Words Matter: Managing Vocabulary Resources to Support Productivity

WA-RD 896.5

Leni Oman
Michelle Morgan
Christy Granquist
Steve Kollmansberger

Riley Woodward-Pratt
Andy Everett
Kathy Szolomayer

September 2018

Washington State Department of Transportation

Office of Research & Library Services

WSDOT Research Report

WSDOT Practical Solutions Accelerated Innovation Deployment Project

Words Matter:
Managing Vocabulary Resources to Support Productivity

Final Technical Report

Prepared for:

The FHWA AID Demonstration Program Project:
Deploying Practical Solutions using Lean Techniques and Knowledge Management

Prepared by Washington State Department of Transportation (WSDOT):

Knowledge Services, Development Division
Data Management Services, Information Technology Division
Research & Library Services, Development Division
WSF Libraries, Washington State Ferries

September 2018

Deploying Practical Solutions Using Lean Techniques and Knowledge Management (PS AID Project)

Project Management Staffing

Leni Oman, Knowledge Strategist, WSDOT
Steve Hanson, Project Control Manager, WSDOT
Riley Woodward-Pratt, Information Specialist, WSDOT
Paul Houle, Senior Lean Consultant, Washington State Department of Enterprise Services
Frances Harrison, Spy Pond Partners, LLC
Denise Bedford, Georgetown University

PS AID Project Vocabulary Management Team

Dr. Denise Bedford, *Georgetown University (Lead Taxonomy Task)*
Andy Everett, *Metadata and Taxonomy Analyst, Information Architect, Data Curator – WSDOT Data Library, Data Management Services, Information Technology Division (Lead Metadata Task)*
Christy Granquist, *Vessel Engineering Librarian, Washington State Ferries (Lead – Vocabulary Governance Task)*
Steve Hanson, *Project Control Manager, WSDOT Knowledge Services (Lead – Thesaurus Task, Co-lead - Glossary Task)*
Kathy Szolomayer, *WSDOT Librarian, Research & Library Services (Co-lead – Glossary Task)*
Riley Woodward-Pratt, *Information Specialist, WSDOT Knowledge Services, (Co-lead – Glossary Task)*
Michelle Morgan, *Assistant Director, Data Management Services, WSDOT Information Technology Division*
Gordon Kennedy, *Information Resources Manager (retired), Data Management Services, WSDOT Information Technology Division*
Steve Kollmansberger, *Information Resources Manager, Data Management Services, WSDOT Information Technology Division*
Leni Oman, WSDOT *Knowledge Strategist, Development Division*

1. Report No. WA-RD 896.5	2. Government Accession No.	3. Recipient's Catalog No.	
4. Title and Subtitle Words Matter: Managing Vocabulary Resources to Support Productivity		**5. Report Date** September 12, 2018	
		6. Performing Organization Code	
7. Author(s) Leni Oman, Michelle Morgan, Riley Woodward-Pratt, Andy Everett, Christy Granquist, Kathy Szolomayer and Steve Kollmansberger		**8. Performing Organization Report No.**	
9. Performing Organization Name and Address Washington State Department of Transportation 310 Maple Park Ave, SE Olympia, WA 98504-7371		**10. Work Unit No.**	
		11. Contract or Grant No. K992 FHWA Program code M37A (Delphi Code 15X0445060	
12. Sponsoring Agency Name and Address Washington State Department of Transportation 310 Maple Park Ave, SE Olympia, WA 98504-7371		**13. Type of Report and Period Covered** Final Research Report, December 2016 through July 2018	
		14. Sponsoring Agency Code	

15. Supplementary Notes

Conducted in cooperation with the U.S. Department of Transportation, Federal Highway Administration and Accelerating Innovation Deployment Program through the grant award titled *Deploying Practical Solutions using Lean Techniques and Knowledge Management*

16. Abstract

This document examines Washington State Department of Transportation (WSDOT) practices for developing, applying, and maintaining metadata, taxonomies, glossaries, and thesauri. These are collectively called vocabulary management for the purposes of this report. These vocabulary resources support search, navigation, and filtering of data and information. Current practices for managing these resources were analyzed for their ability to support multidisciplinary engagement throughout the Practical Solutions life cycle. This information was used to identify strengths and weaknesses and develop recommendations.

17. Key Words Knowledge management, Practical Solutions, resource model, capabilities, learning culture, information management, metadata, taxonomy, glossary, data management.		**18. Distribution Statement** No restrictions. This document is available through the National Technical Information Service, Springfield, VA 22161.	
19. Security Classif. (of this report) None	**20. Security Classif. (of this page)** None	**21. No. of Pages** 41	**22. Price**

Form DOT F 1700.7 (8-72)   Reproduction of completed page authorized

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Washington State Department of Transportation or Federal Highway Administration. This report does not constitute a standard, specification, or regulation.

Table of Contents

List of Figures

Executive Summary

Imagine that you are:

- looking for information and are able to narrow your search to a particular geographic area, business function, or other facet of the subject.
- discussing a subject with someone from another discipline and you realize you aren't sure of the meaning of a word. You are able to go to the agency glossary to find the definition.
- searching for information and you are offered the option to look for related terms or to auto-complete the search field once you start typing.
- searching for historical resources and can find them using the terms we use today, even if those terms were not in use at the time of publication.
- searching for information and you find what you are looking for in a timely manner.

These are the types of outcomes that vocabulary management can help the department achieve.

What is the problem?

During meetings and workshops for the Practical Solutions AID Project (PS AID Project), participants often discovered that a particular word was used in different ways and different words were used for similar things. This creates confusion when working within and across disciplines and makes it difficult to find relevant information and data.

Surveys of WSDOT employees often include feedback about the difficulty of finding information or data resources. This is common in many organizations and studies have shown that people spend up to 35% of their time looking for information (Feldman, 2004). If they are not an information specialist, they only find the information they seek about 50% of the time. However, there are things that we can do to improve findability of information and data resources in the department (Menzano, 2009). And that's what this report is about.

What is vocabulary management?

Vocabulary management is a term we are using to cover the development, implementation, and maintenance of metadata, taxonomies, thesauri, and glossaries. These are classification tools that information scientists use to manage findability, relevance, discoverability and navigability of content, and support integration of data and information. For the most part, information users don't need to know or understand classification methods, but in order to be effective these resources need to be actively developed and maintained. Most of this work is done by information professionals behind-the-scenes with periodic input from subject matter experts.

Information science has evolved to allow the text of digital content to be searched. Full text searching and machine learning are important tools to support managed vocabularies but by themselves, they often lack precision and cannot provide a historical understanding of terms. In a large organization with very specialized work units, term use often varies by profession, making it difficult to narrow results to desired content with full text searching alone. Term use also varies between WSDOT and external parties (e.g., luminaires vs street lights). This is complicated further by the changes in terms over time, of which current users of the system may or may not be aware. The messiness of language is something that remains a problem even for the most advanced artificial intelligence systems. Implementing

effective tools for improving the findability of documents at WSDOT will require combining automation with the best practices from the fields of library and information science.

Language can change over time for reasons such as modifications in policy direction, changes in organizational focus areas, or the emergence of new disciplines and increasing specialization. Vocabulary management documents the relationships of key terms to provide a bridge between current usages and historical usages and across disciplines and specializations. This allows us to increase the findability of historical information and data and to support future use of the information we create today. It also provides a bridge across differences in language used by different information consumers.

It is not possible to go back and hand enter metadata for and assign taxonomy classifications to all digital information resources. However, techniques are now available that can partially automate this process by leveraging available vocabulary resources.

A team of WSDOT's library and information science professionals have been meeting to examine vocabulary management practices currently in use in the department and to develop practices that would improve our ability to find relevant information and data. Our goal is to limit the amount of effort needed for vocabulary management by subject matter experts in non-information management fields and to apply successful information management practices to vocabulary management within the department.

What is the current state of vocabulary management at WSDOT?

Glossaries

The department has many glossaries managed independently by business units. Some terms are used by multiple organizations. Some duplicative terms have specific use within a specific context. Many 'definitions' are or include descriptions of use of the term without a clear statement of what the term means. Examples of the current variability in glossaries are provided below:

Similar definitions that could be reconciled
- ADA (Design Manual): An abbreviation for the Americans with Disabilities Act of 1990. The ADA is a civil rights law that identifies and prohibits discrimination based on disability. Title II of the ADA requires public entities to design new pedestrian facilities or alter existing pedestrian facilities to be accessible to and usable by people with disabilities
- ADA (Local Agency Guidelines): The Americans with Disabilities Act of 1990 which mandates sweeping changes in building codes, transportation, and hiring practices to prevent discrimination against persons with disabilities, not just in projects involving federal dollars, but all new public places, conveyances, and employers.
- Americans with Disabilities Act of 1990 (ADA) (Summary of Public Transportation): Federal civil rights law that assures persons with disabilities get equal opportunity to fully participate in society, the ability to live independently, and the ability to be economically sufficient.

Same word, conflicting definitions
- Ramp (Design ADA/Pedestrian): A walking surface with a running slope steeper than 20H:1V (5%).
- Ramp (Bridge Design): Any directional route of pedestrian travel having a running slope between 2% and 8.3%.
- Sand (Geotech): Particles of rock that will pass through a 0.19 in. (4.75 mm) opening, but will not pass a 0.003 in. (0.075 mm)

- Sand (Roadside): A mineral soil particle between 0.05 and 2.0 mm in diameter. A soil textural

class.

Description rather than definition

- Coastal Zone Management (Local Agency Guidelines): Applicants for federal permits or licenses must certify that their project will comply with the State Coastal Zone Management Program (Shoreline Management Act -RCW 90.58 - applies to projects within 61 m (200 feet) of a shoreline).

Practices for defining terms are varied.   Some consider legal definitions and others reference more generic definitions. As multidisciplinary work increases there is a higher demand to understand what organizations mean by certain terms and an agency glossary is a good resource to address this need.

The Vocabulary Management Team has begun to collate agency glossaries in published documents and is developing a method to post this to the Internet. The consolidated glossary will allow us to see current uses of terms and to identify gaps and opportunities for improvement. The collated glossary will include multiple definitions for the same term. It may be appropriate to maintain separate definitions depending on business use or it may be appropriate to select preferred terms or create one definition.

Taxonomies

The department has many taxonomies managed independently by business units. The diversity of taxonomies increases the challenges of finding related and relevant information. As multidisciplinary work increases, there is a higher demand to understand connections between organizations and the management of connections between taxonomies helps facilitate those connections in the digital environment.

A separate report titled *Enterprise Architecture: An Overview* describes strategies for an information architecture that supports core organizational objectives. One of the strategies recommended is the development of a particular type of taxonomy called a Business Function Classification Scheme (BFCS). A BFCS represents the essential work of the department independent of the official organizational structure and provides a framework for the development of enterprise information resources. The Vocabulary Management Team has begun to draft a business function classification scheme to help clarify information resources that have common interest across the organizations involved in the Practical Solutions lifecycle.

Thesaurus

The department does not use a thesaurus at this time. A draft thesaurus of Practical Solutions terms has been developed. The Vocabulary Management Team is exploring opportunities to use the thesaurus to support search and navigation of the WSDOT Web environment and other agency digital resources.

Metadata

WSDOT uses metadata to support search, navigation, and documentation of digital resources. Previous attempts to apply metadata to repositories have had limited success because of the number of metadata fields and an inconsistent implementation across the organization. The practices we are recommending identify a very limited number of required metadata fields with taxonomies providing consistent values to aid in search and navigation. The set of metadata may be extended based on subject or type of information resource.

How can WSDOT improve vocabulary management?

Advances in technology and information science provide new ways of managing vocabularies, minimizing the need for authors and stewards to develop and collect metadata by hand. Table 1 describes the goals and recommendations for each vocabulary product, and the impact the improvement will have on the agency, authors, and information stewards.

WSDOT has made significant progress in clarifying vocabulary management needs and strategies for improvement and implementation through the Practical Solutions Accelerated Innovation Deployment (PS AID) Project. The following recommendations will help WSDOT realize the value of improved vocabulary management. Additional details about these recommendations are provided in Table 1.

1. Endorse the role of the Vocabulary Team, and develop and implement term governance.

2. Publish and operationalize the metadata framework. Continue to build out metadata fields.

3. Develop a strategy and priorities to build out taxonomies and implement the strategy.

4. Publish an integrated glossary and establish a glossary management practice.

5. Develop a method to integrate the WSDOT Thesaurus into agency search tools. Develop a strategy and priorities for continued development of term relationships.

Summary

Words, their definitions, and their relationships are building blocks for effective coordination and collaboration in both human conversations and digital navigation and automation. WSDOT's vocabulary components have evolved to address specific business needs. As the expectation for collaboration and coordination has increased, there is greater need to work across traditional business lines. This is increasing the need for both common and context specific definition of terms. The field of information science has proven tools for managing term (word) associations. This document recommends a vocabulary management practice that builds on available tools and methods to enable and sustain common understanding and improved access to relevant information. A mature vocabulary management practice will improve access to business critical knowledge, help identify potentially duplicative or conflicting information, and provide a foundation for a nimble and navigable information resource for employee and community engagement.

	Glossaries	Taxonomies	Thesaurus	Metadata
Goal	Manage an accessible enterprise glossary that defines key terms used at WSDOT. Definitions will be created using information standards, regulatory considerations, and contextual usage. Multiple definitions may be maintained for the same term if contextually important.	Manage taxonomies that support search and navigation of digital data and information resources. Taxonomies will be created using information standards, regulatory considerations, and contextual usage. Hierarchical term relationships will be managed to help narrow search requests and improve relevance of search returns.	Develop and manage an enterprise thesaurus that documents term relationships in support of the agency glossary, agency taxonomies, and improved search and navigation of digital resources. The thesaurus will be created using information standards, regulatory considerations, and contextual usage.	Apply the metadata framework to digital resources to improve search and navigation. Limit effort needed by the author to capture core metadata. Automate capture of metadata as soon as feasible.
Recommendation	• Publish the enterprise glossary. • Develop an implementation strategy and procedures to incorporate more glossaries used across WSDOT, improve alignment of terms and definitions, provide for modification of terms, and improve access to glossaries for employees and stakeholders. Identify priority needs to improve agency glossary resources.	• Develop an implementation strategy for development and application of taxonomies scaled to the business need. • Analyze existing taxonomies to gain a better understanding of vocabulary needs and opportunities for reuse of existing resources. • Build the enterprise taxonomies outlined in the WSDOT Metadata Core Schema.	• Develop a method to integrate the WSDOT Thesaurus into agency search tools. • Develop a strategy, procedure, and priorities for continued development of the enterprise thesaurus. Pilot the use of the thesaurus and monitor the impact to the findability and relevance of agency information resources.	• Publish and operationalize the metadata framework. • Develop an implementation strategy for Core Metadata that applies to all digital resources. The implementation strategy should address application to legacy resources as well as current and future resources.
How will this affect agency work?	Provides a resource of definitions, promotes common understanding of terms, and establishes a practice for integrating and disseminating new terms or term use and raising concerns with existing definitions.	Improve search experience on WSDOT Web resources and navigation between agency digital data and information repositories. As vocabulary management evolves you will be able to submit feedback on terms that you would like to use for search or navigation.	The thesaurus will be managed by the vocabulary team. The primary impact this will have is to improve search and navigation of WSDOT's digital resources.	The primary impact this will have on your work is to improve your search experience of WSDOT's digital resources.
How will this affect authors or information stewards?	A glossary references existing definitions, and is a resource to help you integrate new terms and modify definitions or use of current terms.	Review existing taxonomies for fit to your need. If the key terms are not available to support search and navigation of your subject area, work with the Vocabulary Team to develop a new taxonomy and document term relationships. Taxonomies require an initial intensive effort and then limited maintenance.	You will be able to review thesaurus contents to see if the terms you would like to add already exist and if existing relationships are appropriate. The thesaurus will map existing term relationships and will allow you to identify how proposed terms should fit in to existing relationships or if changes are needed.	You may be required to fill in a limited number of fields about digital resources that you develop.

Table 1. Summary of Vocabulary Goals, Recommendations, and Impacts

Vocabulary Management Glossary

Browsing: The process of users following paths through a site that results in the retrieval of specific content objects. Source: *Information Architecture Glossary*, Argus and Associates (2000).

Classification Scheme: A method of organization according to a set of pre-established principles, usually characterized by a notation system and a hierarchical structure of relationships among the entities. NOTE: A classification scheme often also includes an index. Source: *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*, ANSI/NISO Z39.19-2005 (R2010).

Controlled Vocabulary: An organized arrangement of words and phrases used to index content and/or to retrieve content through browsing or searching. It typically includes preferred and variant terms and has a defined scope or describes a specific domain. Controlled vocabularies include Taxonomies, Thesauri, Controlled Lists, etc. Source: *Introduction to Controlled Vocabularies*, The Getty Research Institute.

Domain: An area of knowledge with an explicit scope established for the purpose of creating a knowledge organization system. Subject domains are characterized by engagement with a specific topic. Functional domains encompass an area of work that involves the same or closely related tasks. Discipline domains encompass an area of work performed by workers with a similar disciplinary background and who employ a consistent approach or set of approaches. Source: WSDOT Vocabulary Governance Team.

Findability: a) The quality of being locatable or navigable. b) The degree to which a particular object is easy to discover or locate. c) The degree to which a system or environment supports navigation and retrieval. Source: *Ambient Findability,* Peter Morville.

Glossary: A list of words with brief definitions, usually related to a specific topic or function. Definitions of terms may vary based on the context in which they are used. Glossaries assist in establishing common understanding of terms within and across business areas. Source: WSDOT Glossary Team.

Information Governance: A set of people and processes that shape the creation and maintenance of information. Formally defined as "the specification of decision rights and an accountability framework to ensure appropriate behavior in the valuation, creation, storage, use, archiving and deletion of information. It includes the processes, roles and policies, standards and metrics that ensure the effective and efficient use of information in enabling an organization to achieve its goals". Source: Gartner.

Lexical: Pertaining to words or the vocabulary of a language as distinguished from its grammar and construction. Source: *Information Technology --Metadata Registries (MDR) - Part 5: Naming and identification principles*, ISO 11179-5:2003.

Metadata: Data describing context, content, and structure of documents and records and their management through time. Often described as data about data. Source: *Information and documentation -- Records management -- Part 1: Concepts and principles,* ISO 15489.

Natural Language: A language that has evolved naturally as a means of communication among people. Source: *Collins English Dictionary*.

Navigation: The means by which users get from page to page on a website, while searching and browsing for content. Derived from: *Information Architecture Glossary*, Argus and Associates (2000) and Usability.gov Glossary.

Preferred Term: One of two or more synonymous keywords or keyword phrases chosen as the most relevant term for inclusion in a controlled vocabulary, for describing a concept or aspect of a content object. Derived from: *Information Architecture Glossary*, Argus and Associates (2000) and ANSI/NISO Z39.19.

Public Record: Any paper, correspondence, completed form, bound record book, photograph, film, sound recording, map drawing, machine-readable material, compact disc meeting current industry ISO specifications, or other document, regardless of physical form or characteristics, and including such copies thereof, that have been made by or received by any agency of the state of Washington in connection with the transaction of public business, and legislative records . . . Source: RCW 40.14.010.

Record: Data or information in a fixed form that is created or received in the course of individual or institutional activity and set aside (preserved) as evidence of that activity for future reference. Source: Society of American Archivists.

Search: The process of users entering terms into a system that results in a selection of content objects. The system can search the full text of content objects or attributes of those content objects. Search can be limited to one part of a site. Users who search may have more definite ideas of their information needs than those who browse. Source: *Information Architecture Glossary*, Argus and Associates (2000).

Semantics: The branch of linguistic science that deals with the meanings of words. Source*: Information Technology --Metadata Registries (MDR) - Part 5: Naming and identification principles*, ISO 11179-5:2003.

Syntax: The relationships among characters or groups of characters, independent of their meanings or the manner of their interpretation and use; the structure of expressions in a language, and the rules governing the structure of a language. Source: *Information Technology --Metadata Registries (MDR) - Part 5: Naming and identification principles*, ISO 11179-5:2003.

Taxonomy: A collection of controlled vocabulary terms organized into a hierarchical structure. Each term in a taxonomy is in one or more parent/child (broader/narrower) relationship(s) to other terms in the taxonomy*.* Source: *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*, ANSI/NISO Z39.19-2005 (R2010).

Term: One or more words designating a concept. Source: *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*, ANSI/NISO Z39.19-2005 (R2010).

Thesaurus: A type of controlled vocabulary that shows hierarchical (e.g. parent-child), associative (e.g. related) and equivalent (e.g. synonymous) relationships among terms. Source: *Information Architecture Glossary*, Argus and Associates (2000).

(Controlled) Vocabulary Governance: The coordination of essential business terms and alignment with organizational objectives so that staff and business partners can find and navigate the data and information they need. Source: WSDOT Vocabulary Governance Team.

Vocabulary Management: The process of keeping a vocabulary in line with changes in term usage, encoding practice, and understanding, as well as with the needs in the community. Good management practices should be based on openly available policies that enable distributed and configurable notifications for known users of the vocabulary; see also "Controlled Vocabulary". Source: *Issues in Vocabulary Management*, NISO TR-06-2017.

Introduction

Findability and access to reliable data and information are crucial to business processes at the Washington State Department of Transportation (WSDOT). Business functions produce and use data and information throughout the department, creating a rich environment of information resources. However, employees frequently report difficulty in finding data and information and ask for "Amazon-like" search capabilities to aid findability, selection, and access. Information-rich businesses rely on robust language management practices in order to provide search and navigation services that meet customer needs. For example, retail outlets that provide ways to narrow search selections by type of item, color, or size can efficiently guide customers to desired products.

Information resources are stored in many repositories throughout WSDOT. These repositories are organized mainly around organizational unit, subject, or type of content. The content of these systems can be characterized as unstructured or structured data. Unstructured data include narrative documents, images and other artifacts intended primarily for human comprehension. Structured data include tabular and geospatial data that are machine-readable

For the purposes of this document, structured and unstructured data are referred to as content.

A grant to WSDOT from the Federal Highway Administration (FHWA) Accelerating Innovation Deployment (AID) Demonstration program provided funding for a project titled, *Deploying Practical Solutions Using Lean Techniques and Knowledge Management (PS AID Project).* The PS AID Project examined data and information flow across the Practical Solutions lifecycle. Practical Solutions is based on a holistic approach to managing the multimodal transportation system. To efficiently support this more comprehensive and integrated approach, information resources also need to be managed holistically. Managing information resources with a holistic or enterprise view helps ensure that data and information resources are accessible to those who need them, when they need them, in the form they need. Holistic information resource management also eliminates confusion about authoritative data sources and highlights gaps and potential redundancies as areas for improvement.

Information management is guided by multiple disciplines of library science, records management, data management, and usability design. The Practical Solutions AID Project sought to find information management strategies that build on and blend the sound practices of these disciplines, in order to:

- Provide improved support for WSDOT employees in performing their duties.
- Increase findability, relevance, discoverability, navigability, and integration of content.
- Streamline and clarify responsibilities for management of content (which office/individual has the authoritative source? who must agree?).
- Improve confidence in repositories as reliable and timely sources of authoritative information.
- Enhance curation of content for improved records management, publication version control, web management and data management.
- Increase the reuse of physical and digital information resources for multiple purposes (instead of recreating them).

Policy, procedure, and governance are common needs across information management disciplines. A separate report titled *Improving Information Management for Practical Solutions at WSDOT* addresses these broader information practices. Vocabulary management is a unique discipline and need within information management. This report addresses policy, procedure, and governance specific to vocabulary management. These practices are intended to align with and enrich the broader agency information management practices.

Discussions about vocabulary management tools can seem either boring or scary to many. However, these are critical tools for successful information management, and have been implemented according to recognized standards in many organizations in both the private and public sectors.

Just as the *Manual of Uniform Traffic Control Devices* (MUTCD) sets standards that help guide travelers to their destinations, metadata and taxonomies provide directional guidance for the content within our information systems so that data and information get to knowledge workers so they can deliver intended business products. Applying categorization schema such as metadata, taxonomies, thesauri, and key words to content substantially improves the ability to search for and find information, and refine search result relevance to improve productivity.

What about machine learning?

Information science is rapidly evolving. Machine learning tracks user patterns and associates terms to improve future search activities. Machine learning is useful but does not yet have the capabilities to address complex connections. Machine learning can help Vocabulary Managers see the interests of system users. A Vocabulary Governance Team can use this information to redirect to more connections that are appropriate and to enrich a thesaurus of related terms. The enriched thesaurus can be used by system search engines to improve the user experience.

Vocabulary management is at the heart of successful content management. Using common strategies across physical and digital information repositories promotes a seamless structure that allows users to find all sources of content.

This report provides an overview of vocabulary management activities that would enhance findability, relevance, discoverability, navigability, and integration of WSDOT data and information resources. Improving vocabulary management will also facilitate WSDOT's ability to collate similar content and assess cost efficient strategies for management.

This document summarizes the vision for language management, technical elements that are part of language management, status and need, and makes recommendations for improvement.

Improving Vocabulary Management

Language changes. New terms are established, uses of terms are modified, and relationships among terms evolve. A vocabulary management practice is needed to ensure employees and stakeholders can efficiently find and navigate the data and information they are seeking, and to support automation of workflow and reporting.

Vocabulary management does not mean controlling the words people use and is not a one-size-fits-all solution. Vocabulary management is about documenting terms and managing term relationships over time so relevant and related information can be found.

Business Case for Change

Managing vocabulary resources is a fundamental step towards organizing data and information to enhance findability. Organized and accessible data and information are assets that help support WSDOT's business functions in a number of ways, forming the basis for the following goals:

- Reducing costs through time savings, avoiding mistakes, providing better access to information, reducing duplicative work, and enabling more effective communication;
- Speeding delivery of projects and products through faster decision-making and better coordination;
- Improving decision-making through easier access to authoritative data and information;
- Improving performance through faster error detection, improved customer service, greater anticipation of customer need, and faster responsiveness;
- Minimizing risk through earlier recognition of risks, more rapid response to risks, and faster recovery from adverse events;
- Improving feedback through faster needs identification, better coordination of responses, and better identification and propagation of lessons learned; and
- Supporting innovation through improved access to agency experts and knowledge resources, enabling collaboration, and connecting activities.

The Challenge

Current practices for classifying data and information are largely unique to each business area. While this decentralized approach has worked in the past, it is not meeting current business needs for data and information sharing and integration. The diversity of classification methods creates challenges for enterprise use of these resources. Challenges include the inability to find authoritative content, inconsistent usage of terms by different business areas, and ambiguity around the relationships between various data and information resources.

A Vocabulary Team was established within the Practical Solutions AID Project to develop recommendations for improved vocabulary management. The Vocabulary Team included individuals with library and information science expertise, supported by the PS AID Project Team, WSDOT Data Management Services, and periodic input from the Web Management Team and Records Management. The focus on library and information science expertise allowed the Vocabulary Team to work from a common understanding of best available practice for classification and retrieval.

The Vision

The Vocabulary Team developed the following vision statement to guide its work:

> WSDOT manages language to help employees and stakeholders predictably find trusted information to support business decision-making and to foster common understanding, connections, and collaboration across professional disciplines and among users.

The Vocabulary Team studied five interrelated aspects of vocabulary management: Vocabulary Governance, Metadata, Taxonomies, Glossaries, and Thesauri. These each provide specific benefits to coordinating agency vocabularies to address the goals listed in the Business Case. This report proposes to manage them in a complementary manner.

- **Vocabulary Governance** ensures data and information sources align with organizational needs.
- A comprehensive, uniform **Metadata** structure is key to making new WSDOT-produced materials easily findable from the start.
- The development of **Taxonomies** supports more accurate search and improved understanding of the data environment to facilitate better information management strategies.
- A departmental **Glossary** provides a human-readable view of the key terms managed through the governance structure, fosters common understanding of terms, and incorporates the vocabularies of offices with unique language usage.
- Creating and implementing a **Thesaurus** allows for more advanced information management techniques, including more refined keyword searching, and easier access to legacy materials.

Figure 1. describes how the components relate to each other and how they are used by people and computers.

Recommendations

WSDOT has made significant progress in clarifying vocabulary management needs and strategies for improvement and implementation through the Practical Solutions Accelerated Innovation Deployment (PS AID) Project. The following recommendations will help WSDOT realize the value of improved vocabulary management.

1. Endorse the role of the Vocabulary Team, and develop and implement term governance.
2. Publish and operationalize the metadata framework. Continue to build out metadata fields.
3. Develop a strategy and priorities to build out taxonomies and implement the strategy.
4. Publish an integrated glossary and establish a glossary management practice.
5. Develop a method to integrate the WSDOT Thesaurus into agency search tools. Develop a strategy and priorities for continued development of term relationships

The remainder of this document describes the five interrelated aspects of a vocabulary management practice for WSDOT and provides more detail for the recommendations described above.
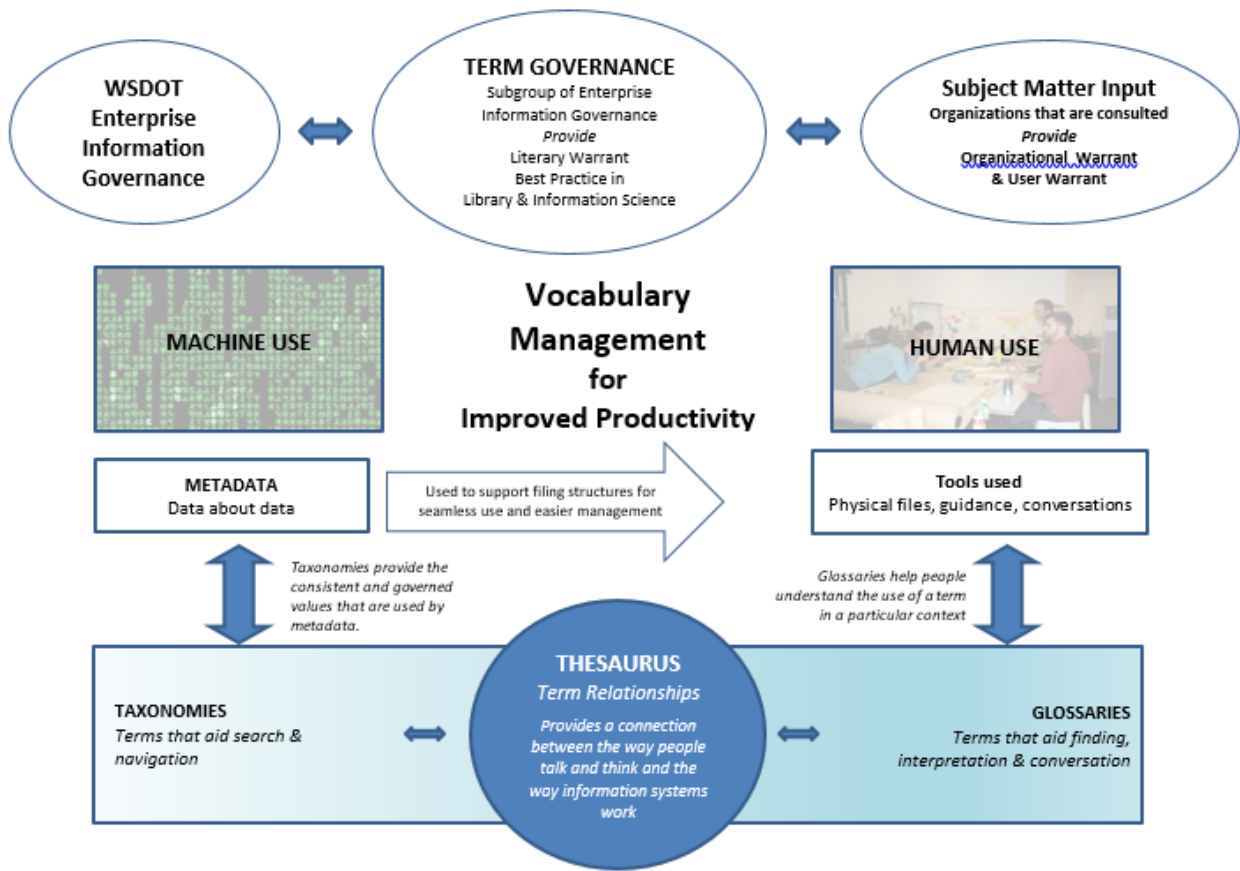
Figure 1. The relationship of vocabulary components and their use by people and computers.

Vocabulary Governance

WSDOT has an executive-level Enterprise Information Governance Group (EIGG) with responsibility for governance of data and information repositories, processes, business needs and technology and applications. The EIGG is currently in hiatus as the agency seeks to clarify governance expectations, roles and practices.

Vocabulary management, in order to have an enterprise-wide impact, requires oversight. A vocabulary governance practice is needed to maintain consistency in managing essential business terms to ensure employees, stakeholders, and computers can efficiently find and navigate the data and information they are seeking.   Because vocabulary management is a unique discipline, a governance group for terms and term use should be established as an independent group working under the umbrella of the EIGG.

To shape recommendations for vocabulary governance, policies and procedures from the World Bank, a Kent State University Student Study (Kennison, 2014, unpublished), SeaVox (moderated by the British Oceanographic Data Centre), and others were reviewed along with the draft proposed glossary governance guidance discussed with the Practical Solutions Working Group. These were found to be specific about governance process (adding and changing terms), but less specific about roles, responsibilities and selection of personnel to fulfill those roles. Figure 2 identifies the governance objectives for WSDOT's EIGG proposed by the Kent State Student Study.
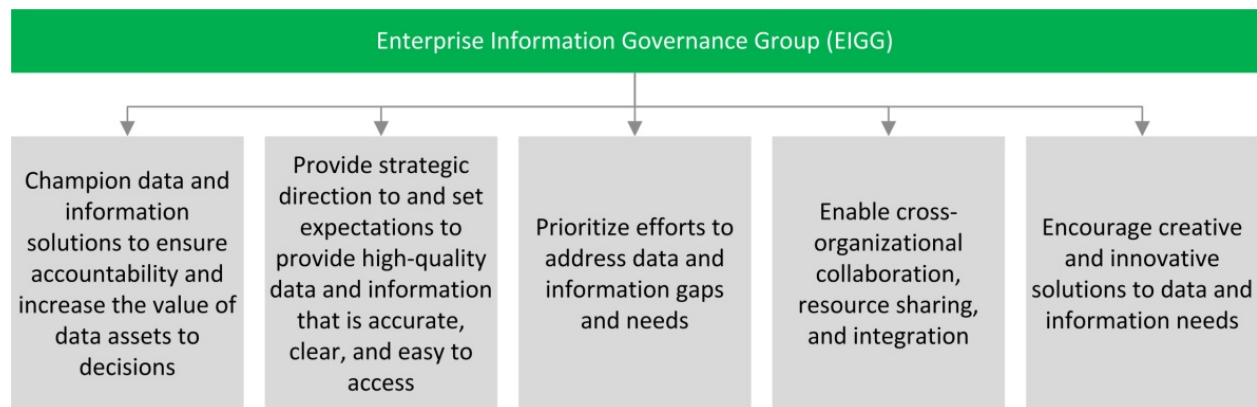

Figure 2. Objectives of the WSDOT Enterprise Information Governance Group. Kent State Study on Governance

The Vocabulary Team established through the PS AID Project is intended to guide governance efforts and promote the use of industry standards for implementing WSDOT's vocabulary management. Vocabulary Team members have technical expertise in classifying information (i.e. Library and Information Science) and are the current members of the vocabulary governance working group submitting this recommendation. It is expected that Vocabulary Team membership will evolve as practices are tested and improved, while retaining a strong core in library and information science in order to take advantage of that discipline's wealth of knowledge and successful practice for finding and retrieving relevant resources.

The governance practice will establish roles and responsibilities for decision-making for new terms and definitions, modification of terms, and implementation of management practices. The Vocabulary Team

will work with domain stewards and subject matter experts to determine best "light touch" practices for incorporating their input. A community of interest may be developed as a forum for discussion with employees interested in vocabulary management.

Domain stewards and practitioners within subject domains, utilizing and generating various specialized kinds of information and data, will advise the Vocabulary Team. Working together, they will achieve these basic elements of term management:

- Document currently used terms for key concepts
- Identify and designate preferred terms for concepts
- Strengthen definitions, distinguishing them from descriptions
- Determine and document term relationships

Vocabulary governance provides a department-wide structure but it does not necessarily mandate a single set of definitions across WSDOT. The governance practices should identify common governance needs and whether there are unique governance aspects for each of the elements of language management. The objective is to capture terms for concepts that are critical for agency-wide understanding and to allow specific subject domain vocabularies to be integrated into a broad-reaching and universal structure that represents all facets of business across the department.

One governance process will be used for terms used in the agency metadata, glossary, taxonomies, and thesaurus. Organizational structure and application of the terms may vary based on the way it is applied. The WSDOT Thesaurus will serve as the hub for the process, collecting terms and their relationships used in the agency metadata, glossary, and taxonomies.

Recommendations

- Endorse the Vocabulary Team to steward agency practices for vocabulary management in accordance with industry best practices.
- Establish a workflow for term submission, vetting, and selection.
- Establish business domain groups that comprise representatives from all organizations with responsibilities and accountabilities for the domain. The domain groups should provide subject matter expertise on the topic to vet policy, procedure, information and data resources, roles, and terms and definitions. When possible, identify existing groups that can serve this function. Strive for multimodal and multifunctional groups as appropriate for the breadth of the domain.
- Establish baseline metrics on performance and monitor changes as vocabulary improvements are made.
- Establish a maturity model based on available resources.
- Develop a change management and communication plan for vocabulary management.

Metadata

Metadata are a cornerstone for navigating digital resources. "Data about data" are critical for efficiently sifting through large volumes of information. They are a foundation in aiding search, navigation, and classification. WSDOT has previously established metadata standards for individual information resources and these implementations have some consistency, but there are also many gaps in application. A Publication Abstract metadata set was developed and applied to WSDOT Research Reports. The impact of the change was tested with a user group and found to improve findability of requested research reports. Due to the positive results, the Publication Abstract metadata set was applied to WSDOT's Engineering Publications.

Through the PS AID Project, the Vocabulary Team developed a framework for WSDOT's metadata. This metadata framework defines a standard for core metadata that should be applied to all information resources within the department. The framework is also extensible, with additional modules that capture the unique attributes of specific categories of information, such as Geospatial, Open and Reference Data. The WSDOT Metadata Framework builds on lessons learned from previous efforts and current practices.

National and international metadata and description standards from various disciplines were compiled and compared by the Vocabulary Team. The Dublin Core Metadata Element Set, an internationally recognized standard, was selected as the foundation for WSDOT's metadata schema. The team also referenced two previously developed in-house metadata projects: the Kent State Metadata project and the Publications Abstract metadata set. The Kent State study prepared a metadata schema to support discovery for public records requests. The Publications Abstract is a WSDOT metadata set created to improve findability on our public website for engineering and administrative manuals, and for research reports. Additionally, the team examined the Washington State Geospatial metadata, Encoded Archival Description (EAD), Machine-Readable Cataloging (MARC) and Project Open Data recommendations. The Vocabulary Team sought solutions that limited the number of metadata fields required while enhancing search and navigation capabilities. This review lead to the development of a WSDOT Core Metadata Framework.

The metadata strategy involves metadata elements that highlight specific aspects of the content being described, augmented by taxonomies. Taxonomies are described in the section below. They provide consistent and governed values for metadata elements. The WSDOT Metadata Framework provides a roadmap for taxonomies that need to be developed.

The proposed metadata scheme is built in modules: Core Basic and Core Complete

- The Core Basic Metadata identifies seven elements that apply to all agency information resources. While the primary application will be to support digital navigation, the WSDOT Metadata Framework can be used to support organization of paper files and other artifacts. Using the same structure for these physical resources will improve the seamless navigation between all resources.
- Core Complete identifies elements that improve relevance of information retrieval for a broad set of digital resources within and across shared business interests.

Additional metadata that are specific to different content and digital resources have been created or leveraged from existing sources.

- The Federal Geographic Data Committee (FGDC) Content Standard for Digital Geospatial Metadata (CSDGM) has been used to describe geospatial data at WSDOT for purposes of data sharing. Additional metadata to describe ESRI Feature Classes has been developed by Data Management Services to refine the information about a feature's location. A sample is described in Appendix A.

- The WSDOT Open Data Committee adopted a metadata schema based on the White House-developed Project Open Data Metadata Schema to describe any data set within WSDOT. The metadata schema can be found in WSDOT's Open Data Plan submitted to Washington State's Office of the Chief Information Officer.

- Transportation Asset metadata applies to information about physical transportation system assets.

Other business specific metadata extensions will be developed as needed to help improve the findability and relevance of information so that it can be retrieved and used. Standards will be updated as new ones emerge and are adopted by WSDOT.

The WSDOT Metadata Framework elements are provided in Appendix A. A separate document titled "WSDOT Metadata Core Documentation" provides a more detailed description.

Some attributes of the framework are supported by existing reference data lists, other reference data lists need to be developed, and some attributes will be supported by taxonomies. The WSDOT Core Metadata Framework is being applied to some information resources that are in development. In addition to the development and maintenance of new reference data and taxonomies, future efforts will investigate opportunities to digitally harvest metadata.

Recommendations

- Publish and operationalize the metadata framework

- Develop an implementation strategy for Core Metadata that applies to all digital resources. The implementation strategy should address application to existing resources as well as resources in development. Determine resourcing needs and implement as feasible.

- Devise a prioritization schedule for developing taxonomies to support metadata fields. Determine resourcing needs to implement the strategy and initiate implementation.

Taxonomies

Taxonomies help the user see related information and select specific distinctions within that grouping of terms. Retail websites such as Amazon, where users can progressively narrow their search by selecting nested categories of information within the search return, use robust taxonomies that group like information and classify the distinct groupings within that material. The result, in a transportation example, is to have the ability to search *assets* and narrow the selection to distinguish major asset categories (e.g., highways, ferries, …), asset class (e.g., facilities, IT, etc.), or asset groups (e.g., buildings, sites), asset subgroups (e.g., storage building, office building, etc.), the asset itself (e.g., the specific storage building), or component of the asset (e.g., substructure, services, etc.).

Librarians have traditionally built taxonomies. As personal computing and Internet resources have expanded, other information professionals have developed taxonomies to guide system or site users to the content they are seeking. Computer applications have also been developed to build taxonomies based on the terms and patterns of users. Automated and manual techniques can be combined to create meaningful taxonomies. Website users can also tag sites (folksonomy) to guide other users to a resource. There are benefits and limitations of each of these methods.

Taxonomies can aid information sharing within and across business units. They should align with the objectives and scale of the agency's enterprise architecture. For example, topics that collate information from a variety of organizations, such as Practical Solutions, may benefit from an enterprise scale taxonomy. Other topics, such as safety, may be managed as a cross-organizational matrix; still more may be unique or highly standardized, such as purchasing or legislative review activities. A special type of taxonomy, a business function classification scheme (BFCS), helps organize essential work of the department and promotes cross-organizational connections. A business function classification is described further at the end of this section.

Taxonomies are not one size fits all. It is very important to understand whether a topic is used by the whole agency or defined groups or individual offices to support business needs. At present, a limited number of taxonomies exist in the department and the majority of these address localized business needs.

Taxonomies provide the values for metadata elements that would benefit from consistent, repeatable, and managed values. This allows for certain metadata elements to have controlled values such as master data lists or terms organized by hierarchical relationships., Managed values help eliminate findability errors caused by misspellings and providing consistency in indexing information objects. Taxonomies mapped to a metadata element allow us to highlight and classify different aspects of the content of a document. Some controlled vocabularies already exist as agency reference data such as WSDOT Region and WSDOT Organization. The WSDOT Metadata Schema becomes a roadmap to develop taxonomies.

Recommendations
- Develop an implementation strategy for development and application of agency taxonomies. The strategy should address enterprise taxonomies, group taxonomies, and unique taxonomies,

and include processes to register all agency taxonomies and vet terms appropriate to the scale of use. Determine resourcing needs and implement as feasible.

- Analyze existing taxonomies to gain a better understanding of vocabulary needs and opportunities for reuse of existing resources.

- Build the enterprise taxonomies outlined in the WSDOT Metadata Core Schema.

- Pursue development of Tier 1 and 2 of a business function classification scheme and an implementation strategy for continued development.

A Specific Type of Taxonomy: The Business Function Classification Scheme

A separate report titled *Enterprise Architecture: An Overview* describes strategies for an information architecture that supports core organizational objectives. One of the strategies recommended is the development of a particular type of taxonomy called a business function classification scheme (BFCS).

A business function classification scheme**:**
- Provides a framework for organizing information to meet core business requirements.
- Presents a stable and persistent view of what the organization does – historically, currently and into the future. This taxonomy may change when the nature of the business changes, but not when organizational structures, the accounting structures, strategic plans or the technologies change.
- Enables categorization of information assets from the "business process" perspective.
- Pulls together content from multiple sources and systems into a business-function view.
- Enables searching for information, data, services, content, and people from a business-function perspective.
- Supports workflow across applications by providing a common framework.
- Supports software interoperability by providing an enterprise view into the different functional applications.

The *World Bank Guiding Principles for Creating, Managing and Maintaining an Enterprise Business Function Classification Scheme (BFCS)* describes the BFCS in the following way:

The purpose of the Business Function Scheme is to provide a structure that describes 'how the Bank does its work.' This focus is different from Topical or Organizational schemes, which describe the subject areas in which the Bank works, or the way the Bank organizes itself to accomplish its work…

The World Bank organized their BFCS into five Tiers
- Business Areas represent the organization's high level strategy and performance goals. Business Areas are defined by the strategic direction of the organization. Business areas describe the Bank's business at its highest level of aggregation. Business Areas are represented as Level 1 in the BFS.
- Bank's Lines of Business (LOB). The second level categories represent the specific Bank's Lines of Business (LOB). A line of business is defined in terms of lines or groups of products or services which are produced by a set of business processes. Lines of Business reflect an organization's strategic choices and levels of business risk, and are closely aligned with the organization's performance management. Lines of Business are represented as Level 2 in the BFS.

The Business Function Classification - *continued*

- Business Processes. The third level categories represent business processes. A business process is defined as a set of one or more linked procedures or activities which collectively realize a business objective or policy goal, normally within the context of an organizational structure defining functional roles and relationships. Each Business Process can be further decomposed into multiple Business Subprocesses or directly into business activities. Business Processes are represented as Level 3 in the BFS.
- Business Subprocess is a process that is enacted or called from another (initiating) process, and which forms part of the overall (initiating) process. Business Subprocesses are represented as Level 4 where they exist.
- Business Activity is a description of work that forms one logical step within a process. An activity is the smallest unit of work which is scheduled in a process, and may result in multiple work items being assigned to a participant or actor. Business Activities are represented as Level 4, except where a Subprocess exists. Where a Subprocess exists, they are represented as Level 5.

The WSDOT Vocabulary Team reviewed organizing structures used within WSDOT but did not find a similar classification scheme in use in the agency. They examined options for organizing agency business functions and began developing and drafting a BFCS. The first draft used the WSDOT Chart of Accounts as the basis. The draft was vetted with a broader group of information managers and the Transportation Data GIS and Modeling Office. This review indicated that the BFCS was too granular to provide an enterprise-wide framework and aligned better with deeper tiers of the BFCS (Tiers 3, 4, 5). The Vocabulary Team reviewed several more resources related to state and organizational mission and goals including: the Practical Solutions Lifecycle, WSDOT organizational charts, transportation goals established by the Washington State Legislature, Results Washington Goals, and Results WSDOT Goals. Based on this review, five information objectives were identified that support implementation of the Practical Solutions Lifecycle:

1. **Multimodal Transportation System Information**: Information about the multimodal transportation system is core to all business functions and must be managed as an enterprise resource.

2. **Multimodal Transportation Services:** We must be able to know and provide information to travelers about the current status of transportation services on the state system.

3. **Multimodal System Management:** We must be able to access and share information on the transportation system infrastructure, system performance, and local context in order to efficiently and collaboratively operate, maintain, preserve, and improve the state transportation system.

4. **External Relations:** We must manage information on customer relationships and commitments in order to ensure consistent information and support inclusive practices.

5. **Resource Management**: We must be accountable and demonstrate that our practices are efficient and effective.

If WSDOT pursues use of a business function classification scheme the Vocabulary Team will need to work with subject matter experts to refine the draft tier descriptions, align the classification with agency practices, finalize Tier 1 and 2, and develop an implementation strategy. Deeper levels of the BFCS may align with products developed through business analysis and value stream mapping. Review of these products should be included in the implementation strategy.

Enterprise Glossary

One of the most common vocabulary tools in use at WSDOT is the glossary: a list of words with brief definitions, usually related to a specific topic or function. Currently, multiple offices within WSDOT independently produce and manage their own glossaries, frequently associated with technical manuals. These provide an important tool for users within specific business areas, but they are not created to a common standard or with cross-organizational coordination. In some cases, definitions help support compliance with legal requirements and inconsistent definitions can lead to reporting challenges. The separate vocabularies include identical terms that mean different things. Different meanings are sometimes necessary for a particular discipline or use but add complexity to understanding and navigation. An enterprise glossary helps users understand the different uses and provides information for subject navigation.

Consolidating glossary creation and management will not only reduce redundancies where multiple offices share the same definitions, but provide an opportunity to document divergent language usage and establish common understanding of terminology within and across business areas.

For instance, the term 'ramp' is commonly used both in relation to roadways, to mean a short approach or connection, and in relation to pedestrian spaces, to mean a walking surface with less than 5% slope. Similarly, in different contexts, 'buffer' might refer to a curb setback or to an area of land designated to protect aquatic resources. While 'ramps' and 'buffers' are fairly simple examples, distinguishing between other terms is more difficult, such as 'certification,' the meaning of which varies significantly between the *Consultant Services Manual* (a release of state jurisdiction to local government) and the *Right of Way Manual* (a prerequisite for construction). A well-constructed glossary documents all these uses, allows users to distinguish between them, and, in the case of concepts like pedestrian ramps or different forms of certification, details the specific criteria used to define them.

Glossaries are important because they are already a recognized format within the department for managing selected terms and their meanings. Glossaries help capture the natural language of an organization and support socialization of terms beyond a business unit. A unified glossary will be the foundation of vocabulary governance at WSDOT. Accordingly, it must be accessible, understandable and current in order to facilitate successful communication and information management.

The Enterprise Glossary identifies terms that are commonly used throughout the department and are critical for understanding strategic initiatives and essential activities. It defines terms used by business areas in their work. There may be more than one definition for a term if the context for use varies.

Terms relevant to Practical Solutions were collected through meetings associated with business process map development, training courses, and requests. These terms were collected into a spreadsheet and research was conducted to identify definitions for these terms in federal and state law, agency manuals, data definitions, and other information resources. This research identified some conflicting definitions, gaps in term definitions, and inconsistency in the approaches and rigor used to develop definitions. Further work is needed to develop a reader-friendly version of the glossary and to vet definitions.

The ISO standard for definitions was reviewed and guiding principles for glossary management were proposed to the Practical Solutions Working Group. A vocabulary governance model needs to be established in order to provide a process for the collaborative development of glossary material between SMEs and the Vocabulary Team. A consolidated glossary of terms referenced in agency manuals has been developed and work is underway to publish this on WSDOT's Internet site.

Preferred Terms

Best practices in library and information science rely on analyzing vocabularies to support cataloging and retrieval of information. Analysis includes:

- identifying the crucial ideas and things under consideration (concepts) and the names we use for them (terms)
- examining usage in literature, practice, and organizations (warrant) to determine which term will be the primary point of reference in the organizational system (the 'preferred term').

Technically, all concepts get a preferred term, but this only becomes significant when multiple terms are used for the same concept as, for example, crash/collision/ accident. We use 'crash' as the main entry in the vocabulary system, but the synonyms 'collision' and 'accident' are also present and reference 'crash' as the preferred term. A preferred term that connects all of them allows us to make the search process much more effective.

This does not mean that the preferred term is 'correct' and that only it can be used. Identifying preferred terms and their relationship to related terms makes it easier to find information. When an information system uses this practice, a person searching for a document with 'accident' can find the document even though it uses 'crash' or 'collision' instead of accident.

Preferred words may change over time but term relationships will persist. Managing vocabulary allows a person searching for historical information with contemporary language to find relevant information.

Recommendations

- Publish the enterprise glossary.

- Develop an implementation strategy and procedures to incorporate more glossaries used across WSDOT, improve alignment of terms and definitions, provide for modification of terms, and improve access to glossaries for employees and stakeholders. Determine resourcing needs and implement as feasible

- Identify priority needs to improve agency glossary resources.

Enterprise Thesaurus

A thesaurus provides information about a term and its relationships to other terms within the same thesaurus. In addition to clearly specifying which terms can be used as synonyms or antonyms, a thesaurus also indicates which terms are more specific (narrower terms), which are broader, and which are related terms. A thesaurus is best managed at the enterprise scale to support topical organizational connections.

A thesaurus, therefore, provides a map of how natural language is used. This map of term relationships allows automated search to provide more comprehensive search returns and improve precision of the information found. A basic search engine can only interpret queries as a string of characters, but a well-developed and implemented thesaurus allows for the aggregation not only of matching character strings, but also of identified synonyms and related terms. Additionally, this helps retain the connection of legacy information resources to current terminology that would be uneconomical to update with full metadata.

Thesaurus development is based on the principles outlined in ANSI/NISO Z39.19-2005 *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*. A WSDOT Thesaurus is under development to formally structure the concepts and terms describing WSDOT's topical knowledge domains and areas of expertise: the "what-we-do" and "what-we-know" aspects of our work. The WSDOT Thesaurus provides an enterprise-wide, application-independent framework for describing all of the areas of expertise and knowledge domains represented in WSDOT material, current as well as historical. It will represent the vocabularies used by domain experts and domain novices, WSDOT staff, and clients.

The WSDOT Thesaurus is comprised of three basic elements:

- controlled vocabulary
- terms
- semantic relationships

Recommendations

- Develop a method to integrate the WSDOT Thesaurus into agency search tools.

- Develop a strategy, procedure, and priorities for continued development of the enterprise thesaurus. Determine resourcing needs and implement as feasible.

- Pilot the use of the thesaurus and monitor the impact to the findability and relevance of agency information resources.

Summary

Words, their definitions, and their relationships are building blocks for effective coordination and collaboration in both human conversations and digital navigation and automation. WSDOT's vocabulary components have evolved to address specific business needs. As the expectation for collaboration and coordination has increased, there is greater need to work across traditional business lines. This is increasing the need for common or context specific definition of terms. The field of information science has proven tools for managing term (word) associations. This document recommends a vocabulary management practice that builds on those successful practices to enable and sustain common understanding and improved access to relevant information. These activities support machine learning by providing "rule based" guidance to enhance navigation. Machine developed taxonomies (natural language processing) can inform the work described above.

A mature vocabulary management practice will improve access to business critical knowledge, help identify potentially duplicative or conflicting information, and provide a foundation for a nimble and navigable information resource for employee and community engagement.

Appendix A – Metadata Schema

WSDOT Core Basic

The WSDOT Core Basic is a minimum required set of metadata that describes resources within WSDOT.

- Title
- Description
- Keyword
- Content Type
    - Content Type Classification Scheme (in progress)
    - Assists in identifying different metadata extensions to the WSDOT Metadata Core
- Contact
    - Contact Refinement
        - Contact Name
        - Contact Role
        - Contact Job Title
        - Contact Email
        - Contact WSDOT Region
        - WSDOT Organization (either Office or Division)
        - Contact Organization (non WSDOT)
        - ORCID (a unique ID for each individual that persists throughout job changes)
- Create Date
- Date Modified

WSDOT Core Complete

The WSDOT Core Complete extends the Core Basic to include terms to describe all information resources at WSDOT. The Core Complete improves the context specificity of an information resource, improving the relevance of search returns. The Vocabulary Team recommends using Core Complete whenever possible.

- WSDOT Core Basic
- Topic
    - Business Topic Classification Scheme (to be completed)
- Business Function
    - Business Function Classification (In process)
- Disposition Authority Number
    - Existing Controlled Reference (Records Management)
- WSDOT Region
    - Existing Controlled Reference (Data Management Services)
- WSDOT Organization
    - Existing Controlled Reference (Data Management Services)
- Physical Location
    - In cases where resource is Physical, a narrative to where the resource can be found.
- Format
    - Existing Controlled Reference (Internet Assigned Naming Authority)
- Service URL
    - Uniform Resource Identifier where the resource can be found for use or download

Open Data/Data Set Metadata Schema

A metadata schema adopted by the WSDOT Open Data Committee to describe Open Data Sets and used to describe other data sets within WSDOT, as well. The schema is based on the federal Project Open Data Metadata Schema and includes elements from Washington State OCIO's Open Data Metadata Standard Data Set Metadata. This metadata schema is comprised of two parts that 1) describe the data set and 2) describe the distribution of the data set.

Data Set Metadata:

- Core Basic
- Posting Frequency
- Originator
- Published Date
- Data Sharing Agreement Flag
- Period of Time
- Theme
- Access Level
- Geospatial Metadata URL

Distribution Metadata:

- Title
- Description
- ArcGIS Service Category
- ArcGIS Service Geometry
- Service URL
- Described By
- Described By Type
- Format

Geospatial Feature Class Metadata Schema

A set of metadata created by Data Management Services to describe all new data layers originating within WSDOT referred to as the Feature Class Template. There are three templates based on Geometry: Line, Polygon and Point. Point is shown below. The set is in addition to the Content Standard for Digital Geospatial Metadata (CSDGM) from the Federal Geographic Data Committee, used to describe geospatial data for sharing.

Note: While none of the following metadata elements are required, they are recommended when describing geospatial data.

Feature Class Template Point:

- Route Identifier
- State Route Number
- Related Route Type
- Related Route Qualifier
- State Route Milepost
- Begin State Route Milepost
- End State Route Milepost
- Ahead Back Indicator
- Begin Ahead Back Indicator
- End Ahead Back Indicator
- Accumulated Route Mile
- Begin Accumulated Route Mile
- End Accumulated Route Mile
- Label
- Symbol Code
- Inventory Direction
- Azimuth
- Elevation
- Easting
- Northing
- Latitude
- Longitude
- City Name
- County Name
- Region Name
- Maintenance Area Name
- Program Item Number
- Work Item Number
- Record Create Date]
- Record Create User ID
- Record Update Date
- Record Update User ID
- Publish Date
- Snapshot Date
- LRS Date
- Begin Effective Date
- End Effective Date
- SHAPE

Project Metadata

The proposed set of metadata for project related resources is based on a Kent State University Student Study (Lee, 2015, unpublished). The study investigated common public disclosure requests for project information and proposed metadata to support improved search returns for responsive documents. This has not been completely built out and vetted.

- Project Identifier
- Project Category
    - In development
- Project Phase
    - Existing Controlled Reference (Data Management Services)
- Funding Type

Transportation Asset Metadata

A proposed set of metadata for transportation system assets based on a Kent State University graduate student Capstone project to draft a taxonomy for multimodal transportation system assets. "*Physical objects of value that have a transportation purpose or function that often deteriorate or become obsolete over time." (Transportation Assets Definition created by Capstone Project Advisory team members)* Note: Both Thesaurus and Classification Scheme have only been vetted by the Advisory team. Broader vetting needs to occur.

- Transportation Asset Classification Scheme (Taxonomy)
    Draft scheme developed
- Transportation Asset Thesaurus
    - Draft scheme developed.

Appendix B – Draft Governance Framework for Glossary & Thesaurus Terms

The following is a governance model to facilitate the development of glossary and thesaurus content. The overall structure will be consistent with ANSI/NISO Z39.19-2005 *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*.

Term Selection

Term Selection will follow the guidelines as outlined in ISO 704-2009 *Terminology Work – Principles and Methods*

The process of selecting terms for inclusion in controlled vocabularies involves consulting various sources of words and phrases as well as criteria based on:

- • The natural language used to describe content objects (literary warrant),
- • The language of users (user warrant), and
- • The needs and priorities of the organization (organizational warrant)

Term Selection and Preferred Terms

Term selection in building taxonomies and thesauri is based on analyzing a domain such as transportation on several merits or warrants:

- Literary – What language is contained in books, manuals, documents within the transportation domain and specifically WSDOT?
- User – What language do people within the domain use to describe a concept?
- Organizational – what are the priorities and needs of the organization?

Term selection is using these warrants to select the most suitable term or preferred term. Preferred terms are the terms that are used to tag or index a piece of content (document, image etc.). A Non-Preferred Term or Synonym is used to map the searcher or user's language to the preferred term to improve search results.

Example:

- **Crash** is WSDOT's preferred term for the concept of "An occurrence of injury or damage involving at least one Motor Vehicle or Pedalcycle for which a Police Traffic Collision Report (PTCR), Vehicle Collision Report (VCR) or a report via SECTOR is filed."
- **Collision, Accident, Wreck** may be other words that people may use to refer to the same concept.

Definition Selection

The Practical Solutions AID Project team has created a hierarchy for how definitions are selected for each word or term. The definitions from sources in the higher levels of the hierarchy will receive precedence over those from lower levels of the hierarchy. If a higher-level hierarchy term exists and the Vocabulary Team opts not to use it, the reasons will be documented.

Hierarchy of Term Definition Selection
1. Legal Definitions (Federal and State)
2. Industry Standards
3. WSDOT Manuals

4. WSDOT Data Catalog (DOTS) or other vetted WSDOT resources
5. Other WSDOT resources (not broadly vetted)
6. Other state DOTs or key stakeholders
7. General Dictionary Definitions

Relationship of Terms

Where possible, we tried to provide the relationship that exists between terms (preferred term, broader terms, narrower terms, related terms, and synonyms). We will rely on stakeholders, TRT (Transportation Research Thesaurus) and other thesauri to help us identify these relationships. Note: UF – Used For; BT – Broader Term; RT – Related Term

Example: (based on ANSI/NISO Z39.19-2005 display guidelines)

Multimodal

UF Multimodal Systems,

BT Transportation

RT Intermodal Transportation

Term Review

The Vocabulary Team will guide implementation of best available content management practices and facilitate subject matter expert review.

Term selection will be based on sound information science informed by subject matter experts. The Vocabulary Team will work with agency management to clarify organizational responsibilities for Business Function and Subject Domain groups. It is anticipated that agency management will use the RASCI model to distinguish the following five roles:

- R = Responsible – responsible for action, owns the problem / project
- A = Accountable – approves or signs off on work before it is finalized
- S = Supports - provides resources or plays a supporting role in implementation
- C = Consulted - has information and/or step necessary to complete the work
- I = Informed - must be notified of results, but need not be consulted

The Vocabulary Team will seek input from WSDOT organizations that are responsible for a subject or function. It is expected that those responsible for a subject or function will coordinate with other organizations that are accountable for a subject or function.

The Vocabulary Team will review alternate definitions with subject matter experts to develop an understanding of variances in definitions and seek opportunities to reconcile similar uses. Alternate definitions will be captured if they are essential to the business process.

The Vocabulary Team will use this input to guide the selection of terms and define term relationships.

<u>Change Management and Process for Enterprise Vocabulary Management Stewards</u>

Initiating a request for a new entry or a change to an existing vocabulary product will be done by:

- Employees – may suggest terms to enhance findability or clarify term relationships.
- Domain or Managerial Stewards – especially in response to data stewards and staff- can request additions or changes to an existing vocabulary.
- Vocabulary Team- may suggest changes based on periodic review of terms as used in agency.
- Review of commonly used Public Disclosure Request terms, search terms and terms associated with library reference requests.

Events that trigger changes in vocabulary products (glossaries, taxonomies):

- Emergent subject matter or process
- Retiring a function, process, or subject
- Regulatory change
- An expansion in coverage of the topic domain
- A shift in focus which changes the preferred/non-preferred status or relationship between terms

Changes are evaluated primarily through a comparison of the request, its impacts on use, a review of the information resources to which it will be applied, and an analysis of other resources to which it may need to be applied

Change process must also consider cost of implementing the change:

- Impacts to systems consuming the scheme and coding changes required;
- Updating or creating reference maps
- Updating automatic classification tools
- In close collaboration with the requester and those identified as having responsibility or accountability for the data effected
- In collaboration with other Domain and Managerial Stewards who may be impacted by the change
- Evaluation of the impact on enterprise use

Appendix C – Examples of Glossary Improvement Opportunities

Similar definitions that could be reconciled

- ADA (Design Manual): An abbreviation for the Americans with Disabilities Act of 1990. The ADA is a civil rights law that identifies and prohibits discrimination based on disability. Title II of the ADA requires public entities to design new pedestrian facilities or alter existing pedestrian facilities to be accessible to and usable by people with disabilities

- ADA (Local Agency Guidelines): The Americans with Disabilities Act of 1990 which mandates sweeping changes in building codes, transportation, and hiring practices to prevent discrimination against persons with disabilities, not just in projects involving federal dollars, but all new public places, conveyances, and employers.

- Americans with Disabilities Act of 1990 (ADA) (Summary of Public Transportation): Federal civil rights law that assures persons with disabilities get equal opportunity to fully participate in society, the ability to live independently, and the ability to be economically sufficient.

Same word, different uses

- Amendments (Bridge Design): are revisions to specific sections of the Standard Specifications, which are approved and enacted during the two year period that a specific edition of the Standard Specifications is in force. Amendments are published normally three times during a calendar year – April, August, and December.

- Amendments (Consultant Services): A written document used to modify the contents of an existing task order document previously issued under a master on-call agreement. An amendment may be used to add new elements, make up for a deficiency, or extend or strengthen the task order.

Similar meaning, different terms

- Area of Influence (Design Manual): The area that will be directly impacted by the proposed action: freeway main line, ramps, crossroads, immediate off-system intersections, and state and local roadway systems.

- Area of Potential Effect (APE) (Environmental Manual): The geographic area or areas which an undertaking may directly or indirectly cause alterations in the character or use of historic properties. The APE is three dimensional including auditory, visual and ground disturbing activities. The APE is influenced by the scale and nature of an undertaking and may be different for different kinds of effects caused by the undertaking. The APE should be defined before historic properties are identified and not on land ownership (36 CFR 800.2(c)).

Same word, conflicting definitions

- Ramp (Design ADA/Pedestrian): A walking surface with a running slope steeper than 20H:1V (5%).

- Ramp (Bridge Design): Any directional route of pedestrian travel having a running slope between 2% and 8.3%.

- Sand (Geotech): Particles of rock that will pass through a 0.19 in. (4.75 mm) opening, but will not pass a 0.003 in. (0.075 mm)

- Sand (Roadside): A mineral soil particle between 0.05 and 2.0 mm in diameter. A soil textural class.

Description rather than definition

- Coastal Zone Management (Local Agency Guidelines): Applicants for federal permits or licenses must certify that their project will comply with the State Coastal Zone Management Program (Shoreline Management Act -RCW 90.58 - applies to projects within 61 m (200 feet) of a shoreline).

Obsolete Definition

- Planning and Programming (P&PSC) (Local Agency Guidelines): WSDOT's branch responsible for coordinating with local agencies on planning issues.

Incorrect and descriptive

- HPA (Local Agency Guidelines): Hydraulic **Power** Approval permit is issued by the Washington Department of Fish and Wildlife. The Hydraulic Code (RCW 75.20.100-160) requires that any person, organization, or government agency wishing to conduct any construction activity in or near state waters must do so under the terms of a permit (the Hydraulic Project Approval – HPA, to be exact) issued by the Washington State Department of Fish and Wildlife. State waters include all marine waters and fresh waters of the state.

Inconsistent presentation of definitions based in RCW or WAC

- Water Quality Standards (Highway Runoff): The minimum requirements for water purity for uses like drinking water supply, contact recreation (such as swimming), and aquatic support (such as fishing). The Washington State Department of Ecology sets water quality standards for Washington State. Surface water and groundwater standards are established in WAC 173-201A and WAC 173-200, respectively.

- Water Quality Standards (Roadside): The discharge of polluted waters to waters of the state is a violation of the Washington State Water Pollution Control Law (RCW 90.48.080). Ecology establishes surface water quality standards (WAC 173-021A) as required by the Environmental Protection Agency (EPA). These standards are intended to protect surface waters of the state for beneficial uses including: public health, enjoyment, and the propagation and protection of fish, shellfish, and wildlife. State water quality criteria are defined for classifications of waters and include turbidity, fecal coliforms, dissolved oxygen, temperature, and pH. All water discharged from construction sites shall meet the water quality standards for turbidity at the discharge point into the surface water resource. There is not a "dilution" or "mixing zone." Turbidity shall not exceed 5 NTUs over the water resource's turbidity when the water resource's turbidity is 50 NTUs or less, or have more than a 10 percent increase in turbidity when the water resource's turbidity is more than 50 NTUs

References

Bedford, Denise. *Enterprise Information Architecture: An Overview.* 2018. Forthcoming.

Content Standard for Digital Geospatial Metadata. FGDC-STD-001-1998. Washington, D.C.: Federal Geographic Data Committee, 1998.

Feldman, Susan. *The High Cost of Not Finding Information.* KMWorld 13(3), March 2004. KM World (http://www.kmworld.com/Articles/Editorial/Features/The-high-cost-of-not-finding-information-9534.aspx)

Geospatial Metadata Standard. Standard No. 161.02. Olympia, WA: Office of the Chief Information Officer, Washington State, 2014, OCIO Policy (https://ocio.wa.gov/policy/16102-geospatial-metadata-standard).

Gill, Tony, Gilliland, Anne J, Whalen, Maureen, and Woodley Mary S., *Introduction to Metadata Version 3*. Ed. by Murtha Baca. Los Angeles, CA: Getty Research Institute, 2016. Getty Research Institute (http://www.getty.edu/publications/intrometadata).

Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies. ANSI/NISO Z39.19-2005 (R2010). Bethesda, MD: National Information Standards Organization, approved 2005, reaffirmed 2010. ANSI/NISO Standard (http://groups.niso.org/apps/group_public/download.php/12591/z39-19-2005r2010.pdf).

Guiding Principles for Creating, Managing and Maintaining an Enterprise Business Function Classification Scheme (BFS). The World Bank, 2007.

Hagedorn, Kat. *The Information Architecture Glossary*. Argus Associates, 2000. The Information Architecture Glossary (http://argus-acia.com/white_papers/ia_glossary.pdf).

Harpring, Patricia. *Introduction to Controlled Vocabularies: Terminology for Art, Architecture, and Other Cultural Works*. Ed. by Murtha Baca. Los Angeles, CA: Getty Research Institute, 2010. See Chapter 2: What Are Controlled Vocabularies? Introduction to Controlled Vocabularies (http://www.getty.edu/research/publications/electronic_publications/intro_controlled_vocab/what.pdf).

Harrison, Frances. *Improving Information Management for Practical Solutions at WSDOT.* 2018. Forthcoming.

Information and documentation -- Records management -- Part 1: Concepts and principles. ISO 15489-1:2016(E). Geneva: International Standards Organization, 2016. Geneva International Standards Organization (https://groups.niso.org/apps/group_public/download.php/18410/NISO_TR-06-2017_Issues_in_Vocabulary_Management.pdf) .

Information technology -- Metadata registries (MDR) – Part 5: Naming and Identification Principles. ISO/IEC 11179-5:2015(E). Geneva: International Standards Organization, 2005.

Issues in Vocabulary Management. NISO TR-06-2017. Baltimore, MD: National Information Standards Organization, 2017.

Kennison, Maria. *Washington State Department of Transportation, Recommendations for Enterprise Information Governance Policy.* 2014. Kent State University Student Study, unpublished.

Lee, Kenneth. *Public Disclosure Project*. 2015. Kent State University Student Study, unpublished.

Menzano, Ralph. *Here Comes E20.* Presentation at AASHTO IS 2009 Conference, Seattle, WA: May 5, 2009.
AASHTO Information Systems Conference (http://www.wsdot.wa.gov/partners/aashtois2009/Agenda.html). (See E20 Part 2)

Morville, Peter. *Ambient Findability*. Sebastopol, CA: O'Reilly Media, 2005.

Pearce-Moses, Richard. *A Glossary of Archival and Records Terminology*. Chicago, IL: Society of American Archivists, 2005. A Glossary of Archival and Records Terminology (http://files.archivists.org/pubs/free/SAA-Glossary-2005.pdf).

Project Open Data Metadata Schema v1.1. Washington, D.C.: US CIO Council, 2015, OCIO Open Data metadata (https://project-open-data.cio.gov/v1.1/schema/).

Semantics of Business Vocabulary and Rules Version 1.4. Object Management Group, 2017 Object Management Group (https://www.omg.org/spec/SBVR/1.4/PDF).

Terminology Work -- Principles and Methods. ISO 704:2000(E). Geneva: International Standards Organization, 2000.