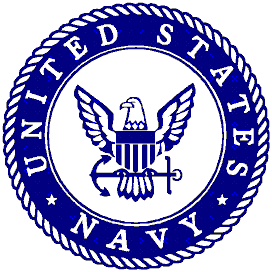
****

Date: 21 November 2022

**Data Governance, Maintenance, and Configuration Control Recommendations**

**Ticket Id – DRAFT Governance document for Enterprise Vocabulary**

**For**

**Some Organization**

Version 1.0

|  |
| --- |
|  |

Table of Contents­­\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

[Revision History 3](#_Toc496530471)

[Executive Summary 4](#_Toc496530472)

[Introduction 5](#_Toc496530473)

[U.S. Navy Policy 7](#_Toc496530474)

[Governance 8](#_Toc496530475)

[Goals 8](#_Toc496530476)

[What are we maintaining? 9](#_Toc496530477)

[Advantages of Governance 9](#_Toc496530478)

[Technical Debt 10](#_Toc496530479)

[Metadata Management 11](#_Toc496530480)

[Management 12](#_Toc496530481)

[Phase 1 – Enterprise Vision 13](#_Toc496530482)

[Organization Structure 13](#_Toc496530483)

[Phase 2 – Initiative Vision 15](#_Toc496530484)

[Phase 3 – Definition/Construction (Management & Maintenance) 15](#_Toc496530485)

[Configuration Management (CM) Plan 16](#_Toc496530486)

[Maintenance 19](#_Toc496530487)

[Phase 4 - Implementation 19](#_Toc496530488)

[Phase 5 - Support 20](#_Toc496530489)

[Phase 6 - Retirement 20](#_Toc496530490)

[Implementation Recommendations 22](#_Toc496530491)

[Approach 22](#_Toc496530492)

[Tools Implementation 23](#_Toc496530493)

[Abbreviations/Glossary 24](#_Toc496530494)

[References 25](#_Toc496530495)

[Navy Policy 25](#_Toc496530496)

[Historical References 25](#_Toc496530497)

[Standards 25](#_Toc496530498)

[Additional References 25](#_Toc496530499)

[Implementations 26](#_Toc496530500)

[Agency Strategy 26](#_Toc496530501)

[Configuration Management Concepts & Plans 26](#_Toc496530502)

[Meta-data 27](#_Toc496530503)

[Figure 1 - Data Governance Concept of Operation (CONOPS) 5](file:///C:\Users\christopher.wood.AD\Documents\SpiderOak%20Hive\ActiveWork\CTG807\NOe-ControlledVocabulary_Governance_BASELINE.docx#_Toc495824006)

[Figure 2 - Management and Maintenance (Becker, 2009) 12](file:///C:\Users\christopher.wood.AD\Documents\SpiderOak%20Hive\ActiveWork\CTG807\NOe-ControlledVocabulary_Governance_BASELINE.docx#_Toc495824007)

# Revision History

|  |  |  |  |
| --- | --- | --- | --- |
| Date | Version | Description | Author |
| 2017/09/22 | 0.1 | Enterprise Vocabulary Governance, Maintenance and Configuration Control Recommendations, inception. |  |
| 2017/10/13 | 1.0 | DRAFT Enterprise Vocabulary Governance, Maintenance and Configuration Control Recommendations |  |

# Executive Summary

The scope of the following recommendations for the Some Project (SomeProj) Enterprise Vocabulary Governance, Maintenance, and Configuration Control is submitted as a DRAFT and provides seed material intended to foster thought and potentially act as a blueprint for a path forward in support of the subject.

Organizations use governance to establish strategic direction that provides purpose, helps set expectations, inspires goals, and drives the actions necessary to guide a business pursuit and is aligned with business objectives.

The value of a Controlled Vocabulary is quickly outweighed by a lack of controls for the subject vocabulary. All too often, initiatives are formed with the intention to deliver a high value product but never materialize due to a lack of controls that help ensure relevance and accuracy of the subject material.

This document seeks to provide an outline of what Data Governance is and to define the minimal requirements necessary to manage a Some Project Controlled Vocabulary.

# Introduction

A Data Governance initiative involves the management of data for the purpose of delivering data that is trustworthy, timely and relevant to the community of interest (COI). Data governance is not about the data’s content but the business vision, processes, decisions, and interaction between participants/stakeholders to oversee and control the management and usage of data, data-related resources and information within the enterprise.

Data Governance requires cross-organizational cooperation to deliver quality results. Data Governance is achieved by respecting and partnering Management with Maintenance (Operations). (Wheatcraft, 2017)

Figure - Data Governance Concept of Operation (CONOPS)



*Program* management differs fundamentally from *project* management. Project management focuses on the achievement of immediate tasks with specifically allocated resources and time. Program management, on the other hand, manages multiple related tasks, each of which makes its own contribution to overall strategic goals. Program management allows the data governance team to use work from earlier projects in later projects, avoid duplicated effort, and ensure that all the program’s projects work smoothly together in support of desired strategic goals. In addition, data governance depends on management support. It demands the vision, leadership, and cooperation at the top enterprise as well as the community level. The commitment of the leadership team is essential for the success of a data governance program. (Enterprise Data Management Data Governance Plan, 2007)

SOME ORGANIZATION has gathered information from a multitude of sources to define a Vocabulary. The Vocabulary in question is referred to as a Controlled Vocabulary (CV), meaning changes of any type undergo a specific set of steps and is reviewed by previously-identified Subject Matter Experts (SMEs) who serve specific roles within those steps. A CV should help achieve the SOME ORGANIZATION enterprise’s goal of using semantic technologies to rationalize BUSINESS terminology and meaning, thus bringing data together from many different systems.

Data Governance defines the rules of “*what* will happen”. Data Management (DM) and Information Technology (IT) adhere to the rules which represents “*how* things will happen”. Data Management defines the needs and requirements for the IT infrastructure. IT supplies and maintains that infrastructure per those requirements. (Wheatcraft, 2017)

# U.S. Navy Policy

**DoD Instruction 8000.01**, Management of the Department of Defense Information Enterprise (DoD IE) indicates that:

“Information is considered a strategic asset to DoD. It must be safeguarded, appropriately secured and shared, and made available to authorized DoD personnel and mission partners to the maximum extent allowed by law, DoD policy, and mission requirements, throughout the information life cycle.

Functional processes are to be examined, and if possible streamlined or improved, in order to improve effectiveness and reduce cost before investment is made in information technology.”

**DoD Instruction 8320.02**, Sharing Data, Information, and Information Technology (IT) Services in the Department of Defense states:

“Data, information, and IT services are considered enablers of information sharing to the DoD. Data, information, and IT services will be made visible, accessible, understandable, trusted, and interoperable throughout their lifecycles for all authorized users. Authorized users include DoD consumers and mission partners, subject to law, policy, data rights, and security classifications.”

**DoD Instruction 8320.07**, Implementing the Sharing of Data, Information, and Information Technology (IT) Services in the Department of Defense requires:

“Data, information, and IT services will be visible to authorized users by creating and associating metadata (“tagging”), including discovery metadata, for each asset. In accordance with Reference (a), DoD metadata standards will comply with applicable national and international consensus standards for metadata exchange when possible.”

And that:

“…For existing systems, high-value data and content will be made available through Web application programming interfaces (APIs) and apply metadata tagging, as appropriate. To the greatest extent possible, the DoD’s common information sharing standards, protocols, and interfaces will be compatible and interoperable with those of other federal departments, agencies, and mission partners.”

# Governance

Governance is an industry-recognized best practice when managing resources for any level of the Enterprise. The benefits of Governance include reduction or elimination of technical debt accrued in implementing IT solutions without an explicit set of standards, guidance, and verification. A Controlled Vocabulary is a living entity which requires management, the process of controlling resources, and maintenance, the support at some periodicity of the existing state of the process.

Governance requires a formal declaration of how the subject will be managed and maintained. A governance initiative should have outlined goals. For governance to be successful, there must be a thorough understanding of the subject matter for which processes, standards, and management practices will be created and employed.

## Goals

The goals of governance are to standardize, harmonize, and integrate data across the Enterprise. The following points describe how to achieve these goals for governance:

* Adopt formal policies and procedures to ensure data consistency, data standardization, data reuse, and data exchange.
* Create a formal decision-making structure to standardize data across the enterprise.
* Provide a central mechanism for communicating data-related initiatives across the enterprise.
* Serve as a liaison between technical and business groups, both internal and external.
* Define and enforce best practices in data standardization and data quality.

Governance of a subject requires established goals, which should be categorized, achievable, and managed.

Some high level business goals for SomeProj Controlled Vocabulary should be:

* Align BUSINESS terminology with wider organization (DoN) and agency (DoD) capabilities and services.
* Support development efforts by using consistent terminology and integrating subject vocabulary into newly burgeoning or upgraded IT capabilities.
* Support efforts to improve Content Discovery & Retrieval (CDR) through data tagging.

High level technical goals for the SomeProj Controlled Vocabulary should be:

* Manage the SomeProj Controlled Vocabulary in accordance with industry and DoD best practice.
* Maintain formal rules for the management, release, and promulgation of the vocabulary.
* Seek touch points and interfaces with other agency vocabularies that deal with BUSINESS concepts.
* Host, communicate, and advertise SomeProj Controlled Vocabulary in support of integration with other agencies.
* Establish a Plan of Action and Milestones (POA&M) for BUSINESS Service and Capabilities integration.
* Establish formalized rules of use for the SomeProj Controlled Vocabulary within the community and standards for determining if a service and/or capability warrants use of the vocabulary.

## What are we maintaining?

Ultimately, this governance will ensure a consistently managed set of terms and definitions structured in such a way to support semantic technology. The governance itself carries its own data and meaning, in the form of metadata, which are necessary to help provide history, accountability, and a deeper understanding of how the managed language changes over time.

Semantic technology is based upon data stored in a graph, referred to as graph data, and a description of what that data means. As the name implies, semantic technology stores meaning with the data. It also removes the need to define structure, or data models. There is no predefined structure to which the data must be semantically bound. This provides tremendous flexibility in what information can be stored, and it enables information to be combined and used rapidly in ways that are not possible with relational or traditional XML data structures.

Semantic technologies are “meaning centered” technologies that, when implemented, can make Operational Oceanography information more readily available. Capabilities such as auto-recognition of information topics and concepts, information meaning and extraction, and categorization of information will lead to a future where information is both available and easily discoverable through online catalogs. These catalogs will provide comprehensive, focused results to user queries. An Enterprise Controlled Vocabulary (ECV) is a key semantic component to achieve this goal. The vision is to develop, maintain, and utilize a comprehensive ECV that has appropriate scope and detail to accurately describe Operational Oceanography information. The ECV will be available to metadata tools and catalog interfaces and will enable a more intelligent, capable, relevant, and responsive system interaction.

## Advantages of Governance

* Visibility into how data is flowing through the environment. The first step toward managing the environment is to have a clear visual map of how the data is moving between data sources and targets.
* Impact analysis and root cause analysis. Impact analysis enables IT staff to see the impact of a proposed change to the environment before it is implemented and to understand the complex cross-dependencies involved. (For example, changing the definition of “Due Date” requires the ability to see what other applications, reports, and analyses will be impacted by the change.) Root cause analysis enables business users to drill back from a term in a report to understand the source of the data along with how it was moved and transformed as it was added to the report.
* A common business vocabulary that standardizes terminology. The resulting business glossary enables clear communication among business units and between business and IT.
* Accountability into who is responsible for business terms and definitions. One of the core concepts of data governance is accountability. If an issue is discovered relating to a business term, it is important to be able to identify who owns the term and is responsible for remediating any issues. It is also important to be able to see who has made changes to the term and when it was changed.
* Audit trail for auditability and compliance. Metadata management should be able to tell you who changed what and when it was changed. It should also be able to provide timely answers to questions about the flow of data in the environment. (For example, the audit trail should be used to easily determine from where a particular number originated and how it was calculated.) The transparency is especially important to ensure compliance with many data-driven regulatory edicts such as Sarbanes-Oxley and Basel II.

## Technical Debt

Technical Debt is a concept in programming that reflects the extra development work that arises when code that is quickly implemented in the short term sacrifices a solution’s design quality and robustness with speed. Technical debt can be applied to any effort where adherence to standards or quality is lacking which is typically due to governance not being enforced.

* For organizations that undergo a great deal of personnel flux, such as the Department of Defense, it is impossible to maintain anything without clear, documented guidance. All too often new team members are provided material they are expected to manage and/or maintain only to have the burden of understanding what it is they are overseeing as well as establishing any semblance of order. This lack of order causes delays or failure on other initiatives that depend on the subject or can lead to the decay of the subject due to lack of interest.
* The mission and requirements of a subject can easily change if not documented due to a multitude of factors. In a military organization, this scenario is exacerbated because of personnel turnover. Without formalized documentation providing mission, requirements, scope, and direction, an initiative such as SomeProj Controlled Vocabulary can easily change. The cost of that change in mission, requirements, or scope typically has a ripple effect into all other efforts that actively interface with or will eventually interface with the subject. Most IT project schedules slide due to external dependencies which in turn can be sub-categorized into poorly-defined interfaces or scope.
* What goes around comes around. This concept focuses on the fact that poorly or non-governed efforts typically become stale only to resurface years later. This resurfacing may involve starting the project over from scratch or significant overhauling and creates a large burden on resources, in contrast to a steadily maintained and well-documented effort that has less overall impact on resources.

## Metadata Management

Metadata management is an important capability, enabling IT to manage change while delivering trusted, secure data in a complex data integration environment. The benefits are compelling when business stakeholders systematically engage in this process and accept ownership of the reference frame around data. At that point, enterprises can link the business metadata with the underlying technical metadata, providing a vocabulary and context for collaboration across the company.

For example, when a business user asks an IT counterpart to show “net revenue” in a report or analysis, it will no longer prompt the question, “which net revenue—financial, sales, or marketing?” Good metadata management contributes to data governance by helping to pre-empt such important questions, among several others:

• What does this business term mean?

• Which (of several similar) business terms should I use?

• What is the source(s) of the data behind that term?

• How was the data transformed as it was moved from source to target?

• Who is accountable for the definition, documentation, and management of that term?

• Who has changed this term? How and when was it changed?

• What policies and rules are applicable to that term? (Examples include data quality rules, security masking rules, archiving rules, and data retention policies)

• What other data objects would be impacted by a change to a specific data object in the environment?

• How long will it take to implement a change to the environment, without causing issues to other reports and analyses that may use the same data objects?

•What happens if the CV is deployed, and a term is deprecated or removed, and relationships/links have been made to the term?

All these questions have bearing on the overall quality and scope of a proposed governance initiative and should be asked and have documented responses as part of the Use Cases previously referenced.

# Management

The following phased approach enumerates how governance should be defined at the macro level with details provided in each phase to help facilitate the beginnings of Data Governance. Phases 1 and 2 are focused on the Management of the data. Phase 3 is a blend of Management and Maintenance (Operations). Phase 4 is about the maintenance/operations of the data. Phases 5 and 6 are a blend of Management overall and execution (maintenance/operations) of that management policy.

Figure - Management and Maintenance (Becker, 2009)



## Phase 1 – Enterprise Vision

During Phase 1 – Enterprise Vision, the organization creating Data Governance should ensure a vision for the data they are seeking governance over has a vision and requirements to support that vision. A Data Governance charter should specify:

* Vision, Purpose, Approach
* Scope
* Goals/Objectives
* Deliverables
* Quality Objectives
* Organization structure and Responsibilities
* Plans for Support Activities (supported in phase 2)
* Risk (continuously updated)
* Schedule (finalized in phase 2)

Will the Vocabulary grow beyond its current state? Will integration with other agencies be sought out? How much penetration will the SomeProj CV have on IT initiatives and how soon will that occur? What is the Scope of the SomeProj CV? Are all Echelon commands below SOME ORGANIZATION responsible for the support, use and integration of the SomeProj CV?

Consider using the Appendix E in the Federal Student Aid Data Governance Plan (see Additional References) as the baseline for your governance charter. Document the charter, circulate amongst stakeholders, revise and release to the community.

### Organization Structure

The organization of Data Governance should be at four tiers of scope: Executive, Strategic, Tactical, and Stakeholder. Executive Data Governance Council (Executive Council) provides strategic direction, ensuring that data governance efforts address all relevant and mission-critical needs of the enterprise. The Admiral, Technical Director of SOME ORGANIZATION, possibly the Technical Directors of each Echelon command below SOME ORGANIZATION, and Architect for the SomeProj should comprise the Executive Data Governance Council.

The Strategic Data Governance Steering Committee (Strategic Committee), also known as the NAVBUSINESSCOM Information Management Working Group (IMWG), carries out plans and policies to implement guidance from the Executive Data Governance Council. It prioritizes data governance efforts and communicates with stakeholders, users, and other communities of interest. It identifies staff (data stewards) to oversee areas of data (data domains). The SOME ORGANIZATION Operations department is recommended as the governing body with overall responsibility for the SomeProj CV, and will designate an Operations Information Manager (OIM) charged with overall execution and monitoring of proper implementation by NMOC Organization Information Managers (NOIM’s) across the enterprise. The IMWG would be chaired by the OIM and the NOIM’s for any Data Governance modifications that would have influence on the SomeProj CV.

Figure - Data Governance Organization



The Tactical Data Governance Working Group (Tactical Group), a primary role of NOIM’s, is to ensure the relevance and accuracy of the SomeProj CV as related to their command-specific mission needs. All changes proposed to the SomeProj CV by any NMOC stakeholder would go through their respective NOIM and must follow the proper change request procedures detailed by the Data Governance.

Nearly every member of the NAVBUSINESSCOM community is a stakeholder. Every person that interacts with any NAVBUSINESSCOM customer or fellow community member could have valid input to the content within the SomeProj CV. Stakeholders should have a NOIM as their representative who would help act as the first engagement for change and guide the change through the Configuration Management (CM) Plan.

## Phase 2 – Initiative Vision

During Phase 2 – Initiative Vision, the organization should make the Data Governance charter more concrete by:

* Gathering/Generating Requirements – Map your Goals and Objectives cited in the Data Governance Charter to requirements and Use Cases.
* Defining Deliverables – Formalize deliverables that your Data Governance will produce such as: membership/distribution lists, updated policies/procedures with minutes, metadata defined in support of management, etc.
* Planning – Establish a strategic plan to achieve your Objectives cited in the Data Governance Charter and put that plan into the preliminary/draft schedule established in Phase 1. Consider things like integration with other agencies and IT initiatives to help ensure promulgation of the Vocabulary.
* Identifying Personnel – Identify personnel that have expertise, time, and are willing to be participants in the roles you have identified. Establish Memorandums of Agreement (MOA’s) with each participating Organization to ensure the participants are aware of their duties both in responsibility as well as time.
  + Expertise – Ensure the personnel selected have either sufficient background in: Management, Governance, policies/procedures, IT initiatives, and/or Subject Matter Expertise (SME).
* Supporting Activities – Ensure resources are planned and budgeted for support of this effort in future budgetary forecasts. If licensing for the implementation selected for maintenance of the SomeProj CV is known, ensure those funds are reserved as well as IT manpower to support upgrades and Information Assurance compliance.

## Phase 3 – Definition/Construction (Management & Maintenance)

Continue development of the Data Governance plan by additional refinement by management in response to execution of the plans drafted:

* Tools Selection – map requirements, from Phase 2, to desired features of a solution that can manage your data and meet the: quality, metadata, and deliverables of your governance plan. Execute a Decision Analysis and Resolution (DAR) to ensure your needs are met by the tool that fits within your CONOPS and organization’s compliance requirements.
  + Selection of a tool or tools should be completed before this phase is finished.
  + Potential procurement and an implementation plan should be complete before this phase is finished.
* Formalize teams and personnel identified in Phase 2.
* Formalize standards, policies, and procedures for the management and maintenance of SomeProj Controlled Vocabulary. Circulate those standards, policies, and procedures with the participants to ensure buy-in. Construct a Configuration Management Plan.
* Specify how construction of data will be approached:
  + Use of existing semantic content.
  + Utilize automated methods of data capture against known data sources (NEP-Oc/Catalog Service, Enterprise Architecture Framework, Flight Weather Briefer, etc.) to capture as many terms as possible and compare that data set to the existing SomeProj CV to ensure any potential gaps are covered. This strategy correlates to Section 11.1.3.3 of the [ANSI/NISO Z39.19-2005 (R2010), Guidelines for the Construction, Format, and Management of Monolingual Controlled](http://www.niso.org/apps/group_public/download.php/12591/z39-19-2005r2010.pdf).

### Configuration Management (CM) Plan

A configuration management plan identifies the procedures that will drive how change occurs on a system, a set of connected things or parts forming a complex whole. A CM plan must be formalized, reviewed and accepted, as well as published to ensure all participants of an initiatives have guidance on how to interact with the system.

Terms will be evaluated for inclusion in the vocabulary based on the IMWG/Tactical Working Group documented scope of the vocabulary. Only terms relevant for indexing/cataloging and retrieval/selection of BUSINESS information will be considered for inclusion.

Details regarding the specifics of a: add, update, or delete should be documented in the CM plan and verified by the IMWG/Tactical Working Group. An individual record will be created for every term registered within the SomeProj CV as soon as it is admitted into the controlled vocabulary. The details regarding the density of data captured per term should be an appendix within the CM plan specifically for the controlled vocabulary, thus allowing this governance to extend to other concepts related to semantic content. The data captured for the vocabulary should be defined first to help assist in the selection of tools to support it. The following are examples of how potential change processes could be defined specifically for SomeProj CV. With minimal effort higher level process definitions could be abstracted from these recommendations to handle any data under this governance. Specificity should be applied, when germane, to specific content to ensure a consistently generated output.

**Add**: Any member of the NAVBUSINESSCOM community can submit a term for potential inclusion into the SomeProj CV. Submission by any stakeholder starts with the NOIM which represents that portion of the community or category of content. Upon identification of a new term for SomeProj CV inclusion, the NOIM will submit the candidate term IMWG. Once submitted the term will be have all quality standards applied to the term prior to IMWG review. Section 11.1.5 & 11.1.6 of [ANSI/NISO Z39.19-2005(R2010)](http://www.niso.org/apps/group_public/download.php/12591/z39-19-2005r2010.pdf) provides guidance on how inclusion of terms should be considered. Section 11.3.1.1, Addition of Terms, speaks to guidance for general maintenance purposes. If the term meets the quality standards then subject term will be included with all other compliant terms for IMWG review. All IMWG members will review the term based on agreed standards of acceptance, make comments, suggest any modifications, and vote to approve or reject the term. At the following scheduled IMWG meeting, the term will undergo final review, and after any required discussion, a final vote will be taken to include, reject, or modify the term. Final adjudication authority rests with the OIM. Once adjudicated, the term(s) will be updated in the SomeProj CV, and released with the next scheduled version release.

**Update**: Updates to a term will be handled in a process similar to the addition of new content. Proposed changes would be highlighted showing review members the before and after as well as research conducted to ensure the subject change doesn’t overlap with another term. Investigation into other common/related vocabularies will also be conducted to ensure subject modification is to the benefit of the SomeProj CV.

**Delete**: The removal of a term would originate with a stakeholder or the IMWG itself, however consideration for impact on existing releases and external dependencies on the vocabulary, IT systems, will be taken into account. The vocabulary’s metadata should include the ability to either deprecate and/or delete a term, however the implementation of the vocabulary has to be sensitive to the aforementioned external dependencies. Part of the Data Governance initiative should be rules set forth that dictate how the vocabulary will be used to include treatment of data that is deprecated and/or deleted. The ensure referential integrity in data implementations using the SomeProj CV deleted terms will not be removed from the actual vocabulary itself but masked out of view. The presentation of a vocabulary is discussed extensively in the [ANSI/NISO Z39.19-2005(R2010)](http://www.niso.org/apps/group_public/download.php/12591/z39-19-2005r2010.pdf) as well as other texts. Consideration for how terms that are no longer valid will be documented.

**Release**: The release of the SomeProj CV will be based on a scheduled determined by the IMWG. The release schedule will be published and the notification process for stakeholders exercised to ensure all parties are cognizant of the release. Individual IT assets utilizing the SomeProj CV should be notified before the community to ensure lead times for integration into the target platform. Section 11.2 of [ANSI/NISO Z39.19-2005(R2010)](http://www.niso.org/apps/group_public/download.php/12591/z39-19-2005r2010.pdf) provides guidance on how to Test and Evaluate Controlled Vocabularies. The CM Plan should include the required steps of test, evaluation, and verification of the SomeProj CV prior to any release.

In the Configuration Management Concepts & Plans section of the References a list of potential plans are provided as guidance for the SomeProj CM Plan. Past CM Plans for Joint BUSINESS Configuration Management System (JMCMS) should also be referenced to help ensure corporate knowledge is not lost.

#### Reasons for Configuration Management

Good Configuration Management ensures that the current design and build state of the system is known and recorded; and doesn’t rely on the tacit knowledge of the team. An accurate historical record of system development is very useful – not only for project management and audit purposes, but for integration and development activities. Effective use of CM tools enables team participants to formally (and safely) record all change details. Not only what was changed in the product but why it was changed.

#### CM Technical Debt

Not establishing CM can lead to the following:

* The wrong requirements being accepted
* The wrong design being implemented
* The wrong tools being used for development
* The wrong software being tested
* The wrong test suite being used
* The wrong version of software being released

Any of these could lead to large amounts of wasted effort, wasted money, late deliveries and seriously dissatisfied customers.

# Maintenance

Maintenance is the process of preserving the existing state of a system or entity. Subject state should be known (documented) and there should be parameters the maintainers utilize to help ensure that preserved state. How do you maintain the goals associated with governance of the SomeProj Controlled Vocabulary?

1. Establish goals, document subject goals, and ensure the governing body agrees and embraces them.
   1. Meet periodically to both verify, add, and update goals.
   2. Examples
      1. Identify which services could make use of SomeProj CV and start collaboration with their respective stakeholders.
      2. Integrate vocabulary into x% of BUSINESS web services.
2. Ensure Requirements for the vocabulary exist, are maintained, and that Use Cases are created and maintained to demonstrate the requirements’ fulfillment.

3. Configuration Management is a must as is the definition of a community development and contribution process.

4. Establish a public community mailing list where stakeholders and interested individuals/groups can follow development.

6. Persistent investigation into other agencies approaches and potential integration of external agencies weather-oriented data.

7. Defining and documenting a release management procedure ensuring that new contributions and developments are made available through formal public open-source releases.

The following “phases” are part of maintenance and have guidance that indicates how to ensure the maintenance of the data this governance document references.

## Phase 4 - Implementation

Phase 4 is focused on the actual implementation of the Data Governance for SomeProj CV and future potential for all semantic data management for the BUSINESS community.

* Policies and Procedures – In Phase 3, policies, procedures, and standards were formalized, meaning they were selected, circulated, and agreed upon. Subject content should now be published in an easily read and accessible medium. Explicit care towards following this guidance should be taken to ensure consistency and allow for modification when errors/discrepancies are found.
* Data – Data should be reviewed, verified for quality as well as standards’ compliance, baselined, and published.
  + Exercise policies and procedures for change.
  + Release as per policy and schedule.
* Teams – Finalized and scheduled to have their first meeting on the calendar.
  + Manage/maintain objectives and goals.
  + Maintain data as per policy and procedure.
  + Adhere to schedule established and provide feedback towards schedule updates.
* Tools – Execute the implementation plan from Phase 3. Note that your tools could be a series of steps towards a final/desired solution or could be procurement of the most desired solution first.

## Phase 5 - Support

Phase 5 should be focused on support of the continued funding and resource allocation necessary to support the Data Governance for SomeProj CV. The following is a list of resources, estimated, that would be required for continued support:

* Tools
  + Licensing
  + Server Cost (hosting solution)
  + Manpower Resources
    - Upgrades of core solution
    - Upgrades of support capability
    - Information Assurance concerns (CTO, IAV\*, STIG)
* Team/Participants
  + Manpower Costs to support the effort at all levels

The following is an example equation for calculating manpower costs for different participants of the SomeProj Data Governance:

Data Governance Lead - 1 person x 2 hours/week x 50 weeks = 100 hours

SME Resources – 8 people x 1.5 hours/week x 24 weeks = 288 hours

IT Resources – 1 person x 8 hours/week x 52 weeks = 416 hours

Architect/Metadata Specialist – 1 person 5 hours/week x 50 weeks = 250 hours

Maintenance Resources – 1 person x 24 hours/week x 50 weeks = 1200 hours

Total estimate manpower cost per year are 2, 254 hours at various pay grades and rates.

## Phase 6 - Retirement

Thought should be given towards retirement for any Data Governance at the onset. Two potential strategies for determining if/when retirement of the SomeProj CV Data Governance effort is appropriate are:

* Establish a time period, ten (10) years out, that the strategic thinkers, architects, and upper management will conduct a review of the program as a whole. Based on amount of use, penetration into IT initiatives, and other measurable criteria present the case of either keeping or discontinuing the effort.
* Create parameters for continued use that are applicable at any time. For example:
  + Lack of integration of SomeProj CV in any IT initiative within one year of Data Governance Charter launch.
  + No updates applied to the SomeProj CV within a year time period.
  + %X of all changes hung in committee not moving within a specific period of time.

While it’s possible to ignore the Retirement phase of a Data Governance initiative, history shows that some IT initiatives linger when they should in fact be discontinued. The DoN Policy cited in this text indicates the need to maintain relevant IT capabilities and judiciously fund those IT capabilities that are current and planned.

# Implementation Recommendations

There are a multitude of tools that support Vocabulary management and the dissemination of content. A DAR process execution is recommended to verify the recommendations made by GDIT in the past as well as a phased approach towards development. Scope changes, governance implementation, technology, and budgetary constraints mandate the need for an up-to-date analysis of options.

The working group should establish those capabilities most desired to fulfill their documented requirements. A DAR sufficiently constructed will provide a mathematical answer to the best implementation. A phased approach allows for the working group to establish a presence and begin the work all while working towards a solution that benefits all. Considerations for sharing other agencies capabilities should also be considered as a cost savings benefit both in manpower as well as licensing could occur.

## Approach

1st and 2nd Sprint – (Phase 1) Initiate and close out Phase 1 where initial scope, vision, organizational structure, quality standards, and stakeholders are identified. Specifically detail your goals and requirements necessary to achieve those goals, write Use Cases to support examples of the fulfillment of those requirements, complete and circulate the governance charter. Start seeding long term planning requests with upper management.

3rd and 4th Sprint – (Phase 2) Finalize your first cut of requirements and ensure 100% mapping to your goals. Identify deliverables that are part of governance, propose a strategic timeline, start identifying personnel and making assignments, and consider budgetary options for subsequent years.

5th and 6th Sprint – (Phase 3) Initiate a DAR for tools that meet your needs since you’ve finalized your first cut of requirements. Ensure teams are formed and subject teams have been briefed by the OIM and they know their role. Start teams working by formalizing your standards and procedures for the management and maintenance of SomeProj CV. The Tactical Group would discuss procedures for changing terms as well as review a proposed CM Plan. The Strategic Steering committee would ensure management of the governance is being finalized. The Executive Council would ensure they concur with the strategic plan. Increase interface with IT stakeholders to start incorporation of the SomeProj CV into their systems.

7th and 8th Sprint – (Phase 4, 5, 6) Select the tool(s) that will support your efforts, consider budgetary concerns, and discuss retirement expectations by upper management. Phase 4 focuses on day to day operations, the tools that support operations, release of the data, and integration with external systems.

## Tools Implementation

A software package doesn’t provide governance, it helps ensure your governance is upheld by providing conveniences to the different parties involved in maintaining and managing your data. It is essential to ensure core requirements of your plan are met by tools that help stakeholders facilitate controlled and measured change. Examples of useful functions that will likely be requirements of a Data Governance plan: process identification, notifying stakeholders, release management, term record management, merging/researching external vocabularies, etc...

Due to the size of the SomeProj CV and complexity of managing any sufficiently complex dataset, a vocabulary management application is required to effectively manage content. There are some external organization dependencies/requirements that must be met to ensure DoN compliance for the SomeProj CV such as:

1. Application(s) must be DON Application and Database Management System (DADMS) approved.
2. Presentation in multiple formats such as: human-readable in MS Word and HTML, XML

At present, TopQuadrant’s TopBraid, Enterprise Vocabulary Network (EVN) meets many requirements of both management as well as presentation. Additionally, the use of TopBraid EVN allows for the ability to access the system remotely for term review and management, a factor that makes it well-suited for the remotely separated nature of NMOC Organizations.

Consideration for meeting current requirements, costs, sustainability, the DoD IT environment, and future requirements (specified by the Goals, Objectives, and Vision in the Data Governance Charter) should be made. A full DAR should be considered at this time. Regardless of selection, lessons learned integrating SomeProj’s architecture should be applied to a preliminary Implementation Plan that takes the tool selection into account prior to purchase. By vetting the selection against the DoN IT constraints that are currently present as well as local compliance requirements verification that the selection is a viable solution is more assured.

There are other examples of repository management that are effective, perhaps not as automated, or efficient, such as NASA’s GCMD Keywords Community Forum which clearly uses an Atlassian Jira and Confluence instance to manage their controlled vocabulary if contrast is needed between different implementations.

# Abbreviations/Glossary

**SOME ORGANIZATION** is the official acronym for Commander, Naval Meteorology and Oceanography Command.

**Configuration management** (**CM**) is a system engineering process for establishing and maintaining consistency of a product's performance, functional, and physical attributes with its requirements, design, and operational information throughout its life. A CM schema should include scope, references, Configuration Item (CI) nomenclature, baselines, change control, verifications & audits, and release management.

**Controlled vocabularies (CV)** provide a way to organize knowledge for subsequent retrieval. They are used in subject indexing schemes, subject headings, thesauri, taxonomies and other forms of knowledge organization systems. Controlled vocabulary schemes mandate the use of predefined, authorized terms that have been preselected by the designers of the schemes, in contrast to natural language vocabularies, which have no such restriction.

**Enterprise** is a project or undertaking, typically one that is difficult or requires effort. For the context of this document the Some Project starting at the SOME ORGANIZATION level down its chain of command. SomeProjstands for Some Project.

**Enterprise Controlled Vocabulary (ECV)** are Controlled Vocabularies managed by a specific Enterprise based on Policy with governance to help mandate the management and maintenance of the vocabulary.

**Governance** is all of the processes of governing, whether undertaken by a government, market or network, whether over a family, tribe, formal or informal organization or territory and whether through the laws, norms, power or language of an organized society. It relates to "the processes of interaction and decision-making among the actors involved in a collective problem that led to the creation, reinforcement, or reproduction of social norms and institutions. In the context of this document the processes, standards, and rules of interaction amongst stakeholders identified as having responsibility for the maintenance and management of the SomeProj Controlled Vocabulary.

**Metadata** a set of data that describes and gives information about other data. Metadata is what gives your data, in this case a controlled vocabulary, a frame of reference.

**Policy** is a course or principle of action adopted or proposed by a government, party, business, or individual. For this document U.S. Navy Policy is used as the basis of why governance is necessary.

**Vocabulary** are the body of words used in a particular language or sphere. For purposes of this text vocabulary has relevance as it is the subject of management and is the means of unifying disparate functions within the SomeProj to support better integration of services.

# References

## Navy Policy

**DoD Instruction 8000.01**, “Management of the Department of Defense Information Enterprise (DoD IE)”, July 27, 2017

**DoD Instruction 8320.02**, “Sharing Data, Information, and Information Technology (IT) Services in the Department of Defense”, August 5, 2013

**DoD Instruction 8320.07**, “Implementing the Sharing of Data, Information, and Information Technology (IT) Services in the Department of Defense,” August 3, 2015

## Historical References

## Standards

**American National Standards Organization (ANSI) and National information Standards Organization (NISO)** ANSI/NISO Z39.19-2005 (R2010) Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies (<http://www.niso.org/apps/group_public/download.php/12591/z39-19-2005r2010.pdf>) [Accessed 27 Sep. 2017].

## Additional References

<http://www.taxonomies-sig.org/about.htm>

<http://www.getty.edu/research/publications/electronic_publications/intro_controlled_vocab/what.pdf>

<http://www.doncio.navy.mil/TagResults.aspx?ID=47>

<https://confluence.di2e.net/display/MMT/DCGS+Controlled+Vocabularies>

<http://federalstudentaid.ed.gov/static/gw/docs/ciolibrary/ECONOPS_Docs/DataGovernancePlan.pdf>

<http://pmiclg.org/images/downloads/Presentations/pmi_presentation_data_governance_lou_wheatcraft_052517.pdf>

### Implementations

**TopBraid Enterprise Data Governance** (<https://www.topquadrant.com/products/topbraid-enterprise-data-governance/>)

**Informatica** (<https://www.informatica.com/>)

**PoolParty** (<https://www.poolparty.biz/taxonomy-thesaurus-management/>)

### Agency Strategy

**Navy Manpower, Personnel, Training and Education (MPTE) Enterprise Information Management (EIM)** (<http://www.doncio.navy.mil/mobile/ContentView.aspx?ID=4558&TypeID=21>)

**NASA** Digital Strategy (<https://www.nasa.gov/agency/digitalstrategy/index.html>)

**NASA Common Metadata Repository (CMR)**

* <https://earthdata.nasa.gov/about/science-system-description/eosdis-components/common-metadata-repository>
* <https://earthdata.nasa.gov/about/gcmd/global-change-master-directory-gcmd-keywords>
* <https://cdn.earthdata.nasa.gov/conduit/upload/5182/KeywordsCommunityGuide_Baseline_v1_SIGNED_FINAL.pdf>

**Federation of Earth Science Information Partners (ESIPFed)**

* <http://wiki.esipfed.org/index.php/Semantic_Technologies>
* <https://github.com/ESIPFed/sweet>

### Configuration Management Concepts & Plans

Configuration Management (<https://en.wikipedia.org/wiki/Configuration_management>)

Configuration Management Plan Concepts (<http://www.techrepublic.com/blog/tech-decision-maker/itil-what-goes-in-a-configuration-management-plan/>)

Project Management Docs CM Template (<http://www.projectmanagementdocs.com/project-planning-templates/configuration-management-plan.html#axzz4vCbcWkyV>)

**United States Climate Reference Network (USCRN) CM Plan** (<https://www1.ncdc.noaa.gov/pub/data/uscrn/documentation/program/X033FullDocumentD0.pdf>)

**NASA IV&V CM Template** (<https://www1.ncdc.noaa.gov/pub/data/uscrn/documentation/program/X033FullDocumentD0.pdf>)

### Meta-data

**NASA Unified Metadata Model (UMM)** (<https://earthdata.nasa.gov/about/science-system-description/eosdis-components/common-metadata-repository/unified-metadata-model-umm>)