ISAO 300-2

Automating Cyber Threat Intelligence Sharing

Draft Document—Request for Comment

v0.1

ISAO Standards Organization

July 18, 2018

Acknowledgments

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

Dr. Gregory B. White
ISAO SO—Executive Director
Director, Center for Infrastructure Assurance and Security, UTSA

Allen Shreffler
ISAO SO—Deputy Director
LMI

Tommy McDowell
Senior Director
Retail Cyber Intelligence Sharing Center

Working Group Three—ISAO Information Sharing

Kent Landfield
Director, Standards and Technology Policy
McAfee LLC

Roger Callahan
Consultant
FS-ISAC

The ISAO SO leadership would also like to acknowledge those individuals who contributed significantly to the development of these guidelines:

This material is based upon work supported by the U.S. Department of Homeland Security under Grant Award No. 2015-PD-128-000001.

Disclaimer: "The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security."

Table of Contents

Figures

Revision Updates

Item	Version	Description	Date
1	V1.0	Initial version	7/16/2018

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 life-cycle 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.

2 INTRODUCTION

The "ISAO 300-1 Introduction to Information Sharing"[2] document published by the ISAO standards organization 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. Shown below is Figure 1 from the referenced document. It depicts the overall context for information sharing discussed in the ISAO 300-1 document.

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

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

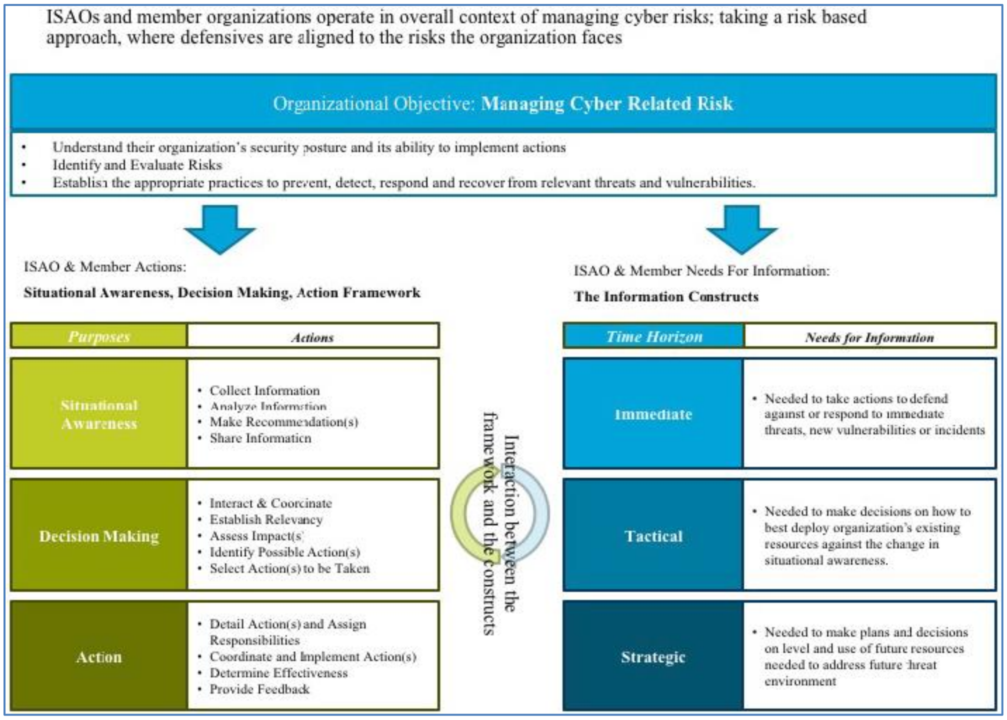


Figure 1. Context for Information Sharing

Further, 300-1 noted, "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 is focused on tactical considerations that organizations should be addressing as recipients of threat intelligence information. This document does not directly provide guidance on the important aspect of how they can also be potential sources (publishers) of threat intelligence that can be shared with others through the application of more automation.

2.1 FRAMING CONCEPTS

To support 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 the Information Life Cycle Model Section 2.1.1.

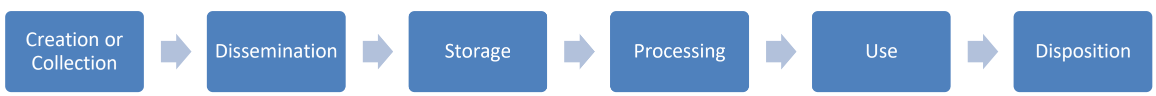
2. *The notion of structured and unstructured data and how that impacts the ability to automate processes associated with it*: This is described in the Structured and Unstructured Data Section 2.1.2.
3. *What do we mean by automation?*: This is described in the Levels of Automation Section 2.1.3.

2.2 AN INFORMATION LIFE-CYCLE MODEL

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

One common example of useful threat intelligence is “Indicators of Compromise (IOCs),” which generally are a piece of information that if observed on a network or operating system will indicate with high confidence a computer intrusion. To use such information, you first must collect it and provide it to systems that can process these as IOCs as part of an intrusion detection system.

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.[3]



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. **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.

[3] The information life cycle is taken from “OMB Circular A-130, Transmittal Memorandum #4” and is further described in the second draft of NIST SP 800-150 (though not in the final version).

2.3 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 are associated with a predefined data model, whereas unstructured data may consist of a narrative.

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

Technologies do exist for supporting the transformation of unstructured data into more structured and machine-readable information—for example, the technology that unpins 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 data sets, artificial intelligence and other specific technologies can provide levels of analysis that would not otherwise be available through other means.

2.4 DIFFERENT TYPES OF AUTOMATION

The third framing concept is what do we mean by automation in the context of threat intelligence sharing.

To help organizations think about automation and assess where automation can be used, we define five levels of automation for information sharing.

Level 1: No automation	• Communication, processing, decision making, and actions all require human involvement. • Tools such as email, telephone, 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.
Level 2: Manual process supported	• Communication, processing, decision making, or action is supported by technology that automates some element, but other elements still require human action to complete the process.

through automation	• Example: Threat intelligence is automatically published by one organization to an email distribution list. The email is read by the threat intelligence officer, who decides whether the intelligence is applicable to the organization and manually updates the threat detection tools with information provided in the email.
Level 3: Semi-automated process	• Communication, processing, decision making, and action are automated, but it requires human review and approval at some stage in the process before action is taken. • Example: The technology suggests changes to firewall rules and is also capable of making the changes automatically, but it requires human approval before changes are made.
Level 4: Automated process with human involvement	• Communication, processing, decision making, and action are automated, but there remains active human oversight. • 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 change.
Level 5: Full automation	• Communication, processing, decision making, and action are automated and human oversight is minimal or non-existent. • 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.

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 process.

3.1 ESSENTIAL CONSIDERATIONS FOR AUTOMATING CYBER THREAT INTELLIGENCE SHARING

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

- Comprehension of the ecosystem where information sharing takes place

- Determination of who the stakeholders are
- Agreement of goals and purpose for information sharing
- Determination of what information is meaningful to share
- Agreement on meaning of information
- Agreement on standards
- Agreement on protocols for exchange
- Determination of how information will be shared and used.

3.2 CYBER THREAT INTELLIGENCE ECOSYSTEM

The cyber threat intelligence ecosystem is formed by companies, governmental entities (such as the Automated Indicator Sharing 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 electronic 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

Information sharing involves multiple stakeholders both within your organization and external to it. Stakeholders can be at the governmental, regulatory, organizational, departmental, and individual level. Some or all stakeholders may need to be engaged when automating information-sharing processes.

3.3.1 PARTIES OR ROLES OF PARTIES THAT HAVE THE NEED AND AUTHORITY TO EXCHANGE SPECIFIC INFORMATION

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).[4] Safety, security, and privacy must be designed into the foundation of information-sharing environments and specifications. 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 framework,[5] as an example, identifies areas that should be addressed in such policies.

In support of safe and secure information sharing, the identity of the parties that information may be shared with is required in support of the authorization of those parties to participate in specific exchanges and/or to access kinds of information (based on its semantics). The Health Insurance Portability and Accountability Act is an example of a set of requirements in the medical community that specifies what kind of information (the information semantics) may be shared with what parties under what conditions.

There are multiple identity and authorization technologies. These technologies tend to provide either identity, role based, and/or attribute-based access control. Typical technologies include Security Assertion Markup Language, Web Services Security, and Web Authorization (OAuth).

Identity and authorization technologies are frequently combined with encryption technologies to keep communications safe and private.

3.4 AGREEMENT 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, especially where any programs to automate the sharing have substantial cost implications for the parties involved.

3.4.1 AGREEMENTS FOR AUTOMATING CYBER THREAT INTELLIGENCE INFORMATION SHARING

Information sharing can be a process that’s human to human, machine to machine, or machine to human. For both humans and machines, there must be

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

[5] Ibid, https://www.first.org/iep/.

some agreement as to what exchanged data means, how it is to be communicated, and with whom. 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 the following:

- 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.

We say 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, useful and secure information sharing is impossible, regardless of how it is realized. With those agreements in place, resources can be allocated by each party to enable communications based on the 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—we will discuss the advantages and disadvantages of this below.

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.

Fortunately, many of these agreements come “prepackaged” in industry-standard, 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 understand 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.

The concept of 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 also interact suggests the need for communities, agreements, and standards that include but go beyond cyber threats.

The smallest information-sharing community is two specific parties that have agreed to share some specific information in a specific way. This “point-to-point” kind of sharing is typical of many legacy systems and processes. The issue with point-to-point sharing is that it is very costly and anti-agile. Every point-to-point interaction must be agreed, 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 are 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.

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 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 is 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, 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 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 the database management system (DBMS) also have schema, frequently representing the same semantics as what is share; 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 PROTOCOLS FOR EXCHANGE

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

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), HyperText Markup Language) (HTML), and REpresentational State Transfer (REST-Query).
2. **Broadcast:** The broadcast pattern is provider driven. The provider "broadcasts" information determined to be relevant to some group or community able to and authorized to receive the broadcast. The syntax and semantics of the broadcast must be mutually understood. Think of this like email to a group or a radio station. Typical technologies include message queuing protocols such as 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 predetermined 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 are 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 prepackaged and based on industry standards.

4 PART 2: DESIGN

This section contains information that organizations can use to help design automated processes for capitalizing on information sharing.

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

As with any initiative, the following processes should reference the mission and goals or the organization.

For example, if the mission of the organization includes providing support and services during a crisis, then the data feeds and surrounding processes need to be sufficiently resilient so they continue to operate during crisis situations.

4.1.1 DATA CONSUMER REQUIREMENTS

Start by understanding who and what within the organization requires data—these are referred to as data consumers. Data consumers may be the end user of the data, or they may require the data to process it and then send it to another data consumer.

It is important to record all of the data consumers, though it may be possible to consolidate the list to avoid duplicates (e.g., a process and the team that performs that process could be consolidated into a single data consumer).

The names and descriptions of the data consumers should be recorded in the data requirements document.

The following are examples of who and what requires data:

- Teams

- Specific individuals
- Customers
- Members
- Reports
- Products
- Systems
- Processes
- Auditors.

The following discusses information on the requirements of each data consumer.

4.1.2 CONSUMER NEEDS

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

- 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
- Trust worthiness
- Potential value of the data
- Applicability of the data
- Cost to acquire data
- Ease of filtering or searching
- Whether relationships to other data elements are already established (e.g., is the data in a graph database?)
- Need for associated metadata (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 the agreed requirements.

4.1.4 IDENTIFY DATA SOURCES

4.1.4.1 DETERMINE POTENTIAL SOURCES OF DATA THAT CAN MEET THE DATA REQUIRES

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

Data sources can come from many areas:

- Public and commercially available intelligence feeds
- U.S. government agencies
- Governmental sources in foreign countries
- Members
- ISACs and ISAOs
- Formal and informal affinity groups of subject matter experts and re-searchers.

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

Using the data requirements document, data providers 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, and quality of the data it pro-vides.

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

- *Reputation.* The data provider is used by many other organizations, which can attest to its trustability.
- *Controls and processes that the data provider has in place*. Do the data providers have processes to ensure the ongoing quality of their data?
- *Contractual agreements and SLAs with the organization*. Is it possible to enter into a contractual agreement that defines the level of service that will be provided?
- *Communication to set clear expectations*. Can trust be established by each party communicating its needs, its ability to provide services, and 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

Some data providers may place conditions on the organization if they are to send or share their data, such as ensuring that the organization or data consumer has a sufficient clearance level, as with classified information.

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.

If a data provider is selected that has certain requirements on the organization, the organization will need to create and ensure the data provider that controls are in place and that they are operating effectively. Often a form of TLP is defined to express the requirements or expectations for data sharing.

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 provide data that meets the data consumer's requirements. Where this is the case, the organization will need to determine what processes should be put in place to transform the data into something that meets the needs of the data consumers.

The following are examples of transformation processes:

- 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 the following:

- 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 the 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 the potential data provider would have on the organization ingesting the data

- An assessment of where there are any gaps between the requirements and what can be provided, 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 stakeholder who may favor certain vendors or stakeholders' requirements over others.

4.1.4.7 DEFINE DATA INGESTION ARCHITECTURE

A data ingestion architecture consists of the following:

- The data model
- The data policies, procedures, and controls
- Processes, technologies, and other factors that support data quality and the needs of the data consumers and the organization.

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 the data requirements document). The following are example elements of data quality:

- 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 GOVERNANCE

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

4.1.5 DEFINING 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.

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. Following are the goals of mapping the data flow:

- Understanding what processes used the data
- Understanding what technologies and 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 or 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 technics 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; communities don’t have to reinvent those wheels. This agreement is then made available in technology stacks—web servers, enterprise service buses, 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 prebuilt 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.

4.2.1 FEDERATING AND TRANSLATING SHARED INFORMATION

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), the networking and information services it uses, along with those of critical partners interconnect 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 service, 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 receipt processes (ingesting), correlating applicability, and incorporating it to mitigate risks is 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 is 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 *__o__bserving*, *__o__rienting*, *__d__eciding*, and *__a__cting*. 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 that 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, let’s group enterprises into just three categories to consider the operational use of cyber threat intelligence. The three categories are shown in Figure 2.

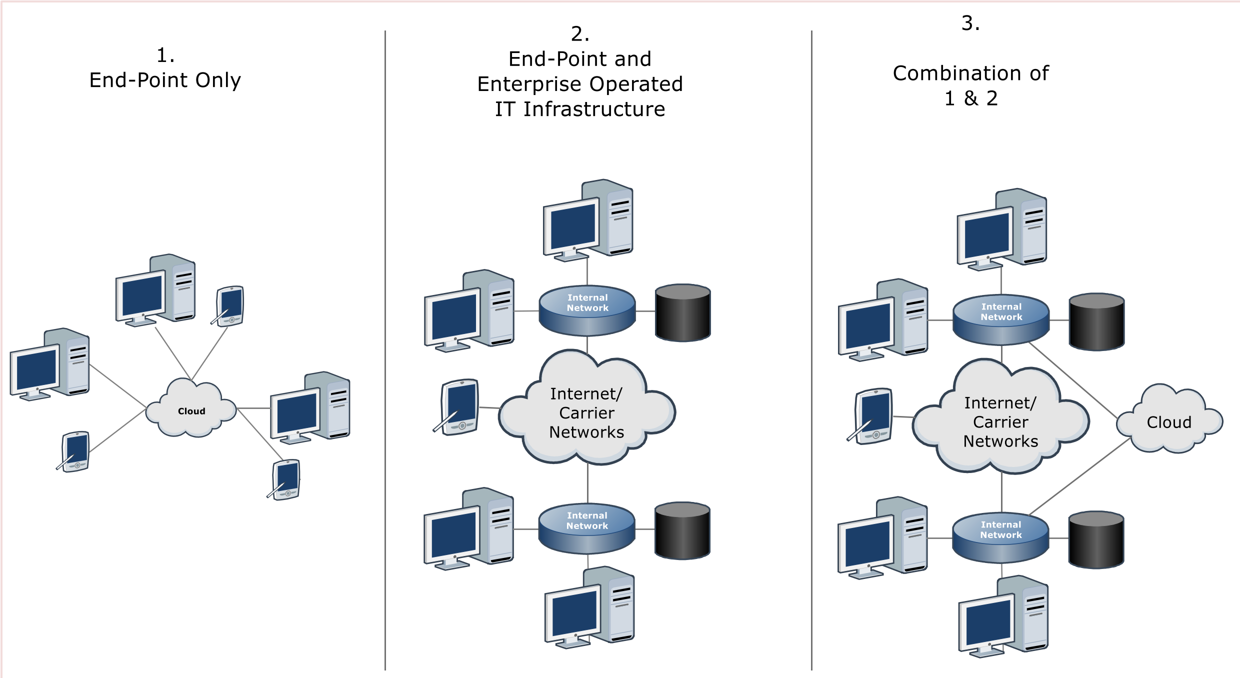


Figure 2. Three Categories of Enterprises

In category 1, 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 2, 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 geographical dispersed operations structure.

Today, and more so in the future, most organizations will have operations that fall into category 3, that is, a range of cloud-based applications and 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 category 3.

Another useful set of information is the identification of the key business or purpose objectives of the organization and a current risk management assessment of cybersecurity practices; hopefully, using the National Institute of Standards and Technology (NIST) Cybersecurity Framework 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. It is likely that today's process has many manual steps. Throughout this spectrum of operational considerations, the timeliness of the threat intelligence can materially affect its effective use. The application of automation to the receipt processes (ingesting), correlating applicability, and incorporating it to mitigate risks is 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 recent 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.[6]

There are many ways an organization can automate the use of machine-readable threat intelligence within its 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 a Security Incident and Event Management (SIEM) system, Next Generation Firewall (NGFW), Intrusion Detection/Prevention System (IDS/IPS), Threat Intelligence Platform (TIP), and Endpoint Detection and Response (EDR) solution.[7] 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, such as SIEM, IDS/IPS, NGFW, and EDR. 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

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

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

network devices using APIs, depending on the device and source of the feeds.[8] Organizations at a higher risk of cyber attacks—for example, the financial or manufacturing industries—and with larger budgets and more personnel may be more likely to implement 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."[9]

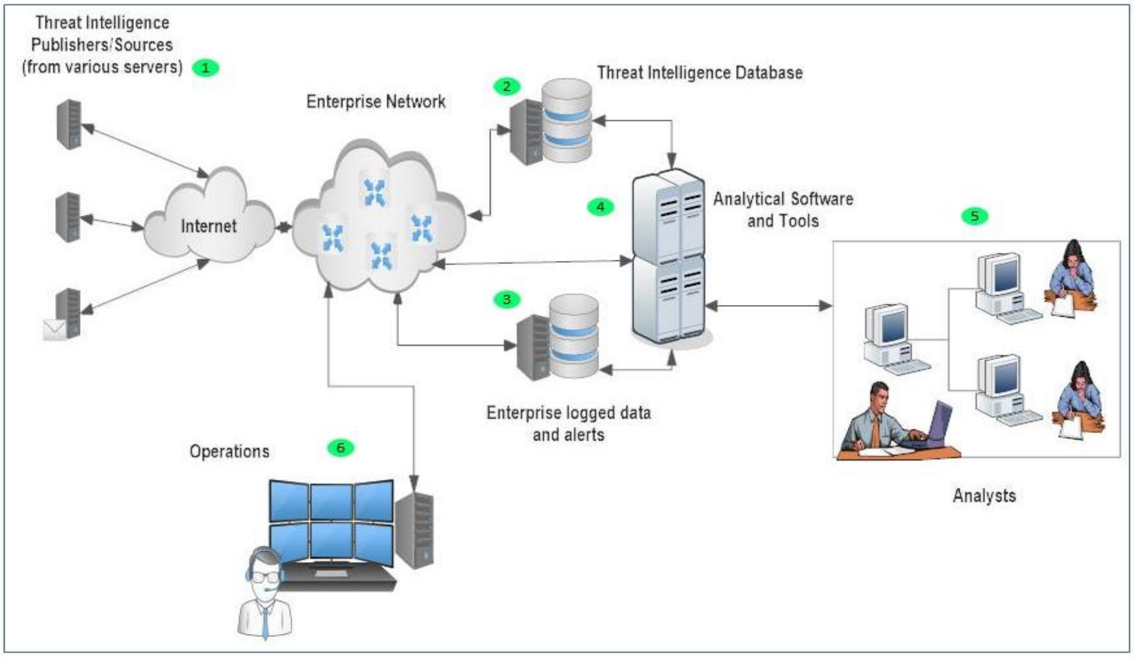
In these cases, organizations may use a product designed to manage threat intelligence, including a TIP that would allow personnel to analyze external threat information, correlate this activity with internal network activity, and respond to threats through automating incident response "playbooks." Benefits 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 direct integration with network defenses.[10]

The high-level model depicted in Figure 3 shows key elements of any cyber threat automation effort with an enterprise.

[8] Ibid.

[9] Gartner, "Innovation Insight for Security Orchestration, Automation and Response," November 30, 2017.

[10] Ibid.



RCallahan_20170824

Figure 3. Draft Diagram for Describing Aspects of Automating Cyber Threat Intelligence Sharing

Item 1—represents sources of electronic cyber threat intelligence, for example, open-source or commercially available threat feeds, indicators received through Trusted Automated eXchange of Indicator Information (commomly referred to as TAXII), or servers or other automated means of information sharing.

Item 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 the capabilities in Item 4.

Item 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.

Item 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, for example, SIEM, NGFW, IDS/IPS, or EDR. The capabilities in Item 2 and Item 4 are often found to varying degrees in TIPs.

Item 5—those operational capabilities taking advantage of the analytical processes or capabilities.

4.5 DATA INGESTION PROCESSES

The data ingestion process defines the processes that are needed to effectively ingest the data and ensure that 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.

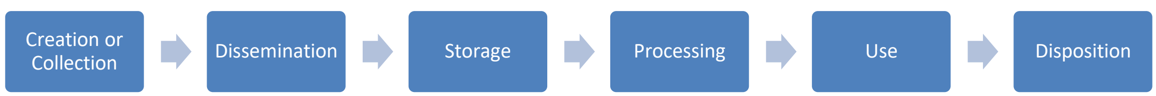
The following are design considerations needed for the data ingestion process:

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 to transport the data from source to where it is ultimately used is required.

The information life-cycle process steps discussed earlier can be used to assist in collecting this information.



Required at each step in the process is 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 the transporting of data should be able to communicate with each other.

Considerations should be made for any transformation, cleansing, or enrichment of data needed in the process. As part of these considerations, an understanding should be reached as to the level of automation and the level of human involvement needed, or desirable, at each step.

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 unsure 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.

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

As discussed in Section 3.8, there are technical standards and protocol that the data to be ingested should adhere to. Consideration should also be given to how the information falls into the standards and classification of the consuming organization.

What can be overlooked is determining how information that is being ingested falls into an organization's data classification. The Traffic Light Protocol has been described elsewhere in this document, along with rules on how shared information can be handled. Data can also come from sources that require levels of official governmental clearance to access. It is important to also consider how the shared information falls into the organization's own data classification, as this will impact how the information is used and which systems are allowed to support the ingestion of the data.

To understand how the data to be ingested falls into the organization's data classification, it is important to know what type of information is being shared. 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.

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

What would happen if the following occurred?

- The process that ingests data stops working.
- The provider of the data increases the volume of data 100-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.

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.

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 if errors occur in the data ingestion process, events will be 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 following:

- 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:[11]

- 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—the data is valid if it conforms to the syntax (format, type, range) of its definition.
- Accuracy—the degree to which the 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.

[11] Adapted from "The Six Primary Dimensions for Data Quality Assessment," DAMA UK. See https://www.whitepapers.em360tech.com/wp-content/files_mf/1407250286DAMAUKDQDimensionsWhitePaperR37.pdf.

4.6.2 DATA ENRICHMENT

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

- 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

To make data readable or processible by one system may require that format of the data to be converted. It may also be necessary to transform unstructured or semi-structure 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. Also, if the organization's systems require a non-standard format of data, custom scripts or other methods will need to be deployed to convert data in a usable 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

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 that all data that has been digested has been categorized in line with 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.

Using the data mapping, it should be determined which suppliers of data are critical for the operations of the organization. These suppliers need to be managed, with SL and so on, in accordance with how critical they are to the organization's operations.

4.9 DEFINING ROLES AND RESPONSIBILITIES

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

The roles and responsibilities should align with effective data management practices and the policies and procedures of the organization.

4.10 DEFINING DATA ASSESSMENTS AND FEEDBACK PROCESSES

Feedback processes can be useful for both the consumers and producers of data. By providing the producers of data with information about how and if that data was used, and the usefulness of the data, the producers of data either better tailor the data they provide or improve the quality of data they produce in general. The feedback from other consumers of the same data can also support the organization in determining whether a piece of data is worth consuming.

Data consumers can provide feedback in qualitative and quantities 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 him or her 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, such as the following:

- 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 IMPLEMENTATION

5.1 AN IMPLEMENTATION GAME PLAN

From the earlier discussion in this document, improving an organization's effectiveness in the use of cyber threat intelligence will require a broad commitment of many parts of the organization. As such, the decision to support the effort must have broad executive level support and buy-in from the involved organizations. Developing the plan should be led by the operational and security organizations addressing the current state and how increased automation will be phased into the organization. Some practical considerations are described in Appendix A, and efforts from the Department of Homeland Security, the National Security Agency, and John Hopkins University Applied Physics Laboratory on an Integrated Adaptive Cyber Defense framework initiative[12] offer additional considerations.

Implementing the plan will be a multi-year effort and should be considered a major initiative in any organization. Because the effort will engage and require cross-organizational support and commitments, a single point of responsibility should be well documented, and the roles and responsibilities of others fully addressed. Processes used in an organization for major initiative regular reviews will be required, and sponsorship by the chief operating officer may be most appropriate.

5.2 IMPLEMENTATION OF THE TECHNOLOGY INFRASTRUCTURE TO CONSUME AND MANAGE DATA

5.2.1 VENDOR SELECTION

In addition to the selection of data source vendors, it may be necessary to select a vendor that 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. It 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

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

placed on the organization by data providers, it should be possible to conduct a fact-based approach to selecting vendors.

5.2.2 SCALABILITY, ELASTICITY, AND CAPACITY OF APPLICATIONS AND INFRASTRUCTURE

The organization should also consider its future needs and those of the data consumers. Organizations can grow, shrink, and change in unexpected ways. To take this into account, the organization should select applications and hosting infrastructure that enables it to scale up or down to meet future needs. The following are areas where the organization may need flexibility:

- 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

Having analysts who can understand and interpret the output of various systems that provide logged information, alerts of anomalous activity, suspicious events, and/or behavior indicating possible or actual intrusions is essential. However, a reliance on manual correlations with various threat intelligence among this significant volume of data is highly impractical for most enterprises.

Besides selecting threat intelligence sources that permit information 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's 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.

The 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 business and enterprise needs in pilot initiatives can confirm that 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

Practioners and analysts performing these “cyber hunting” efforts are often challenged to deal with a high level of false positives from analytical systems being identified as suspicious activity, which require analysts to investigate forensically to resolve its relevance.

Often ramping up the number of the required analysts is not possible. Therefore, the analytical system 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 system 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 the “Cyber Kill Chain®”[13] can be most critical. This requires that the enterprise will need 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? Examples include interrupting communication to specific domains or preventing 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 performance and security that’s acceptable to an enterprise, that is an approach that should receive evaluation. 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 must be started early to involve the human resources and legal counsel organizations when employee issues are a focus.

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

5.2.6 MANAGEMENT REPORTING AND PERFORMANCE METRICS

As discussed in Section 3.4, stakeholders must agree on what success factors to achieve 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.

Reporting capabilities must be very responsive to management inquiries that will arise from operational problems or incidents, prominent news reporting, or impacts suffered by others, especially those with an organization in the same business sector. 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 LEARNINGS AND COMMUNICATING TO PARTNERS

Over time, there will be accumulated valuable information about the operations of the enterprise IT environment that can inform risk management practices and areas of valuable investment strategies. This offers an opportunity for learnings that can also contribute to the identification of potential operational problems and where improved efficiencies may be warranted. Specific processes should be incorporated to communicate these learnings to partners across the enterprise and to instill experienced 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 are key benefits of automation 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 a “crawl, walk, and run” approach to automation. The following process is suggested to determine where automation can be used, the benefits or 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 cycle. 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 that is currently used, or available for use, at each stage of the life cycle. The levels of automation can be used here.
   e. Identify any constraints, “pinch points,” or “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, and 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 remediations 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 (e.g., the ability to use the same software platform for the automated process of multiple information sources) across potential solutions and information sources.
4. Create a short list of potential options that offer the most benefit.
5. Review these options with stakeholders to help determine which ones you will choose to investigate in greater depth.

APPENDIX B—GLOSSARY

Selected terms used in this 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 using 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 or 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.

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.

Malware: 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 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 (e.g., incident analysis, identification of specific tactics, techniques, procedures, or threat actors).

Real-Time Information Sharing: See “Automated Cybersecurity Information Sharing.”

Secure Portal: A web-enabled resource providing controlled secure access to and interactions with relevant information assets (information content, applications, and business processes) to selected audiences using web-based technologies in a personalized manner.

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.

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 C—ACRONYMS

AIS	Automated Indicator Sharing
IACD	Integrated Adaptive Cyber Defense
IEP	Information Exchange Policy
IDEF-0	Integration Definition Schema and Function Modeling
IOC	Indicator of Compromise
ISAC	Information Sharing and Analysis Center
ISAO	Information Sharing and Analysis Organization
NIST	National Institute of Standards and Technology
OAuth	Web Authorization
TIP	Threat Intelligence Platform
TLP	Traffic Light Protocol
XML	Extensible Markup Language