DATA MANAGEMENT STANDARD

GEIA-859™	REV. B
Issued 2004-08 Revised 2022-08 Superseding GEIA859A	

(R) Data Management

RATIONALE

Revising document to include “what” language and moving the “how” language to the handbook revision, removing the “past” language to reflect “current” language, restructure of content and providing a framework to understand data management through the life cycle maturation.

NOTICE

This document contains a major rewrite. Content changed to reflect updates removing material to incorporate in the GEIA-HB-859 Implementation Guide for Data Management, language updates to current views, and clearer connection to the life cycle of data management.

FOREWORD

The scope of data management (DM) has grown significantly in the past decade—and this standard has played a significant role toward that progress. DM is a valued and critical partner in enterprise architecture design and application and has found a strong, meaningful, and acknowledged role with digital technology.

The identification, definition, preparation, control, archiving, and disposition of data all require dedicated resources, supporting systems, and time. High-quality DM makes certain that the enterprise ensures compliance, avoids rework and reduces cost. A well-designed DM process ensures that the enterprise and customers receive the correct data when required, in the form required, and of the appropriate quality.

The third release of this standard provides updates to reflect current processes, and scope of DM. The methods of DM have undergone significant changes as paper documents transitioned to digital data and continue to evolve.

Use this standard when establishing, performing, or evaluating DM processes, applicable to both electronic and non-electronic data, in both the commercial and government sectors. It introduces the data life cycle and the tasks associated with the management of data describing DM methods using neutral DM terminology. The associated Handbook GEIA-HB-859 contains “how-to” information for implementation of this standard in a variety of environments.

TO PLACE A DOCUMENT ORDER: **Tel: 877-606-7323 (inside USA and Canada)**
Tel: +1 724-776-4970 (outside USA)
Fax: 724-776-0790
Email: CustomerService@sae.org
SAE WEB ADDRESS: **http://www.sae.org**

For more information on this standard, visit
https://www.sae.org/standards/content/GEIA859B/

TABLE OF CONTENTS

1. INTRODUCTION

1.1 Scope

Data is information that has been recorded in a form or format convenient to move or process. It is important to distinguish between data and the format. The format is a structured way to record information, such as engineering drawings and other documents, software, pictures, maps, sound, and animation. Some formats are open source, others proprietary. Regardless of the format, there are three broad types of data. Table 1 lists these types of data and provides examples.

Table 1 - Types of data

Type	Examples
Product	Cost, schedule, and performance data. Scientific data such as written notes and observations . Engineering analysis, studies, tests, drawings, and models, parts catalogs, software application and their components, manufacturing operations, operational and maintenance instructions, and training materials. Services such as consulting, verification, auditing.
Business	Plans and schedules, financial information, software applications, inventory status, medical, facility, contracts, service level agreements, program and project information, administrative, and human resource information.
Transactional	Orders, issues, receipts, bills of lading, usage data, and invoices.

DM, from the perspective of this standard, consists of the disciplined processes and systems utilized to plan for, acquire, and provide management and oversight for product and product-related business data, consistent with requirements, throughout the product and data life cycles. Thus, this standard primarily addresses product data and the business data required for stakeholder collaboration extending through the supply chain during product acquisition and sustainment life cycle. This standard has broader application to business and transactional data recognizing the data addressed by this standard is subject to data administration, metadata management, records management, and other processes applied at the enterprise level.

Data has many purposes, including stating requirements, providing proof of achievement, and establishing a basis for long-term product support. Deliverable data (customer-accessible information) represents only a small fraction of the project data. In general, the developer/producer continues to own the intellectual property for a vast amount of design, development, fabrication, manufacturing and administrative data unless ownership has been transferred via contractual or other agreement. Further, the value of data is not limited to its use in support of a specific product: data may have a life cycle longer than that of the product it describes. For instance, data from previous projects forms part of the foundation for new product and process design. Data also supports the enterprise in process redesign and quality. Thus, data is essential to competitive position. An enterprise's data, if not properly safeguarded, can also be misused by a competitor to the competitor's advantage. For these reasons, data is an integral part of an enterprise's intellectual assets and overall enterprise knowledge.

1.2 Overview

Two different viewpoints, corresponding to product and data life cycles, are important to DM. The first perspective is the project level. Product data (and related business data) is normally acquired or created as part of the development of a new product or similar initiative. The second perspective is that of the enterprise—and the data that is required by enterprise-level functions.

This standard addresses the functions of DM, but not how to organize for DM. Each enterprise, for valid reasons, locates the functions of DM within enterprise elements that make sense within its own enterprise environment.

Do not use this standard as a compliance document or an evaluation mechanism for DM projects but as a source and reference document. Appropriate application of the functions and principles in this standard enables the user to plan and implement a DM program for a product, project, or enterprise.

1.3 Principles

This standard has eight fundamental DM principles (Figure 1). Principles are high-level statements describing high-quality DM and form a part of the DM life cycle. The degree to which the DM principles apply to a product varies over the product's life cycle. Similarly, they vary in applicability over the data life cycle. Some principles may not apply during every phase of either life cycle. The principles described in this standard also have broader application to business data and operational data generally.

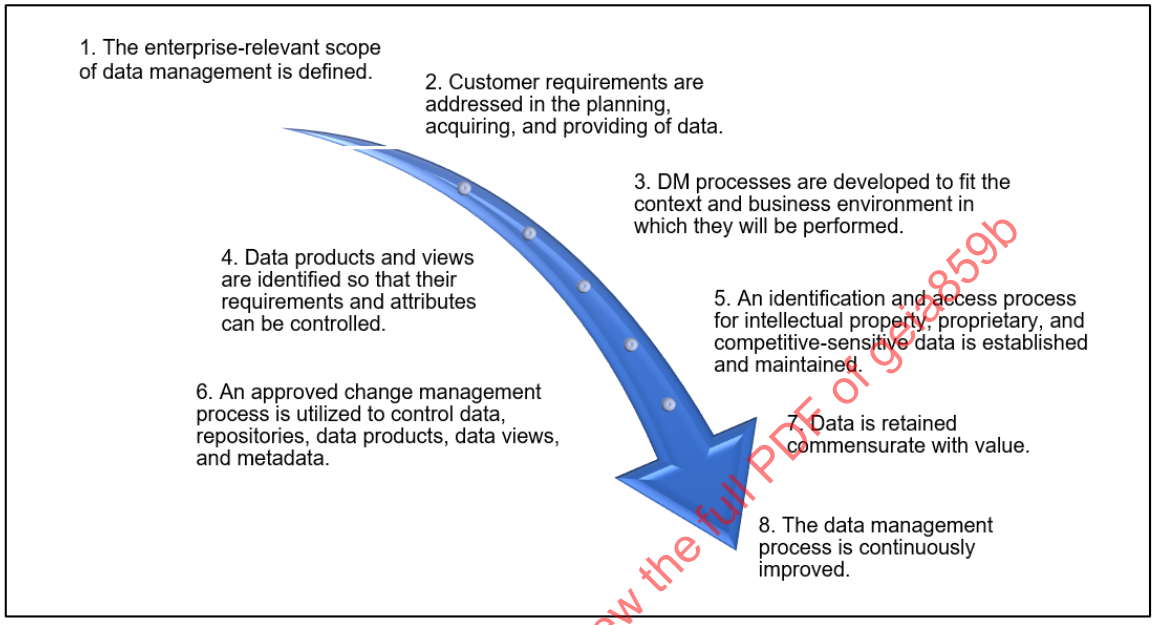

Figure 1 - Data management principles

The DM principles are part of a life cycle for DM (See Figure 2). The DM life cycle gives a structure to the principles to provide a map for implementation and or assessment of a project's needs. This Standard will follow the DM life cycle and incorporate the applicable principle.

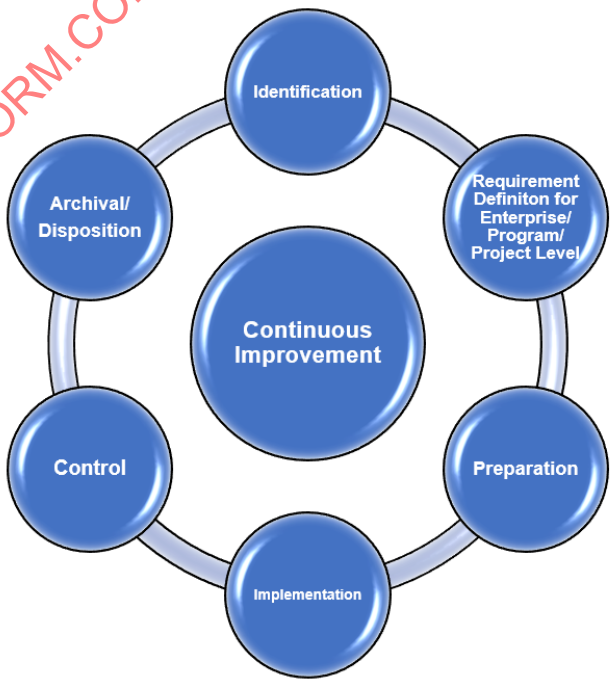

Figure 2 - DM life cycle

1.4 Terminology

During creation of this standard, neutral terms are used wherever possible. Please see the glossary (Appendix A) for the definition of those terms. When planning and documenting a DM program, substitute other aliases in place of the neutral terminology. The terms “program” and “project” vary in usage and scope throughout industry. This standard uses the term “project” for consistency.

Finally, for the purpose of this standard, any “enterprise” or “organization” will be referred to as “enterprise” and is responsible for performing DM. An enterprise may be a commercial business or a government agency. It may also be an organization resulting from a select collaboration between companies, a facility within a company, a program, a large project, or a social group. “The customer” refers to the activity that specifies requirements. A customer may be external to the developing and producing enterprise; may be an internal customer such as marketing, finance, supply chain, management, or the using department; or may even be a supplier in a conventional sense.

2. REFERENCES

2.1 Applicable Documents

The following publications form a part of this document to the extent specified herein. The latest issue of SAE publications shall apply. The applicable issue of other publications shall be the issue in effect on the date of the purchase order. In the event of conflict between the text of this document and references cited herein, the text of this document takes precedence. Nothing in this document, however, supersedes applicable laws and regulations unless a specific exemption has been obtained.

2.1.1 SAE Publications

Available from SAE International, 400 Commonwealth Drive, Warrendale, PA 15096-0001, Tel: 877-606-7323 (inside USA and Canada) or +1 724-776-4970 (outside USA), www.sae.org.

SAE EIA-649 Configuration Management Standard

SAE EIA-836 Configuration Management Data Exchange and Interoperability

2.2 Definitions

See Appendix A.

2.3 Abbreviations

CCB	Change Control Board
CD	Compact Disc
CDM	Configuration and Data Manager
CDRL	Contact Data Requirements List
CM	Configuration Management
COTS	Commercial Off-the-Shelf
CUI	Controlled Unclassified Information
DM	Data Manager/Data Management
DMF	Data Management Framework
EAR	Export Administration Regulation

EIDM	Enterprise Information and Data Management
GSCP	Government Security Classification Policy
IP	Intellectual Property
ISO	International Organization for Standardization
IT	Information Technology
ITAR	International Traffic in Arms
NDIA	National Defense Industrial Association
UK	United Kingdom
XML	Extensible Markup Language

3. DATA MANAGEMENT FRAMEWORK

The data management framework (DMF) establishes the foundation for the scope of DM, developing a structure to obtain and manage data artifacts. Principle 1—Define the enterprise relevant scope of data management—is fulfilled by the DMF. The DMF sets the policies and subsequent processes and procedures, drives and is constrained by funding decisions, and identifies oversight responsibilities to ensure compliance issues are satisfied.

A recommended framework is comprised of a five-level hierarchy as shown in Figure 3. At the foundation Level 1, a policy declares information and data to be valuable assets of the enterprise and managed accordingly. Level 2 of the framework focuses on strategic planning in order to effectively create the data roadmap and data concept of operations, perform a risk assessment, and develop the requirements for data (including security and protection). Level 3 defines the tools and domains needed to ensure data is available to the users based on Level 2 requirements. Level 4 ultimately defines and implements structure(s) with the tools put in place at Level 3. At Level 5, the framework establishes control and consistency of the data artifact while maintaining compliance and traceability. As referenced in Figure 3, this framework builds upon the foundation for control and consistency throughout the entire data life cycle while enforcing the compliance and traceability of all information products.

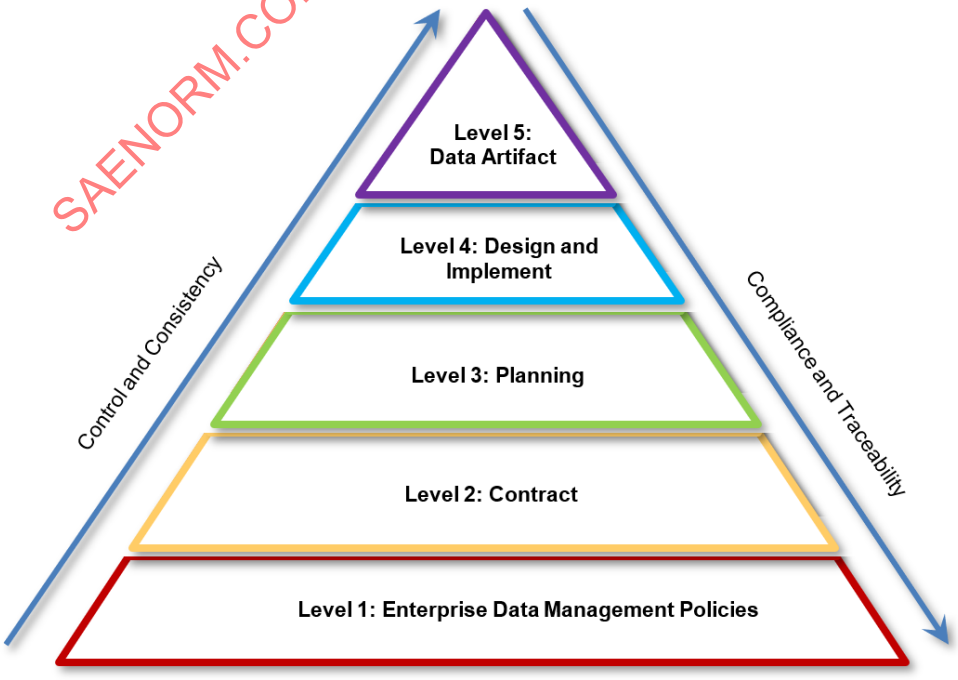

Figure 3 - Data management framework

Without an established, effective DMF in place, data may still be acquired, but not efficiently managed. Poorly managed data can have a significant impact on an enterprises' ability to sustain growth or minimize risk. The DMF avoids potential organizational conflict by providing a focus for consensus and agreement at the beginning of the project. For maximum organizational efficiency, establish the DMF early and at the highest practical level (e.g., enterprise level) in order to align strategy and to prevent information silos.

An organization's projects are dynamic in response to external events and changing conditions. Therefore, an effective DMF provides the structure to identify, assess, and respond to internal and external events and changes. The organization may modify the framework as necessary to accommodate these changing business environments and compliance regulations.

3.1 Level 1: Enterprise Data Management Policies

Policies, set by the enterprise-level authorities, establish primary decisions, guidance, rules, and actions within applicable restrictions. The policies relate to the foundational goals of the enterprise. The policies should be written at a high enough level to enforce, yet allow for flexibility according to project-level requirements. The policies provides management and regulation.

The enterprise data management policies or policies that address DM functionality throughout the life cycle of a project and facilitate the meeting of contractual, legal, and company requirements. The policies also support process verification and audit activities and provide overarching guidance to the organization.

- Determine what enterprise policies apply:
  - Does the enterprise have in place policies that define the authoritative source of truth for management of data?
  - Does the enterprise have in place policies related to retention and disposal of data?
  - Is there policy related to disaster planning and recovery, such as maintaining data backups in more than one physical location?
- Identify what external forces apply:
  - Has the customer requested compliance with national or international standards? If so, develop an understanding of the extent to which changes in the standards, some unforeseeable, create new requirements.
  - Similarly, has the enterprise adopted national or international standards that apply?
  - Are there either national or international legal requirements (e.g., with respect to intellectual property such as patents or copyrights) that have to be considered?

3.2 Level 2: Contract

Level 2 is the project contracts or authorizations which produce data artifact.

The term "contract" is intended to include formal contracts or agreements between two parties (with customer or supplier), internal work authorizations within a company, memoranda of agreement, and any other agreement that describes the duties to be performed. Data Management may apply to internally funded or contractual applications. A contract may also be provided through a standalone contractor, and more generally, as part of a larger contract for goods or services. The data requirements, for example, due dates, methods of delivery and markings, are specified in a contract.

3.3 Level 3: Planning

Level 3 of the framework focuses on planning at the project level to effectively create a concept of operations, perform an opportunity and risk assessment, and develop the requirements for data (including security and protection).

With a strategic business view of the current project, and in compliance with the enterprise policies, select a standard set of tools and processes that satisfy data management requirements.

Benefits of a standard set of tools and processes for use across the enterprise include the focusing of maintenance expenses on the common suite and/or the ability to move people to emerging work without having to retrain them.

A data strategy creates the potential for data-related decisions to contribute to both short- and long-term goals for the project and enterprise by aligning DM with the team environment.

In general, the assessment examines the internal strengths and weaknesses of the project and enterprise, as well as external opportunities, challenges, and risks.

3.3.1 Concept of Operations

Concept of operations describes how data management framework will be employed in daily operations. Determining the project requirements and ensuring enterprise processes and tools are in place and accessible to facilitate the management and/or transfer of data is essential. It is also important to understand what the future expectations are to the enterprise for the data that is developed so that enough assets are provided to meet the needs of both internal and external customers and the organization.

3.3.2 Opportunities and Risk Assessment

Understanding opportunities and challenges external to the project reveals how the external environment can affect DM. Opportunities promote the ability of the project to satisfy the project data requirements. Risks limit the ability of the project to satisfy the project data requirements. The risk management process continues throughout the life of the data management life cycle.

3.3.3 Develop the Requirements for Data

The requirements for data are specified within a contract or agreement. The outcome of the analysis determines when and how data requirements will be satisfied. At this time gaps will be identified and resolved, opportunities are recognized, and potential risks are minimized in order to meet contract/agreement requirements.

3.4 Level 4: Design and Implement

In this level of the framework, the data manager puts into practice the enterprise data management policies, planning assessment outcome, and contract requirements by defining and implementing structure(s) with the tools put in place at Level 3. The information products are produced in support of the project in such a way the data is accessible to the projects consumers as defined by levels 1-3 of the framework.

3.5 Level 5: Data Artifact

The successful execution of the four levels of the framework creates data artifacts throughout the maturation of its life cycle.

Part of the data manager's role is to identify data artifact such as the data dictionary, metadata (attributes of a data element, e.g file size, author) model, specifications, plans, technical data packages, project status, and financials.

The data life cycle encompasses all the following elements: identification, requirements, process, preparation, implementation, control, archival/disposition, and continuous improvement.

4. DATA LIFE CYCLE

Starting with Principle 2 the data requirements are capture. The following sections will identify principles in order and supporting materials defined at the project level.

4.1 Identification and Determination of the Needs for Data

The first step is to identify the data artifact from the contract and business needs that will be required to support the specific project throughout its entire life cycle. On a new project, the process begins with understanding necessary types of data artifact supporting the project life cycle. The data management framework enables the identification of the data artifact.

4.2 Determine the Characteristics of the DM Solution

Consider the broad characteristics of the DM solution and what it must address (but not how to implement the solution). A complete solution includes an analysis of all factors: internal, project specific, and external. Figure 4 illustrates the essential steps, which is a matter of applying standard systems engineering principles to the development of a DM solution. Although the process in Figure 4 is displayed as linear, it generally is iterative, especially requiring some cycling back and forth between development of alternatives and prioritization.

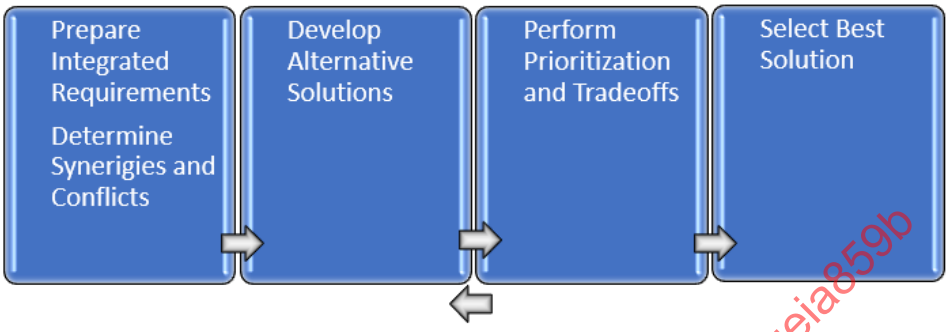

Figure 4 - Process for determining the characteristics of the DM solution

The project should develop a set of alternative solutions for the comprehensive set of requirements. Each solution is a scenario: a particular combination of processes, enterprise, and infrastructure elements. The processes, enterprise, and infrastructure elements do not need to already exist. In fact, it would be a mistake to consider only existing elements, because doing so almost certainly constrains improvement. In particular, do not limit the project with preexisting policies and practices. These policies and practices are essentially the result of previous decisions. Although valuable because they represent enterprise learning, they can also be counterproductive if that learning is not relevant to the problem being studied. Consider the costs (monetary, labor, time, and good will) and benefits of changing policy or practice during the alternatives analysis and evaluation process.

4.3 Determine the Data Management Requirements

Identify the general set of requirements to be addressed. This includes not only the requirements for data, but also the broader requirements that relate to data capabilities and data processes supporting Principle 3. To identify these broader requirements, it is important to understand, at a minimum, the intended use of the data, related business objectives, technology issues, and external constraints. The elements depicted in Figure 5 are a starting point for developing a high level set of requirements.

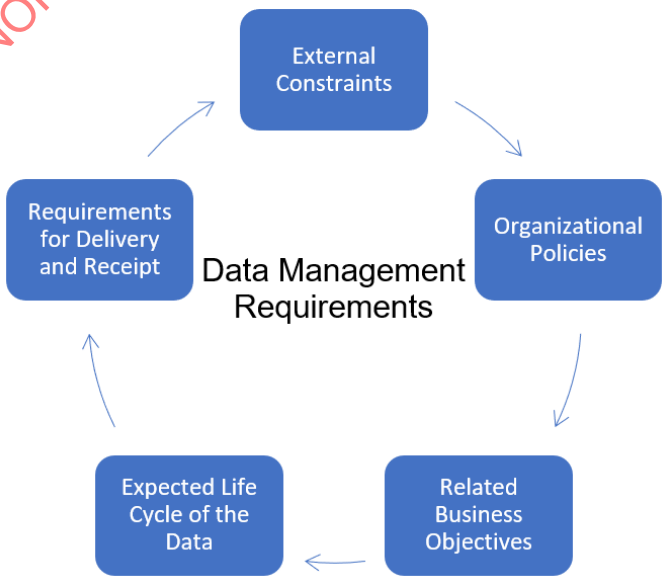

Figure 5 - Data requirements elements

As illustrated in Figure 5, among the essential considerations are the following, although this list is not exhaustive:

- Determine the expected life cycle of the data and expected use of the data. Is data being developed or acquired against a one-time requirement, or is there likely to be a recurring requirement for the data?

- Determine who will create or acquire the data. Many situations can apply such as: customer developed or acquired, developed or acquired for the customer, supplier developed or acquired, or collaboratively developed.

- Determine the expected requirements for access, delivery, maintenance, storage, protection, quality, and disposal over the life cycle. As an example, it is reasonable to expect that data created during the early design phases of a project will be important during later phases. As another example, it is important to determine if the customer is likely to require delivery or access.

- Determine who, over the life cycle, will have access to, be responsible for updating, and be responsible for disposal of data. Assess which of these cases is the most likely, and plan accordingly.

- Determine if there are related business objectives and considerations. The following are examples of questions that should be addressed for data being generated, delivered, or received:

    - Will the data potentially be reused or repurposed?

    - Is there any provision for ensuring the correctness of the data? If so, take this into account and plan for it from both a process and financial standpoint.

    - Is any of the data important from an intellectual property and/or trade secret standpoint?

    - Is there a method and process to ensure accurate data input by users driving consistency/quality?

    - Is there a method to assess quality of data interface, continual or one time?

    - If there is, or is likely to be, a requirement to provide for continued long-term access, what are the provisions for ensuring access if the enterprise ceases to exist in its current form (e.g., as a result of reorganization, a buyout, or a decision to abandon the line of business)?

    - Does the enterprise want to be in the business of long-term retention and delivery (for either electronic or non-electronic data) or will this responsibility be outsourced to a third party?

    - Is the data requirement indicative of an emerging market for the enterprise, a market the enterprise is maintaining, or a market the enterprise intends to withdraw from? The appropriate investments (time, infrastructure, acquisition of staff, training) are different for each case.

The project should assemble the requirements identified in a useful form for designing the DM solution. One way to do this is to list the requirements according to their source and relative priority. The following examples show potential requirements groupings:

- Current external customer contract requirements for specific deliverables or access.

- Current internal customer contract requirements for specific deliverables or access.

- Supplier requirements—i.e., requirements for data to be provided to suppliers rather than external or internal customers. An example might be a set of envelope drawings that a supplier needs in order to proceed with detailed design. Another example is sub tier supplier data receivables being provided to the customer via contract deliverables.

- Derived requirements to satisfy external or internal customer requirements. As an example, even though the contract with an external customer does not require formal logistics support analysis, logistics support analysis could determine sustainment requirements and support designing a supply chain.

- Anticipated near-term new contract or internal requirements for deliverables or access.
- Anticipated longer-term contract or internal requirements. Requirements imposed by enterprise policy and practice.
- Requirements, such as those found in the uniform commercial code or laws, which come from the environment outside the enterprise.

Normally, the requirements from the process just described are not fully independent of each other (Table 2).

Table 2 - Interdependent requirements

Case	Comments
Overlapping requirements	Important to detect and look for ways to combine requirements so they can be satisfied with a single rather than multiple efforts.
Conflicting requirements	Can be simple—such as needs for the same or similar data, but at different points in time. Can be more complex—such as competing needs for data sharing and protection of intellectual property rights.
Synergistic requirements	Example: The data created in response to one requirement, when integrated with or augmented by data created in response to a second requirement, provides a solution to a third requirement. Almost always introduces time dependencies; important to capture interdependencies in a process chart, network chart, or by similar suitable means.

The method for portraying integrated requirements varies from circumstance to circumstance. In some cases, a simple database is sufficient. In others, particularly where process and capability considerations are important, a narrative report may be required. A strong DM solution depends on a clear understanding of the relationships.

In some cases a contract will require a process that differs from existing solutions and an alternate solutions are considered to meet the contract.

4.4 Develop a List of Data Artifacts and Assign Data Requirements to the Project Areas Responsible for Data Generation and Distribution

The process begins by identifying the project area (or areas) responsible for generating each of the data artifacts. In the development of a data product, information may be required from multiple authors to provide sources in the area of final responsibility to finish the data product. In addition, clearly define and document the data requirements, including format, and marking requirements (see Principle 5). Communicate requirements to any internal stakeholders or outside subcontractor responsible for preparing a data product. Provide the project with the schedule and the supporting information, make sure they understand what is required, and secure a commitment. The data manager will manage the list of data artifact throughout the life cycle of the project.

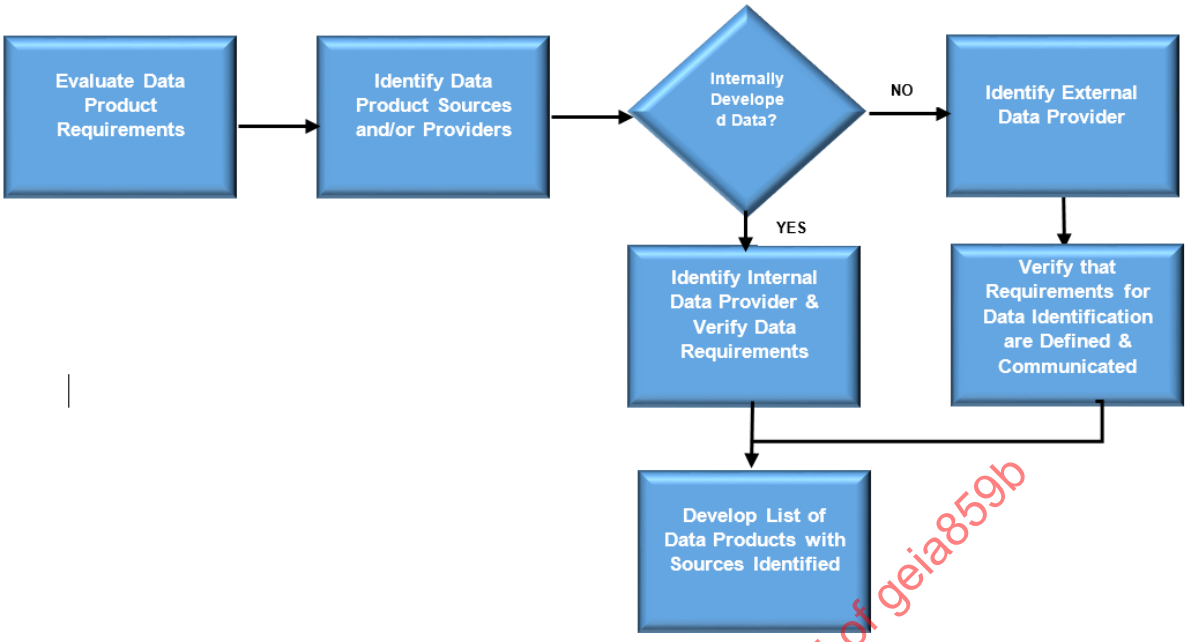

Figure 6 - Relate data requirements to the functional areas responsible for generating the data

4.5 Defining the Data Delivery Schedule

The schedule will be defined for external customers by the contract requirements, milestones, or events that drive internal need dates, based on data/hardware product and functional areas. The internal schedule consists of the required need dates, e.g., setbacks, events etc., for delivery between functional areas supporting on-time delivery.

The data manager is responsible for creation and management of schedule and communicating the forecast of the data product schedule to the stakeholders. Figure 7 shows the steps included in this process.

If the data is required to be updated, or is periodic in nature, determine the necessary frequency for any updates. The data manager will frequently review for contract and data requirement updates and change the schedule based on those factors.

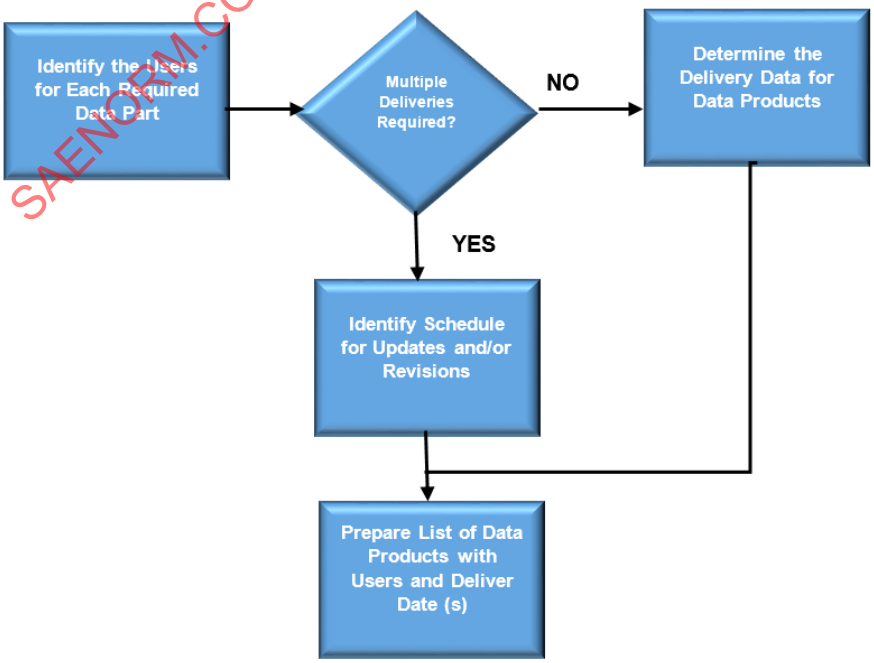

Figure 7 - Identify users of the data artifact and establish when data will be needed

Delivery of contracted data can take many forms, including the following:

- Providing delivery of data at a specified time or in conjunction with a specified event (by pushing data to the customer.)
- Data access under agreed-to provisions (for example, period of time, individuals who may access, purposes of access, and limitations). This approach does away with delivery, per se. “Delivery,” when it is needed at all, is effected by notification that the data is available to be accessed (by the customer pulling the data when notified).
- Deferred delivery: A method for delaying the delivery times for specified data. This approach is used when it is in the buyer’s interest, for example, when design is still evolving and what is desired is technical data that corresponds to the final design.
- Deferred ordering: A method to establish the right to obtain data that may be needed in the future but for which a specific requirement is not identified at the current time. This approach is used when a firm requirement for a particular data product has not been established before contract award but there is a potential need for the data.

5. PREPARATION

5.1 Value of Data Artifacts and Consistent Identification

Data provides value to the enterprise only if users can locate and access it. Maximize data usefulness by ensuring it can be discovered, retrieved, managed, and leveraged across organizational project and system boundaries. Data—including product definition, intellectual property, and other information—are vital assets of an enterprise.

Metadata, or information about documents and other data artifacts, is essential for identifying, cataloging, releasing, locating and retrieving data. Create and use a standard process for selecting metadata with a consistent, uniform, repeatable approach tailored to specific business requirements. This uniform process helps to minimize the time and cost associated with locating information. Adopting a standard approach across multiple projects or programs of an enterprise encourages data reuse and shared costs.

5.2 Data Identification

5.2.1 Process Map

Ensure data is properly managed throughout its life cycle by use of data identification supporting Principle 4. Create, protect, and retain data by selecting metadata for effective identification, manage, storage, and retrieval. Coordinate the process for selecting metadata with users or other stakeholders to ensure compatibility among organizations who will exchange data.

First, identify stakeholders and review their requirements to identify data to be developed or procured. Assure that the data structure includes the various affected projects and functions across your enterprise. Identify stakeholders from all affected functions who create, update, exchange, or enter data into a repository, or who search for data.

Provide consistent, harmonized methods for describing data. Use an organized approach to avoid confusion that arises from using redundant terms for one data element, such as “author” versus “person.” Create a systems architecture to help develop a controlled data taxonomy for use across multiple projects, functions, and systems. Identify data owners and users in the process, along with any requirements associated with metadata.

Processes must be consistent and will map the flow of data throughout the data life cycle.

5.2.2 Data Structure

Consistent use of metadata enable effective communications between enterprises exchanging data as well as within each enterprise. Standardize the process for establishing metadata, even if the types of data to be managed vary among enterprises and projects. Data managers should be aware of applicable standards for the development of metadata. Some examples are:

- The Dublin Core Metadata Initiative, ISO 15836 (https://www.iso.org) for developing interoperable online metadata standards for different purposes, is useful in different business settings.

- The World Wide Web Consortium works on metadata activities via its Semantic Web Activity (http://www.w3.org/2001/sw/).

- Domain-specific standards exist, so data managers should search for and be aware of the standards for the domains in which they work, such as the Content Standard for Digital Geospatial Metadata, and the Standard for the Exchange of Product Model Data (ISO 10303) (http://www.iso.org/).

Attributes are the properties that uniquely characterize the data, such as document number, title, maturity status, export regulation, and data type. A metadata record consists of a set of attributes necessary to describe the data. Consider also the relationships between artifacts and attributes. There may also be a need to create superior and subordinate relationships, or parent-child relationships, among data artifacts. Identify these relationships when selecting metadata attributes.

Although identification of attributes initially occurs during the early stages of planning, it repeats throughout the data life cycle. New types of data may evolve, requiring different ways of storing and retrieving. Changes to metadata should support all currently existing storage and retrieval approaches that are in use, while maintaining the integrity of existing attributes.

Consider the cost of creating, entering and maintaining metadata attributes, against the potential benefits. If users are required to complete numerous metadata entries when placing a document in a repository, documents may be entered with missing or erroneous entries or documents may not be entered into the repository at all. Evaluate potential attributes based on whether tracking and locating data would add value. Keep the set of required attributes as small as possible to enable a user to create simple descriptive records and provide for effective retrieval. Tailor existing metadata standards to meet the needs of the users and stakeholders.

Business rules for metadata must describe data consistently throughout the life cycle. Select attributes from a controlled vocabulary, a limited set of consistently used and carefully defined terms. Without basic terminology management, inconsistent metadata diminishes the quality of search results. Both field names and values benefit from standardization. Ideally, the controlled vocabulary is not project-specific but is created at the enterprise level in the form of a standard data dictionary, standard ontology, or similar means and applied consistently to all projects.

5.3 Establish Relevant Attributes to Refer to and Define Data

Catalog, store, and retrieve data dependent on the requirements of the data format. Electronic files are managed differently than hard-copy paper or other physical media, so the format characteristics should be considered when establishing attributes. File format, or the software application used to create or view the file, is relevant for retrieval of electronic files but not for data stored only in hard-copy format. The storage medium and file formats influence readability and reproducibility of the content. All data formats may have retrieval and readability limitations over time. Selection of attributes to support identification of storage medium is useful in planning for storage facilities. For example, identifying the file size of data to be stored electronically helps identify resource allocation.

5.4 Alignment of Projects with the Enterprise

Once metadata structure and processes are developed and tested, train users in using the metadata and templates to identify the data artifacts. The purpose, expected results, and business rules should be in place to assist users with developing standardized data artifacts. Allow for project adaptation of the enterprise controlled data vocabulary, to minimize proliferation of project-specific terms.

5.4.1 Involve Applicable Enterprise and Project Stakeholders to Ensure Data Interoperability

When selecting metadata attributes, the enterprise should identify relevant stakeholders who potentially create data, update data, exchange data, enter data into a repository, or search for data. Contact relevant stakeholders to obtain input and coordinate requirements. Alternatively, each stakeholder maps to a neutral standard. In any event, flow down and explain customer-invoked standards.

The enterprise should identify relationships and their importance relative to other data elements in order to efficiently identify and manage related objects. For example, the attribute for document revision may be related to another attribute for date of revision. Also consider relationships among artifacts and data. Create superior and subordinate relationships, or parent-child relationships, among data artifacts as necessary. As an example, a prime contract is divided up into multiple delivery orders with attributes for prime contract, delivery orders, and down to contract data requirements list (CDRL). Identify these relationships when selecting metadata attributes.

Identification methods, applicable to data, are typically identified at the enterprise level. Data and data requirements are defined and identified with unique identifiers. At the project level, document the identification methods if deviating from an enterprise policy or an enterprise policy does not exist.

5.5 Metadata Maintenance

Metadata attributes may change over time due to evolving requirements throughout the life cycle. These changes include changes to the data repository (e.g., facility or system upgrades) as well as obsolescence of data or of systems that can open it. Part of the overall DM process includes periodic reviews of metadata attributes.

5.6 Attributes

5.6.1 Standard Taxonomy

5.6.1.1 Uniqueness

Attributes of a data product serve to identify it for retrieval purposes. Uniquely identify each data product. Use multiple attributes when necessary to uniquely identify a data product, such as document number and revision level.

5.6.1.2 Allowable Values

Control allowable input values and formats within an attribute (e.g., list of valid values) to avoid unintended variations of a single value or data field. For example, use automation to collect attributes from the actual artifact.

5.6.1.3 Information Security

Data is controlled to preserve its integrity and to protect its value. Change management requirements described in Principle 6 should be adapted to the type and maturity of the information, and are necessary to guard the integrity of data from unauthorized alteration. Access control is used to protect the value and ownership of information from unintended disclosure beyond the boundaries of a program, a company or a country.

Topics in this section include determining data sensitivity, marking data, and controlling data. The need to control comes from one or more of the following: assertion of data rights, export controlled by government regulations, U.S. DoD classified, U.S. DoD controlled unclassified information (CUI), and need to know. Apply proper protections to data received, data created, and data distributed. Control access to data using appropriate attributes that identify requirements for special handling or access limitations. For more information, see Principle 5. Provide access to data only when attributes of the person (e.g., team, citizenship, license) are aligned with the attributes of the information.

5.6.2 Establish and Maintain a Management Process for Intellectual Property, Proprietary Information, Competition-Sensitive Data, and Data Rights

Intellectual property (IP) is a term used to describe real but intangible assets, embodied in such items as patents, copyrights, trademarks, and trade secrets. IP is at the center of an enterprise's competitive position and ultimately contributes to financial success.[1] Typically, intellectual property rights belong to the organization which invested in its creation or as documented in an agreement. For this reason, protection of IP is necessary to maintain an enterprise's competitiveness. In many cases, it is also necessary to comply with legal obligations to trading partners, including suppliers and customers.

[1] From a process standpoint, protection of classified data and protection of intellectual property are more alike than different. Here the emphasis is on intellectual property. The rules for management of classified data can be found in applicable government documents.

IP assets come from a variety of sources. In addition to internally developed data, IP is received from suppliers, subcontractors, trading partners and customers. All of this data is identified and tracked for protection based on data rights. The management process for IP and its relationship with other DM principles follows the steps:

1. Establish process for access levels and definitions.
2. Identify data source (vendor, subcontractor, trading partner, customer, internally developed).
3. Utilize established IP identification method.
4. Utilize established IP control method.
5. Determine protection obligations and requirements.
6. Ensure marking compliance.
7. Validate data security and access protection.
8. Disposition data (see Principle 7).

Determine how IP is managed depending on the rights obtained from the provider through documented agreements, such as statements of work, license agreements, and contract negotiations. These documents also define the ability to deliver the information obtained from a supplier to a third party, as well as the obligations and requirements to limit access and use.

Enterprise policies for IP management provide a standardized way to categorize, mark, and identify the information; control and track ownership; manage rights to use and sell; control access; distribute; and dispose of IP within the enterprise.

In some instances, an enterprise needs to sell, purchase, or license IP for purposes such as establishing standards, developing business relationships, creating new and larger markets, and realizing strategic goals. Under certain circumstances it may even be in the best interest of the enterprise to place IP in the public domain. Regardless of the reasons, these transfers should take place under predetermined conditions and be carefully controlled, to protect the rights of the data originators and owners. Regardless of the category or source of IP, it should be managed as an asset of the enterprise. Failure to successfully manage IP can have personal, enterprise, national, and international implications.

There are several types of proprietary data. Some examples include:

- General business information (not available to the general public).
- Financial and other information to be used only within the enterprise (internal use only).
- Information developed by the enterprise that has monetary value (enterprise proprietary information).
- Enterprise-developed information that has been officially registered with a legal authority (registered proprietary information).

The enterprise should evaluate internally developed and funded data upon creation and upon update to assess the viability as a patent, trademark, or copyright. If appropriate, patents and trademarks should be registered with the appropriate government agency, such as U.S. Patent and Trademark Office (http://www.uspto.gov/), and copyrights should be similarly registered, such as with the U.S. Copyright Office (https://www.copyright.gov/).

Data not specifically categorized as IP, but which might be construed as providing an enterprise advantage, is considered to be competition-sensitive data.

Carefully review asserted data rights, whether sending information to a customer or supplier, or receiving information from them. Once data rights have been negotiated and agreed by affected parties, data must be controlled in accordance with those legal rights. Use audits of access and evidence of control to ensure compliance.

Changes to the data may or may not impact documented agreements for data rights. The enterprise should review documented agreements to assess the impact of the change. Areas of particular concern exist where the right to use the updated item is not part of the original agreement. In those instances, new agreements must be negotiated. Establish review and disposition methods for IP changes based on the business needs.

5.6.3 Establish and Maintain a Process for Export Controlled and Government Classified Data

In addition to the data protection requirements above, which are subject to civil penalties, government data security breaches carry criminal penalties or could result in loss of life on the battlefield. Controls can be imposed by government agencies, for example U.S. Commerce Department Export Administration Regulation (EAR), U.S. State Department International Traffic in Arms (ITAR), U.S. Department of Defense Classified, U.S. DoD Controlled Unclassified Information (CUI), UK Government Security Classifications Policy (GSCP), other government security documents.

5.6.4 Establish and Maintain a Change Management Process for Reviewing and Reducing Controls

Data is controlled so that changes to data, including its sensitivity, are reviewed and authorized by the appropriate governance body and results are provided on a need-to-know basis. Because data protection carries a cost to the enterprise, change management must include reducing controls when appropriate and allowed. Details of the change process are defined in presented in section 6.

Over time, rights to data may expire or may lose value to the enterprise. There should be an enterprise retention policy, because there is often a legal or contractual obligation to maintain the data. The enterprise therefore must retain the IP information, including all documented agreements that define the data rights. Principle 7 provides guidelines for data retention and storage.

Validate the security of the data periodically as part of an audit process. An audit can address the following, among other things:

- IP is properly identified by type and source.
- IP rights granted are current and enforced.
- IP distribution is reviewed.
- IP disposition schedules and methods are followed.
- Data is properly marked and tracked.
- User access rights are reviewed.

The enterprise should review the types and varieties of controlled data that are to be addressed, and should create a method of controlling access and distribution. Regardless of the type of environment, manage data through markings and enforced controls such as limited access facilities and role-based access control.

Depending on the variety of information to be protected (e.g., IP, pricing, or classified data) it may be useful to prepare guidance in the form of a decision tree that leads employees through the government regulations and business rules that must be followed to properly categorize and control data artifacts.

When information is provided to a person who is a citizen of a different country, export control requirements need to be satisfied prior to disclosure. This includes printed, electronic, or verbal disclosure of information.

5.6.5 Ensure Markings are in Compliance with Conventions and Requirements

Once IP, government classification, export jurisdiction and control code, etc., has been identified, it should be marked appropriate to its type or variety, both as metadata and on the document. Information provided to a government entity is marked using government notices or legends. Disclosure of proprietary information in any other context requires an agreement establishing the limits on disclosure.

To summarize information security:

- Data sensitivity must be determined, marked and registered.
- Asserted rights of data exchanged must be respected.
- Secure data appropriate to its sensitivity.
- Validate data security.

5.7 Data Delivery, Distribution, and Access

Preparing a data distribution package can include identifying the data requirement and due date being satisfied, one or more data artifacts, and an intended distribution list. For data delivered to military customers, identifying the data rights or assertions may be appropriate, based on negotiated agreements. Distribution includes both incoming and outgoing data artifacts.

There are several means of providing data. One way is delivery of data to the customer, another is providing the customer access to the data. Access privileges are governed by rules and roles, whereas delivery is an event.

When access is provided, an authorized user can retrieve data that has been organized to meet their needs, which is referred to in this standard as a “data view” and may include one or more files. It is important to define and control the metadata, particularly where the data views are complex and when it is important to ensure that the same view is provided each time it is accessed. Access expiration should be per documented agreements.

The enterprise should document the access constraints for the various types and varieties of data. Once this process is defined, it can be applied to all sources of data at all levels of the enterprise. A project may document further constraints, as required.

5.7.1 Process Health Metrics

Consider the requirements for tracking and reporting metrics when selecting attributes. Metrics are typically used to monitor throughput, to ensure the process is operating as intended (e.g., timeliness, quality), and resources are properly allocated. Use predictive or forward-looking metrics (e.g., forecast report) to avoid missing requirements. For more information, see Principle 8.

5.8 Data Migration

When historical data must be migrated from a legacy system to a new or updated data management solution, mapping of equivalent fields is necessary to retain data relationships. To maximize the benefit of migrating historical data into a new data structure, use the opportunity to clean up the metadata to eliminate unintended variations. Cost and benefit analysis will determine if migration of all the data and shutting down the originating system is warranted or if the originating system will be maintained.

When making changes to attributes, the enterprise should consider the impact on legacy data. In a large repository, it may not be feasible to update the metadata of existing data collections, and it may be necessary to develop translation tables or similar mechanisms.

5.9 Train and Certify Users

A mandatory training plan for data managers, authors and data consumers will support compliance with enterprise policies and procedures. The enterprise should have representation on the SAE EIDM Committee.

Recurring training should provide refresher and updated direction.

There is an NDIA Certified Configuration and Data Manager (CDM) course and exam available. Some universities and colleges have data management classes. Companies may develop their own training, including usage of systems, within your SAE enterprise.

6. IMPLEMENTATION

Earlier, in the requirement definition for enterprise/program/project level and preparation phases, the organization determined what processes, policies, and changes in organization and infrastructure are needed for successful DM operations. Those DM requirements were constrained by the factors shown in Figure 8.

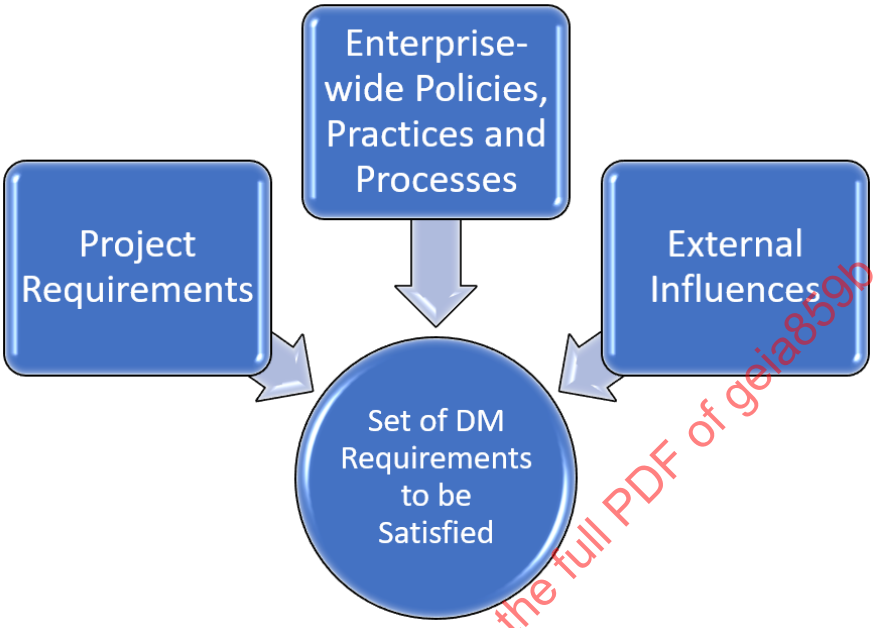

Figure 8 - Factors in determining DM requirements

The enterprise-wide policies, practices, and processes will include enterprise standards with which the data management solution must comply. For example, some enterprises have adopted a standard software solution for DM. In that case, the data manager must adapt either the DM processes to fit the software, and/or configure the standard software to support their DM processes. In the final design of the DM solution, the data manager must determine the best approach to meeting DM requirements while complying with enterprise standards.

Project requirements are a major driver for DM requirements. For example, a customer may contract for the delivery of product data as a three-dimensional model in a particular format. In that case, the data manager must restrict DM software choices to those that can export data in the format the customer required, with no loss of data integrity. External influences, such as a regulatory change, can contribute requirements for extra steps and metadata in the management of product data.

The set of DM requirements drives the design of the DM system—the processes, roles and responsibilities, staffing, and tools to manage the project's data. The implementation of the DM system is based on Level 4 of the DMF. In this phase, the DM staff works with the information technology (IT) group to determine how to satisfy the DM requirements. There are many options for this, from using software already in use elsewhere in the enterprise, to obtaining commercial off-the-shelf (COTS) software and configuring it to meet requirements, to developing custom software. Depending on the option most likely to be used, the design activities will differ.

Once a new tool has been selected, data is entered into the system via some combination of data conversion, import, or manual entry. Once the new system is fully operational, the prior system is disabled and old data is dispositioned based on business decisions. The data may be migrated, archived or deleted. If data is to be archived considerations for obsolete tools and hardware need to be addressed along with security requirements.

Once the system is implemented and functioning smoothly, it should be in operations/maintenance mode, and the continuous improvement activities should begin. These topics are addressed in Section 9.

6.1 Maintain Data Assets and an Index of Enterprise Data Assets

To ensure that data assets are well identified, use appropriate metadata—identifying data pertinent to the items—to enable retrieval of those assets. Metadata may include date, contract number, author, title, general topic key words, owning enterprise document number, document serialization, default retention date, and data steward. Effectively identified data assets enable timely retrieval and destruction as appropriate.

When evaluating how to store data, the following questions may assist with decision making:

- Will the data ever be used as direct source information in the creation of new material?
- How long will this electronic data likely have value to the enterprise? If valuable for a long period of time, electronic data migration might be considered more heavily than if its value is for a short duration.
    - To mitigate electronic data loss due to shelf-life limitations of storage media, the enterprise should perform periodic refreshes and data validation.
- How long will this paper data likely have value to the enterprise, and how frequently might it be accessed? If a long duration, electronic conversion may be warranted.
    - Hard-copy data, while not as susceptible to some of the issues electronic media face, has its own vulnerabilities.
- How difficult will it be to retain compatible (and potentially obsolescent) equipment and software in order to provide for irretrievability and readability of the files over their anticipated lives?
    - Use of neutral formats to store data will generally simplify life cycle storage issues with long-term preservation.
- Is duplicate data needed?
    - To minimize retention of duplicate media, the enterprise should review duplicate data assets and determine the need to store multiple media beyond that needed for disaster recovery.

To protect data assets from unauthorized viewing, the enterprise should place security requirements on data assets as appropriate (Principle 5). Enterprise data assets may be of a proprietary, government classified, or other sensitive nature, warranting protection against unauthorized viewing.

6.2 Assess the Current and Potential Future Value of Data

The frequency for review of holdings can be easily managed in an electronic environment. Many enterprises have computer reporting of potentially obsolete data by virtue of a metadata sort, for example, date-driven reports that flag original metadata entries as suggesting they are obsolete. At such time, the enterprise reevaluates the continued value of such data. Upon reevaluation, the enterprise should readjust default retention dates or, if the data is determined to be non-value added, should arrange for disposal. When evaluating the current and future value of data, some questions that may be asked are as follows:

- What data is currently stored and for what purpose?
- Is the data accurate and up-to-date?
- What are the criteria for the storage life of data?
- What are the costs associated with data retention?
- Is there potential for this data to satisfy contractual or other legal requirements (see section 7.5)?
    - Conversely, Is there data being stored past its policy or contractual lifespan that could jeopardize the enterprise's position in a legal matter?

7. CONTROL

7.1 Introduction

This section provides guidance that will ensure the integrity and timeliness of data, data elements, data structures, and data views by adapting the principles of configuration management (CM) that are described in EIA-649.

DM and CM are two disciplines critical to the success of any project. They are strongly related and interwoven in their scope, application, and elements. Both are disciplines whose ultimate purpose is ensuring the accuracy of the products they support. One of the functions of each of these disciplines is to control change or, for some types of data, to protect the data from change. Not all data requires formal change management, or the same level of control; it is a matter of balancing cost and benefits. This section addresses Principle 6 and the body of data for which some level of control is appropriate.

An important factor to consider is when a data product is ready to be placed under formal data management control. This process implies a transfer of control (or stewardship) from the author or originating integrated product team to the data management control process.

The data product needs to be in a state of maturity that makes control both meaningful and intentional for the scope of the project. Determining when a data product should come under the data management control involves the evaluation of several factors: the completeness of the data product to fulfill its end use, whether it is ready (or needed to assert IP, for example) to fulfill that intended use and project constraints like schedule and need. It is at this point that the author (or originating team) delivers the data product for use. Consideration of the following factors is essential:

- The format and media are in alignment with the requirements.
- The data is accurate and at an appropriate level of completeness.
- The timing of the transfer is required against the data product's intended use (too early is just as critical as too late).
- The data product has been reviewed and internally approved by recognized level of authority (e.g., engineering manager, IPT lead). DM receives the control-ready product from predetermined sources.
- In large programs/projects, it may be necessary to develop a formal process to cover this transfer as part of the general change management process. This would be important when dealing with projects creating IP.

The change management functions and principles defined in EIA-649 are appropriate for DM. While the CM process is defined for the control of product data, it is valid for the formal control of data assets as well. It provides orderly discipline for change and also provides the traceability required to maintain the history of the approved revisions of the data assets. Knowledge of the change history over the life of the product is critical to reconstruct the exact configuration of the data asset. Additionally, there may be contractual or regulatory requirements that demand this traceability.

For these reasons, what is discussed here borrows heavily from CM and describes how the change management process applies within DM.

The levels of control, which can be formal or informal, are defined by the requirement, and must be identified and communicated at the beginning of a project. The following discussion applies to data under formal change management.

7.2 Determine the Impact of Change to Include Associated Products, Data, Data Elements, Data Structures, and Data Views

The selection of the reviewers is critical to the successful accomplishment of the impact assessment. Reviewers are subject matter experts who will use a set of criteria by which they perform the assessment. The criteria will cover cost, schedule and performance within the technical area associated with the proposed change.

The impact assessment should include all areas affected by the change by documenting the extent and significance of the impact. Areas to consider include:

- Requirements (for example interfaces, data usage, metadata within automation systems).
- Specifications.
- Customer-furnished information.
- Supplier data.
- Contract.
- Product performance.
- Cost and schedule.

The impact of a change can be significant, but obtaining a precise statement of the impact and its potential consequences may be difficult, reinforcing the need for careful selection of reviewers.

The review process entails ensuring balance between thoroughness and the timeliness of the review. Once this review is conducted and any impacts (and their potential consequences) are determined, create a concise written statement and forward it to the designated approval authority for action.

7.3 Approve/Disapprove Change

After completing the processes for conducting reviews of proposed changes, the process for determining change disposition follows logically. The formality of this process (e.g., a structured CCB or a single person authority) is dependent on the size, scope, and contractual requirements placed on the DM process.

The change authority dispositions the change in one of three ways:

- Approve the change and forward the change to the proper authority for implementation and ultimate closeout.
- Disapprove the change. In this case, note the disapproval in the CCB record and notify the change originator/sponsor.
- Defer the change. This could occur for a variety of reasons. For example, additional information may be required to make an informed decision, there may be a flaw in the supporting documentation submitted, or there are unresolved funding issues. Return the change to the originator or sponsor for further correction, strengthening, or clarification.

Regardless of the disposition, a notification is prepared and issued to those affected by or otherwise interested in the disposition. It is good practice to document not only the disposition but also the position (for or against, with reasoning) of each party participating in the decision making process. The notification also provides essential information for updating status accounting records, the “official” records of change dispositions.

7.4 Control the Integrity of Data, Data Elements, Data Structures, and Data Views

DM ensures that data artifact satisfy requirements. Doing so, in part, requires that the integrity of the data artifact (see Principle 2 for definition) and associated data elements are maintained using a consistent change management process and those changes are traceable and baselined and are approved by an authorized approval authority.

Data retains value commensurate with its accuracy, timeliness, and relevancy to the business. The value added by the DM processes is the preservation of this worth. Relevancy of the data will vary throughout the life cycle of the project. Data can vary in maturity and therefore importance. Data within a project undergoes a life cycle from origination, iterative development of working data through maturity, e.g., created, submitted, approved, delivered, and archived data. At each of these stages, data possesses different levels of value and importance. Recognition of these relative stages is important with regard to the level of control that is imposed on the data. Some data is placed under change management at project inception, some is considered for formal control at a later date, and some may never be brought under any formal control.

A given project may also require different levels of control, depending on life cycle stage, customer imposed requirements, enterprise requirements, or maturity of the project. Make decisions early in the process outlining which data "objects" (products, views, queries, etc.) will require configuration control. It is also important to consider when control needs to be imposed, whether changes to data assets need to be traceable and baselined, and therefore what level of control is necessary.

Determination of the level of change management is a policy decision that should be driven by an information governance system (See Principle 1). The over-application of controls is just as ineffective as too little control, and the application of control too soon is as ineffective as too late. A change management process ensures efficient and effective processing of change requests without impeding design development, production, or operational readiness.

7.5 Establish a Change management process that Imposes the Appropriate Level of Review and Approval

Data management control of data within a project is as important as configuration management control of the product's design. In fact, accurate, controlled data is an essential support requirement for control of the product design. Figure 9 represents a basic change management process which can be tailored to meet the requirements of a particular project.

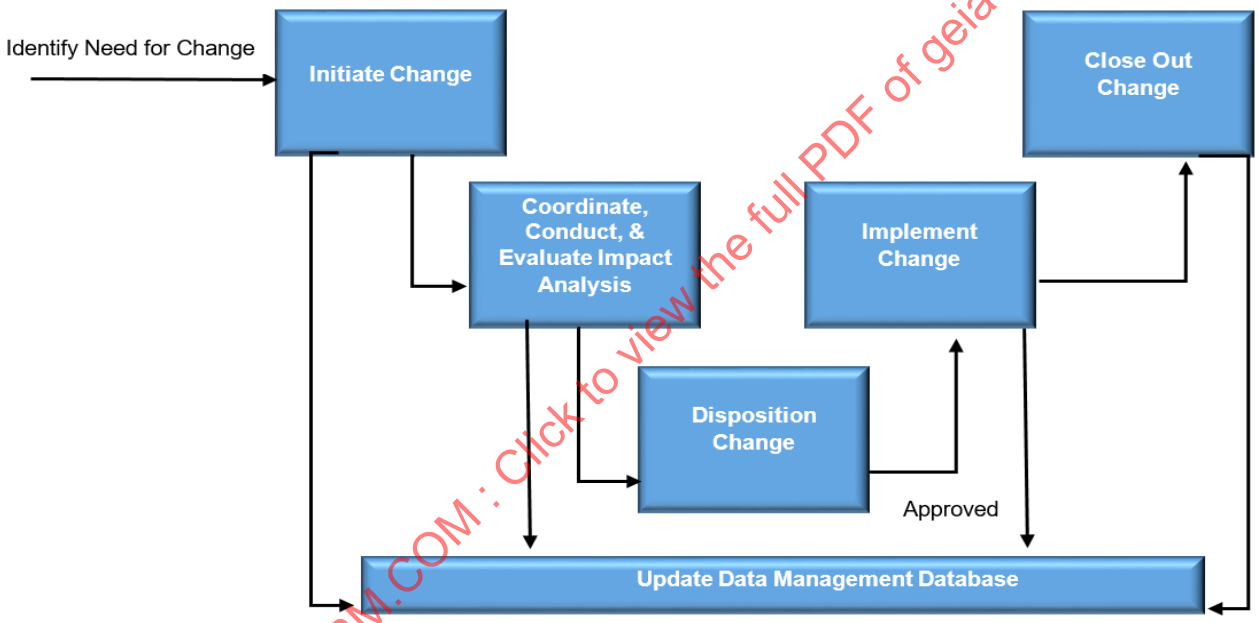

Figure 9 - Example change management process

7.6 Provide a Systematic Review of Proposed Changes within the Change Process

One of the most important reasons for an organized change management process is the thorough review that is applied to a proposed change. This process provides critical information for the status accounting of the change, the change history, and the ultimate disposition of the change. Although it is not always necessary to exactly mimic the CM change process, the basic principles of good CM should be adopted.

The change authority plays a vital role in the change management process. This authority evaluates requests for change based on information developed as a result of an administrative and technical review to approve (and implement) the change, disapprove the change, defer the change, or return the change request to the originator for rework. This change authority may take the form of a single project-designated individual or be implemented as a more formal Change Control Board (CCB).

7.7 Management of Data Objects and Artifacts

Data and other project-related information should be accounted for and tracked. This data may include items delivered to the contractor by subcontractors, items delivered to a customer from a contractor, project plan data, performance and financial data, project review charts and more. Recording metadata relative to a project and its owners whether they be people or companies can be helpful when categorizing data, conducting data mining activities and for reporting. Examples of the types of metadata that may be recorded are listed below (for more examples, refer to GEIA-HB-859):

- Unique identifier.
- Title of data product.
- Source of the data product.
- Date created.
- Date delivered.
- Author.
- Contractual due date.
- Contract (or subcontract) reference.
- Media type (hard copy, disk, CD, file).
- Destination (who received a copy or who was notified of receipt).
- Storage location of original.
- Security classification (if applicable).
- Export/import information (if applicable).
- Data markings (ex. contractual, policy, and distribution).
- Record of authority (if applicable).

The process for maintaining metadata, pictured in Table 3, defines how the metadata is generated, gathered, and populated into the database. The process largely defines the information that is stored in the database and sets the database requirements.

Project requirements, availability of technical database expertise at affordable cost, and reuse opportunities drive the architecture of the database. These elements have a bearing on whether the database is standalone, network centric, internet centric (web-based), a simple flat file, relational, homegrown, or a commercial product. Table 3 lists examples of the types of functions that should be considered.

Table 3 - Example elements of database functionality

Administrative Functions	User Interfaces (UI)	Data Relationships and Functions
Database security User permissions Administrative overrides Administrative rights	Layouts Sign-on View Input Output Reports Ad hoc User generated	Data fields Data formats Data relationships Metadata requirements Metadata definitions Search mechanisms Search criteria

7.8 Establish and Maintain an Internal Validation Mechanism for Metadata

Several key validations are required within the DM process. The metadata creation process itself, the data stored in a repository (if used), and the data contained in the change database.

The validation process is best accomplished by the element of the enterprise charged with DM. There are several reasons for this:

- The self-validation process helps to build lessons learned in a more meaningful fashion when conducted by the process owners.
- The enterprise element charged with DM is most familiar with the DM processes.
- Correction of deficiencies is normally more expedient when responsibility is centralized.
- Self-examinations can be conducted prior to, or in conjunction with, formal quality validations, thus expediting and adding integrity to a formal validation.

Integrity of data in a repository needs to be maintained. A self-validation can assess the completeness and uniqueness of the data product within the repository, adequacy of metadata, and similar essential characteristics. The validation should also ensure that each data product is appropriate for continued storage and retrieval. Validations are a convenient time to reassess the continued value of data and to dispose of or archive it if appropriate. The validation should also examine the adequacy of protections for intellectual property.

7.9 Establish and Maintain an Effective Data Control Process

7.9.1 Establish and Maintain Control Methods

Processes should exist that establish control methods for enterprise data. Data is controlled so that changes to data are reviewed and authorized by the appropriate personnel and results are provided on a need-to-know basis. Details of the change process are defined in Principle 6. For IP, control systems are different based on owners and use of data and include appropriate approval methods and updated documented agreements for data rights. This provides another layer of control for IP to ensure that the data is handled in accordance with IP policies and legal obligations.

The enterprise should evaluate internally developed and funded data to assess the impact of the change on a patent, trademark, or copyright. If appropriate, reregister patents and trademarks with the U.S. Patent and Trademark Office (http://www.uspto.gov/), and re-register copyrights with the U.S. Copyright Office (http://www.copyright.gov/) or applicable governing office.

Changes to the data may or may not impact documented agreements for data rights. Areas of particular concern exist where the right to use the updated item is not part of the original agreement. In those instances, new agreements must be negotiated. Establish review and disposition methods for IP changes based on the business needs.

7.9.2 Establish Mechanisms for Tracking and Determining Status of Data

The mechanism for tracking IP continues when tracking changes to IP. When changes occur, the ability to trace users of IP data assists in determining the distribution for approved updates. As with other IP issues, changes need to be tracked and the data rights reviewed before distribution.

At some point, rights to data expire or are no longer of value to the enterprise. There should be an enterprise retention policy, because there is often a legal obligation to maintain the data. The enterprise therefore must retain the IP information, including any and all documented agreements that define the data rights. Principle 7 provides guidelines for data retention and storage.

8. ARCHIVAL AND DISPOSAL

8.1 Retain Data Commensurate with Value

The purpose of this principle is to define methods for ensuring adequate retention and preservation of data assets which are valuable to the enterprise.

Retain any data assets that are of potential business, project, operational, or historic value until the information is no longer needed. Data of sustained value to the enterprise should be retained and evaluated on an ongoing basis. Timeliness of the decision-making process with value-added data decreases the company's risk. The right data at the right time is cost effective and enables decision making and business processes. Non-value-added data should be removed from the enterprise's inventory.

Active data needs to be readily available to the enterprise for regular reference. Inactive data still retains value but is not considered to have regular, continuing need. Such data needs to be retrievable, but can be stored less centrally. The data retention program should include a periodic review to determine whether data is no longer being actively used, at which time it can be classified as inactive and can be moved to an archival location. Relocation of such inactive data frees up physical and digital storage space. Supplier off-site storage may be an option for storing historical hard copy that is space consumptive.

8.2 Disaster Preparedness

Supplier off-site storage may be an option for storing historical copies of data assets. If additional copies of these archived files are available, storage of the additional copies at a separate location in the local area or even in a different region of the country is prudent to overcome risk of local disasters such as hurricanes, tornados, or earthquakes. Conversely, if these archived files have no second copy, they are likely of greater value to the enterprise with local storage, because it facilitates faster retrieval.

8.3 Plan to Ensure Data Are Available to Meet Future Needs

Data assets should be declared a record and have planned retention requirements identified and documented. It is important to ensure record keeping processes cover archive formats, frequency of reviews, purge planning, disposition funding, and related activities. Clearly defined methods for data retention help ensure that the data will be available when and if needed. One such method is to develop an enterprise policy on data retention.

Effective control of data is best accomplished through defined process ownership and accountability. To ensure proper planning for eventual disposition of data assets, establish disposition dates early in the data life cycle. Manage physical custody of enterprise assets to ensure, as a minimum, that electronic data with wide applicability is stored in retrievable storage media, that inactive data is archived, that hard copies are protected, and that data is identified and catalogued for retrievability. Any movement of data assets and any copies of the assets must be coordinated with stakeholders. Further, ensure that planning is in place to control data assets near the end of their useful lives so the enterprise does not store items that no longer retain value.

Data management practices ensure data is stored at authorized locations, and backup copies of vital records are kept at locations separate from the masters for best disaster recovery practice. They ensure that backup copies are not maintained or stored at unapproved locations, such as personal residences. They make certain that the physical location of data assets is known and that those assets are easily retrievable by those who need them.

Maintaining control of the repository and the associated processes or data holdings is necessary and may be a DM function. The data manager should pay special attention to changes of stewardship. These changes can result from a number of factors, including the following:

- Major organizational change—such as enterprise policy changes, corporate mergers, or corporate divestitures.
- Personnel changes resulting from changes in position responsibilities, attrition, or similar actions.
- Changes in management during the data life cycle—for example, from an on-site location in the early life cycle to an off-site location when data is archived.