Metadata Strategies

An Information Management Perspective

Agenda

❖ Information management challenges

❖ Enterprise metadata strategy

Challenge #1: Integrating data and content management

	Structured	Less structured		Collection	Management	Publication		Identification	Access control	Description	Analysis	Visualization	Link to source	Link to related
Data Management	✔			✔	✔			✔	✔	✔	✔	✔	✔	
Content Management		✔		✔	✔	✔		✔	✔	✔			✔	✔

- ❖ 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.
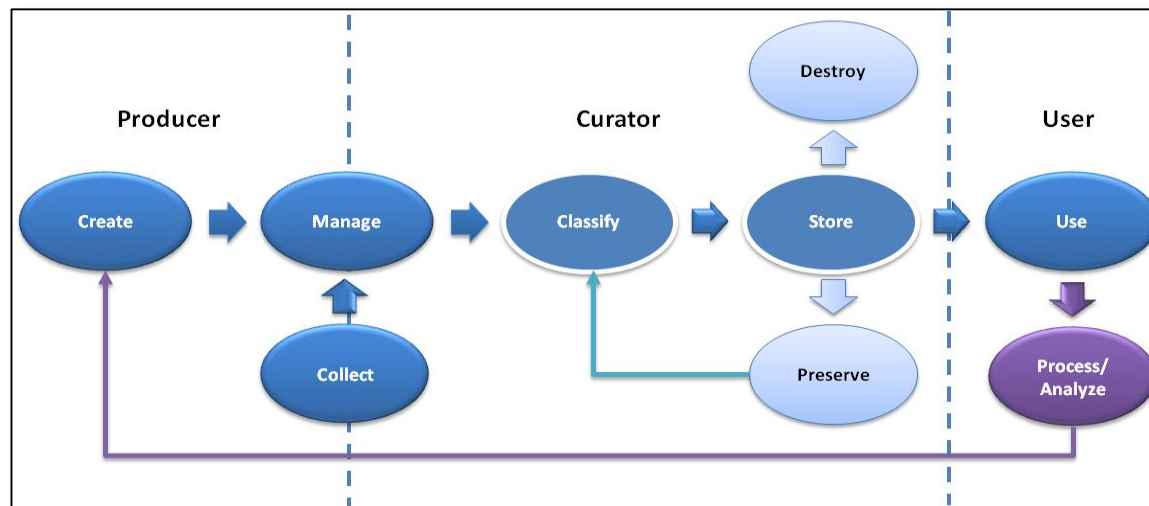
Agenda

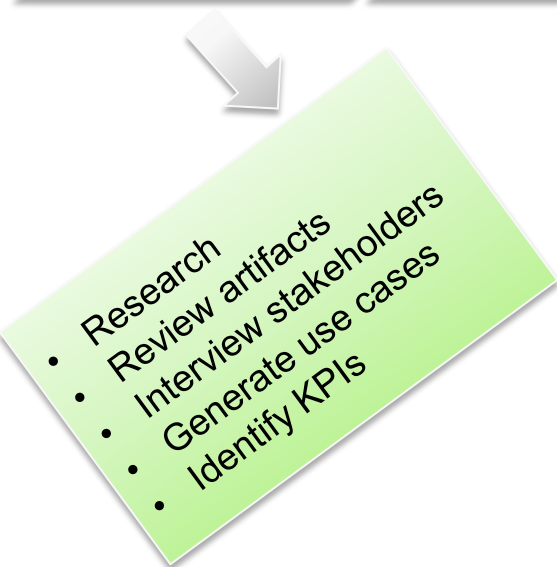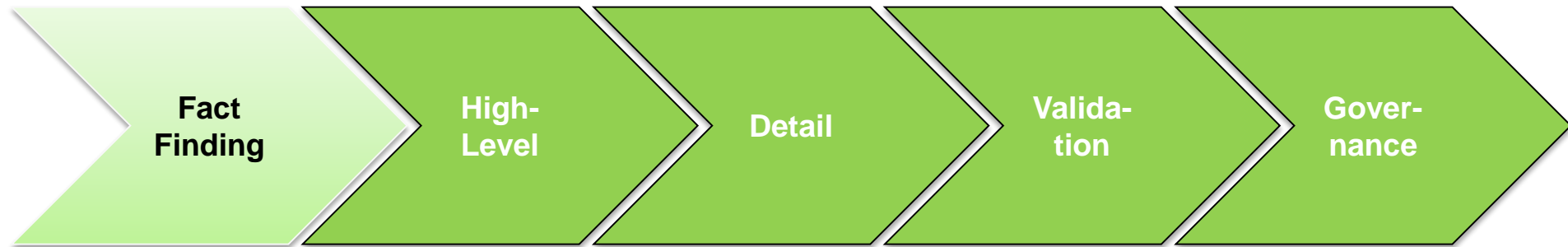❖ Information management challenges

❖ 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 → **High-Level** → **Detail** → **Valida-tion** → **Gover-nance**

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

❖ Background research
  ▪ Industry standards and best practices
  ▪ Competitor and peer practices
  ▪ Organization policies and procedures
❖ Quantitative inputs – **review analytics**.
  ▪ Application statistics
  ▪ Content use statistics
  ▪ Search query logs
❖ Qualitative inputs – **ask stakeholders**.
  ▪ One-on-one interviews
  ▪ Focus groups
  ▪ Surveys

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.
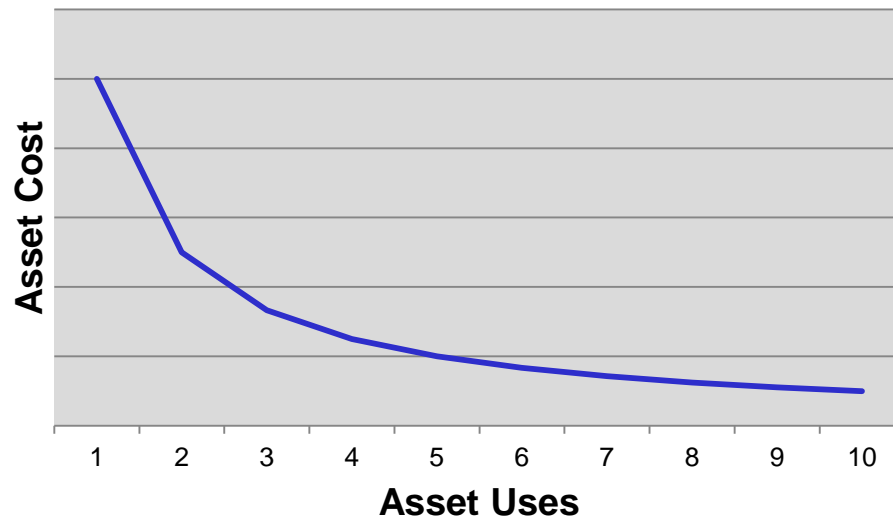
Metadata capital*

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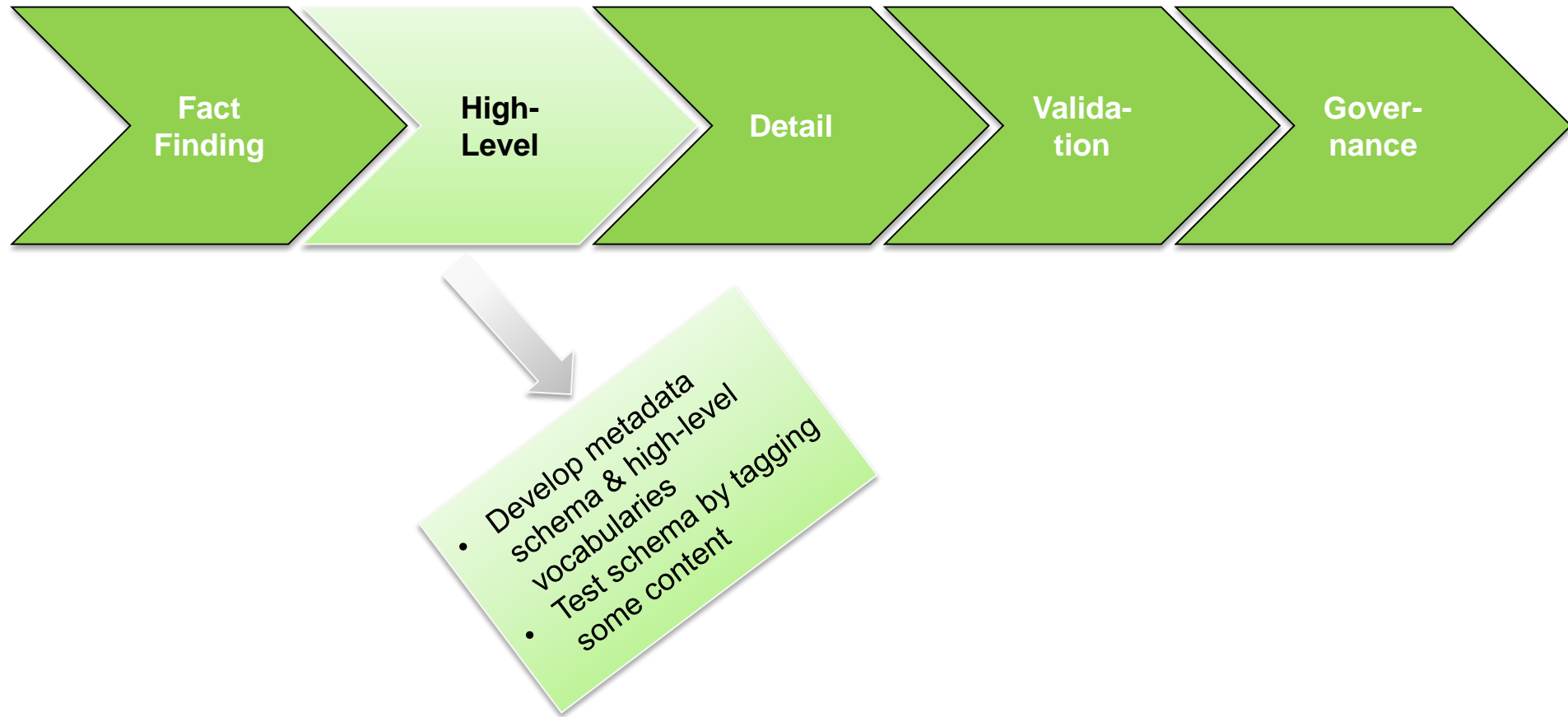
❖ Every re-use increases the asset value.

---
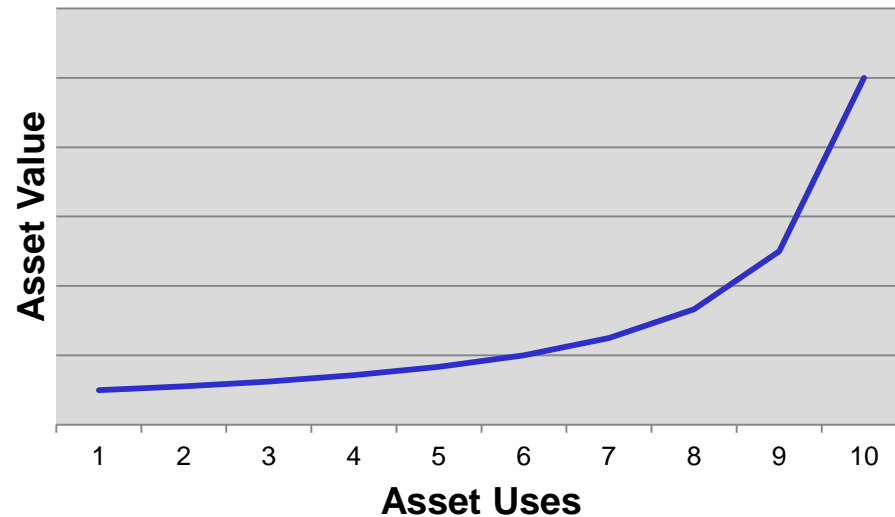
* "Metadata capital" is a term recently coined by Dr. Jane Greenberg, Director of the Metadata Research Center at the University of North Carolina at Chapel Hill.

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

Valida-tion

Gover-nance

- 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

Element	Description
Title	The name given to the resource.
Description	An account of the resource.
Type	The nature or genre of the resource.
Subject	The topic of the resource.
Coverage	The spatial or temporal topic of the resource.
Creator	The entity primarily responsible for making the resource.
Owner	The organizational entity that is responsible for the lifecycle of the resource.
Approver	The entity that is responsible for approving the publication of a resource.
Retention Schedule	The retention schedule that applies to the resource.
Dates	The dates associated with a resource lifecycle event.

Identify and manage key resource lifecycle events/transactions

Event	Metadata	Actor	Description	Default
Creation	Date Created	Author	Date content originally created.	Today
Approval	Date Approved	Approver	Date content approved	Not approved
Publication	Date Published	System	Date content was published.	Today
Expiration	Date of Expiration	Author	Date when content is to be deleted.	+365 days
Review	Date of Next review	System	Date content must next be reviewed.	+365 days
Modification	Date Modified	Author or Approver	Date content was last modified.	Today
Deletion	Date Deleted	Author, Approver, NPI Team	Date content was deleted.	+365 days

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

TRT

Red = Facets
Black = Categories

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

❖ <u>Dublin Core</u> (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.

❖ <u>SKOS</u> 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

Blue = Dublin Core
Red = Vocabularies

Element	Scheme	Value
dc.identifier		http://www-fars.nhtsa.dot.gov/Main/index.aspx
dc.title		FAR Encyclopedia
dc.description		The FARS Encyclopedia offers an intuitive and powerful approach for retrieving fatal crash information.
dc.type	DITA	Reference resource
dc.subject	TRT	Traffic crashes, Fatalities
dc.coverage		1994-2011
dc.creator	US Govt Manual	National Highway Traffic Safety Administration
dc.date		2012

Metadata elements vs. metadata values

Blue = Dublin Core
Red = Vocabularies

<!--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.creator" scheme="US Government Manual" content="National Highway Traffic Safety Administration" />

<meta name="dc.date" content="2012" />

<meta name="dc.format" content="text/html; charset=utf-8" />

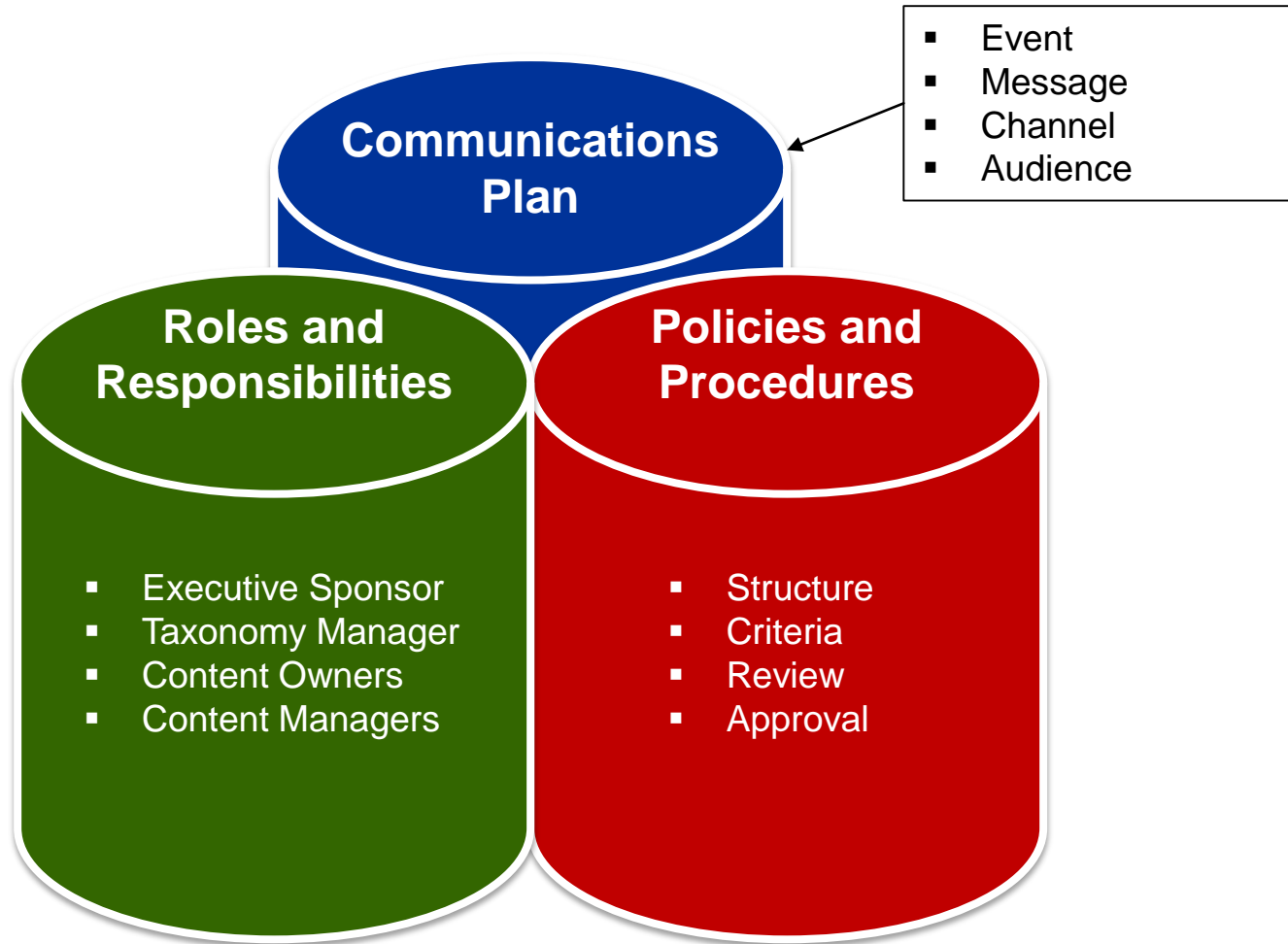<meta name="dc.language" scheme="DCTERMS.RFC1766" content="EN-US" />

<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" />

Three pillars of enterprise metadata governance

Communications Plan

- Event
- Message
- Channel
- Audience

Roles and Responsibilities

- Executive Sponsor
- Taxonomy Manager
- Content Owners
- Content Managers

Policies and Procedures

- Structure
- Criteria
- Review
- Approval

❖ Each pillar has four elements.
❖ Fully mature Governance Models implement all elements.

Resources

- ❖ Cambridge Systematics. NCHRP Report 754: Improving Management of Transportation Information. 2013. http://www.trb.org/Publications/Blurbs/169522.aspx

- ❖ Dublin Core (ISO Standard 15836:2009) http://dublincore.org/documents/dces/

- ❖ J. Greenberg, S. Swauger, E.M. Feinstein. Metadata Capital in a Data Repository. Proceedings of International Conference on Dublin Core and Metadata Applications 2013. http://http://dcevents.dublincore.org/IntConf/dc-2013/paper/view/189/172.

- ❖ Simple Knowledge Organization System (W3C Recommendation 18 August 2009) http://www.w3.org/TR/2009/REC-skos-reference-20090818/

- ❖ Transportation Research Thesaurus. http://trt.trb.org/

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

QUESTIONS?