Metadata Strategies

An Information Management Perspective
Agenda

- Information management challenges
- Enterprise metadata strategy
Challenge #1: Integrating data and content management

<table>
<thead>
<tr>
<th></th>
<th>Structured</th>
<th>Less structured</th>
<th>Collection</th>
<th>Management</th>
<th>Publication</th>
<th>Link to source</th>
<th>Link to related</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Management</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Content Management</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Data management applications are typically structured to provide a view at the present time.

- For example, **What** is the current balance in a program fund account?
- But it is more difficult to answer questions like— **Why** have there been fluctuations in a program fund over the life of the program?

Content management applications focus on versioning and formatting narrative content for presentation and publication.

Enterprise content management (ECM) encompasses both data and content management.
Challenge #2: Linking to source data

- Enterprise content is heterogeneous.
- Narrative content is often based on structured data sets and includes visualizations of that data.
  - For example, a research report on highway safety includes tables of data, charts and maps.
  - Can further analysis of the same data set readily be replicated, or new analyses performed?
- It is no longer sufficient to manage narrative content as a static content item. It is becoming necessary to link narrative content to source data.
Challenge #3: Managing the content lifecycle

- Content evolves over time through drafts and versions, and annotations and commentary are associated with it.
  - For example, a PowerPoint report on material properties of highway surfaces is developed through many drafts and versions for different audiences such as engineers and budget analysts.
  - Which version is the most current? Which one is the official approved document of record? etc.

- It is a requirement to manage and synchronize multiple versions of overlapping sets of heterogeneous sources.
Agenda

- Information management challenges
- Enterprise metadata strategy
Characteristics of enterprise metadata strategy

- **Strategy:** Reflects overall program goals of organization.
- **Integration:** Framework for organizing, finding and presenting assets from disparate systems.
  - Capability to leverage available tools to pull related information from multiple applications to 1) manage the enterprise, and 2) communicate with the stakeholders.
- **KPIs:** Provide common way to measure and report performance.

Objectives of enterprise metadata strategy

- Compliance with regulations.
- Measure and optimize performance
  - Content findability and use.
- Support operations/procedures.
Metadata strategy development methodology

Fact Finding
- Research
  - Review artifacts
  - Interview stakeholders
  - Generate use cases
  - Identify KPIs

High-Level
- Background research
  - Industry standards and best practices
  - Competitor and peer practices
  - Organization policies and procedures

Detail
- Quantitative inputs – review analytics.
  - Application statistics
  - Content use statistics
  - Search query logs

Validation
- Qualitative inputs – ask stakeholders.
  - One-on-one interviews
  - Focus groups
  - Surveys

Governance
Generate use cases

- DOT information is created to support a business function or activity.
- Anticipate and envision future secondary and potential tertiary uses of that information.
  - Information is generated as part of an immediate operational activity such as accessioning assets which is part of the DOT asset management function.
  - Later that same information may be analyzed to produce an asset maintenance plan.

- How should information be structured to maximize its potential uses?
- When is it appropriate to archive and/or purge information from an active collection?
Metadata ROI

- Assets are expensive to create so it’s critical that they can be found, so they can be used and re-used.
- Every re-use decreases the asset creation cost.
Metadata capital*

- Asset reuse is contingent on the creation and accessibility of complete and consistent metadata.
- Every re-use increases the asset value.

* "Metadata capital" is a term recently coined by Dr. Jane Greenberg, Director of the Metadata Research Center at the University of North Carolina at Chapel Hill.
Identify key performance indicators (KPI’s)

- Number of content assets added/edited during the period.
- Number of content assets used and re-used during the period.
- Revenue from content assets during the period (conversion and lift)
- Number of content assets with metadata (completeness)
- Number of metadata inconsistencies (consistency)
- Number of times each category has been used.
- Number of cross-references.
- Number of end user reviews, subscriptions to feeds, etc. (channel engagement)
- Number of new content asset alerts generated.
- Number of end user subscriptions to feeds.
- Number of change requests handled (new categories, synonyms, notes, etc.)
- Number of information products/applications added (aggregation, search, filtering, personalization, multi-channel, etc.)
- Number of content asset/service recommendations.
- End user satisfaction (survey)
Metadata specification development methodology

Fact Finding

High-Level

Detail

Validation

Governance

- Develop metadata schema & high-level vocabularies
- Test schema by tagging some content
Enterprise metadata standardizes

- Resource description (title, summary and subject categories)
- Resource function/purpose in the record retention context (typology of use cases).
- Resource types
  - E.g., Darwin Information Typing Architecture (DITA)
    - Task Resources. Describe how to accomplish a task, listing a series of steps that users follow to produce an intended outcome.
    - Concept Resources - Definitions, rules, and guidelines.
    - Reference Resources - Detailed, factual material.
- Geographic locations generally, and facilities specifically.
- Roles (employee, manager, partner, supplier, etc.) and/or audience/persona.
# Adopt an enterprise core metadata standard

<table>
<thead>
<tr>
<th>Element</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>The name given to the resource.</td>
</tr>
<tr>
<td>Description</td>
<td>An account of the resource.</td>
</tr>
<tr>
<td>Type</td>
<td>The nature or genre of the resource.</td>
</tr>
<tr>
<td>Subject</td>
<td>The topic of the resource.</td>
</tr>
<tr>
<td>Coverage</td>
<td>The spatial or temporal topic of the resource.</td>
</tr>
<tr>
<td>Creator</td>
<td>The entity primarily responsible for making the resource.</td>
</tr>
<tr>
<td>Owner</td>
<td>The organizational entity that is responsible for the lifecycle of the resource.</td>
</tr>
<tr>
<td>Approver</td>
<td>The entity that is responsible for approving the publication of a resource.</td>
</tr>
<tr>
<td>Retention Schedule</td>
<td>The retention schedule that applies to the resource.</td>
</tr>
<tr>
<td>Dates</td>
<td>The dates associated with a resource lifecycle event.</td>
</tr>
</tbody>
</table>
Identify and manage key resource lifecycle events/transactions

<table>
<thead>
<tr>
<th>Event</th>
<th>Metadata</th>
<th>Actor</th>
<th>Description</th>
<th>Default</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creation</td>
<td>Date Created</td>
<td>Author</td>
<td>Date content originally created.</td>
<td>Today</td>
</tr>
<tr>
<td>Approval</td>
<td>Date Approved</td>
<td>Approver</td>
<td>Date content approved</td>
<td>Not approved</td>
</tr>
<tr>
<td>Publication</td>
<td>Date Published</td>
<td>System</td>
<td>Date content was published.</td>
<td>Today</td>
</tr>
<tr>
<td>Expiration</td>
<td>Date of Expiration</td>
<td>Author</td>
<td>Date when content is to be deleted.</td>
<td>+365 days</td>
</tr>
<tr>
<td>Review</td>
<td>Date of Next review</td>
<td>System</td>
<td>Date content must next be reviewed.</td>
<td>+365 days</td>
</tr>
<tr>
<td>Modification</td>
<td>Date Modified</td>
<td>Author or Approver</td>
<td>Date content was last modified.</td>
<td>Today</td>
</tr>
<tr>
<td>Deletion</td>
<td>Date Deleted</td>
<td>Author, Approver, NPI Team</td>
<td>Date content was deleted.</td>
<td>+365 days</td>
</tr>
</tbody>
</table>
Develop an enterprise taxonomy

❖ Standard vocabularies used to describe what the content is about and why it is important—the “subject”.
  ▪ 8-12 facets—discrete aspects of the “subject” such as who, what, where and why.
  ▪ 2-3 levels deep.
  ▪ < 20 categories per level.
  ▪ 1500 total categories.
  ▪ Captures synonyms, abbreviations, acronyms, translations, and other, term variations (such as regional variations) as well as notes that explain how the term has been determined, and how the term should be used.

❖ Example: TRT (Transportation Research Thesaurus)
TRT high-level taxonomy example

**Disciplines**
- Agriculture
- Architecture
- Civil
- Drilling
- Economics
- Electrical
- Environmental
- Geography
- Geology
- Geotechnical
- Hydraulic hydrology
- Industrial
- Information management
- Materials science
- Mechanical
- Planning
- Process
- Safety processes
- Simulation
- Structural
- Surveying geodesy
- Telecommunications

**Populations**
- Age groups
- Gender
- Educational level
- Families & households
- Race & ethnicity
- Language backgrounds
- Socioeconomic levels

**Organizations**
- Businesses
- Industries
- Service agencies
- Governments
- Military organizations
- Membership organizations
- Forms of business or industry

**Facilities**
- Structures
- Specialized facilities
- Facilities & structures by use
- Transportation, hydraulic & utility facilities
- Structural deterioration & defects

**Vehicles & equipment**
- Vehicles
- Vehicle characteristics
- Vehicle dynamics
- Vehicle components
- Equipment
- Equipment characteristics

**Materials**
- Classes of materials
- Properties of materials

**Physical phenomena**
- Laws of physics
- Sound
- Chemical reactions & processes
- Electricity
- Electromagnetism
- Mechanics
- Light
- Nuclear phenomena
- Surface phenomena
- Heat
- Systems

**Topics**
- Transportation
- Transportation operations
- Management & organization
- Communication & control
- Planning & design
- Construction & maintenance
- Testing
- Safety & security
- Environment
- Economic & social factors

**Locations**
- States A-Z
- Regions
- Geographic features

**Content Types**
- Monographs
- Serials
- Dictionaries
- Registers
- Correspondence
- Patents
- Reviews
- Reports
- Specifications
- Data sets
- Forms
- Visual presentations
- Multimedia
- Legal documents
- Maps
Metadata elements vs. metadata values

- **Dublin Core** (ISO 15836:2009) is a vocabulary of fifteen properties for use in resource description
  - Dublin—Originated at a 1995 workshop in Dublin, Ohio.
  - Core—Elements are broad and generic, usable for describing a wide range of resources.

- **SKOS** Simple Knowledge Organization System (W3C Recommendation 18 August 2009) is a data model to identify, define and link concept vocabularies.
# Metadata elements vs. metadata values

## Table: Metadata Elements and Schemes

<table>
<thead>
<tr>
<th>Element</th>
<th>Scheme</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>dc.title</td>
<td></td>
<td>FAR Encyclopedia</td>
</tr>
<tr>
<td>dc.description</td>
<td></td>
<td>The FARS Encyclopedia offers an intuitive and powerful approach for retrieving fatal crash information.</td>
</tr>
<tr>
<td>dc.type</td>
<td>DITA</td>
<td>Reference resource</td>
</tr>
<tr>
<td>dc.subject</td>
<td>TRT</td>
<td>Traffic crashes, Fatalities</td>
</tr>
<tr>
<td>dc.coverage</td>
<td></td>
<td>1994-2011</td>
</tr>
<tr>
<td>dc.date</td>
<td></td>
<td>2012</td>
</tr>
</tbody>
</table>

### Diagram: NHTSA FARS Encyclopedia Map Features

- **Blue = Dublin Core**
- **Red = Vocabularies**

**Map Features** - Click here for information.

**VMT changes** - Click here for information.

---

**2012**
Metadata elements vs. metadata values

<!--Each page must contain this info, per OMB-->
<meta name="dc.identifier" content="http://www-fars.nhtsa.dot.gov/Main/index.aspx" />
<meta name="dc.title" content="FAR Encyclopedia" />
<meta name="dc.description" content="The FARS Encyclopedia offers an intuitive and powerful approach for retrieving fatal crash information." />
<meta name="dc.type" scheme="DITA" content="Reference resource" />
<meta name="dc.subject" scheme="TRT" content="Traffic crashes, Fatalities" />
<meta name="dc.coverage" content="1994-2011" />
<meta name="dc.date" content="2012" />
<meta name="dc.format" content="text/html; charset=utf-8" />
<meta name="dc.language" scheme="DCTERMS.RFC1766" content="EN-US" />
<meta name="keywords" content="FARS, Fatality Analysis Reporting System, PAR, Police Accident Reports, statistics, data, facts, car, truck, motorcycle, vehicle, pedestrian, street, road, highway, interstate, accident, injury" />

Blue = Dublin Core
Red = Vocabularies
Three pillars of enterprise metadata governance

Each pillar has four elements.

Fully mature Governance Models implement all elements.
Resources

QUESTIONS?

Joseph A Busch, Principal
jbusch@taxonomystrategies.com
twitter.com/joebusch
415-377-7912