



OFFICE OF THE DIRECTOR OF NATIONAL INTELLIGENCE

The Intelligence Community

Data Management Lexicon

J A N U A R Y 2 0 2 2

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 information and data across the IC Information Environment. The IC Data Management Lexicon (IC DML) establishes definitions for over 100 data management terms. The defined data management terms contained in this document, including 'data', 'metadata', 'provenance', and 'lineage', ensure the IC can communicate more effectively with its partners.

The IC DML includes terms to support common interpretation when communicating and collaborating among IC Elements. Terms are selected for inclusion by the IC Chief Data Officer (CDO) Council when clarification beyond the definition approved by Data Management Association (DAMA) International is required, or 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.

LEXICON EDITION	DESCRIPTION OF CHANGE
January 2022 (Version 1.1)	Original version

ROLE/TERM	DEFINITION
Analytic Developer	A person (e.g., software developer or analyst) who designs, codes and/or tests software for the exploration and processing of data to discover and identify meaningful information and trends.
Analytic Production Steward	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.
Analytics	The systematic computational analysis of data or statistics to discover and identify meaningful information and trends.
Authoritative Data	Data provided by an authoritative source.
Authoritative Source	A source of data or information that is recognized by members of the 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 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.
Authorized IC Person	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. 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.
Bulk Data	Data or datasets acquired or collected without the use of discriminants, typically the result of a bulk collect activity. The resulting data is handled in accordance with applicable law and policy.
Bulk Collect	Data acquisition activities that support mission requirements, which due to technical or operational considerations do not target a specific person or entity.
Business Data	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. Does not include data collected or generated principally for mission (e.g., intelligence, defense, law enforcement) purposes.
Catalog	A curated collection of metadata about resources (e.g., datasets, data services in the context of a data catalog), usually arranged systematically.
Cataloging	The process of curating (gathering, organizing, maintaining, presenting) a collection of metadata about resources.
Chief Data Officer	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.
Collection (noun)	Any information [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 [data] obtained directly from its source, regardless of whether the information [data] has been reviewed or processed.

ROLE/TERM	DEFINITION
Collection (verb)	Acquisition of information [data] to meet an intelligence requirement.
Collection Steward	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.
Commercially Available Information	<p>Any information [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 [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 [cooperative source].</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 the information [data] is appropriately acquired, processed, and disseminated.</p>
Data	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.
Data Access	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.
Data Acumen	The ability to sufficiently understand, analyze, reason, communicate, and make decisions and judgements with and about data in context.
Data Analyst	<p>Someone who produces reports, briefings, and actionable insights that are informed from data.</p> <ul style="list-style-type: none"> ▪ Data analysts leverage data-driven tools and algorithms to create actionable insights. ▪ 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.
Data Architect	<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"> ▪ Data architects design the eco-system (e.g., procedures, governance, and architectures) to hold, manage, process, and preserve or dispose of data. ▪ 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.
Data as an IC Asset	Data that may be relevant to one or more IC elements for intelligence purposes.
Data Asset	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.

ROLE/TERM	DEFINITION
Data Attribute	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.
Data Categorization	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.
Data Category	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).
Data Centricity	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.
Data Cleansing	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.
Data Conditioning	The controlled processes used to transform data (e.g., cleansing, metadata assignment, format and content normalization, data model mediation, enrichment) to make it useable for a particular purpose at any point in its lifecycle.
Data Consumer	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.
Data Curation	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).
Data Custodian	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.
Data Element	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.

ROLE/TERM	DEFINITION
Data Engineer	<p>Someone who conditions data to fit within the data architecture and transforms it to be exploitable.</p> <ul style="list-style-type: none"> ▪ Data engineers transform data into usable and computationally accessible forms. ▪ They condition data through Extraction, Cleansing, Transforming, and Loading (ECTL), also known as data munging. They implement data systems which separate data from application and scale as required.
Data Entity	<p>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.</p>
Data Entity Tag	<p>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").</p>
Data Fabric	<p>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.</p> <p>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, data hubs and databases.</p>
Data Governance	<p>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.</p>
Data Governance Council	<p>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.</p>
Data Ingest	<p>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.</p>
Data Ingest/Wrapping Point of Contact	<p>An IC Element employee responsible for the instrumentation of formatting, labeling and wrapping of data in preparation for ingestion into the IC Cloud.</p>
Data Interoperability	<p>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.</p>
Data Lake	<p>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.</p>

ROLE/TERM	DEFINITION
Data Lifecycle	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.
Data Lifecycle Management	Establishment and execution of policies and interconnected processes for managing data throughout the data lifecycle to support data management functions, such as data governance.
Data Lifecycle Phase 1: Plan & Task	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.
Data Lifecycle Phase 2: Acquire & Assess	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"> ▪ Ensuring source vetting; ▪ Validating and verifying data; ▪ Evaluating preliminary data quality; ▪ Identifying filtering and PII minimization and data volume reduction opportunities; and ▪ Documenting data impact assessments on all data sensitivities, handling, use, protection, and disposition requirements.
Data Lifecycle Phase 3: Process & Transform	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.
Data Lifecycle Phase 4: Discover & Access	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.
Data Lifecycle Phase 5: Analyze & Exploit	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.
Data Lifecycle Phase 6: Preserve or Dispose	Activities related to final data disposition. 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.
Data Management	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.

ROLE/TERM	DEFINITION
Data Management Plan	<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> <ol style="list-style-type: none"> Description of the data to be collected/created; Authority under which the data is collected; Standards/methodologies for data collection and management; Ethics and Intellectual Property concerns or restrictions; Plans for data sharing and access; and, Strategy for long-term preservation of the data.
Data Modeler	<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"> ▪ 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. ▪ They create data models for domain specific data, support and advise domain scientists/researchers during the whole research cycle and data management lifecycle
Data Object	<p>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.</p>
Data Owner (deprecated term)	<p>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.</p>
Data Preparation	<p>This term is synonymous with Data Conditioning.</p>
Data Producer	<p>This term is synonymous with Data Provider.</p>
Data Protection	<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"> Monitoring unexpected events, including security violations, and suspicious activity or inappropriate access; Protecting data from improper alteration, deletion, or addition; Encrypting data; Masking or obfuscating data; Protecting data-at-rest and data-in-motion; Using credential security; and, Applying data security restrictions.
Data Provider	<p>Organization or person who initially creates or provides the data on behalf of the Originating Element. This may be the Originating Element, Collection Steward, Analytic Production Steward, Data Custodian, or an external data source functioning on behalf of the Originating Element.</p>

ROLE/TERM	DEFINITION
Data Quality	The degree to which data is accurate, complete, timely, consistent with all requirements and business rules, and relevant for a given use.
Data Quality Analysis	The evaluation of data quality deficiencies and its causes against data quality issues (e.g., identification of being inaccurate, incomplete, inconsistent).
Data Quality Audit	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.
Data Quality Dimension-Accuracy	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.
Data Quality Dimension-Completeness	The degree to which all required data is present and can be measured at the dataset, record, or column [element] level.
Data Quality Dimension-Conformance	The degree to which data follows agreed upon internal policies, standards, procedures and architectural requirements.
Data Quality Dimension-Consistency	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.
Data Quality Dimension-Integrity	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, the degree to which data has verifiably not been changed unexpectedly by a person or NPE.
Data Quality Dimension-Timeliness	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.
Data Quality Dimension-Reasonability	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.
Data Quality Dimension-Validity	The degree to which data conforms to domain or syntax values (e.g., format, type, range) and defined mission and business data rules.
Data Quality Dimension-Uniqueness	Assessment of key values to ensure no entity (thing) exists more than once within a defined domain (e.g., within a dataset).
Data Repository	<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>
Data Scientist	<p>Someone who creates repeatable means to draw key insights and signals from data.</p> <ul style="list-style-type: none"> ▪ Data scientists invent, perfect, or apply algorithms to extract insights from data. ▪ 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.

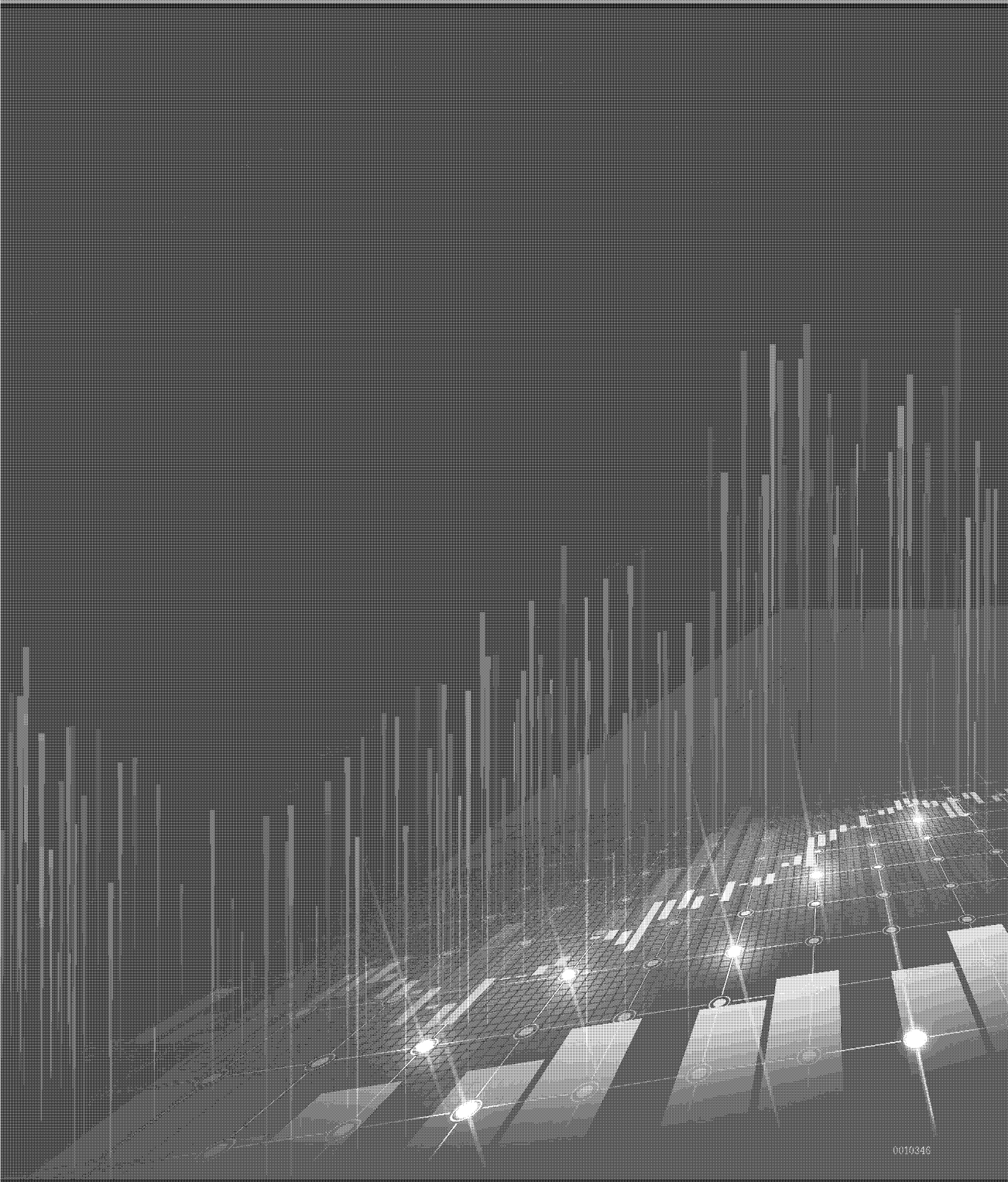
ROLE/TERM	DEFINITION
Data Security	<p>The ability to protect data resources from unauthorized discovery, access, use, modification, and/or destruction. Secure data sharing relies on several key functions: data identification, categorization, and labeling; entitlement management; and policy establishment.</p> <p>Note: Data Security is a component of Data Protection.</p>
Dataset	<p>One or more data objects that share common properties and characteristics, and are managed as a unit.</p>
Data Standards	<p>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.</p>
Data Steward	<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> <ol style="list-style-type: none"> a) Establishing protection, sharing, and governance guidelines for data and datasets within an assigned subject area; b) Maintaining data names, business definitions, data integrity rules, and domain values within an assigned subject area; c) Compliance with legal and policy requirements and conformance to internal and IC data policies and data standards; d) Ensuring application of appropriate security controls; e) Analyzing and improving data quality; and f) Identifying and resolving data related issues.
Data Store	<p>A place where data assets, including structured and unstructured databases, files, and text documents, are kept, protected, and maintained while at rest.</p>
Data Structure	<p>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.</p>
Data Tag	<p>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).</p>
Data Tagging	<p>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.</p>

ROLE/TERM	DEFINITION
Data Transformation	The process of converting data from one format or structure to another. Because 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.
Data Type	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, datetime.
Discovery	The act of obtaining knowledge of the existence, but not necessarily the content, of information [data] collected or analysis produced by any IC element. Discovery as defined and applicable under ICD 501.
Dissemination	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 person and NPEs outside the IC Element.
Evaluated Data	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).
Evaluated Information	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).
Functional Manager	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.
Information	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"> a) The business or mission meaning of data elements and related terms; b) The format in which the data is presented; c) The timeframe represented by the data; and, d) The relevance of the data to a given usage.
Information Environment	The aggregate of individuals, groups, organizations, communities, technical systems, and information technology capabilities, that collect, process, share, disseminate, or act on information [data].
Information Sharing and Safeguarding Executive	The senior official charged with overseeing information sharing and safeguarding efforts for the 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.

ROLE/TERM	DEFINITION
Information Space	An aggregation of data, kept 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 [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.
Lineage	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.
Master Data	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 consistent and uniform set of identifiers and extended attributes that describe the core entities.
Master Data Management	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.
Metadata	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.
Mission Data	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.
Ontology	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.
Originating Element	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.
Pedigree	The description of the Data Quality (e.g., accuracy, correctness, completeness, and timeliness) assessment for data, its compliance with established standards, and the processing steps performed to derive the data. Pedigree information is used to augment Lineage.
Provenance	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.

ROLE/TERM	DEFINITION
Publicly Available Information	<p>Information [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 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 [data] 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>
Public Domain (Legal)	<p>Information [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.</p>
Public Domain (Non-Legal)	<p>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.</p>
Record (Information and Records Management Context)	<p>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.</p>
Reference Data	<p>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.</p>
Reference Data Management	<p>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.</p>
Semi-structured Data	<p>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) and other markup languages, 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)).</p>
Structured Data	<p>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.</p>

ROLE/TERM	DEFINITION
Support Data	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. 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.
Unevaluated Data	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.)
Unstructured Data	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.



0010346