Enterprise Information Architecture: An Overview

WA-RD 896.4

Denise A. D. Bedford

September 2017

Washington State Department of Transportation

Office of Research & Library Services

WSDOT Research Report

WSDOT Practical Solutions Accelerated Innovation Deployment Project

Enterprise Information Architecture: An Overview

Technical Report

Prepared by:

Dr. Denise Bedford, Georgetown University

Prepared for:

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

The Washington State Department of Transportation

September 14, 2017

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, Department of Enterprise Services
Frances Harrison, Spy Pond Partners, LLC
Denise Bedford, Georgetown University

PS AID Project Taxonomy Team

Andy Everett, *Metadata and Taxonomy Analyst, Information Architect, Data Curator – WSDOT Data Library, Data Management Services, Information Technology Division*
Christy Granquist, *Vessel Engineering Librarian, Washington State Ferries*
Steve Hanson, *Project Control Manager, Knowledge Services*
Kathy Szolomayer, *WSDOT Librarian, Office of Research & Library Services*

Acknowledgements

The work described in this report was conducted in association with Spy Pond Partners, LLC and the project titled *Deploying Practical Solutions using Lean Techniques and Knowledge Management.* The metadata and classification scheme products were developed by Dr. Denise Bedford, who brought substantial expertise and experience in these areas from her work at the World Bank and with other organizations. Dr. Bedford worked closely with the WSDOT taxonomy team. As the Project Manager, I am grateful for the support, knowledge and contributions of: the Federal Highway Administration's Accelerated Innovation Deployment demonstration program for their financial assistance; Dr. Denise Bedford, Georgetown University for sharing her expertise, information, guidance, and continued encouragement; Steve Hanson, Andy Everett, Kathy Szolomayer, and Christy Granquist, expertise, and thoughtfulness in the development of the classification resources for application within WSDOT; Gordon Kennedy, WSDOT, and Frances Harrison, Spy Pond Partners, LLC, for their encouragement, advice, and editorial support.

1. Report No. WA-RD 896.4	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Enterprise Information Architecture: An Overview		**5. Report Date** September 14, 2017 **6. Performing Organization Code**
7. Author(s) Denise A. D. Bedford, Ph.D.		**8. Performing Organization Report No.**
9. Performing Organization Name and Address Denise A. D. Bedford, Ph.D. Communication Culture and Technology Georgetown University School of Continuing Studies 640 Massachusetts Ave NW Washington, DC 20001		**10. Work Unit No.** **11. Contract or Grant No.** K1008 FHWA Program code M37A (Delphi Code 15X0445060
12. Sponsoring Agency Name and Address Federal Highway Administration 1200 New Jersey Avenue, SE Washington, DC 20590		**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 introduces the practice of enterprise architecture, which has been implemented by many organizations to better align their data, applications, technology and staff with organizational goals and objectives, and improve organizational agility. Enterprise architecture is built out in layers that provide different viewpoints (typically including layers for goals and objectives, lines of business, people and knowledge, information and data, applications, and technology). As WSDOT continues advance the Practical Solutions approach, techniques from enterprise architecture can provide models for defining a desired future state and aligning workforce, data and information, and technology management in support of that future state. Resource Models provide the connective tissue across the different layers of an enterprise architecture, defining resource needs (e.g. data, information, applications and people/skills) to business capabilities.

17. Key Words	18. Distribution Statement
Enterprise architecture, information architecture, data management, information management, metadata, taxonomy	No restrictions. This document is available through the National Technical Information Service, Springfield, VA 22161.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
None	None	31	

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

ENTERPRISE INFORMATION ARCHITECTURE: AN OVERVIEW 1

1. Introduction 1

2. Traditional and Accepted View of Information Architecture 1

2.1 Views of Information Architecture 1

2.2 Placing Information Architecture in the Enterprise Architecture Model 3

2.3 Landscape of the Information Architecture 7

3. Metadata – Core and Extended 11

3.1 Metadata Defined 11

3.2 Extended Metadata 13

3.3 Specifications and Data Models for Metadata 13

3.4 Public and Standard Metadata Schemes 15

3.5 Creating Enterprise Metadata Leveraging Local Metadata 15

3.6 Generating Metadata for Information Assets 15

3.7 Metadata Repositories 15

4. Master Data 16

5. Controlled Reference Sources 17

5.1 Knowledge Organization Systems 17

6. Governance Processes and Tools 20

7. Developing an Information Architecture 21

7.1 Architecture Processes 21

7.2 Organizational Readiness for Enterprise Information Architecture 22

8. Summary 23

LIST OF FIGURES

Figure 1. Enterprise Information Architecture Perspectives 2

Figure 2. Enterprise Architecture Layer Cake Reference Model 5

Figure 3. Technical Architect's View of EIM – Example Domains & Subdomains 7

Figure 4. Information Systems View of EIM 9

Figure 5. Search System Components and Use of Taxonomies 11

Figure 6. Core Metadata Conceptual Model 13

Figure 7. Examples of Metadata Specifications 14

Figure 8. Creating a Metadata Repository Using Local Metadata Extraction 16

Figure 9. Faceted Taxonomy 17

Figure 10. Flat Taxonomy 18

Figure 11. Hierarchical Taxonomy 18

Figure 12. Network Taxonomy 19

Figure 13. Ring Taxonomy. Entry into the ring at any point produces the same result 19

Figure 14. Developing an Information Architecture 21

LIST OF TABLES

Table 1. Metadata Attributes by Purpose 11

Table 2. Characterization of Metadata Attributes as Taxonomy Structures 20

Table 3. Readiness for Enterprise Information Architecture (September 2017) 22

ENTERPRISE INFORMATION ARCHITECTURE: AN OVERVIEW

1. INTRODUCTION

This is one of a series of deliverables produced in support of Washington State Department of Transportation's (WSDOT) Accelerated Innovation Deployment (AID) project, "*Deploying Practical Solutions with Lean Techniques and Knowledge Management"* (PS AID Project.) Its purpose is to provide a high-level overview of Enterprise Information Architecture (EIA) and describe how the products of the PS AID Project fit within this context. This is intended to provide WSDOT with an understanding of possible next steps for defining and building out an EIA. In this document, the terms "Enterprise Information Architecture" and "Information Architecture" are used interchangeably.

This document is organized into seven sections:

- Section 1 describes the organization of this document.
- Section 2 discusses different perspectives on information architecture and clarifies how they converge in the enterprise view. It also provides a high-level description of a popular enterprise architecture reference model – the architecture layer cake – and explains where enterprise information architecture fits in that model.
- Section 3 provides a deeper dive into the role of Metadata in information architecture.
- Section 4 defines Master Data and describes its role.
- Section 5 discusses Controlled Reference Sources, and differentiates the different types of structures that support these sources.
- Section 6 addresses the role of Governance.
- Section 7 describes good practices for building out an enterprise information architecture. This section also discusses WSDOTs readiness to undertake an information architecture project.

2. TRADITIONAL AND ACCEPTED VIEW OF INFORMATION ARCHITECTURE

2.1 VIEWS OF INFORMATION ARCHITECTURE

The first step in developing an enterprise information architecture is developing a common understanding of the concept. It is important to ensure that everyone's view is incorporated and that all those doing information architecture in some form understand how their work contributes to achieving the enterprise solution.

"Information Architecture" is a term used in different contexts, and means different things to different people, depending on their role and the domain in which they work. A scan of the open and peer-reviewed literature produces five common and different uses of the term that are listed below and illustrated in Figure 1.

- Information user or usability engineering perspective,

- Information lifecycle management perspective,
- Systems or application engineering perspective,
- Information discovery perspective, and
- Taxonomy or knowledge organization structure perspective.

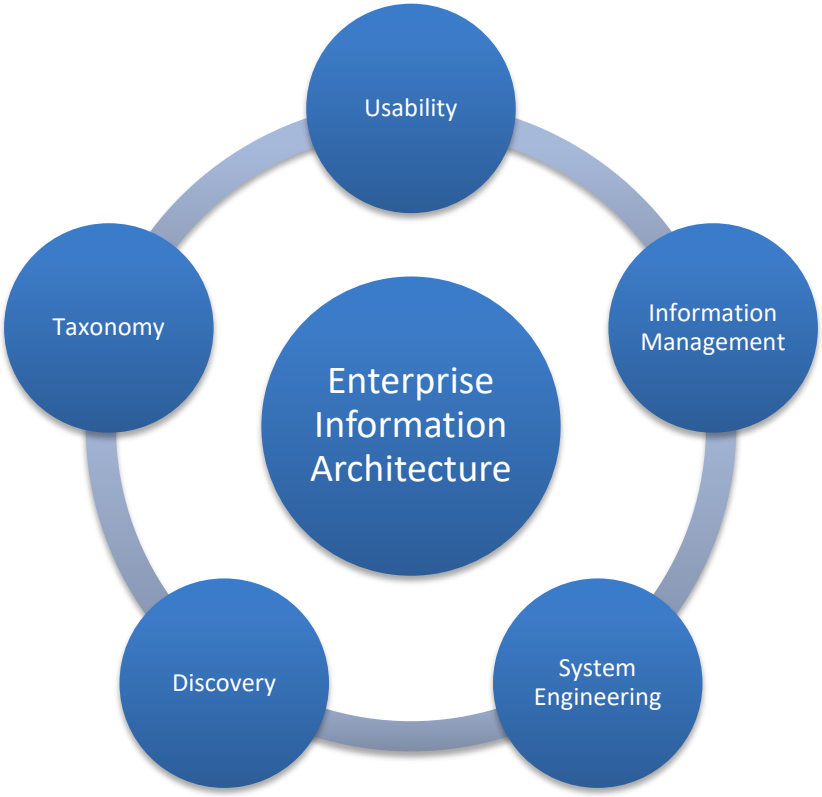

Figure 1. Enterprise Information Architecture Perspectives

Each information architecture perspective has its own body of knowledge and practice. All five perspectives of information architecture fit within a well-designed enterprise architecture. Each of the five Information Architecture perspectives is summarized below.

Information Architecture Perspectives

The first view of information architecture is that used by usability experts and engineers. In this context, information architecture focuses on the design of information structures and systems so that they are usable by both current and potential users. Usability engineering recognizes variations across users with respect to goals and abilities and seeks to optimize the efficiency and effectiveness of information use and management tasks. While usability engineering is often thought to focus on the interface or surface level of information systems, effective engineering begins with good design of the components of enterprise information architecture.

The second view of Information architecture is the context of information management. This perspective introduces the concept of information life cycles. While there are many information life cycle models, the model elaborated in AIIM's Technical Report 48 [1] is considered the most comprehensive and detailed. This model identifies several different stages from information creation to end of life processes such as archiving and preservation. Information architecture in this context includes all of the functions and processes that support each stage. The different stages of the life cycle are tied together through the enterprise information architecture.

A third perspective is the system or application engineering perspective. This perspective takes as a starting point the set of applications that contain the organization's information assets. These may include human resource management systems, financial systems, workflow and decision-making systems, dedicated information repositories, archival systems, web content management systems, transactional data repositories, and so on. In this context, information architecture often focuses on the flow of information across these systems. The goal is to ensure that information and data is available to other applications without human intervention.

A fourth perspective is that of information discovery. While this perspective has gained greater exposure in the past ten years, the relevant knowledge base dates back to the early 1980s. Information discovery focuses on the design of search systems, clustering engines, automated indexing and classification. These are core functions identified in information life cycle models, but they are often viewed in isolation. When information architecture is used in this context it often refers to the design of search systems. In this context, information architecture is designed to support the process of searching and discovery, but also the process of learning about the information sources while searching.

Finally, the fifth perspective is that of taxonomy. Taxonomies serve as knowledge organization structures that are at the core of any enterprise information architecture. In fact, the term taxonomy is sometimes confused with information architecture. The challenge with this perspective, though, is to understand the range of structures needed to support enterprise information architectures.

Enterprise information architecture ties all of these perspectives together. It focuses on effective and appropriate design of all the functions and components that support information management. This document addresses several of these perspectives, with an emphasis on information discovery and taxonomy.

2.2 Placing Information Architecture in the Enterprise Architecture Model

There are several authoritative enterprise architecture reference models. A common business-friendly, non-technical perspective of enterprise architecture is the "architecture layer cake" (Figure 2). The layer cake metaphor illustrates layers and slices:

[1] AIIM/ARMA TR48-2004, Framework for Integration of Electronic Document Management Systems and Electronic Records Management Systems, Link: http://www.aiim.org/Resources/Standards/AIIM_TR48-2004

- The *layers* of the "cake" represent different views: business goals & strategies (performance outcomes), business capabilities, people/knowledge, information/data, application, and technology. Each layer in the architecture cake has its own principles, data models, business rules, and governance processes.
- The *slices* of the cake represent individual business capabilities. For the PS AID Project, the slices are the Practical Solutions life cycle process steps. Each vertical slice of the cake looks at how the different layers support the business of the organization.

There is another way to look at the architecture cake - to see what is core to the enterprise (e.g., what everybody does, what everybody uses, where there is a single view of the "truth"), what is common tobusiness areas across the organization (e.g., goals, standards, types of information or applications that are used by only some of the business areas but which are essential to those areas); and what is unique to some business areas (e.g., no other area uses or needs these but they are critical to a single business area). Core, common and unique represent critical architecture design, planning and procurement views. Core, common and unique apply to all layers of the cake. For example, well designed organizations will have a single email platform, a single financial system, and a single personnel management application. These are "core" applications – everyone uses them. GIS applications may be common applications for several business areas, though, not to all. Finally, a transportation maintenance unit may need a pavement monitoring application and repository to support their work. No other area needs that application, but they cannot do their work without it. WSDOT has chosen to call 'core' elements enterprise, 'common' elements 'group, and 'unique' remain unique.

Information architecture is the middle layer of the architecture layer cake. It serves as a critical connection point in the overall design and structure. The layers of the architecture layer cake are described below.

Business Goals & Strategies

The starting point to build an enterprise architecture is the organization's strategy and business goals, reflected in a set of desired performance outcomes. The business goals and strategies layer describes the intent of the enterprise architecture; it establishes a shared vision that all other layers support. For the top layer, WSDOT should examine the transportation goals set by the Washington State Legislature, the Strategic Plan, performance framework, and other resources and determine how to represent persistent business goals and strategies. ,. This sets the vision for what is to be accomplished and supported by the other layers. For example, one priority outcome such as: "Inform strategic investment decisions in corridors through a broad understanding of system performance and community goals and values." Achieving this outcome requires business capabilities for planning, community engagement and analysis; it requires data and information about system performance and community perspectives; applications for managing and accessing this information; and a technology infrastructure supporting applications and communications.

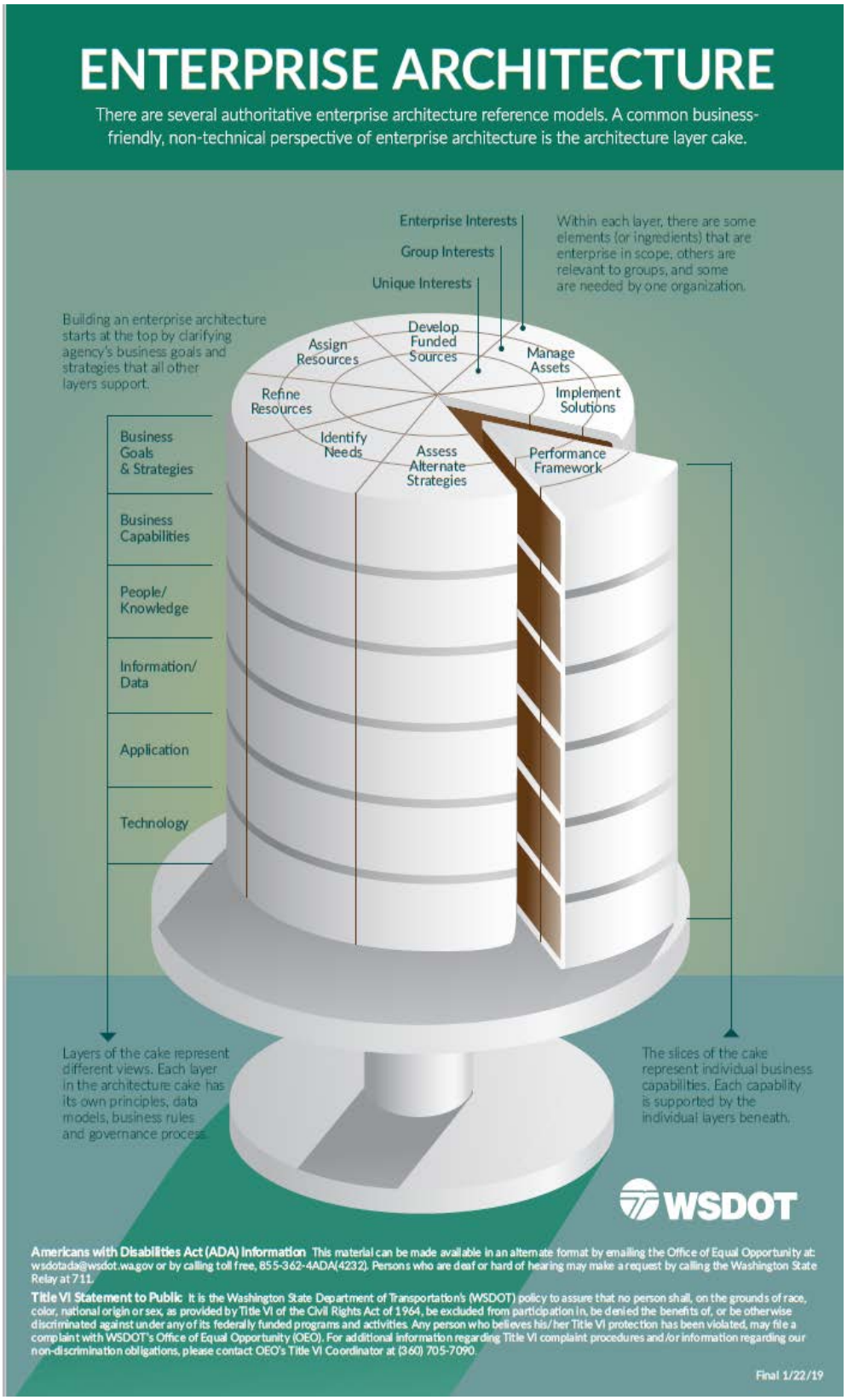

Figure 2. Enterprise Architecture Layer Cake Reference Model

Business Capabilities

The Business Architecture layer identifies the high-level set of capabilities that define what the organization does to deliver value to its stakeholders. Basing a business architecture around capabilities that describe *what* activities/services are accomplished (rather than processes that describe *how* they are done) provides a stable framework that does not require modification as process changes and improvements are made. At a deeper level, the business layer may also include business process flows – "how the organization does what it does" – but business architecture should begin with the "what" rather than the "how." An over-emphasis on process flows reinforces "stakes" in existing processes and methods. This may prevent present people from seeing opportunities for making substantial changes that can improve organizational effectiveness.

People/Knowledge

The people/knowledge layer represents the human resources of the organization – it encompasses organizational culture, structure, roles, personas, and relationships; as well as the knowledge, skills and capabilities needed to provide the capabilities defined in the business layer.

Information/Data

The information/data architecture layer connects the business layer with the application layer. It describes all of the information assets generated or used by the organization, and all of the functions that support the life cycles of those assets. Within this layer there will be design components that support enterprise-wide capabilities (e.g., records management, web content management, document management, enterprise search, etc.) as well as design components that support functionally-oriented and unique business needs.

Application

Application architecture focuses on managing applications to support enterprise functions as well as unique business functions. The focus of application architecture is ensuring that there is appropriate and efficient application support for as many functions as possible. The goal is to ensure that the architecture is also efficient; that there are no redundant applications that generate contradictory or duplicate information, or that operate on different business rules. It is also the goal of this layer to ensure that applications are right-sized for the business needs that they support.

Enterprise application architecture is a critical layer where an organization allows procurement decisions at the business unit level. In this scenario, an organization is likely to see a proliferation of information systems which may not consistently apply enterprise information principles, structures and business rules. This increases the challenge of achieving an effective and efficient enterprise information architecture. It also tends to increase information redundancies and variants – ultimately a management problem for the organization. Efficiency is improved when applications are managed at the enterprise level through categorized inventories that facilitate discovery of redundancies, gaps and unused or poorly optimized applications.

Technology

Finally, the technology architecture focuses on the information technologies that enable use of the information resources that support an organization. This layer is distinct from the application architecture layer to recognize the importance of managing the hardware as a distinct asset. Combining technology and application architectures often obscures an organization's ability to assess application choices which are driven by technology platforms. Technology architecture has its own set of principles, data models, processes and business rules, and inventories.

Enterprise Information Architecture has important connections to the business application layer through the definition of business capabilities. As noted above, the business capabilities (part of the business layer) define the slices of the cake. The information/data layer defines the information assets that support or are produced by those capabilities. The information/data layer is, in turn connected to the application layer through specifying the applications that manage and provision various information assets. All of these connections are detailed in what architects refer to as a Resource Model. Resource Models are developed for individual business capabilities.

2.3 LANDSCAPE OF THE INFORMATION ARCHITECTURE

There are two common views of information architecture – a technical architect's view that defines functional domains and subdomains (as illustrated by the example shown in Figure 3.), and an information systems view that looks at integration across existing content management systems and applications to support improved enterprise information discovery and access.

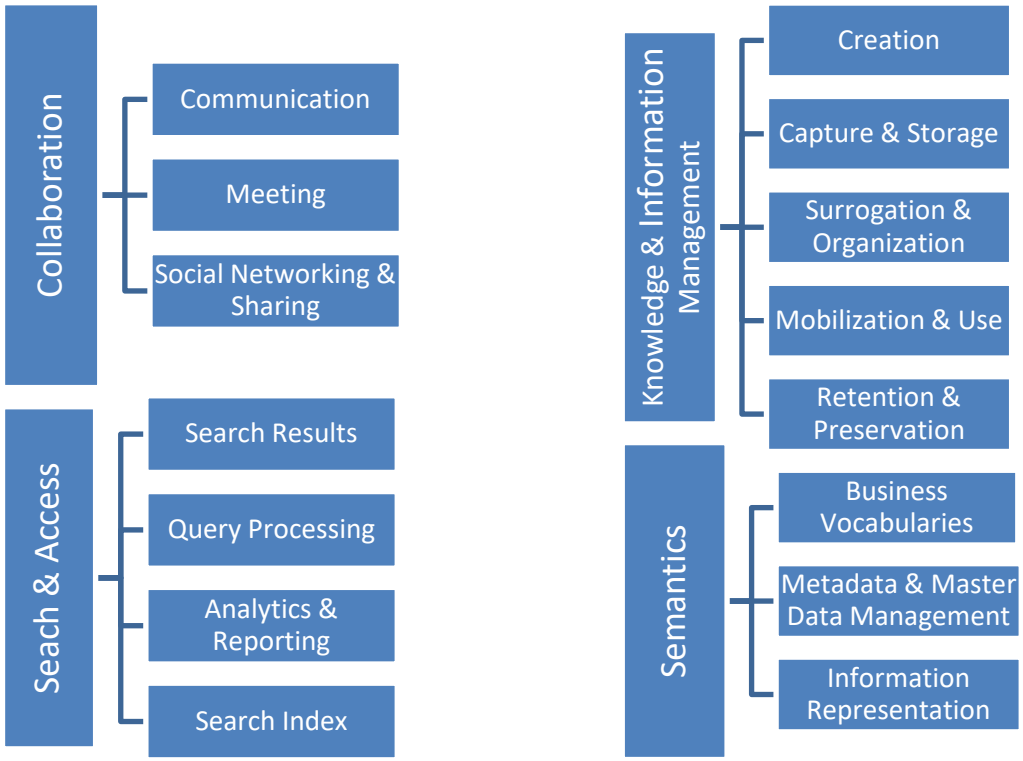

Figure 3. Technical Architect's View of EIM – Example Domains & Subdomains

The technical view provides a breakdown of information management functions that an organization may wish to put in place. Example domains shown in Figure 3. include:

- A ***Collaboration*** domain including subdomains for internal and external *Communication* (email, instant messaging, text messaging, voice mail, etc.), meetings (calendaring, web conferencing, video conferencing, meeting room reservation, etc.), and social networking and sharing (blogs, wikis, collaboration spaces, discussion groups, RSS feeds, FTP sites, contact lists, social bookmarking, ratings, etc.)
- A ***Search & Access*** domain including subdomains for *Search Results Management* (results page design, results de-duplication, results sorting and refinement, recommender engines, search suggestions, results dissemination, etc.), *Query Processing* (query expansion, relevance algorithms, personal search profiles, syndication profiles, etc.), *Search Analytics and Reporting* (results feedback, search logs, index monitoring) and *Search Index* (enterprise search index, index transformation pipeline, indexing rules, etc.)
- A ***Knowledge & Information Management*** domain including subdomains for *Information Creation* (scanning OCR systems, authoring systems, content generation, versioning, forms management, etc.), *Information Capture & Storage* (document management systems, records management systems, team space repositories, network drives, etc.), *Information Surrogation & Organization* (descriptive cataloging, information classification, information abstraction, information indexing, etc.), *Information Mobilization & Use* (security classification, authentication, encryption, digital rights management, etc.), and *Retention & Preservation* (records declaration, physical & digital shredding, records disposition, archives management, etc.)
- A ***Semantics*** domain including subdomains for *Business Vocabularies* (business glossaries, topic schemes, business function classification schemes, keyword thesauri, etc.), *Metadata and Master Data Management* (metadata models and profiles, metadata repositories, master data models, master data repositories, employee master data, etc.) and *Information Representation* (translation and interpretation resources, content models and templates, RDF schemes, etc.)

Not all of these functions are required, and an incremental approach to building out the various functions can be taken.

Figure 4. illustrates the information systems view of information architecture. This view provides an example approach to standardizing and then leveraging core metadata harvested from existing systems to power enterprise search and other services for connecting users to relevant information. In this approach, existing metadata schemes used within individual applications do not need to be changed. Analogous to a classic data warehouse model, metadata is extracted, transformed, and then loaded into a central repository.

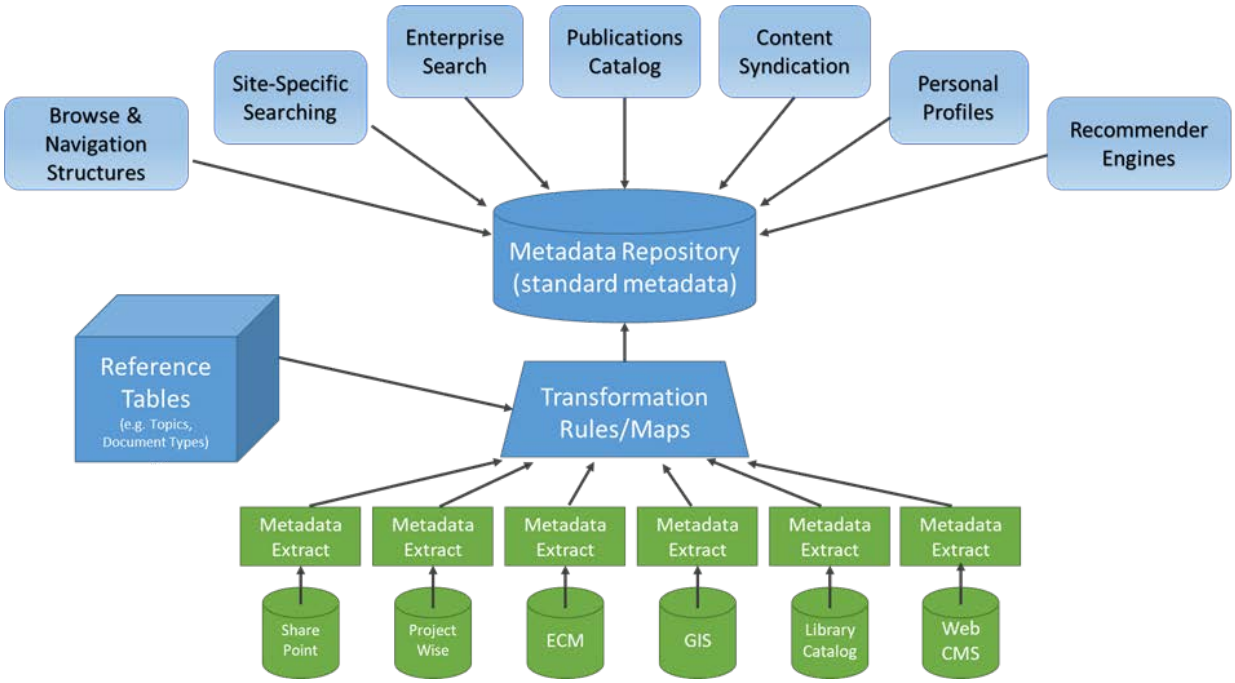

Figure 4. Information Systems View of EIM

Additional background on information architecture elements is provided below.

Core and Extended Metadata

Core metadata are those metadata that are required for all information assets. Extended metadata are those metadata that pertain to particular information assets often associated with specific business functions. For example, project data, human resources information, financial data often have extended metadata that is critical to managing and access those information assets.

Enterprise Metadata Repositories

Enterprise Metadata Repositories are centralized stores of metadata extracted from local systems. Where local systems do not have the capacity to generate or manage metadata, the enterprise repository may serve as the primary source.

Local Metadata in Local Business Applications

Local metadata are those metadata defined by individual applications to support business functions. These metadata can be extracted through automated processes, and (depending on the attribute), mapped to an enterprise metadata scheme. In some cases, this mapping will not be possible. In this situation, if the metadata represents master data (e.g. a linear reference), then the best course of action is to align the local metadata with the enterprise metadata. Otherwise, it is generally not good practice to reverse engineer local metadata to conform to enterprise metadata.

Translation and Mapping Profiles

Enterprise information architecture leverages local metadata through translation and mapping profiles that define the relationship of metadata terms. These profiles may be used to extract and transform

metadata for use in a centralized metadata repository, and they may be used for query translation from enterprise search architecture to the local application. Use of translation and mapping profiles avoids the need to re-create or re-enter metadata for existing information assets.

Master Data

Master Data are metadata with business evidentiary value. These are sources that have a central control authority. Variations from authoritative values are not permitted. Examples of master data at WSDOT are employee IDs and positions, and route numbers and names.

Controlled Reference Sources

Controlled Reference Sources are sources which manage the values of metadata attributes but which are not Master Data. There may be local versions of the values that are mapped to the enterprise values. Controlled Reference Sources support governance of attributes that are critical to information management, discovery, use and access. Examples include content types, subject taxonomies, and security classifications.

Governance Models, Processes and Guidelines

Governance is what holds enterprise information architecture together and ensures the components continue to work together. Each component in the architecture has its own governance model and process.

User-Focused Consuming Applications

User-Focused Consuming Applications may include recommender engines, personal profiles, portal site specific search, an enterprise catalog, and an enterprise browse and navigation structure.

Usability Guidelines

There are authoritative sources of Usability Guidelines, such as www.usability.gov. Organizations generally adapt these guidelines to suit their environments. Usability is supported by strong enterprise information architecture. Metadata, master data and controlled reference sources are important tools for usability engineers.

Enterprise Search Architecture Design and Specifications

Enterprise Search Architecture is a component that is often overlooked or not considered a part of the Enterprise Information Architecture. As shown in Figure 5, enterprise search is powered by the enterprise metadata repository.

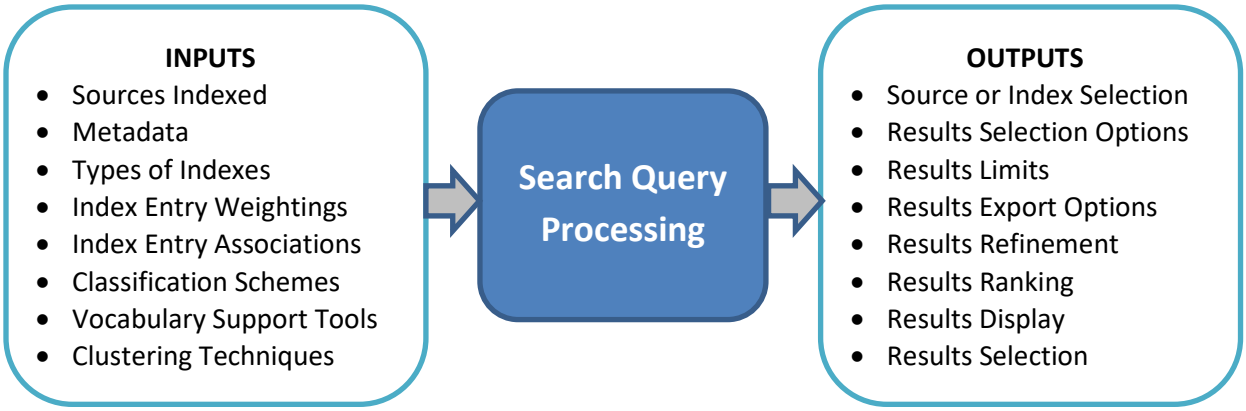

Figure 5. Search System Components and Use of Taxonomies

3. Metadata – Core and Extended

This section focuses on core and extended metadata including: examples of conceptual models to support understanding; specifications to demonstrate what is needed to manage metadata; and, data models that document metadata and support mapping and translation.

3.1 Metadata Defined

Metadata is defined by the Data Administration and Management Association as "data that describes data." In this context, data includes structured and semi-structured data, and unstructured data (e.g., narrative text). There are four reasons why one would create and use metadata including: (1) to identify and distinguish data and information resources; (2) to support search and browsing for data and information; (3) to ensure appropriate access to data and information; and (4) to support administration and management of the information asset over its life cycle. Table 1. below describes commonly used metadata attributes that support these four goals.

Identification & Distinction	Search & Browse	Access and Use Management	Management & Administration
Author	Country	Authorized By	Record Identifier
Title	Region	Rights Management	Disposal Status
Date	Abstract/Summary	Access Rights (Security Classification)	Disposal Review Date
Format	Keywords	Location	Management History
Publisher	Topic	Use History	Retention Schedule/Mandate
Language	Business Function	Disclosure Status	Preservation History
Version		Disclosure Review Date	Aggregation Level
Series Name and Number			Relation
Content Type			

Table 1. Metadata Attributes by Purpose

Each of these metadata attributes should be defined and managed as a separate and distinct attribute. Where fields are combined or linked use is constrained and management will be difficult. Each of these metadata attributes has a distinct set of specifications, a distinct behavior and a particular behavior. Figure 6 illustrates how eight of the metadata attributes listed in Table 1. are supported by different structures. Different structures are needed because they have different values and behave in different ways. For example, within Figure 6:

- "Disclosure Status" tells us whether an information asset is currently disclosed, may be eligible for disclosure in the future, or may never be disclosed to the public. This is a simple flat list of values that are defined by an authoritative unit within the organization. All systems must use the same set of values whenever this attribute is present.
- "Business Function" on the other hand is supported by a hierarchical classification scheme. It defines the functions that an organization currently and historically has performed. This is supported by a classification scheme so that users can understand the overall functional structure, and so that related functions are located in proximity. This structure must be historically accurate and institutionally comprehensive. This means that there must be a governance structure that includes business stakeholders to maintain and manage changes to this structure. Business Function classification schemes are typically narrow and deep – with several levels in the hierarchy.
- "Topic" and "Content Type" attributes are also supported by a hierarchical classification schemes. However, in contrast to the "Business Function" classification scheme, Topic and Content Type schemes should be broad and shallow.
- "Keyword" supports indexing and search. The most appropriate structure for this attribute is a thesaurus. A thesaurus includes controlled vocabularies and relationships among terms to describe how they are used in that domain. Keywords are a different attribute than Topics/Subjects.
- "Author" may be specified as a synonym ring in order to account for different versions of names. If this metadata attribute is used to identify employees who create information assets, then values may be managed as part of the agency's master data.
- "Title" has specifications for values, length, data entry methods, etc. but it is not supported by a controlled reference source.
- "Format" is another field which has a controlled reference source. Most organizations use the list of MIME types that are publicly available and commonly recognized across applications.

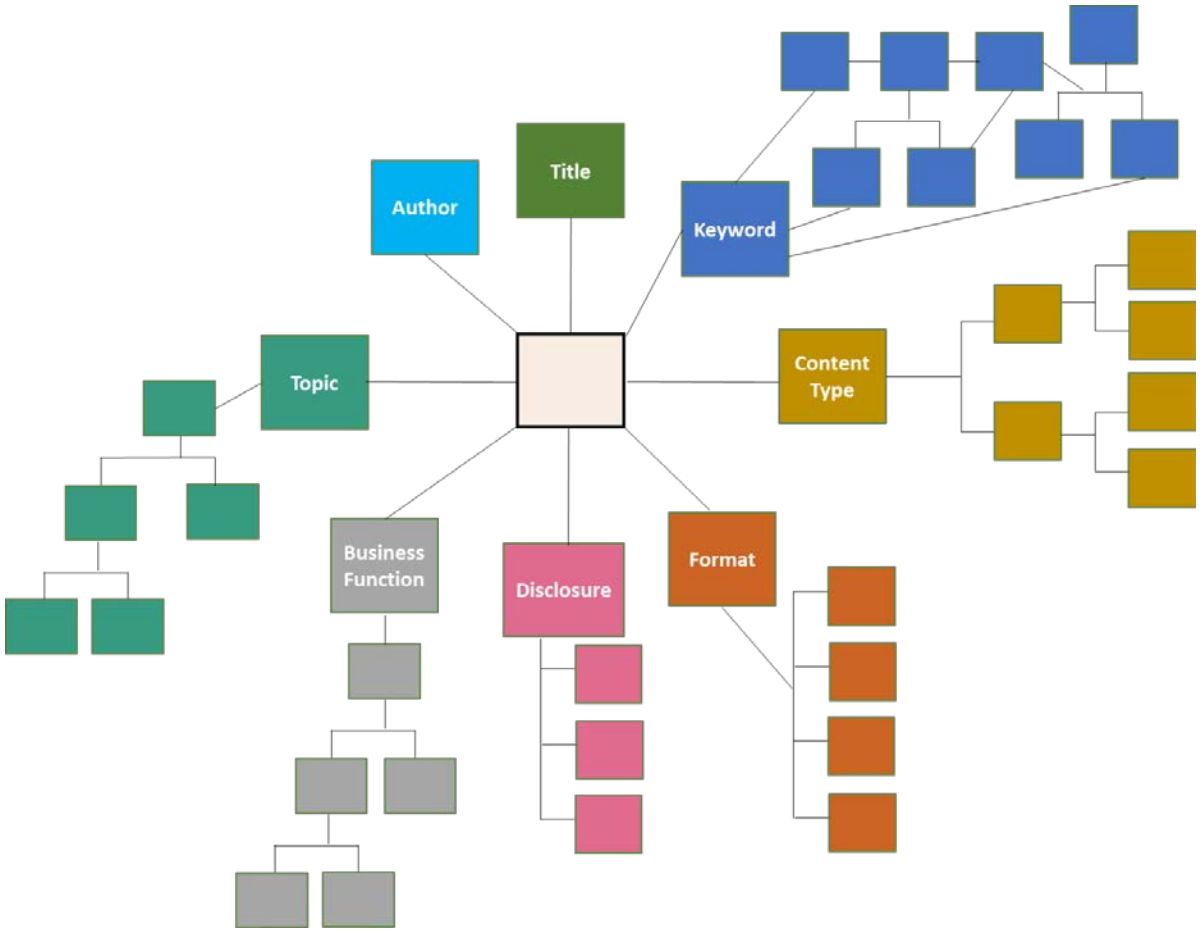

Figure 6. Core Metadata Conceptual Model

3.2 EXTENDED METADATA

An organization needs two kinds of metadata. The first is *core metadata* which define the attributes that must be available to support information management across the organization – and which apply to all kinds of content. The second type is *extended metadata*. Extended metadata are those additional attributes that pertain to a particular kind of content. Wherever that content type is available, the extended metadata must be supported. Often there are kinds of extended metadata that apply to functions – for example, transactional information, human resource information and project information all have attributes which are critical to information management. These attributes may not, however, be important for to managing books, journal articles or photos. The best way to identify extended metadata requirements is to work from the controlled reference source for content types.

3.3 Specifications and Data Models for Metadata

Specifications for metadata attributes should include the following elements:

- Definition
- Obligation (Mandatory or Optional)
- Recordkeeping Obligation (Is it business critical)
- Purpose (Identification & Distinction, Search & Browse, Access & Use Management, Management & Administration)
- Repeatable
- Attribute Length
- Variable Type
- Entry Value (Free or Structured Text)
- Entry State (Simple Edit, Automated Fill, etc.)
- Attribute Syntax
- Status
- Indexed (Yes/No)
- Default Value
- Conversion & Capture Method
- Authoritative Reference Source – Master Data or Controlled Reference Source
- Sub attribute Maps

During the course of the PS AID Project, draft specifications for core metadata attributes were delivered to WSDOT. These were included in a *Metadata Best Practices and Guiding Principles* document. Example specifications from this document are provided in Figure 7. The intention is that WSDOT will review the proposed core metadata, review the specifications and adapt them to work across the Department.

Primary Content Type

Definition: This attribute describes the type of content contained in a resource. It allows searches to be restricted to a finer logical set of resources within the Information Warehouse. There are at present eight valid content types.		
Obligation: Common	Recordkeeping Obligation: Optional	Purpose: Identify & distinguish
Repeatable: Noo	Attribute Length: 70	Variable Type: VARCHAR2
Entry Value: Controlled Text	Entry Value: Drop Fixed	Attribute Syntax:
Status:	Indexed: Yes	Default Value: None
Conversion & Capture: Capture from source systems preferred.		
Authoritative Reference Source: There are eight authoritative primary content types maintained in CDS+, as defined below. The reference source consists of (1) authority records for the eight content types; (2) authority records for all secondary content types with the additional reference to the referring primary content type. Secondary content types are inherited from source systems and mapped to Primary Content Types, as the guiding principle is to introduce no new changes to business rules in source systems		
Primary Content Type Values: Documents & Reports, Publications, Data, Knowledge, Communications, People & Communities, Services, Collections		
Alternative Reference Sources Mapped:		
Subattribute Maps:		

Geographical Region Coverage

Definition: Indicates the regions that are the object of the resource. Where the focus of the resource is on a region, a region code is used. Where the focus is on more than one region, codes are added in repeatable or string delimited fields. Where the resource is not limited to a particular region, a global geographic coverage code is used.		
Obligation: Essential	Recordkeeping Obligation: Optional	Purpose: Search & Browse
Repeatable: Yes	Attribute Length: 40	Variable Type: VARCHAR2
Entry Value: Controlled Text	Entry Value: Drop Fixed	Reference Source:
Status:	Indexed: Yes	Default Value: None
Conversion & Capture: Capture from source systems preferred. Rule based interpretation or assignment possible in future.		
Authoritative Reference Source: CDS Codes		
Alternative Reference Sources Mapped: This attribute uses as the authoritative source the World Bank Geographic Regions maintained in CDS. Other incoming metadata may use other coding structures. A reference map structure from ISO or other region definitions and codes to the CDS structure only where it is imperative for the incoming metadata to use a variant structure for other metadata quality/submission requirements. Where these requirements do not exist, the CDS source will be used.		
Subattribute Maps:		

Figure 7. Examples of Metadata Specifications

Data models are important elements of the specification for every metadata attribute. Properly specified data models are important for application engineers designing metadata into existing and new applications. They are the primary tool for making sure existing metadata practices can be extended as new systems come on line.

3.4 Public and Standard Metadata Schemes

Every organization must define the metadata attributes that are critical to managing their information assets. There is no one scheme that fits all sizes. Organizations should be cautious about using popular metadata schemes whose design is intended to support general publication and access. For example, the Dublin Core Metadata Element Set is designed to provide easy access to general published information. It is not designed to manage information through a full life cycle. It is best to identify the organization's metadata needs and then compare those needs to existing schemes to gauge the goodness of fit. Organizations should also take care to distinguish between metadata schemes and encoding schemes. Encoding schemes do not guide the selection and definition of metadata. Rather they provide a set of protocols for wrapping metadata and values for exchange with other organizations.

3.5 Creating Enterprise Metadata Leveraging Local Metadata

One of the most successful approaches to creating a core and extended metadata model for an organization is to begin with what exists in local applications. Inventory the metadata that exists and map it to the core conceptual schema. Develop a comprehensive view of the behaviors and the values associated with these attributes in the local systems, and use that information to develop mapping and translation profiles, and for building out controlled reference sources. This approach also provides an excellent opportunity to discover variant uses of master data sources.

3.6 Generating Metadata for Information Assets

Metadata is critical to well-formed and managed enterprise information architecture. Humans do a good job of creating metadata, and implementing and maintaining taxonomies. However, human productivity cannot meet the demands for metadata for all of an organization's resources. Applications that help derive meaning from content are known as semantic technologies and they can be configured and trained to generate high quality metadata. However, this requires knowledge of linguistics and natural language processing. The semantic technologies must be suited to the attribute, and trained to behave like a human classifying, selecting values from a list, assigning keywords or making associations. Each of the five taxonomy structures discussed in Section 5.1 below requires a different technology and a different semantic profile. Semantic profiles include the underlying knowledge base for the attribute as well as the business rules and thresholds for selecting and assigning values.

Five types of semantic technologies are required to support the full set of metadata attributes. The five include: (1) rule-based concept extraction; (2) grammar-based concept extraction; (3) rule-based categorization; (4) clustering or stochastic categorization; and, (5) rule-based summarization.

3.7 Metadata Repositories

There are several configurations for managing enterprise metadata. One is a centralized metadata repository. This configuration pulls data from local applications, but leaves the content in its home

repository. As Figure 8 suggests, enterprise information architecture defines extracts and mapping of data from owning applications to the centralized application. The mapping provides translations of attributes and values to ensure that there is good quality control to support enterprise search and other consuming applications.

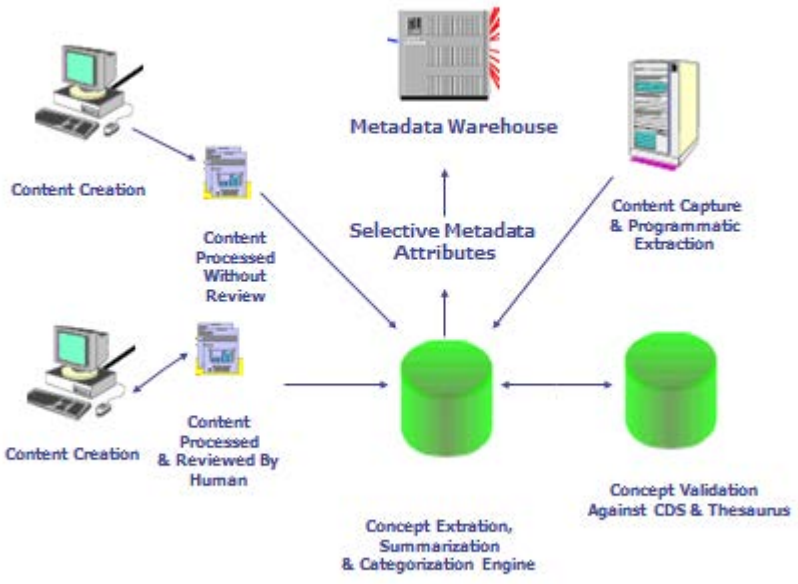

Figure 8. Creating a Metadata Repository Using Local Metadata Extraction

4. MASTER DATA

Master Data are those attributes which have evidentiary and business value to the organization – whose control and oversight are critical to the everyday business operations of the organization. Master data are unique to an organization and are closely aligned with its business capabilities. Master data are called out in and defined in resource models. For WSDOT, Master Data might include: (1) WSDOT regions, (2) WSDOT organizational units, (3) account codes, (4) employee identifiers and names, and (5) route identifiers and names. Each master data source has a publicly designed institutional authority and its own governance board – decision making is owned by one unit – although consultation across business units is generally good practice.

The work to develop core and extended metadata attributes can drive an effort to formalize master data designations and sources. As WSDOT develops Master Data, it will be important to establish an official designation and single authoritative source.

5. Controlled Reference Sources

Controlled Reference Sources are knowledge organization systems which are used to manage the values of a metadata attribute. While Controlled Reference Sources have an authoritative version or an enterprise version, that version does not have business evidentiary value. Its purpose is to facilitate information management, organization, use and discovery. There may be an enterprise version of a Controlled Reference Source as well as local versions. The authoritative version may be achieved by mapping and harmonizing local values. Controlled Reference Sources are used for information management, but they are not designated as having evidentiary value to the organization.

5.1 Knowledge Organization Systems

Knowledge organization systems are any structures that are used to organize information or knowledge, including taxonomies, ontologies, classification schemes, thesauri, and synonym rings. When we describe knowledge organization systems we often use similar words – leaving the reader confused and uncertain about which system to use for which purpose. Rather than use words to describe different types of knowledge organization systems, conceptual models are used. We also provide examples of the types of sources each of the systems supports.

The conceptual models use as a starting point a taxonomy – a simple structure which is used to organize something. There are five types of taxonomies that are often used in information architecture, including: (1) faceted taxonomies (e.g., metadata schemes); (2) flat taxonomies (e.g., controlled vocabularies, picklists); (3) hierarchical taxonomies (e.g., classification schemes); (4) network taxonomies (e.g., thesauri, semantic networks); and (5) ring taxonomies (e.g., synonym management, language variants, query expansion in search)

Faceted Taxonomies

Faceted taxonomies resemble star data structures (Figure 9). Each node in the star structure is linked to and describes the object in the center. The most common representation of a faceted taxonomy is a metadata scheme.

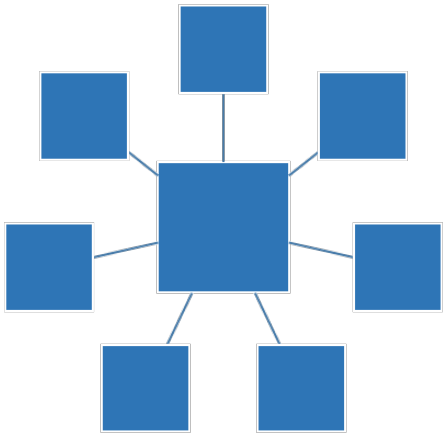

Figure 9. Faceted Taxonomy

Flat Taxonomy

Flat taxonomies are used to support picklists which have a single dimension (Figure 10). When your flat taxonomy begins to develop synonyms, or begins to branch to another level you need to select another structure. Examples of flat lists are lists of languages, lists of security classes, lists of status values, and types of dates.

Figure 10. Flat Taxonomy

Hierarchical Taxonomy

A hierarchical taxonomy is represented as a tree data structure (Figure 11). The tree consists of nodes and links. The links may have definitions and directions. The definition may define a class and a subclass, or a class and types or instances of members of the class. Hierarchical taxonomies support classification schemes. Two of the deliverables – Content Types and Business Function Classification Scheme – are hierarchical taxonomies.

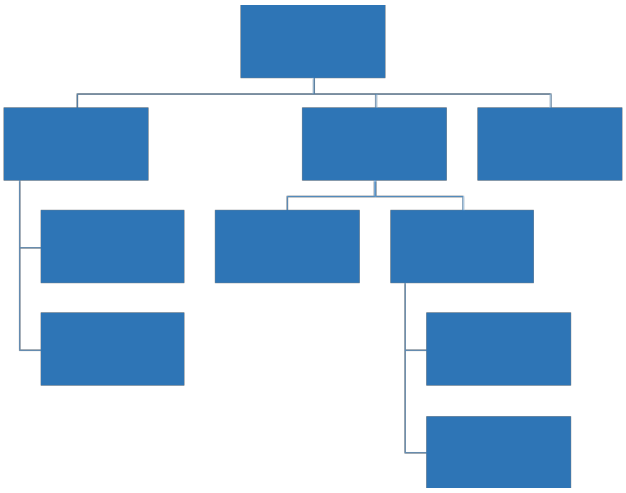

Figure 11. Hierarchical Taxonomy

Network Taxonomies

A network taxonomy is a plex data structure (Figure 12). Each node can have more than one parent. Any item in a plex structure can be linked to any other item. In plex structures, links can be meaningful & different. The most common use of a plex structure is a thesaurus. In a thesaurus, the definition of the links-relationships is as important as the inclusion of the terms (e.g., nodes).

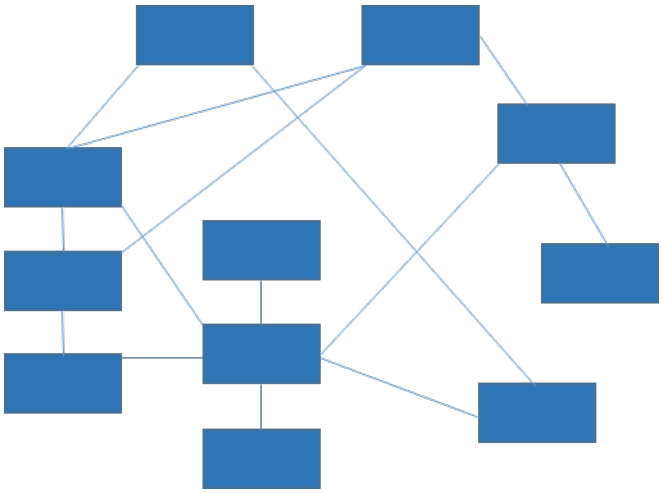

Figure 12. Network Taxonomy

Ring Taxonomies

Ring taxonomies are designed to manage synonyms or things that are equivalent (Figure 13). For example, crashes, collisions, accidents may be synonyms in the transportation field. These may include exact true synonyms, quasi synonyms, lexical variants (e.g., variant spellings, language variants, and regional variants), abbreviations, acronyms, and initialisms. This structure is important for controlled reference sources that support regional language variations, are multidisciplinary in nature, or represent changing perspectives over time.

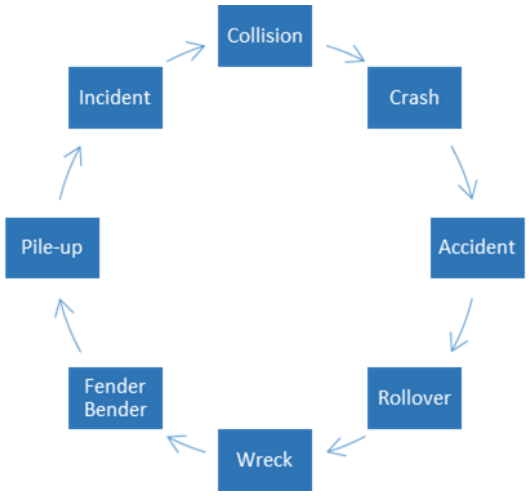

Figure 13. Ring Taxonomy. Entry into the ring at any point produces the same result

Table 2. A high-level characterization of sample core metadata attributes by their underlying taxonomy structures. The variation illustrates the different structures that are required to support metadata.

Identification & Distinction	Search & Browse	Access and Use Management	Management & Administration
Author	Country	Authorized By	Record Identifier
Title	**Region**	**Rights Management**	**Disposal Status**
Date	Abstract/Summary	**Access Rights (Security Classification)**	Disposal Review Date
Format	Keywords	**Location**	Management History
Publisher	Topic	Use History	Retention Schedule/Mandate
Language	Business Function	**Disclosure Status**	Preservation History
Version		Disclosure Review Date	Aggregation Level
Series Name and Number			Relation
Content Type			
Flat Taxonomy	Hierarchical Taxonomy	Network Taxonomy	**Faceted Taxonomy**

Table 2. Characterization of Metadata Attributes as Taxonomy Structures

6. Governance Processes and Tools

Governance is a core process for information management. Each element in the information architecture landscape has its own governance process. The focus in this section is on the governance models for core metadata, business function classification scheme, and content types – the three deliverables which were included in this project. Good governance is supported by:

- Governance roles and responsibilities for business stewards, information and data custodians, technical stewards and business consumers
- Change management processes
- Reference models pertaining to governance decisions
- Guidelines for making decisions, including Guiding Principles and Definitions

Strawman governance models for Metadata, Content Types, and the Business Function Classification Scheme were provided to WSDOT. The expectation is that the strawman models will be discussed and adapted to suit the agency's context.

7. Developing an Information Architecture

The first step is to understand what motivates the organization to manage its information assets, and what information assets are critical to running the business

As stated in Section 2, an enterprise architecture or business capability architecture provide an ideal starting point for the EIA. But organizations may begin by building the enterprise information and use this process to elevate questions about business objectives and capabilities.

Building EIA is a major effort which cannot be done in a scattered or piecemeal way. It is an iterative process of demonstrating what is possible, creating a vision, and achieving incremental improvements. An organization may select components of the EIA to develop over time, but no one of these components will have value or impact unless there is an enterprise information strategy. Strategy does not derive from architecture models or specifications. Rather the strategy drives the development of models. A successful enterprise information project is also solidly grounded in an enterprise information management culture. Everyone has a role to play and this should be understood across the organization. Management support for and buy in to the strategy is critical.

A critical first step is to pull together those who have responsibilities for information management for building a common information management culture, developing a common vision and translating that vision into a common strategy (Figure 14).

Figure 14. Developing an Information Architecture

7.1 Architecture Processes

The process of building out enterprise information architecture is neither simple nor smooth and often does not move in a straight path. However, there is a clear starting point: the first step in is to understand what motivates the organization to manage its information assets, and what information assets are critical to running the business. Ideally, this information is generated as the result of specifying business capabilities and building out the *resource models* for those capabilities.

Resource models provide a blueprint for defining the layer cake slide for a particular capability. They include descriptions, guiding principles, activities and tasks, roles and responsibilities, information assets, and supporting applications.

Identifying supporting applications and information assets for each capability supports identifying controlled reference sources and building out the metadata elements of the architecture:

- As part of developing an inventory of applications, an analysis can be conducted of the metadata that is currently included for each information asset stored in the applications. This analysis can identify common existing metadata and gaps in available metadata. It can also identify lists of values that support individual metadata attributes. This can inform the process of mapping or harmonizing these values across applications to provide enterprise-level controlled reference sources.
- Creating a comprehensive list of instances of information assets can be used to define standard content types. Content types should be managed as a controlled reference source.

This bottom up and middle out approach to constructing enterprise information architecture is more effective than a top down generic approach – particularly for adoption purposes. When business stakeholders understand how their work contributes to the enterprise information architecture, they are more likely to support the initiative.

7.2 Organizational Readiness for Enterprise Information Architecture

Just as there is no single and straight path to establishing enterprise information architecture, no organization is ever perfectly or fully prepared to start the process. You work with what you have, adapt, grow and learn as you go forward. Through the small group discussions, the PS AID Project team developed a high-level description of WSDOT's current state of readiness, as of September 2016 (Table 3.).

EIA Component	Does not Exist	Exists in Pockets	Good Practices to Build Upon	Enterprise Level Practice Exists
Information Landscape		X		
Core Metadata Strategy			X	
Master Data Designations		X		
Controlled Reference Sources:				
Business Function Classification Scheme			X	
Content Type Classification Scheme			X	
Keyword Thesaurus			X	
Asset Classification Scheme			X	
Governance Process and Model		X		
Guiding Principles		X		
Business Rules		X		
Information Management Roles		X		
Performance Metrics	X			

Table3. Readiness for Enterprise Information Architecture (September 2017)

The PS AID Project developed and delivered several important pieces of an Enterprise Information Architecture. Key deliverables included:

- Core Metadata scheme and attribute specifications, which involved expanding the existing model to include attributes related to access and management/preservation;
- Business Function Classification Scheme, including proposed governance model and process, and a first draft build out of the scheme with a high-level mapping to business stewards;
- Strawman structure for a Content Type Classification Scheme which should be mapped to applications used across the Department and adapted by information custodians;
- A framework for future best practices and guiding principles for other core metadata attributes. The framework should be reviewed and adapted by governance groups throughout the Department;
- Inventory of and a classification scheme for applications as a product of the resource models; and
- High level mapping of information resources used in different business capabilities (which were aligned with WSDOT's Practical Solutions Performance-Based Lifecycle steps.)

8. SUMMARY

Organizations benefit from having an enterprise architecture that helps establish a shared understanding of an organization's business goals and strategy, business capabilities, people/knowledge, information/data, applications, and technology. Each element of the architecture has its own body of knowledge and practice but all should align to support organizational productivity.

This report focused on the enterprise information architecture. An enterprise information architecture helps align digital information resources with the business needs of the organization. The metadata and classification scheme products were developed through the author's work with the World Bank and other organizations. With the products of the PS AID Project, WSDOT is well positioned to design and develop an Enterprise Information Architecture. The critical success factor in achieving this goal, though, is management support and staff buy-in. Implementation of an enterprise information architecture is a long, iterative process but the investment in an EIA has potential to help WSDOT streamline information resources, improve productivity, and strengthen the foundation for modernization.

Americans with Disabilities Act (ADA) Information:
This material can be made available in an alternate format by emailing the Office of Equal Opportunity at wsdotada@wsdot.wa.gov or by calling toll free, 855-362-4ADA(4232). Persons who are deaf or hard of hearing may make a request by calling the Washington State Relay at 711.

Title VI Statement to Public:
It is the Washington State Department of Transportation's (WSDOT) policy to assure that no person shall, on the grounds of race, color, national origin or sex, as provided by Title VI of the Civil Rights Act of 1964, be excluded from participation in, be denied the benefits of, or be otherwise discriminated against under any of its federally funded programs and activities. Any person who believes his/her Title VI protection has been violated, may file a complaint with WSDOT's Office of Equal Opportunity (OEO). For additional information regarding Title VI complaint procedures and/or information regarding our non-discrimination obligations, please contact OEO's Title VI Coordinator at (360) 705-7082.