**What is Data: And Why it Matters to JADC2**

# 1.0 Introduction

It has been said that the purpose of information systems is to get the right information to the right people at the right time in the right amount and in the right format. IT systems are integral to every functional area of an organization, to get started it is important to understand some basic terminology.

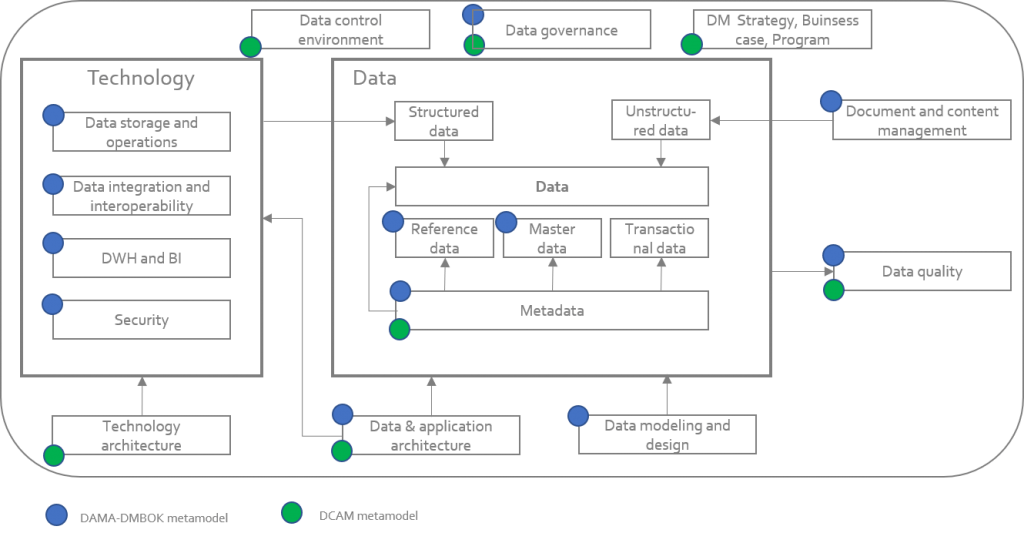
## 1.1 Data, information, and knowledge

Data is a representation of facts or concepts in symbols, usually in digital form suitable for computer processing. Images, video, word processing documents, email, chat, and plain text are all different kinds of data, as is the highly structured data in a database or data warehouse. By itself, data has no meaning – that must be assigned, one way or another, by some knowledgeable person.

For most purposes, information and data are equivalent terms. When the difference matters, it is to emphasize the relevance of the data to some decision and the understanding of the data by the people who decide. In short: information is data that reduces the uncertainty of a decider with respect to a decision.

## 1.2 Information technology infrastructure and information technology architecture.

An organization’s information technology *architecture* is a high-level map or plan of the information assets in an organization. The IT architecture integrates the information requirements of the overall organization and all individual users, the IT infrastructure, and all applications. An organization’s information technology *infrastructure* consists of the physical facilities, IT components, IT services, and IT management that support the entire organization.



# 2.0 JADC2 and Data

It all starts with the data. Data should be of high quality, meaning that they should be accurate, complete, timely, consistent, accessible, relevant, and concise. Unfortunately, however, the process of acquiring, keeping, and managing data is becoming increasingly difficult.

## 2.1 Importance of agile data in JADC2

Data matters for C2 because having the right data supports better decisions, made and put into effect faster. This is a combat force multiplier.

Near-peer adversaries have access to the best information technology, and so competitive advantage will come from making best use of the tools that are available to everyone. This “best use” means understanding what data to collect, process, and provide to C2 deciders at all levels, and then quickly putting those data sharing arrangements into practice. It means rapid development of planned data sharing that supports new and improved doctrine, organization, and training for operations within and between the C2 domains. It also means rapid fulfilment of unanticipated data needs that are discovered in the field during operations. This is agile data sharing.

## 2.1 Importance of data

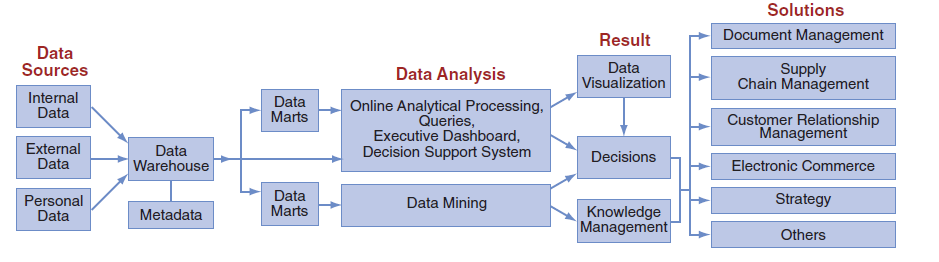
IT applications cannot be performed without using data. Data should be accurate, complete, timely, consistent, accessible, relevant, and concise.

Managing data in the Department of Defense is difficult for various reasons:

(1) the amount of data increases with time;

(2) data are stored in various systems, databases, formats, and languages; and

(3) data security, quality, and integrity are often compromised.



The data life cycle starts with data collection. (Visible and accessible)

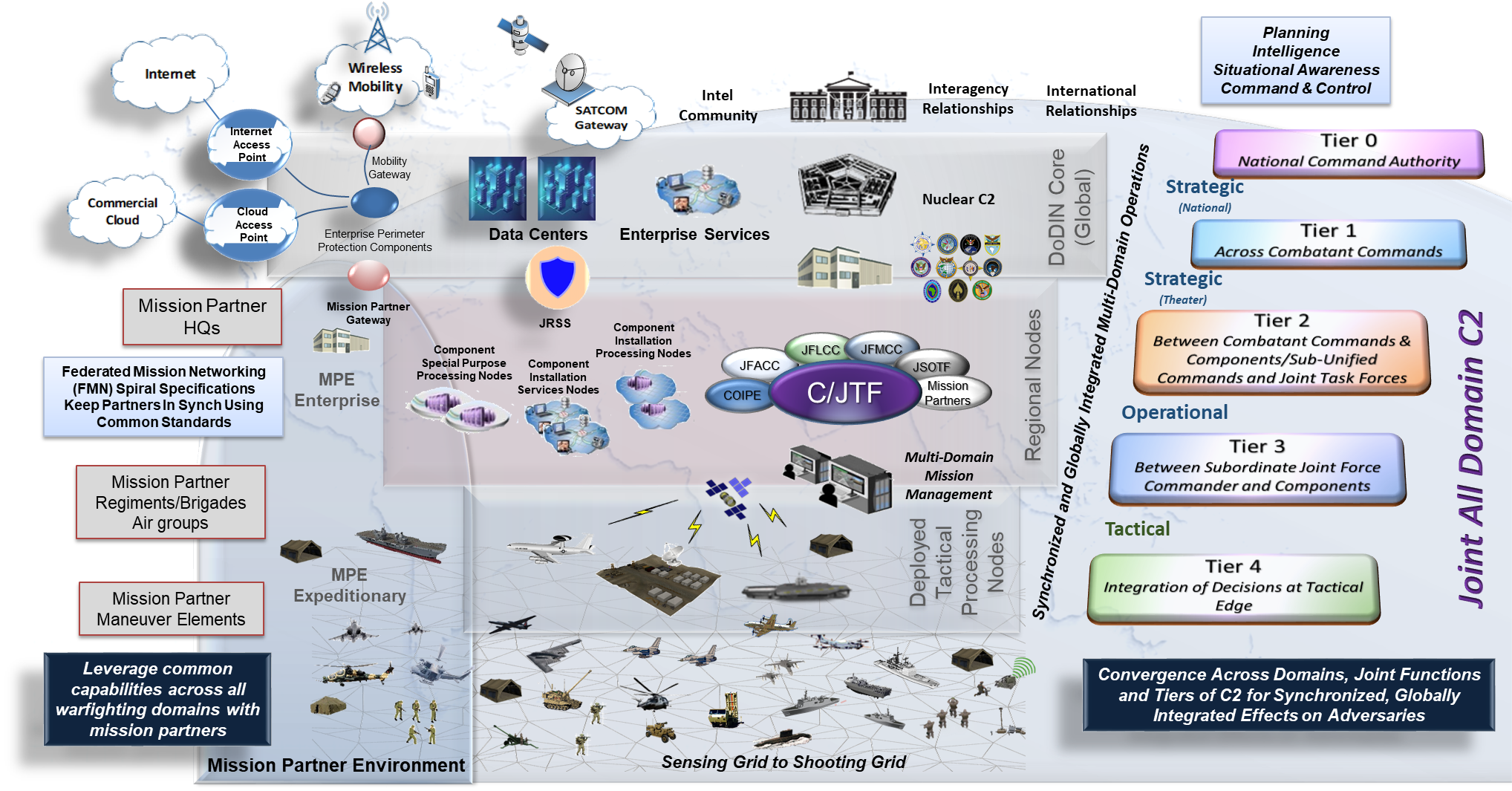
The data are stored in a database(s) and then preprocessed to fit the format of a data warehouse or data marts. (Tagged and understandable [“curated/mediated”])

Users then access data from the warehouse or data mart for analysis. (ICAM)

The result of all these activities is the generation of decision support and knowledge. (Interoperable)

## 2.3 Data Sources

The vision for the JADC2 communication network-of-networks is connecting “every sensor, every shooter” by having data flow across the five domains of warfare: sea, air, land, cyber and space. The challenge is creating data structures that can link interoperable sensors around the world, from the wings of a brand-new fighter jet to a decades-old legacy system on a submarine.



Data come from internal sources (e.g., corporate databases), personal sources (e.g., personal thoughts, opinions, and experiences), and external sources (e.g., commercial databases, government reports, and corporate Web sites). Data also come from the Web, in the form of clickstream data. New sources of data, such as blogs, podcasts, video casts, and RFID tags and other wireless sensors are constantly being developed. Much of these new data are unstructured, meaning that their content cannot be truly represented in a computer record. Examples of unstructured data are digital images, digital video, voice packets, and musical notes in an MP3 file.

To ensure interoperability and make data decision ready, JADC2 Data Enterprise Convergence Plan describes all DOD convergence in five technical areas:

* Metadata tagging
* Standardized data interfaces
* Common data access controls
* Data security
* Data infrastructure

## 2.3 Authoritative Data Sources (ADS)

In accordance with the authority in DoD Instruction (DoDI) 8320.02 Authoritative Data Sources are registered in the DoD Data Services Environment (DSE) (accessible at https://metadata.ces.mil/dse).

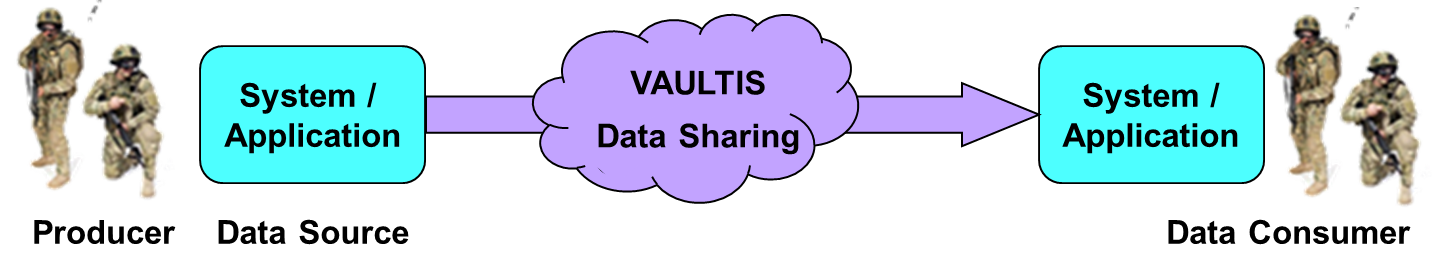
In accordance with the authority in DoD Instruction (DoDI) 8320.02 ADS expose their high-value data and content through Web application programming interfaces (APIs) and apply metadata tagging, as appropriate.

It is preferable that ADS owners expose their native data into a common format that is understood by the enterprise users. This requires either native use of the enterprise standard or translation form a native standard to the common information exchange standard (e.g. National Information Exchange Model).

## 2.4 Agile data and the DoD Data Strategy

The goals of the DoD Data Strategy (September 2020) combine to support agile data sharing for JADC2. These goals are to make data:

* Visible – Can the data consumer discover the data sources he needs?
* Accessible – Can he obtain the content of those sources?
* Understandable – Can he correctly interpret the meaning of that content?
* Linked – Can he exploit connections between data sources?
* Trustworthy – Can he believe what the content says?
* Interoperable – Will his software applications correctly process the data content?
* Secure – Are sources protected from unauthorized use and manipulation?



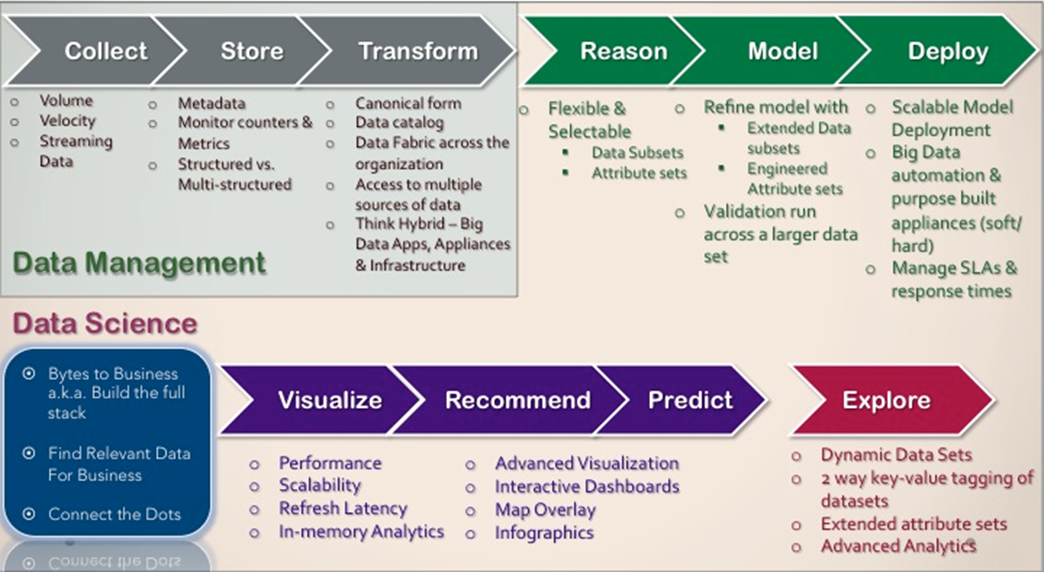
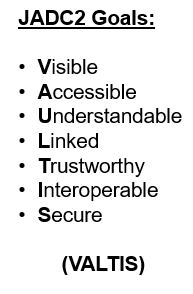
## 2.5 JADC2 Data Enterprise Convergence Plan

The purpose of the JADC2 Data Enterprise Convergence plan is to begin achieving the VAULTIS goals in the near term, while avoiding incompatible stovepipe component (command, service, agency) solutions. The initial objective is a rapid alignment on a small number of enterprise convergence points. Each of these is a combination of standards, services, and processes that (A) together handle an aspect of data sharing, and that (B) can and will be implemented by all systems and applications. Any community of practice (CoP) is free to implement a preferred alternative, so long as they also implement the convergence point approach. The effect is to concentrate the enterprise interoperability effort where the payoff is highest and fastest, while allowing divergence where it is necessary and/or less consequential. The first set of convergence points will be selected in five focus areas:

* Metadata tagging
* Standardized data interfaces
* Common data access controls
* Data security
* Data infrastructure

## 2.6 The Data Pipeline

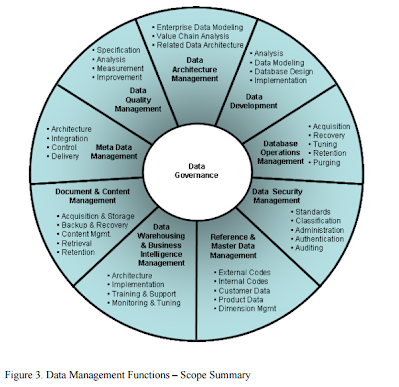
In computing, a pipeline, also known as a data pipeline, is a set of data processing elements connected in series, where the output of one element is the input of the next one. The elements of a pipeline are often executed in parallel or in time-sliced fashion.



* **Define**: Identify agency and stakeholder needs for data of sufficient quality for intended uses
* **Coordinate**: Assess the ability of data resources and infrastructure to meet agency and stakeholder needs
* **Collect**: Organize, plan, and execute data collections and acquisitions to meet agency and stakeholder needs
* **Curate**: Organize, refine, and maintain agency data resources with sufficient quality to meet agency and stakeholder needs
* **Access**: Identify and develop multiple data access methods for agency staff and stakeholders
* **Analyze**: Optimize the ability of staff and stakeholders to use agency data to generate insights
* **Visualize**: Present data insights for consumption by leaders and stakeholders
* **Disseminate**: Provide multiple avenues for release of data and insights
* **Implement & Assess**: Maximize the use of data for decision-making, accountability, and the public good and continuously improving the data process

# 3.0 Government Regulations

In the wake of 9/11 and numerous corporate scandals, the U.S. government passed many new laws, including the Sarbanes-Oxley Act, the USA PATRIOT Act, the Gramm-Leach-Bliley Act, and the Health Insurance Portability and Accountability Act. Organizations must be in compliance with the regulations of these statutes. The process of becoming and remaining compliant is expensive and time-consuming. In almost all cases, organizations rely on IT support to provide the necessary controls and information for compliance.



Additionally, organizations establish **data governance** solutions to managing information across an entire organization. It involves a formal set of business processes and policies that are designed to ensure that data are handled in a well-defined fashion. That is, the organization follows unambiguous rules for the creation, collection, handling, and protection of information. The objective is to make information available, transparent, and useful for the people authorized to access it, from the moment it enters an organization, until it is outdated and deleted.

## 3.1 Data Standards

Public Law 104-113, “The National Technology Transfer and Advancement Act of 1995 (NTTAA),” and Office of Management and Budget (OMB), Circular A-119, “Federal Participation in the Development and Use of Voluntary Consensus Standards and in Conformity Assessment Activities” (2016), direct federal agencies to use voluntary standards in lieu of Government-unique standards whenever feasible. The legislation asks agencies to promote participation by their personnel in standards development to ensure the standards created are usable by both federal agencies and the private sector.

Data standards describe objects, features, or items that are collected, automated, or affected by activities or functions of the organization.

* Benefits
  + More efficient data management (including updates and security)
  + Increased data sharing
  + Higher quality data
  + Improved data consistency
  + Increased data integrity
  + Better understanding of data
  + Improved documentation and information resources
* Levels of standards
  + International
  + National
  + Regional
  + Local

# 4.0 Ethical and Security Issues

Ethics relates to general standards of right and wrong, whereas information ethics relates specifically to standards of right and wrong in information processing practices. Ethical issues are very important because, if handled poorly, they can damage an organization’s image and destroy its employees’ morale. The use of IT raises many ethical issues, ranging from monitoring e-mail to invading the privacy of millions of customers whose data are stored in private and public databases.

Data Security is about protecting the database against unauthorized users. This is to maintain Privacy, to ensure that data is not seen by those who are not entitled to see it, and also to ensure that data is not willfully corrupted. The primary data security mechanisms available include access controls and encryption. Additionally, audit trails may be put in place to track who did what in the event of a breach of security.

* **Access Controls**, such as **ICAM** provide, security features that control how users and systems communicate and interact with other systems and resources. They protect the systems and resources from unauthorized access and can be components that participate in determining the level of authorization after an authentication procedure has successfully completed.
* **Encryption** - Any procedure used in cryptography to convert plaintext (or cleartext) into ciphertext (encrypted message) in order to prevent any but the intended recipient from reading that data.

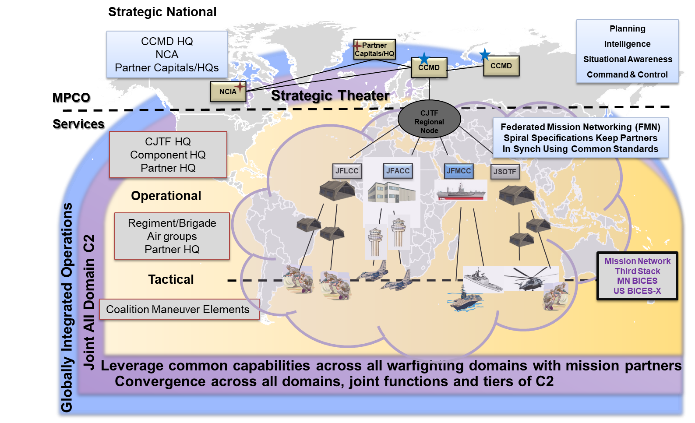
The core goals of security, which are to provide availability, integrity, and confidentiality (AIC triad) protection for critical assets. Each asset will require different levels of these types of protection, as we will see in the following sections. All security controls, mechanisms, and safeguards are implemented to provide one or more of these protection types, and all risks, threats, and vulnerabilities are measured for their potential capability to compromise one or all of the AIC principles.

# 5.0 Global IT Considerations

Regardless of its structure, it is even more complex when the IT systems require interfacing with systems from foreign countries, because of differences in cultures, economies, and politics among parties in different countries. Some countries are erecting artificial borders through local language preference, local regulation, and access limitations. Some issues to consider in designing global systems are:

1. cultural differences,
2. localization,
3. economic and political differences, and
4. legal issues.

The major benefits of global information systems for the Department of Defense are effective communication at a reasonable cost and effective collaboration that overcomes differences in distance, time, language, and culture.



NATO has adopted NIEM under the name NATO Core Data Framework. (NCDF) and is an enabler for Federated Mission Networking (FMN) information sharing (a key MPE implementation and pacing factor). It is a NATO information sharing policy requirement.

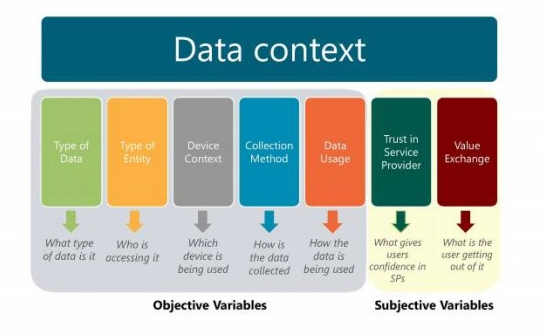
* NCDF is leveraged by emerging NATO programs including the NATO AWACS replacement program (Alliance Future Surveillance and Control (AFSC))
* NCDF provides the pathfinder foundation for a mission partner CJTF-level Data Lake that includes NATO Link 16 feed, national air track feed, maritime OTH-Gold feed, and land SA feeds NATO Allied Command Transformation has requested JS J6 host a NCDF Data Lake instantiation in DDC5I lab spaces as part of a persistent test environment.

JS J6 DDC5I leads the NATO Data Management Capability Team, the FMN Data Management Syndicate, and the FVEY Combine-Communications Electronics Board Data Tagging and Labeling Work Group.

# 6.0 Decision Making

Organizational goals and decisions are achieved through the use of resources (people, money, energy, materials, space, time).

## Systems in Context

****

## 6.1 Decision Support Systems

**Decision support systems (DSSs)** combine models and data in an attempt to solve semi-structured

and some unstructured problems with extensive user involvement. **Models** are simplified

representations, or abstractions, of reality. The DSS is designed to enable interactive access to data, to enable manipulation of these data, and to provide business managers and analysts the ability to conduct appropriate analyses. Decision support systems can manipulate data, enhance learning, and contribute to all levels of decision making.

## 6.2 Geographic Information Systems

A **geographic information system (GIS)** is a computer-based system for capturing, integrating,

manipulating, and displaying data using digitized maps. Its most distinguishing characteristic

is that every record or digital object has an identified geographical location. This process,

called *geocoding*, enables users to generate information for planning, problem solving, and decision

making. In addition, the graphical format makes it easy for managers to visualize the data.

## 6.3 Virtual Reality

There is no standard definition of virtual reality. The most common definitions usually

imply that **virtual reality (VR)** is interactive, computer-generated, three-dimensional

graphics delivered to the user through a head-mounted display. In VR, a person “believes”

that what he or she is doing is real even though it is artificially created.

## 6.4 Intelligent Systems and Artificial Intelligence

**Intelligent systems** is a term that describes the various commercial applications of artificial

intelligence. **Artificial intelligence (AI)**, a subfield of computer science, is concerned

with studying the thought processes of humans and re-creating the effects of those processes

via machines, such as computers and robots.

**Glossary (To be completed and alphabetized TBD)**

**Data items** refer to an elementary description of things, events, activities, and transactions

that are recorded, classified, and stored but are not organized to convey any specific meaning.

Data items can be numbers, letters, figures, sounds, or images. Examples of data items are a

student grade in a class and the number of hours an employee worked in a certain week.

**data warehouse** is a repository of historical data that are organized by subject to support

decision makers in the organization. Data warehouses facilitate business intelligence activities,

such as data mining, decision support, and querying applications.

**Information** refers to data that have been organized so that they have meaning and value

to the recipient. For example, a grade point average (GPA) is data, but a student’s name coupled

with his or her GPA is information. The recipient interprets the meaning and draws

conclusions and implications from the information.

**Knowledge** consists of data and/or information that have been organized and processed

to convey understanding, experience, accumulated learning, and expertise as they apply to a current business problem.

**information technology (IT) architecture** is a high-level map or plan of

the information assets in an organization. It is both a guide for current operations and a

blueprint for future directions. The IT architecture integrates the entire organization’s business

needs for information, the IT infrastructure (discussed in the next section), and all applications.

**information technology (IT) infrastructure** consists of the physical facilities, IT components, IT services, and IT personnel that support the entire organization.

**IT components**are the computer hardware, software, and communications technologies that provide the foundation for all of an organization’s information systems.

**platform** consists of the hardware, software, and communications components that organizations use to

process and manage information.

**database** A group of logically related files that stores data and the associations among them.

**database management system (DBMS)** The software program (or group of programs) that provides access to a database.

**data dictionary** Collection of definitions of data elements, data characteristics that use the data elements, and the individuals, business functions, applications, and reports that use this data element.

**data governance** An approach to managing information across an entire organization.

**data mart** A small data warehouse designed for a strategic business unit (SBU) or a department.

**data model** Definition of the way data in a DBMS are conceptually structured.

**data warehouse** A repository of subject-oriented historical data that are organized to be accessible in a form readily acceptable for analytical processing.

**multidimensional structure** The manner in which data are structured in a data warehouse so that they can be analyzed by different views or perspectives, which are called dimensions.

**normalization** A method for analyzing and reducing a relational database to its most streamlined form for minimum redundancy, maximum data integrity, and best processing performance.

**artificial intelligence (AI)** A subfield of computer science concerned with studying the thought processes of humans and representing the effects of those processes via machines.

**business intelligence** Applications and technologies for consolidating, analyzing, and providing access to vast

amounts of data to help users make better business and strategic decisions.

**data mining** The process of searching for valuable business information in a large database, data warehouse, or data mart.

**artificial intelligence (AI)** A subfield of computer science concerned with studying the thought processes of humans and representing the effects of those processes via machines.

**business intelligence** Applications and technologies for consolidating, analyzing, and providing access to vast

amounts of data to help users make better business and strategic decisions.

**data mining** The process of searching for valuable business information in a large database, data warehouse, or data mart.

**decision support system (DSS)** Business intelligence systems that evolved from decision support systems; they combine models and data in an attempt to solve semi-structured and some unstructured problems with extensive user involvement.

**digital dashboard** A business intelligence system that provides rapid access to timely information and direct access to management reports.

**expert system (ES)** A computer system that attempts to mimic human experts by applying reasoning methodologies or knowledge in a specific domain.

**geographic information system** A computer-based system for capturing, integrating, manipulating, and displaying data using digitized maps.

**neural network** A system of programs and data structures that approximates the operation of the human brain.