

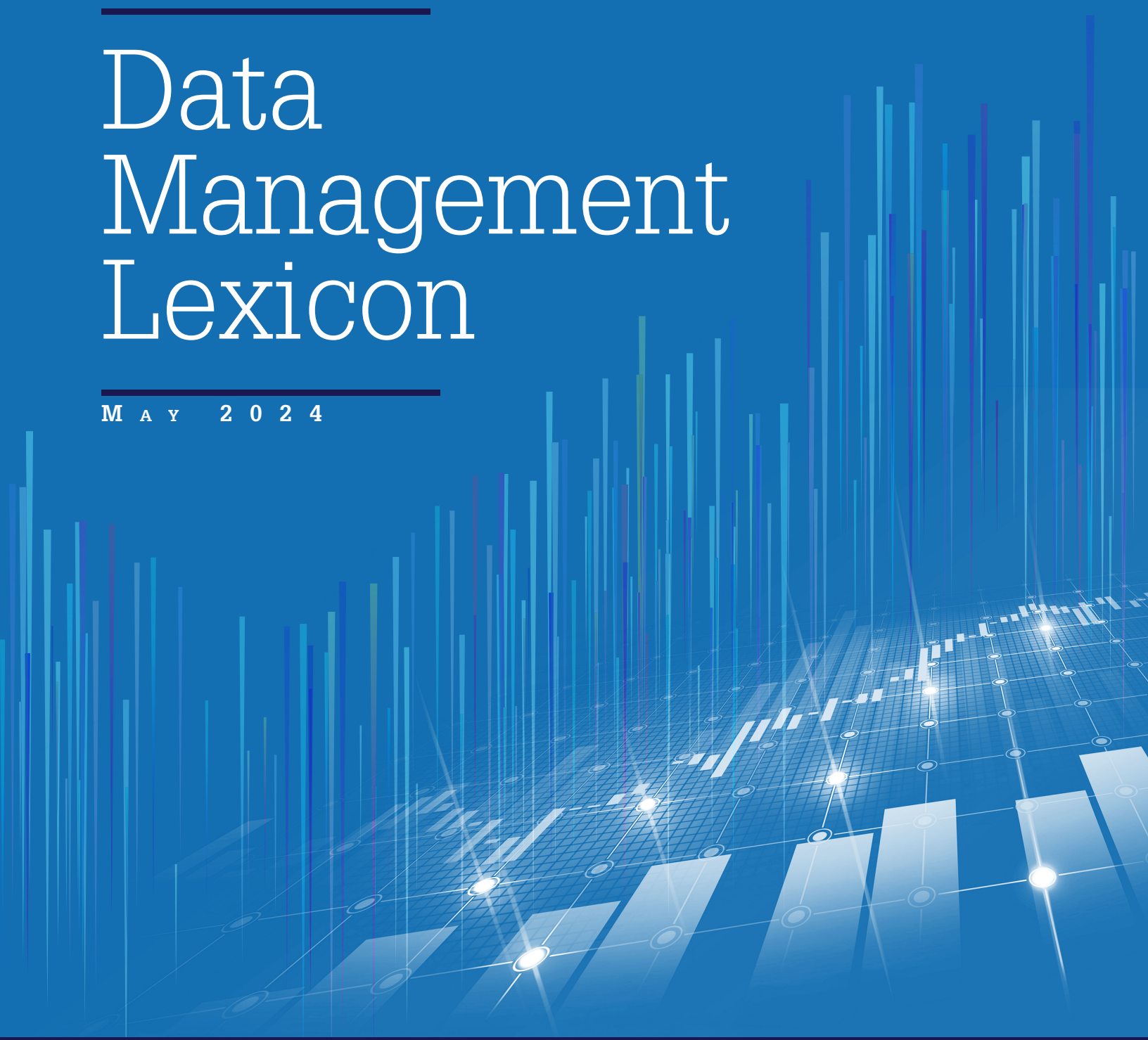


OFFICE OF THE DIRECTOR OF NATIONAL INTELLIGENCE

The Intelligence Community

# Data Management Lexicon

M A Y 2 0 2 4



## Introduction

The use of approved terminology within and external to the Intelligence Community (IC) is fundamental to improving discovery of, access to, and responsible use of our valuable data and information across the IC Information Environment. The IC Data Management Lexicon (IC DML) establishes definitions for over 100 common data management terms in the IC to ensure common understanding and consistent use of this terminology. The terms defined in this document, which include 'Data Acumen', 'Metadata', 'Provenance', and 'Lineage' ensure the IC can communicate more effectively with itself and with its partners. This latest version adds even more modern industry terminology, such as 'Data Culture', 'Data Mesh', and 'Data Pipeline'.

Terms are selected for inclusion by the IC Chief Data Officer (CDO) Council when clarification beyond the definition approved by the Data Management Association (DAMA) international is required, or if a term is not defined by DAMA. All terms and definitions were reviewed and approved by the IC CDO Council. For any terms not included in the IC DML, the IC defers to definitions approved by DAMA, without replicating them as part of this published document.

The IC Data Management Lexicon is hosted at [www.odni.gov](http://www.odni.gov).

ROLE/TERM	DEFINITION
<b>Aggregated Data</b>	Data resulting from processes that combine and summarize granular data from one or more data sources.
<b>Analytic Developer</b>	A person (e.g., software developer, analyst) who designs, codes and/or tests software for the exploration and processing of data to discover and identify meaningful information and trends.
<b>Analytic Production Steward</b>	An appropriately cleared employee of an IC element, who is a senior official, designated by the head of that IC element to represent the analytic activity that the IC element is authorized by law or executive order to conduct, and to make determinations regarding the dissemination to or the retrieval by authorized IC personnel of analysis produced by that activity.
<b>Analytics</b>	The systematic computational analysis of data or statistics to discover and identify meaningful information and trends.
<b>Archived Data</b>	<p>Data that has been identified as being inactive and has been moved out of production systems into a long-term storage repository. Archived data is not immediately available for operational use but can be brought back into service as needed.</p> <p>Note: Archived data is not synonymous with data backups, which are duplications of data typically used for Continuity of Operations purposes.</p>
<b>Authoritative Data</b>	Data provided by an authoritative source.
<b>Authoritative Source</b>	A source of data or information that is recognized by members of a Community of Interest, as defined in Committee on National Security Systems Instruction (CNNSI) 4009, to be valid or trusted because its provenance is considered highly reliable or accurate. An authoritative source may be the functional combination of multiple, separate data sources. During the lifecycle process, the authoritative source (or system of use in which it is housed) can evolve according to use. Subject Matter Experts (SMEs) validate that the data is authoritative, and data management assures that data from the authoritative source is provided to users, and that it is current.
<b>Authorized IC Person</b>	An U.S. person employed by, assigned to, or acting on behalf of an IC element who, through the course of their duties and employment, has a mission need and an appropriate security clearance. An Authorized IC Person (AICP) shall be identified by their IC element head and shall have discovery rights to information collected and analysis produced by all elements of the IC. The term may include contractor personnel.
<b>Bulk Data</b>	Data or datasets acquired or collected, whether classified or unclassified, without the use of discriminants, which is typically the result of a bulk collect activity. The resulting data is handled in accordance with applicable law and policy.
<b>Bulk Collect</b>	Data acquisition activities that support mission requirements, which due to technical or operational considerations do not target a specific person or entity.
<b>Business Data</b>	Data used, gathered, or generated during business actions taken to operate an organization (e.g., IC element, Department of Defense (DoD) element, law enforcement element), including, but not limited to, data concerning communications, payroll, finance, administration, organization-related persons (Human Resources (HR) or Personally Identifiable Information (PII)), physical location, property, security, and business metrics. This does not include data collected or generated principally for mission (e.g., intelligence, defense, law enforcement) purposes.
<b>Catalog</b>	A curated collection of metadata about resources (e.g., datasets, data services in the context of a data catalog), usually arranged systematically.
<b>Cataloging</b>	The process of curating (gathering, organizing, maintaining, presenting) a collection of metadata about resources.

ROLE/TERM	DEFINITION
<b>Chief Data Officer</b>	A designated senior official within each IC element responsible for the management of data as an asset and the establishment and enforcement of data-related strategies, policies, standards, processes and governance.
<b>Classification (Information Security Context)</b>	A set of discrete, exhaustive, and mutually exclusive observations that can be assigned to one or more variables to be measured in the collation and/or presentation of data. In the IC, this is the labeling and characterization of data by the extent of damage to national security reasonably expected to occur as a result of unauthorized disclosure.
<b>Collection</b>	Any information or data, both in its final form, and in the form when initially gathered, acquired, held, or obtained by an IC element that is potentially relevant to a mission need of any IC element. This includes information or data obtained directly from its source, regardless of whether the information or data has been reviewed or processed.
<b>Collection Steward</b>	An appropriately cleared employee of an IC element, who is a senior official, designated by the head of that IC element to represent a collection activity that the IC element is authorized by law or executive order to conduct, and to make determinations regarding the dissemination to or the retrieval by authorized IC personnel of information collected by that activity.
<b>Commercially Available Information</b>	<p>Any information or data that is of a type customarily made available or obtainable and sold, leased, or licensed to the general public or to non-governmental entities for purposes other than governmental purposes. Commercially Available Information (CAI) also includes information or data for exclusive government use, knowingly and voluntarily provided by, procured from, or made accessible by corporate entities at the request of a government entity, or on their own initiative.</p> <p>Note: CAI is not necessarily “Publicly Available Information (PAI)” accessible to the general public. CAI may have privacy, civil liberties, or sourcing restrictions, and must be handled in accordance with applicable law and policy to ensure it is appropriately acquired, processed, and disseminated. Additionally, information or data obtained via legal processes (e.g., subpoenas, warrants) is not considered CAI in an IC context.</p>
<b>Community of Interest (Data Management Context)</b>	A collaborative group of people assembled around a data-related topic that shares resources (e.g., information, services) to address mission and business goals or concerns.
<b>Data</b>	A representation of facts, concepts, or instructions, such as text, numbers, graphics, documents, images, sound, or video, in a form suitable for communication, interpretation or processing, which individually have no meaning by and in themselves.
<b>Data Access</b>	The ability of a human or Non-Person Entity (NPE) to perform one or more operations on data, typically via service endpoints and Application Programming Interfaces (APIs). These operations may include the ability for data to be searched, retrieved, read, created, updated, deleted, manipulated and executed.
<b>Data Acumen</b>	The ability to sufficiently understand, analyze, reason, communicate, and make decisions and judgments with and about data in context.

ROLE/TERM	DEFINITION
<b>Data Analyst</b>	<p>Someone who produces reports, briefings, and actionable insights that are informed by data.</p> <ul style="list-style-type: none"> <li>▪ Data analysts leverage data-driven tools and algorithms to create actionable insights.</li> <li>▪ They work with partners and conduct research to best understand key problems and how to address them with data, they then utilize the work of data scientists and other data professionals to bring data driven insights into connection with subject matter expertise.</li> </ul>
<b>Data Architect</b>	<p>Someone who is responsible for the overall data functional construct of an organization; its data architecture and data models, and the design of the databases and data integration solutions that support the organization.</p> <ul style="list-style-type: none"> <li>▪ Data architects design the eco-system (e.g., procedures, governance, architectures) to hold, manage, process, and preserve or dispose of data.</li> <li>▪ They enable an organization to manage its data as an asset and increase the value it gets from its data by identifying opportunities for data usage, cost reduction, and risk mitigation, making data driven intelligence possible.</li> </ul>
<b>Data as an IC Asset</b>	Data that may be relevant to one or more IC elements for intelligence purposes.
<b>Data Asset</b>	Data maintained and secured as a shared, critical, inexhaustible, durable, and strategic resource with the expectation of future value and benefits. Examples of data assets include databases, documents, data returned as web content, application/system output files, and records.
<b>Data Attribute</b>	Any distinctive feature, characteristic, or property of a Data Object that can be identified or isolated quantitatively or qualitatively by either human or automated means. A Data Object can be made up of one or more Data Elements, and a Data Element will typically have Data Attributes as sub-units.
<b>Data Broker</b>	An individual, group, or service that collects data from one or more sources and sells, licenses, or transports the resulting data sets to new users or organizations. Data brokers can also be used by organizations to supplement or enhance existing data.
<b>Data Categorization</b>	A mechanism for establishing order through the grouping of related data, where members of a grouping bear some immediate similarity within a given context. Example groupings include mission intelligence, subject, data format, language, and context use.
<b>Data Category</b>	A defined data grouping based on a controlled hierarchical taxonomy used to organize data so that it may be located, accessed, processed, analyzed, and protected more efficiently. The utility of any single data category, or list of categories, may not be inherently self-evident, and should be further defined within a given context or scope (e.g., the list of Data Subject Categories for the purposes of cataloging datasets, or the list of Financial Data Categories for the purposes of processing financial data to generate intelligence leads).
<b>Data Centricity</b>	An architectural approach that results in a secure environment separating data from applications and making data available to a broad range of tools and analytics within and across security domains for enrichment and discovery. This environment embraces a more disciplined approach to intelligence integration by ensuring that data is sharable, discoverable, accessible, understandable, retrievable, and protected.
<b>Data Cleansing</b>	A data processing activity to transform data and make it conform to data standards and domain rules; includes detecting and correcting data errors (e.g., removing rows that contain bad values, filling in missing values based on pre-determined rules) to bring the quality of data to an acceptable level. Data Cleansing is a part of overall Data Conditioning.

ROLE/TERM	DEFINITION
<b>Data Conditioning</b>	The controlled processes used to transform data (e.g., cleansing, metadata assignment, format and content normalization, data model mediation, enrichment) to make it usable for a particular purpose at any point in its lifecycle.
<b>Data Consumer</b>	A person or NPE that receives data (e.g., on a screen, in a report, through a query, or via a machine-to-machine interface), uses the data for a specific purpose, and can be affected by its quality.
<b>Data Culture</b>	The collective behaviors and beliefs of people within an organization who value, practice, and encourage the use of data to improve mission and business outcomes. As a result, data centric policies, processes, standards, tools, and techniques are woven into organizational strategies, analysis, operations, and decision making.
<b>Data Curation</b>	The active maintenance of data, throughout its lifecycle, to ensure levels of readiness for current and future use. Data curation activities involve continuously working with data creators and users, enhancing discovery and retrieval, supporting research and data correlation, ensuring data quality, protection and accessibility, and adding value to data (e.g., collection building, adding metadata, providing search mechanisms).
<b>Data Custodian</b>	An IC element that, on behalf of the Originating Element, may perform mission and business data-related tasks such as collecting, tagging, and processing data, and granting individual users access to additional information beyond that of general systems, applications, and file permissions to perform such functions, where appropriate. The Data Custodian does not assume the legal or policy roles of the Originating Element.
<b>Data Domain</b>	A collection of data representing key concepts across a specific mission area and that is usually identifiable via recognizable governance or authoritative bodies.  Note: Data Domain is not synonymous with a data fabric in this context.
<b>Data Element</b>	A discrete unit of data that has a unique meaning within a specific model or schema, and may be comprised of sub-units. Example data elements for a person may include last name, first name, and middle initial.
<b>Data Engineer</b>	Someone who conditions data to fit within the data architecture and transforms it to be exploitable. <ul style="list-style-type: none"> <li>▪ Data engineers transform data into usable and computationally accessible forms.</li> <li>▪ They condition data through Extraction, Cleansing, Transforming, and Loading (ECTL), also known as data munging. They implement data systems which separate data from applications and scale, as required.</li> </ul>
<b>Data Entity</b>	A classification [representation] of objects found in the real world described by the Noun part of speech—persons, places, things, concepts, and events—of interest to the enterprise; usually expressed in singular form.
<b>Data Entity Tag</b>	A data tag that represents a single assertion about a data entity to enable analytic correlation across the enterprise (e.g., tag name of "Person Name" with a corresponding tag value of "Joe Smith").
<b>Data Fabric</b>	A design concept that serves as a federated and integrated layer (fabric) of data, and connecting processes for sharing information through interfaces and services to discover, understand, and exchange data with partners across all applications, domains, echelons, and security levels.  Note: At a minimum, the implementation of the design concept must support cataloging, data event messaging, interface management, and access management capabilities. Data Fabric is not a replacement of traditional data management architectures such as Data Lakes, Data Warehouses, and Databases.



ROLE/TERM	DEFINITION
<b>Data Governance</b>	A discipline comprised of responsibilities, roles, functions, and practices, supported by authorities, policies, and decisional processes (planning, setting policies, monitoring, conformance, and enforcement), which together administer data and information assets across an IC element to ensure that data is managed as a critical asset consistent with the organization's mission and business performance objectives.
<b>Data Governance Council</b>	A decision making and/or policy making council of senior managers, chaired by the CDO, who are responsible for the highest tier of data governance in an IC element. The Data Governance Council (DGC) oversees or manages data governance initiatives (e.g., development of policies or metrics), issues and escalations. The DGC monitors results to ensure the IC elements receive the desired outcomes and business value from data management activities. This may also be called a Data Council, Executive Data Council or Data Executive Council.
<b>Data Ingest</b>	Capabilities and activities that an organization uses to scope, plan and implement extraction, data conditioning, and storage to enable the incorporation of data into managed repositories.
<b>Data Ingest/Tagging Point of Contact</b>	An IC element employee responsible for the instrumentation of formatting, labeling, and tagging of data in preparation for ingestion into the IC Cloud.
<b>Data Interoperability</b>	The ability of systems and services that create, exchange and consume data to have clear, shared expectations (e.g., conventions, standards, policy) for the contents, context, and meaning of that data, across varying platforms and security domains.
<b>Data Lake</b>	A centralized, scalable, and access-controlled repository for structured and unstructured data, no matter the source or format, generally presenting an unrefined view of the data to enable exploration, innovation, and analysis. The data is typically stored in its exact or near exact source formats, along with refined formats to add additional data value for enhanced analytics and data management. In some cases, modern Data Lakes have been used to replace highly structured Data Warehouses.
<b>Data Lifecycle</b>	A conceptualization of a cradle-to-grave value chain for data, which often includes phases such as plan and task, acquire and assess, process and transform, discover and access, analyze and exploit, and preserve or dispose.
<b>Data Lifecycle Management</b>	Establishment and execution of policies and interconnected processes for managing data throughout the data lifecycle to support data management functions, such as data governance.
<b>Data Lifecycle Phase 1: Plan &amp; Task</b>	Activities prior to obtaining data that include how data needs are determined; collection objectives are prioritized; costs, storage, and compute requirements are assessed; collection methodologies or approaches are selected; and decisions are documented with respect to relevant data authorities, permissions, and use and sharing rules.
<b>Data Lifecycle Phase 2: Acquire &amp; Assess</b>	Activities related to procurement, collection, and generation of data, including determining mission-relevant features or business purposes. This phase includes: <ul style="list-style-type: none"> <li>▪ Ensuring source vetting;</li> <li>▪ Validating and verifying data;</li> <li>▪ Evaluating preliminary data quality;</li> <li>▪ Identifying filtering and PII minimization and data volume reduction opportunities; and</li> <li>▪ Documenting data impact assessments on all data sensitivities, handling, use, protection, and disposition requirements.</li> </ul>

ROLE/TERM	DEFINITION
<b>Data Lifecycle Phase 3: Process &amp; Transform</b>	Activities and documentation related to making data fit for purpose (e.g., data conditioning) and fostering data interoperability across systems. This phase includes aspects of data curation to describe data and enhance discoverability.
<b>Data Lifecycle Phase 4: Discover &amp; Access</b>	Activities that ensure data can be found by and made available to any authorized consumer, and protected through policies for access control and need-to-know. This starts dissemination, per Intelligence Community Directive (ICD) 501, for data that is made accessible outside of an IC element.
<b>Data Lifecycle Phase 5: Analyze &amp; Exploit</b>	Activities related to the use of data for mission purposes. These activities ensure the usability of data by specific tools, performance of data gap identification, continued data safeguarding through data handling and usage limitations, and determination of data value. Data value is derived through targeted queries, analytic models, and automated analytic capabilities (e.g., data correlation, data fusion) while preserving provenance, pedigree and lineage. This phase also serves as the foundation for Intelligence dissemination determinations and tradecraft.
<b>Data Lifecycle Phase 6: Preserve or Dispose</b>	Activities related to the final disposition of data. This includes preservation, purge, or deletion performed in accordance with National Archivist approved records schedules, legal hold requirements and lawful guidance such as the Attorney General approved guidelines pursuant to Executive Order 12333 and Presidential Policy Directive 28.
<b>Data Management</b>	The development and execution of plans, policies, programs and practices (4Ps) that acquire, control, protect, and enhance the value of data assets throughout the lifecycle, led or performed by tradecraft professionals following established disciplines and functions.
<b>Data Management Plan</b>	<p>A plan that documents how specific data will be collected, processed, used, and curated in order to facilitate long-term data management decisions and actions. It typically includes topics such as:</p> <ul style="list-style-type: none"> <li>a) Description of the data to be collected/created;</li> <li>b) Authority under which the data is collected;</li> <li>c) Standards/methodologies for data collection and management;</li> <li>d) Ethics and Intellectual Property concerns or restrictions;</li> <li>e) Plans for data sharing and access; and,</li> <li>f) Strategy for long-term preservation of the data.</li> </ul>
<b>Data Mesh</b>	A decentralized organizational and technical approach to share, access, and manage data in large-scale environments within or across organizational boundaries. This approach links disparate sources through centrally managed sharing and governance guidelines. The result is a domain-oriented, federated approach where data is created and consumed as a product.
<b>Data Mining</b>	<p>The process of uncovering patterns, insights, and other valuable information from large data sets via discovery and extraction with methods that combine machine learning, statistics, and exploratory data analysis.</p> <p>Note: Federal agencies must report data mining activities to Congress, and reports shall be made available to the public, per 42 U.S. Code § 2000ee–3 Federal agency data mining reporting.</p>
<b>Data Model</b>	Organized representations of an enterprise's data elements which standardize how elements relate to each other and to the properties of real-world entities divided into conceptual, logical, and physical layers.



ROLE/TERM	DEFINITION
<b>Data Modeler</b>	<p>Someone who is responsible for reviewing and validating data requirements, providing technical data solutions, and designing logical and physical data structures in support of domain specific needs.</p> <ul style="list-style-type: none"> <li>▪ The Data Modeler demonstrates the ability to analyze requirements to develop high-level and detailed data, and access models, conduct business and technical data assessments, and document metadata.</li> <li>▪ They create data models for domain specific data, support and advise domain scientists/researchers during the whole research cycle and data management lifecycle.</li> </ul>
<b>Data Object</b>	An instance of data that is discrete and bounded with an intrinsic, immutable, and unique identity that can persist independently of a system or service. A data object is made up of one or more data elements. For example, a row within a relational database or an image within an image library.
<b>Data Operations</b>	A set of practices, processes, and technologies that facilitates efficient, automated, and secure management of data.
<b>Data Owner (deprecated term)</b>	The data owner is considered a legacy term, since data is an IC asset, in accordance with ICD 501 and the IC Information Enterprise (IE) Data Strategy, and the associated responsibilities have been captured in the definitions for Data Steward, Collection Steward, Analytic Production Steward, and Originating Element.
<b>Data Pipeline</b>	A set of tools and processes to automate or otherwise enable the movement, transformation, and optimization of data from a source to a destination.
<b>Data Policy</b>	Directives that codify principles and management intent into fundamental rules governing the creation, acquisition, integrity, security, quality, and use of data.
<b>Data Preparation</b>	This term is synonymous with Data Conditioning.
<b>Data Producer</b>	This term is synonymous with Data Provider.
<b>Data Protection</b>	<p>The processes, services, and methods used to accomplish the privacy, safety, confidentiality, integrity, availability, and recovery of data. Examples of data protection include:</p> <ol style="list-style-type: none"> <li>a) Monitoring unexpected events, including security violations, and suspicious activity or inappropriate access;</li> <li>b) Protecting data from improper alteration, deletion, or addition;</li> <li>c) Encrypting data;</li> <li>d) Masking or obfuscating data;</li> <li>e) Protecting data-at-rest and data-in-motion;</li> <li>f) Using credential security; and,</li> <li>g) Applying data security restrictions.</li> </ol>
<b>Data Provider</b>	An organization or person who initially creates or provides data on behalf of the Originating Element. This may be a Collection Steward, Analytic Production Steward, Data Custodian, or an external data source functioning on behalf of the Originating Element.
<b>Data Quality</b>	The degree to which data is accurate, complete, timely, consistent with all requirements and business rules, and relevant for a given use.
<b>Data Quality Analysis</b>	The evaluation of data quality deficiencies and its causes against data quality issues (e.g., identification of being inaccurate, incomplete, inconsistent).

ROLE/TERM	DEFINITION
<b>Data Quality Audit</b>	Activities and documentation related to making data fit for purpose (e.g., data conditioning) and fostering data interoperability across systems. This phase includes aspects of data curation to describe data and enhance discoverability.
<b>Data Quality Dimension-Accuracy</b>	The degree to which a data attribute value closely and correctly describes its mission or business entity instance (the “real life” entity) as of a point in time. This assesses the freedom from mistakes or error, the exactness, and the degree of conformity of a measure to a standard or true value.
<b>Data Quality Dimension-Completeness</b>	The degree to which all required data is present and can be measured at the dataset, record, or column [element] level.
<b>Data Quality Dimension-Conformance</b>	The degree to which data follows agreed upon internal policies, standards, procedures and architectural requirements.
<b>Data Quality Dimension-Consistency</b>	The degree to which data values are consistently represented within a dataset and between datasets, and consistently associated across datasets. It can also refer to the size and composition of datasets between systems and across time.
<b>Data Quality Dimension-Integrity</b>	The degree to which data can be trusted due to its provenance, pedigree, lineage and conformance with all business rules regarding its relationship with other data. In the context of data movement, this is the degree to which data has verifiably not been changed unexpectedly by a person or NPE.
<b>Data Quality Dimension-Timeliness</b>	The degree to which data follows: 1) currency - the measure of whether data values are the most up-to-date version of the information, and 2) latency - the length of time between an event occurring and the data representing it becoming available for use.
<b>Data Quality Dimension-Reasonability</b>	The degree to which a data pattern meets expectations within a specific operational context. For example, the expectation that the number of transactions each day does not exceed 105% of the running average number of transactions for the previous 30 days.
<b>Data Quality Dimension-Validity</b>	The degree to which data conforms to domain or syntax values (e.g., format, type, range) and defined mission and business data rules.
<b>Data Quality Dimension-Uniqueness</b>	Assessment of key values to ensure no entity (thing) exists more than once within a defined domain (e.g., within a dataset).
<b>Data Repository</b>	<p>A general term used to describe an environment where data, metadata, data objects, and data collections are ingested or uploaded and are permanently managed, stored, archived long-term, preserved, and made accessible.</p> <p>Note: Organizations, such as DAMA, recommend not using this term because it is used loosely to define any database or file.</p>
<b>Data Scientist</b>	<p>Someone who creates repeatable means to draw key insights and signals from data.</p> <ul style="list-style-type: none"> <li>▪ Data scientists invent, perfect, or apply algorithms to extract insights from data.</li> <li>▪ They are specialists in a range of mathematical, computational, and visualization techniques that allow an organization to draw the greatest benefit from data holdings in terms of insight and decision advantage.</li> </ul>
<b>Data Security</b>	<p>The ability to protect data resources from unauthorized discovery, access, use, modification, and/or destruction. Secure data sharing relies on several key functions: data identification, categorization, and labeling; entitlement management; and policy establishment.</p> <p>Note: Data Security is a component of Data Protection.</p>

ROLE/TERM	DEFINITION
<b>Data Sharing</b>	The practice of providing access to data resources to multiple users, applications, or organizations while maintaining the fidelity and integrity of the data. This includes the technologies, practices, legal frameworks, and cultural elements that ensure data is available to any entity with a need-to-know and proper access permissions, while protecting it from unlawful or improper use.
<b>Data Stakeholders</b>	Any individual or group who has a vested interest in data at any point in its lifecycle.
<b>Data Standards</b>	Specifications, sets of rules, methods, terminologies, or guidance, approved by a recognized body to enable how data is created, stored, exchanged, managed, or processed in a common and repeatable way to facilitate data interoperability. Data standards codify the representation, format, definition, structuring, tagging, transmission, manipulation, use, or management of data.
<b>Data Steward</b>	<p>Within many IC elements, someone whose responsibilities are assigned to specific personnel across a multi-level Data Stewardship hierarchy. Whether represented by a single IC element employee or by responsibilities distributed through an organizational hierarchy, Data Stewards are legally accountable across the data lifecycle on behalf of the Originating Element for:</p> <ul style="list-style-type: none"> <li>a) Establishing protection, sharing, and governance guidelines for data and datasets within an assigned subject area;</li> <li>b) Maintaining data names, business definitions, data integrity rules, and domain values within an assigned subject area;</li> <li>c) Compliance with legal and policy requirements and conformance to internal and IC data policies and data standards;</li> <li>d) Ensuring application of appropriate security controls;</li> <li>e) Analyzing and improving data quality; and</li> <li>f) Identifying and resolving data related issues.</li> </ul>
<b>Data Store</b>	A place where data assets, including structured and unstructured databases, files, and text documents, are stored, protected, and maintained while at rest.
<b>Data Structure</b>	The physical or logical relationships among data elements that represent a specific, pre-defined schema or data model, used for organizing and storing data, and designed to support specific data manipulation functions. Examples include array, file, record, table, tree, queue, linked list, and edge/node.
<b>Data System</b>	The hardware and/or software used to process, access, exchange, analyze, store, and/or retrieve data.
<b>Data System Owner</b>	The organization or entity responsible for the overall procurement, development, integration, modification, or operation and maintenance of a data system.
<b>Data Tag</b>	Metadata applied, through tagging to a data asset to help describe characteristics about the data, such as privacy, security, provenance, source, or other information, and can be used to support automated processing. A “tag” is an assertion describing some aspect of a resource, pairing a semantic label with a value (e.g., a document may have a tag name of “Language” with a corresponding tag value of “English”). The tag values may be known <i>a priori</i> (e.g., controlled vocabulary) or not (e.g., folksonomies).
<b>Data Tagging</b>	The act of associating data tags as metadata to a data object by identifying, labeling, and describing its information. Typically, tagging supports user interpretation and automated processing.

ROLE/TERM	DEFINITION
<b>Data Transformation</b>	The process of converting data from one format or structure to another. Since data often resides in different formats across the enterprise, data transformation is necessary to ensure data is intelligible to multiple applications, services and databases to support data integration and interoperability. Data transformation is a component of data conditioning.
<b>Data Type</b>	A category of logical or physical data structures with common properties, uses, and technically feasible operations (e.g., addition, string concatenation) on values. Example data types include numeric, alphanumeric, packed decimal, floating point, and datetime.
<b>Data Warehouse</b>	A storage architecture designed to hold data extracted from multiple sources (e.g., transaction systems, data stores, external sources). The warehouse then aggregates the data in a form suitable for further analysis.
<b>Database</b>	An organized collection of datasets generally stored and accessed from a computer system that allows the data to be easily searched, manipulated, and updated.
<b>Dataset</b>	One or more data objects that share common properties and characteristics, and are managed as a unit.
<b>Derived Data</b>	Data that is computed or extrapolated from existing data, regardless of its origin
<b>Discovery</b>	The act of obtaining knowledge of the existence, but not necessarily the content, of information or data collected or analysis produced by any IC element.  Note: This is Discovery as defined and applicable under ICD 501.
<b>Dissemination</b>	The act of transmission, communication, sharing, or passing of information collected or analysis produced by an IC element outside of that IC element, either through the ordinary course of business or in response to a request following discovery. Dissemination includes providing any access to information in an IC element's custody to persons and/or NPEs outside the IC element.
<b>Evaluated Data</b>	Data that has been assessed post-collection and determined to meet established criteria related to authorities, United States Person status, or its intended mission purpose (e.g., foreign intelligence, counter intelligence, information assurance). Results of this determination may drive further handling requirements (e.g., retention).
<b>Evaluated Information</b>	Information that has been assessed post-collection and determined to meet established criteria related to authorities, United States Person status, or its intended mission purpose (e.g., foreign intelligence, counter intelligence, information assurance). Results of this determination may drive further handling requirements (e.g., retention).
<b>Functional Manager</b>	A designated senior official, reporting to the Director of National Intelligence, executing Intelligence Functional Manager duties in accordance with Executive Order 12333 and ICD 113. Develops policies, guidance, procedures, and tradecraft standards related to the specific intelligence discipline, and sets related training.
<b>Information</b>	The meaning assigned to data by a known rule or set of rules. Generally, an understanding concerning any objects such as facts, events, things, processes, or ideas, including concepts that, within a certain context and timeframe, have a particular meaning. Information is the interpretation of data based on its context, including the: <ul style="list-style-type: none"> <li>a) The business or mission meaning of data elements and related terms;</li> <li>b) The format in which the data is presented;</li> <li>c) The timeframe represented by the data; and,</li> <li>d) The relevance of the data to a given usage.</li> </ul>

ROLE/TERM	DEFINITION
<b>Information Environment</b>	The aggregate of individuals, groups, organizations, communities, technical systems, and information technology capabilities, that collect, process, share, disseminate, or act on information and data.
<b>Information Sharing and Safeguarding Executive</b>	The senior official charged with overseeing information sharing and safeguarding efforts for an IC element. The Information Sharing and Safeguarding Executive (ISSE) is responsible for information sharing and safeguarding policy including the implementation of enterprise data and information access, information sharing, and information safeguarding policies and business requirements consistent with laws, regulations, and policies. The ISSE coordinates all enterprise data access policy and governance matters with the CDO, if not dual-hatted as the CDO.
<b>Information Space</b>	An aggregation of data, stored and maintained in an organized way, in an information environment and typically made available online. For a specific information environment, it is a content repository that helps users to browse to the information or data they want to use/reuse or the document they need to reference, produced by a set of known procedures, and changed through intentional manipulation of its content.
<b>Lineage</b>	A description of data's pathway from its source to its current location and the alterations made to the data along that pathway, which should be represented as a reproducible ancestry of the data object. Lineage can include traceability between parent and children data objects.
<b>Master Data</b>	Core mission and business data entities used in traditional or analytical applications across an organization, and subjected to enterprise governance policies, along with their associated metadata, attributes, definitions, roles, connections and taxonomies. Master data provides context for mission and business activity data in the form of common and abstract concepts related to activity transactions, along with a consistent and uniform set of identifiers and extended attributes that describe the core entities.
<b>Master Data Management</b>	Processes that control management of master data values to enable consistent, shared, contextual use across applications, of the most accurate, timely, and relevant version of truth about essential mission and business entities. Usually enabled by technology so that mission, business and IT work together to ensure the uniformity, accuracy, stewardship, semantic consistency and accountability of the enterprise's official shared master data assets.
<b>Metadata</b>	Literally, "data about data"; administrative or descriptive data attributes that are consistent across mission and business disciplines, domains, and data encodings, and are used to improve business or technical understanding of data and data-related processes.
<b>Mission Data</b>	Data gathered, acquired, generated, held, or obtained during mission activities by an organization (e.g., IC element, DoD element, law enforcement element) to satisfy mission (e.g., intelligence, defense, law enforcement) needs and which can be shared across systems and organizations working toward the same mission. This data includes, but is not limited to, observations, recordings, images, signals, measurements, and signatures of physical or digital attributes and events.
<b>Ontology</b>	A formal representation of a domain of knowledge. It is comprised of a taxonomy as an integral part, with an underlying vocabulary including definitions of terms representing universals, defined classes, and axioms from which rational arguments can be made.
<b>Originating Element</b>	An IC element or U.S. Government entity that creates or collects information during the course of its business and is legally responsible for it (e.g., records management, classification, and lead for Freedom of Information Act and Privacy Act responsibilities). Responsibilities are executed in accordance with ICD 121.

ROLE/TERM	DEFINITION
<b>Pedigree</b>	The description of the Data Quality (e.g., accurate, complete, timely, and consistent) assessment for data, its compliance with established standards, and the processing steps performed to derive the data. Pedigree information is used to augment Lineage.
<b>Provenance</b>	Description of the origin or source of data, its history of stewardship or custodianship and location(s), which can be used to form assessments about its quality, reliability, or trustworthiness. Within a specific mission context only selective provenance attributes may be considered as relevant.
<b>Publicly Available Information</b>	<p>Information or data published or broadcast for public consumption, is available on request to the public, is intended to be accessible on-line or otherwise to the public, is available to the public by subscription or commercial purchase, could lawfully be seen or heard by any casual observer, is made available at a meeting open to the public, or is obtained by visiting any place or attending any event that is open to the public. DoD extends the definition by stating that it “includes information generally available to persons in a military community even though the military community is not open to the civilian general public.”</p> <p>Note: The extension of PAI is done under the auspices of conducting authorized intelligence activities in a manner that protects the constitutional and legal rights and the privacy and civil liberties of U.S. persons.</p>
<b>Public Domain (Legal)</b>	Information or data that is not or no longer protected by copyright, patent or trademark nor owned or can be acquired by any individual or private entity. Such information can be freely used by any community for public purposes and is available to be used without permission or authorization from its owner.
<b>Public Domain (Non-Legal)</b>	Openly accessible forums that are free for all to use for individual expression, in which different opinions can be expressed, problems of general concern can be discussed, and collective solutions can be developed collaboratively with other individuals.
<b>Record (Information and Records Management Context)</b>	Information and data made or received by an agency of the United States Government under Federal law or in connection with the transaction of public business and preserved or appropriate for preservation by that agency or its legitimate successor as evidence of the organization, functions, policies, decisions, procedures, operations, or other activities of the government or because of the informational value of data in them. Records do not include materials made or acquired and preserved only for convenience for reference or exhibition purposes, extra copies of documents preserved only for convenience of reference, or stocks of publications and processed documents.
<b>Reference Data</b>	Data used to organize or categorize other data (e.g., controlled values), or for relating data to information (e.g., calibration data) both within and beyond the boundaries of the enterprise. Usually consists of codes and descriptions or definitions.
<b>Reference Data Management</b>	Processes that control vocabularies (defined domain values), including control over standardized terms, code values and other unique identifiers, business definitions for each value, business relationships within and across domain value lists, and the consistent, shared use of accurate, timely, and relevant reference data values to categorize data.
<b>Semi-structured Data</b>	Data that has elements of both unstructured and structured data. For example, a Microsoft Word document is generally considered to be unstructured data, but with the addition of metadata tags used to enable discoverability, the data is now semi-structured. Other types of semi-structured data formats include: Extensible Markup Language (XML), JavaScript Object Notation (JSON), email, and formats based on Electronic Data Interchange (EDI) standards (e.g., X12, Electronic Data Interchange for Administration, Commerce, and Trust (EDIFACT), Organization for Data Exchange by Tele Transmission in Europe (ODETTE)).



ROLE/TERM	DEFINITION
<b>Structured Data</b>	Content that conforms to a specific, pre-defined schema or data model, or is tagged or otherwise arranged into database tables (rows and columns). Examples include data in relational databases, data in graph databases, call data records, financial transactions, and system audit logs.
<b>Support Data</b>	<p>Data used to enable or assist a mission or business activity to be performed. This includes, but is not limited to, data concerning mission planning, logistics, reference, schedule, tasking, status, building and maintenance of business and mission support systems (such as algorithms, models, or sensors), and system verification and validation.</p> <p>Note: Unlike other definitions for data (e.g., Business Data, Mission Data, Reference Data), data is considered support data based on how it is used.</p>
<b>Unevaluated Data</b>	Data that has not been assessed post-collection and determined to meet established criteria related to authorities, United States Person status, or its intended mission purpose (e.g., foreign intelligence, counter intelligence, information assurance).
<b>Unstructured Data</b>	Content that does not conform to a specific, pre-defined data model, or is not tagged or otherwise structured into database tables (rows and columns). Examples include documents, presentations, graphics, images, text, reports, videos, or sound recordings.

