MEMORANDUM FOR SECRETARIES OF THE MILITARY DEPARTMENTS
        (ATTN: CHIEF DATA OFFICERS)
        CHAIRMAN OF THE JOINT CHIEFS OF STAFF
        (ATTN: CHIEF DATA OFFICER)
        UNDER SECRETARIES OF DEFENSE
        COMMANDERS OF THE COMBATANT COMMANDS
        (ATTN: CHIEF DATA OFFICERS)
        DEFENSE AGENCIES AND FIELD ACTIVITES DIRECTORS
        (ATTN: CHIEF DATA OFFICER)

SUBJECT: Guidance on Designating Authoritative Data Sets

References: (a) Deputy Secretary of Defense Memorandum, Leadership to Advance Data
        Quality and Trustworthiness, dated September 15, 2022
        (b) DoD CIO [signed by the DoD CDO] Memorandum, Data Stewardship
        Guidebook, dated October 20, 2021

       As directed by the Deputy Secretary of Defense in reference (a), this memorandum provides interim guidance on the designation of an Authoritative Data Set (ADS), which will be reviewed and revised in one year to incorporate lessons learned as Components pilot associated processes. These lessons learned and the guidance below will be utilized to update existing DoD Instructions (DoDIs) (e.g., DoDI 8320.02 and DoDI 8320.03).

       The ADS designation is a function of data stewardship that must be executed across Components and at all echelons. Designation of a data set as authoritative for a specified analytic or operational purpose is a mission decision, which will be officially designated by Data Stewards, as defined in reference (b). Working with the mission leadership for the supported analytic or operational purpose, Data Stewards will leverage the attached guidance on designating an ADS, working with data producers to assess the associated data quality, scope, and purpose as close to the source system as possible, and documenting ADS designations in DoD's federated data catalog.

       My team is available to answer any questions or provide assistance as you use this guidance in your Component. Questions may be submitted to the DoD Chief Digital and Artificial Intelligence Officer's staff at osd.pentagon.dod-cio.mbx.cdo-council-exec-sec@mail.mil.

                                 Dr. Craig Martell

Attachment:
Authoritative Data Set (ADS) Guidance

AUTHORITATIVE DATA SET (ADS) GUIDANCE

PURPOSE. This document provides initial guidance on designating an Authoritative Data Set (ADS), including key elements that need to be specified for each ADS designation. An ADS designation is a mechanism or process to define the intended quality, scope, and purpose of data. (See below for definitions of terms and words used in this Guidance). ADS designations support the DoD Data Strategy goals of Visible, Accessible, Understandable, Linked, Trustworthy, Interoperable, and Secure (VAULTIS) - primarily the goals of Understandable and Trustworthy.

> For the purposes of this memorandum, an **authoritative data set** (ADS) is a data product that has one or more data elements that a data producer, working with their Data Steward, has identified as authoritative for a specific scope and purpose with defined and measured data quality. Multiple sources can be used to "construct" an ADS, and as aggregated, compiled or assembled, "creating" the ADS.

An ADS's scope is typically restricted to one or more specific missions (or communities) and an ADS's purpose is implied (e.g., lacking evidence) or explicit (e.g., metadata indicators attesting to aspects of data quality). The specific scope and purpose of an ADS is used to provide the consumer with insights on each data element and its appropriate use.

Data quality also underpins an ADS designation. Data producers trust their data, with inherent quality, because they understand how the data was generated or collected for a specific scope and purpose. The DoD Data Strategy recognizes the importance of trust and data quality in Objective 3 of Goal 4.5 Make Data Trustworthy: "DoD executes data quality management techniques to assess and enhance data quality." *__The ADS designation and supporting documentation extends the data producer's trust, alongside the other VAULTIS goals, to enable the data consumer's mission.__* Measuring data quality, identifying scope and purpose, for example, supports the consumption of, "quality data that is readily discoverable and understood within the context of its intended use," as outlined in the DoD's Data Strategy, Guiding Principles, 2.2.7 Data *Fit for Purpose*. The ADS designation and supporting documentation, therefore, should provide the data consumer with a similar level of confidence as the data producer as to the data sets' quality, scope, and purpose.

ADS DESIGNATION. This document provides initial guidance on designating an Authoritative Data Set (ADS), including key elements that should be specified for each ADS designation: data set (1) quality, (2) scope, and (3) purpose. Data Stewards will perform data quality, scope, and purpose assessments per ADS designation, using the below questionnaire as initial guidance. The Data Stewardship Guidebook will be updated to incorporate this rubric. Data Stewards will review and update data quality assessments as substantive changes are made to the ADS or at least annually, until such time that data quality measurement is automated. A change to the ADS's originally documented scope or purpose should also be captured in a reasonable time as those ADS changes occur. ADS designations are to be recorded within Department's federated data catalog.

1. DATA SET: ADS DATA QUALITY DIMENSIONS. Data quality (DQ) can be challenging to measure without first understanding why the data is being produced or consumed. Documenting the specific scope and purpose of an ADS is, therefore, critical to then measuring DQ along VAULTIS-related DQ dimensions. An initial list of DQ dimensions is provided in this section. Mission leadership should consider these suggested dimensions of

DQ, *as well as any other dimensions* they deem essential, and establish corresponding benchmarks. In order to justify the degree of DQ sufficient to support the specified scope and purpose, both the relative importance and minimum values for these VAULTIS-related DQ dimensions should be documented in the ADS designation. The following is an initial list of suggested DQ dimensions for Data Stewards to measure an ADS:

- Data values that closely and correctly describe the specified, mission or business, scope and purpose as of a point in time have a high degree of **accuracy**, which can also mean the data set has a low degree of mistakes or error while nearing exactness of measure to a well-defined or true value.
- Data sets that are present as of a point in time and can be measured at the data set, row, or column level are considered nearing **completeness**.
- Data sets that follow agreed upon internal policies, standards, procedures, and architectural requirements demonstrate a high degree of **conformance**.
- Degree to which data values are consistently represented within a data set and between data sets, and consistently associated across data sets; **consistency** can also refer to the size and composition of data sets between systems and across time.
- Assessment of **uniqueness** among key data values ensures that no specified value exists more than once within a defined domain (e.g., within a dataset).
- A data set's pedigree, measuring provenance, lineage, and alignment with all business rules is associated with a data set's **integrity**.
- Data sets that are both current as of the latest version and recent from the point of collection have a high degree of **timeliness**.

DATA QUALITY RUBRIC. The following DQ dimensions provide Data Stewards guidance on what to include in assessments per ADS designation. The below questionnaire is initial guidance and the Data Stewardship Guidebook will be updated to incorporate this rubric:

1.A. ACCURACY. Accuracy measures the structure, content, and variability of data. Structure ensures the data is aligned with the formatting of the rest of the data in its field. Content ensures that the values in the data set match the real-world truth of what the value is describing at that point in time. Limited variability ensures the repeatability of the measured value (i.e., how close we expect the measured value is to the actual value). Precision is often specified in terms of tolerance of accuracy, such as the allowable error between measured and true. Some common questions when assessing accuracy of a data set include:
   1. How frequently do values not align to their assigned format?
   2. How frequently do data values match ground truth?
   3. How is error measured? Is it tolerable for the specified purpose?
1.B. COMPLETENESS. Completeness is important for a data set to perform well in the following three contexts: proportion of data collected to the amount of data required and available to collect; having ample volume of data to support the specified purpose; and proportion of data containing null and invalid values. Completeness in data means consumers can readily determine the content, context, and applicability of the data without needing to estimate gaps in completeness. Some common questions when assessing completeness of a data set include:
   1. Is there known data that add to the dataset to make it more complete?
   2. Does the dataset contain sufficient breadth of information to contextualize the data for its purpose?

3. What fields in the data expect some null values? How often are null values present?

1.C. CONFORMITY. Conformity measures the degree that the data format follows the standards, policies, and procedures agreed upon for that data type. When assessing conformity, ask:
   1. Does the data's format match the specified standard?
   2. Is the data set architecture published and available?

1.D. CONSISTENCY. Consistency measures the degree to which a value does not vary across data sets. If data sets refer to the same real-world entity at the same points in time with different values for like observations, those data sets are inconsistent. When assessing consistency of a data set ask:
   1. Are there other datasets that reference values in this dataset?
   2. Are there discrepancies?

1.E. UNIQUENESS. Uniqueness ensures that there is a one-to-one alignment between each observed event and the record that describes such an event. Uniqueness is also a measure to disambiguate between two or more datasets. In the event of an ADS that has the same function in the same domain, only one should be considered the ADS. Common questions when assessing uniqueness of a data set, include:
   1. Are there other ADSs that serve the same function?
   2. Do I have duplicate records in my dataset?

1.F. INTEGRITY. Integrity measures the degree to which data can be measured as consistent with its provenance, pedigree, lineage, and conformance to business rules, policies, and standards. Data integrity is dependent on the data collection techniques, transmission of the data, cleaning of the data, processing of the data, and storage of the data. Some common questions when assessing integrity of a data set include:
   1. Are there opportunities for data to be tampered with, misreported, data degradation, corruption, data poisoning or otherwise altered during the collection, storage, processing, or transmission processes?
   2. Does the data cleaning process result in data that can be trusted?

1.G. TIMELINESS. Timeliness measures currency and recency. Currency measures whether the data is the most up-to-date version of the data available or the agreed upon update frequency. Recency measures the time between an event occurring and the data representing it becoming available for use. Some common questions when assessing timeliness of a data set include:
   1. How frequently do supported data consumers require updates?
   2. Does the data purpose require reduced latency?

2. ADS SCOPE. The ADS scope should be defined by the collection methods, organization producing the data set, standards or policies under which the data set was collected, and any applicable timeframe pertinent to the data sets existence. An ADS designation should include a domain identification statement, specifying the organizations, echelons, or functional domain which considers the set authoritative.

3. ADS PURPOSE. An ADS's purpose covers the intended and resultant use of the data set from the point of collection to the point of consumption. The purpose statement should

specify the guidance, direction, or requirement that was used to produce the data set and the authority with which that collection occurred.

ADS DESIGNATION. Data Stewards will document the ADS designation in coordination with the data producer and publish the designation in the DoD's federated data catalog, as available. The ADS designation should be predicated on consideration of the above listed dimensions of data quality, determination of corresponding thresholds, as well as identification of appropriate scope and purpose, in coordination with mission leadership responsible for the analytic or operational purpose. An ADS designation is a core component of data governance and the following table describes the ADS responsibilities belonging to each of the roles defined in the Data Stewardship Guidebook. As conflicts may occur over an ADS designation, a dispute resolution process will be updated in the Data Stewardship Guidebook.

Title	Executive Responsibilities	ADS Responsibilities
Chief Data Officer	Develop, promulgate, and oversee the implementation of data-related strategies, policies, standards, processes, and governance, including those measuring data maturity and quality	Track relative maturity of components to ensure core ADS responsibilities and expectations are implemented uniformly across department. Advocate for and seek senior level buy in for ADS designation
xData Officer [1]	Develop, manage, and oversee the implementation of data management policies.	Designate and provide guidance to data stewards. Assist the CDO with the development of ADS policies.
Data Steward	Establish protection, sharing, and governance guidelines	Provide the backbone of ADS nomination and maintenance. Data Stewards should nominate candidate data sets across their component, designate vetted data sets as ADS, and ensure existing ADS are maintaining current department standards
xData Manager [1]	Implement data management policies	Nominate candidate data sets, defining data purpose for nominated data sets, and vetting nominated data sets quality
Data Custodian	Perform mission and business data-related tasks	Perform the daily work of keeping data sets running and supporting users across the department, including ensuring ADS standard compliance on coordination with the xData Manager

---

[1] The "x" is used to provide flexibility in the naming of data officers. The DoD Component is given the flexibility to provide the appropriate term to describe how the DoD Component distributes the responsibility of the role.