activities related to the collection, use, and disposal of information. See section 2.1.1 of this document.

Information Sharing and Analysis Organization - An ISAO is any group of individuals or organizations established for purposes of collecting, analyzing, and disseminating cyber or relevant

information in order to prevent, detect, mitigate, and recover from risks, events or incidents against the confidentiality, integrity, availability and reliability of information and systems.[21]

Malware (malicious software): A program that is covertly inserted into another program or system with the intent to destroy data, run destructive or intrusive programs, or otherwise compromise the confidentiality, integrity, or availability of the victim's data, applications, or operating system.

Malicious Cyber Command and Control: A method for unauthorized remote identification of, access to, or use of an information system or information that is stored on, processed by, or transiting an information system.

Malicious Reconnaissance: A method for actively probing or passively monitoring an information system for discerning its security vulnerabilities, if such method is associated with a known or suspected cybersecurity threat.

Monitor: To acquire, identify, scan, or possess information that is stored on, processed by, or transiting an information system.

Mitigation: The act of reducing the severity, seriousness, or painfulness of a security vulnerability or exposure.

Operational Analysis: Examination of any combination of threats, vulnerabilities, incidents, or practices that results in methods to protect specific data, infrastructure, or functions (for example, incident analysis, identification of specific tactics, techniques, procedures, or threat actors)

Peer-to-Peer Sharing: Where organizations share directly with each other, rather than going through an intermediary.

Point-to-Point Sharing: Sharing information directly between to two firms. This is a form of peer-to-peer sharing.

Real-time information sharing: See "Automated Cybersecurity Information Sharing."

Security Control: The management, operational, and technical controls used to protect against an unauthorized effort to adversely affect the confidentiality, integrity, and availability of an information system or its information.

Security Vulnerability: Any attribute of hardware, software, process, or procedure that could enable or facilitate the defeat of a security control.

Signature: A recognizable, distinguishing pattern associated with an attack, such as a binary string in a virus or a particular set of keystrokes used to gain unauthorized access to a system.

[21] Frequently Asked Questions: What is an ISAO? https://www.isao.org/faq/

Situational Awareness: Comprehension of information about the current and developing security posture and risks, based on information gathered, observation, analysis, and knowledge or experience.

Tactical Intelligence: Intelligence that provides information to assist those actively involved in operational activities. (The context in this document is assisting those defending enterprises from cyber threats.)

Threat: Any circumstance or event with the potential to adversely impact organizational operations (including mission, functions, image, or reputation), organizational assets, individuals, other organizations, or the nation through an information system via unauthorized access, destruction, disclosure, or modification of information, and/or denial of service.

Threat Actor: An individual or group involved in malicious cyber activity.

Threat Source: The intent and method targeted at the intentional exploitation of a vulnerability or a situation and method that may accidentally exploit a vulnerability.

Vulnerability: A weakness in an information system, system security procedures, internal controls, or implementation that could be exploited by a threat source.

APPENDIX D – ACRONYMS

AIS	Automated Indicator Sharing
CIAS	Center for Infrastructure Assurance and Security
CSF	Cybersecurity Framework
CTI	Cyber Threat Intelligence
CVRF	Common Vulnerability Reporting Framework
DBMS	Database Management System
DDS	Data Distribution Service
DHS	Department of Homeland Security
EDI	Electronic Data Interchange
EDR	Endpoint Detection and Response solutions
FIRST	Forum of Incident Response and Security Teams
HTML	HyperText Markup Language
IACD	Integrated Adaptive Cyber Defense
IDEF-0	Integration Definition Schema and Function Modeling
IDS / IPS	Intrusion Detection/Prevention Systems
IEP	Information Exchange Policy
IOC	Indicator of Compromise
ISAO	Information Sharing and Analysis Organization
ISAO SO	Information Sharing and Analysis Organization Standards Organization
IT	Information Technology
JMS	Java Message Service
NGFW	Next Generation Firewalls
NIST	National Institute of Standards and Technology
NSA	National Security Agency
OODA Loop	Observing, Orienting, Deciding, and Acting Loop
REST	Representational State Transfer
SIEM	Security Incident and Event Management systems,
SLA	Service Level Agreement
SOAP	Simple Object Access Protocol
SOAR	Security Orchestration, Automation, and Response
SQL	Structured Query Language
STIX	Structured Threat Information Expression
TAXII	Trusted Automated eXchange of Indicator Information
TCP/IP	Transmission Control Protocol / Internet Protocol
TIP	Threat Intelligence Platform
TLP	Traffic Light Protocol
UTSA	University of Texas at San Antonio
VoIP	Voice over Internet Protocol
XML	Extensible Markup Language