

createsphere

Metadata Deep Dive

Agenda

- ❖ Metadata and interoperability
- ❖ Metadata tools
- ❖ Metadata standards
- ❖ Planning for metadata

What is metadata

- ❖ Metadata provides enough information for any user, tool, or program to find and use any piece of content.
- ❖ Metadata is also used to drive business processes
 - A simple example –publication and expiration dates.

... and business processes may be used to generate metadata (more on this later)

Types of metadata

❖ Asset metadata

- Identifier, Creator, Title, Description, Format, Type, Size, Date, etc.

❖ Subject metadata

- Names of People, Names of Organizations, Names of Events, Names of Products, Names of Things, Topic, Purpose, Expertise, etc.

❖ Use metadata

- Audience, Language, Location, Channel, Rights, Role, etc.

❖ Relational metadata

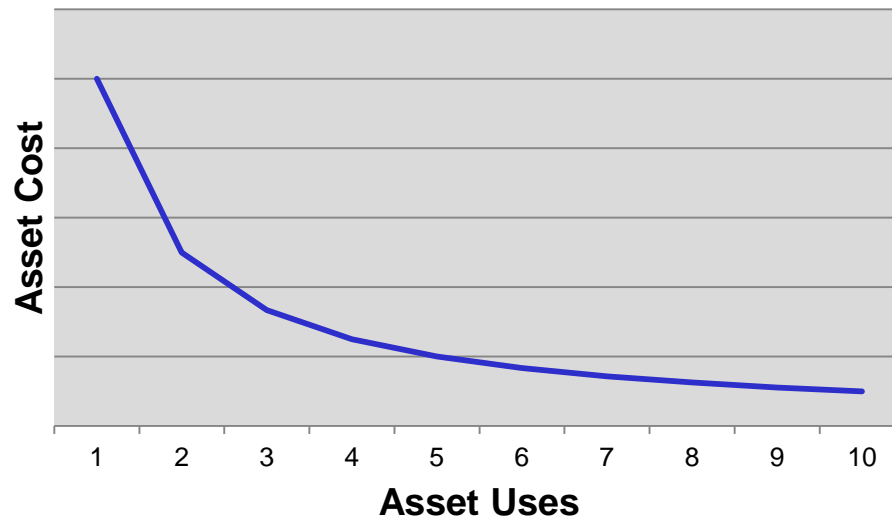
- Source, Collection, Parts, Related to, etc.

Interoperability

- ❖ **The ability of diverse systems and organizations to work together by exchanging information.**
- ❖ Semantic interoperability is the ability to automatically interpret the information exchanged meaningfully and accurately.

Interoperability 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 and increases the asset value.



Interoperability (2)

- ❖ If assets are so important, why can't they be found?
 - They contain no searchable text.
 - They exist in different applications, file shares and/or desktops.
 - ... Other reasons?
- ❖ When they are found why can't assets be reused?
 - When there are multiple versions, it's difficult to choose which one to use.
 - The usage rights may not be clear.
 - ... Other reasons?

Interoperability vision

- ❖ I want to easily find any assets in a particular format that can be used for a specific purpose regardless of where they are located.
- ❖ I want to analyze my collection of assets to identify...
 - Strengths and weaknesses
 - Types of assets
 - Develop new products and services
 - ... other analytics?
- ❖ Challenges:
 - How to align different metadata properties
 - E.g., Title and Caption; Location and Setting; etc.
 - How to align different vocabularies
 - E.g., CA and California; RiM and Research in Motion; etc.

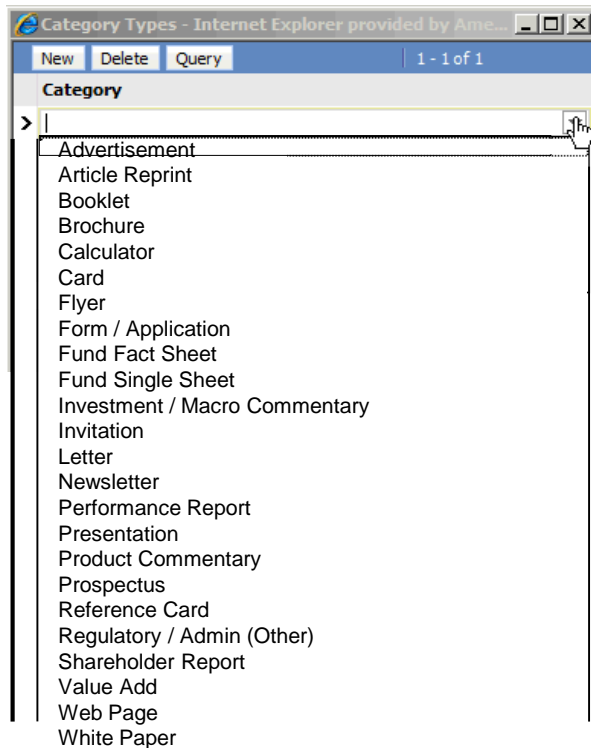
Agenda

- ❖ Metadata and interoperability
- ❖ Metadata tools
- ❖ Metadata standards
- ❖ Planning for metadata

Metadata tools

- ❖ DAMs (and CMSs) provide some capability for metadata capture templates and forms.

Blank metadata capture template, with no values defaulted.



Simple vocabulary in a drop-down list in a tagging template.

Whitepaper: *Title*

Content Types:

Series:

Frequency:

Audience:

Segment:

Channel:

Language:

Region/Country:

Portfolio:

Strategy:

Broad Asset Class:

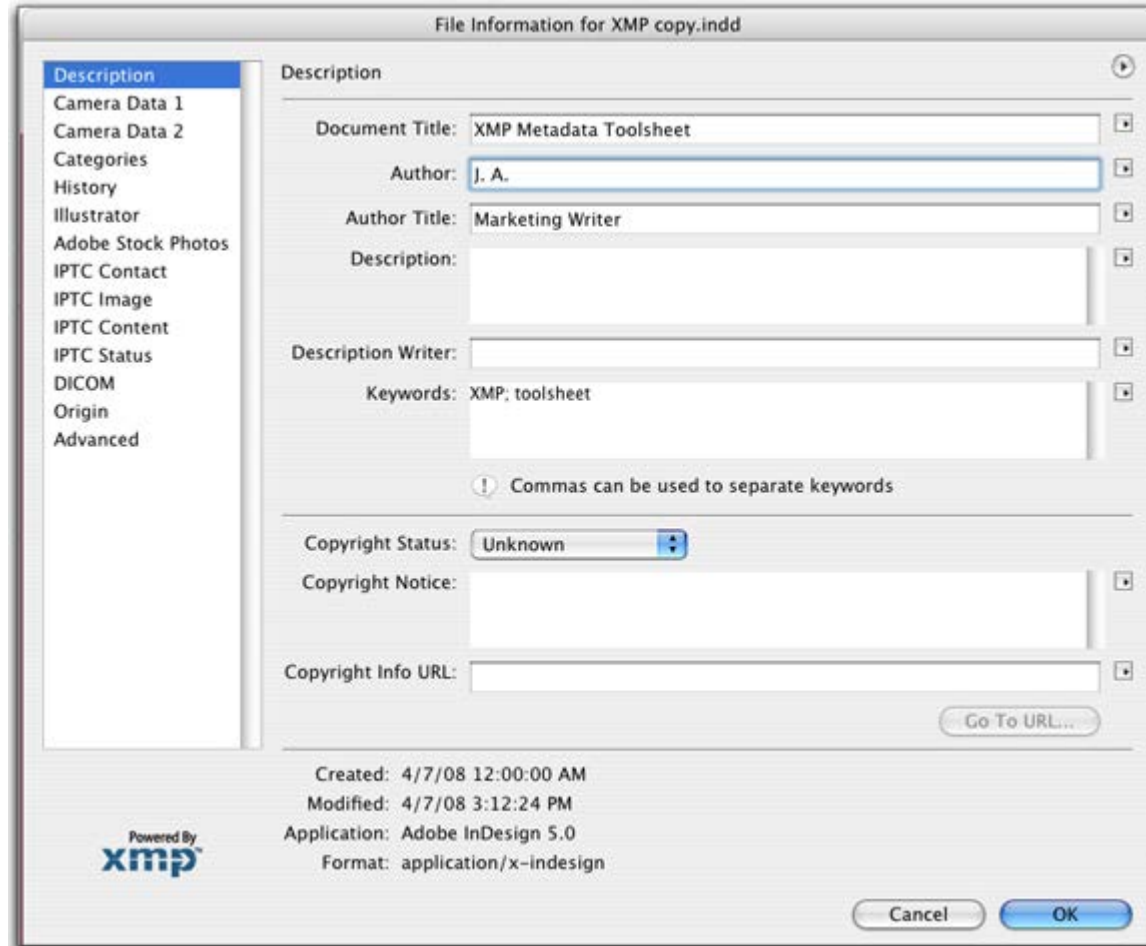
Investment Style:

Risk Level:

Topic:

Metadata tools (2)

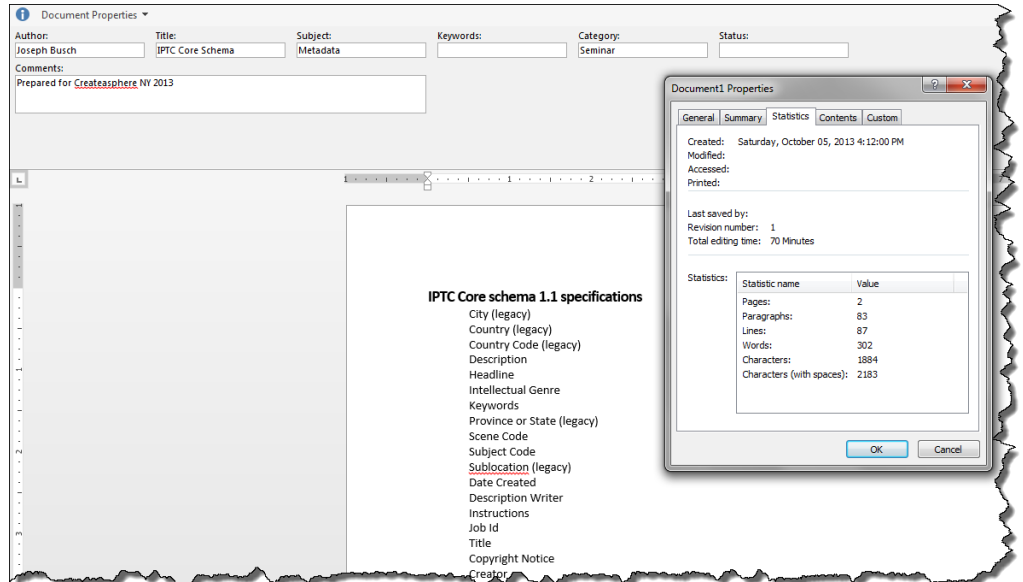
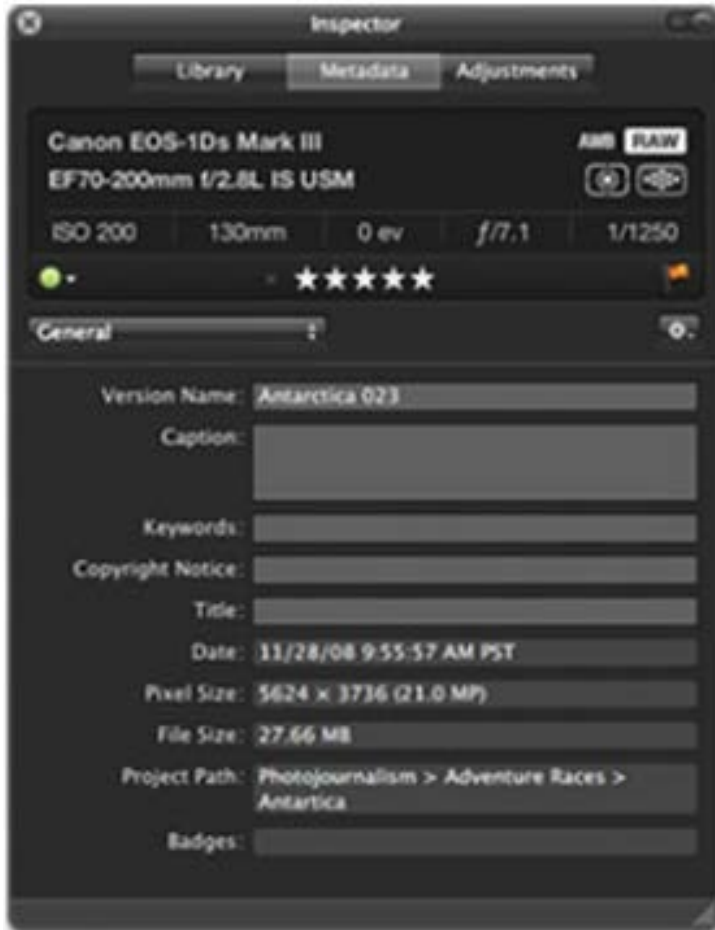
Adobe Creative Suite



North Plains

Metadata tools (3)

Apple Aperture



Microsoft Office 2013

The Tagging Problem

- ❖ How are we going to populate metadata elements with complete and consistent values?
- ❖ What can we expect to get from automatic classifiers?

Cheap and Easy Metadata

- ❖ Some fields will be constant across a collection
 - e.g., format, color, photographer or location
- ❖ In the context of a single collection those kinds of elements may add little value, but they add tremendous value when many collections are brought together into one place, and they are cheap to create and validate.






4 Indexing rules: How to use the taxonomy to tag content

Rule	Description
Use specific terms	Apply the most specific terms when tagging content. Specific terms can always be generalized, but generic terms cannot be specialized.
Use multiple terms	Use as many terms as necessary to describe <i>What the content is about & Why it is important.</i>
Use appropriate terms	Only fill-in the facets & values that make sense. Not all facets apply to all content.
Consider how content will be used	Anticipate <i>how the content will be searched for</i> in the future, & <i>how to make it easy to find it.</i> Remember that search engines can only operate on explicit information.

Automated tagging tools

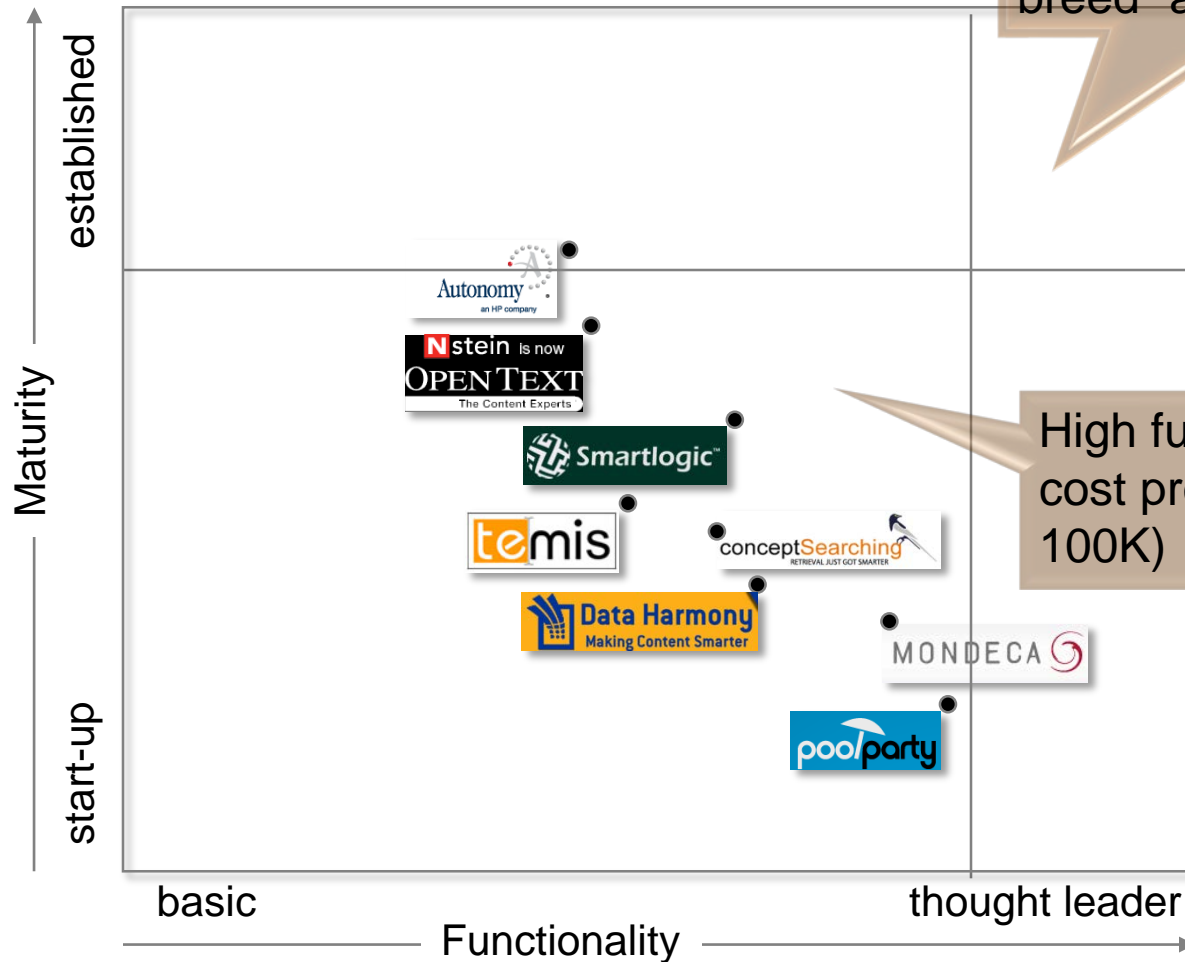
- ❖ Automatic classification tools exist, and are valuable, but results are not as good as people can do.
 - “Semi-automated” is best.
 - Degree of human involvement is a cost/benefit tradeoff.

Tools for tagging

Vendor	Tagging Tools	URL
	Autonomy Collaborative Classifier	www.autonomy.com/content/Functionality/idol-functionality-categorization/index.en.html
	ConceptSearching	www.conceptsearching.com
	Data Harmony M.A.I. TM (Machine Aided Indexing)	www.dataharmony.com/products/mai.html
	Intelligent Topic Manager	www.mondeca.com/Products/ITM
	nStein TME (Text Mining Engine)	www.nstein.com/en/products-and-technologies/text-mining-engine/
	PoolParty Extractor	poolparty.biz/products/poolparty-extractor/
	Semaphore Classification and Text Mining Server	www.smartlogic.com/home/products/semaphore-modules/classification-and-text-mining-server/overview
	Temis Luxid® for Content Enrichment	www.temis.com/?id=201&selt=1

Taxonomy tagging tools

An immature area– No vendors are in upper-right quadrant! No DAM vendors in this list. Tagging is a “best of breed” application



High functionality /high cost products (\$20-100K)

Tagging considerations

- ❖ Who should tag assets? Producers or editors?
- ❖ Taxonomy is often highly granular to meet task and re-use needs, but with detailed taxonomy it's difficult to get complete and consistent tags.
- ❖ The more tags there are (and the more values for each tag), the more hooks to the content, but the more difficult it is to get completeness and consistency.
- ❖ If there are too many tags or tags are too detailed, producers will resist and use “general” tags (if available)
- ❖ Vocabulary is often dependent on originating department, but the lingo may not be readily understood by people outside the department (who are often the users).

Agenda

- ❖ Metadata and interoperability
- ❖ Metadata tools
- ❖ Metadata standards
- ❖ Planning for metadata

Metadata standards

- ❖ The best thing about standards is that there are so many to choose from
 - Dublin Core. Vocabulary of fifteen properties for use in resource description. (<http://www.dublincore.org/documents/dces/>)
 - PRISM. Facilitate content management, aggregation, and search. (<http://www.idealliance.org/specifications/prism-metadata-initiative>)
 - IPTC Photo Metadata. Describe and administrate photographs, and provide the most relevant rights related information. (http://www.iptc.org/site/Photo_Metadata/)
 - IPTC News Codes. Sets of concepts to be assigned as metadata values to news objects like text, photographs, graphics, audio- and video files and streams. (<http://www.iptc.org/site/NewsCodes/>)
 - Schema.org. HTML tags recognized by major search engines including Bing, Google, Yahoo! and Yandex. (<http://schema.org/>)

Dublin Core Element Set

<http://www.dublincore.org/documents/dces/>

Elements

1. Identifier
2. Title
3. Creator
4. Contributor
5. Publisher
6. Subject
7. Description
8. Coverage
9. Format
10. Type
11. Date
12. Relation
13. Source
14. Rights
15. Language

Subject metadata –
What, Where & Why:
Subject, Type, Coverage

Use metadata –
When & How:
Date, Language, Rights

Asset metadata – *Who:*
Identifier, Creator, Title,
Description, Publisher,
Format, Contributor

Relational metadata –
Links between and to:
Source, Relation

DCMI Metadata terms

<http://www.dublincore.org/documents/dcmi-terms/>

Elements	Refinements		Encodings	Types
1. Identifier	Abstract	Is referenced by	Box	Collection
2. Title	Access rights	Is replaced by	DCMIType	Dataset
3. Creator	Alternative	Is required by	DDC	Event
4. Contributor	<i>Audience</i>	Issued	IMT	Image
5. Publisher	Available	Is version of	ISO3166	Interactive
6. Subject	Bibliographic citation	License	ISO639-2	Resource
7. Description	Conforms to	Mediator	LCC	Moving Image
8. Coverage	Created	Medium	LCSH	Physical Object
9. Format	Date accepted	Modified	MESH	Service
10. Type	Date copyrighted	<i>Provenance</i>	Period	Software
11. Date	Date submitted	References	Point	Sound
12. Relation	Education level	Replaces	RFC1766	Still Image
13. Source	Extent	Requires	RFC3066	Text
14. Rights	Has format	<i>Rights holder</i>	TGN	
15. Language	Has part	Spatial	UDC	
	Has version	Table of contents	URI	
	Is format of	Temporal	W3CTDF	
	Is part of	Valid		

PRISM elements by function

<http://www.idealliance.org/specifications/prism-metadata-initiative/prism>

<u>General Description</u>	<u>Identifiers</u>	<u>Provenance</u>	<u>Subject Description</u>
dc:title	dc:identifier	dc:publisher	dc:subject
dc:creator	aggregateIssueNumber	bookEdition	dc:description
dc:contributor	blogURL	distributor	dc:coverage
dc:language	doi	edition	academicField
dc:format	elssn	issueName	corporateEntity
aggregationType	isbn	number	event
alternateTitle	issn	publicationName	genre
blogTitle	issueIdentifier	sellingAgency	industry
byteCount	nationalCatalogNumber	seriesNumber	issueTeaser
contentType	url	supplementDisplayID	keyword
endingPage	productCode	volume	location
issueType	uspsNumber		object
pageCount	versionIdentifier		organization
pageProgressionDirection			person
pageRange			profession
publishingFrequency			sport
samplePageRange			sport
seriesTitle			teaser
startingPage			ticker
subtitle			timePeriod
supplementStartingPage			
supplementTitle			
wordCount			

PRISM elements by function (2)

Times and Dates

dc:date
coverDate
coverDisplayDate
creationDate
dateReceived
killDate
modificationDate
onSaleDate
onSaleDay
offSaleDate
publicationDate
publicationDisplayDate

Relations

dc:relation
hasAlternative
hasCorrection
hasTranslation
isAlternativeOf
isCorrectionOf
isTranslationOf
link
section
subsection

Rights & Use

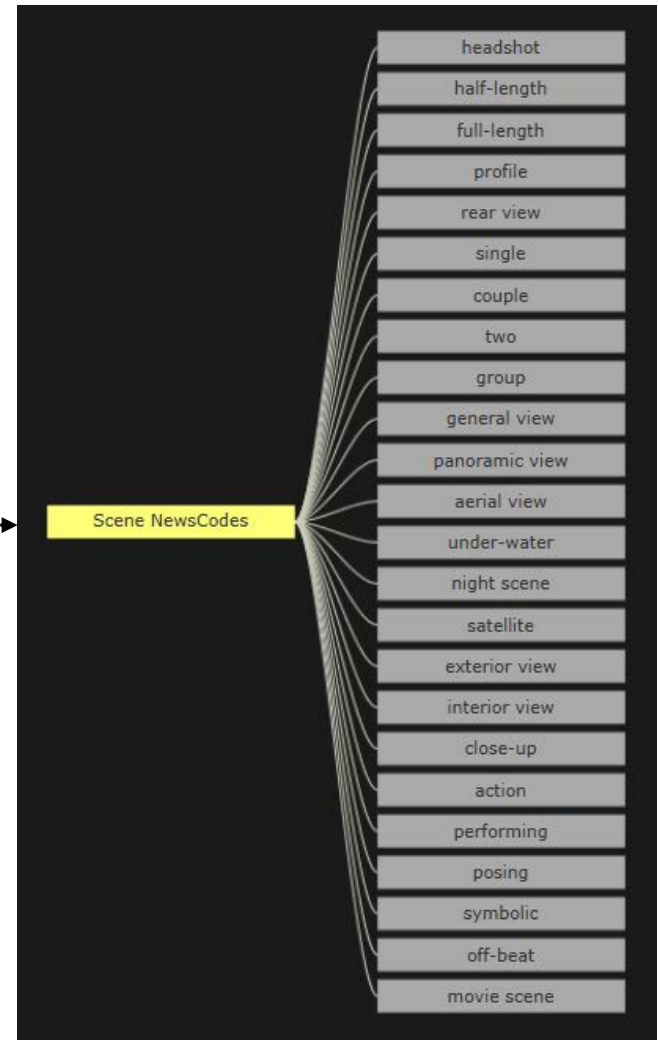
dc:rights
channel
complianceProfile
copyrightYear
device
originPlatform
platform
rating
subchannel

IPTC Photo Metadata

IPTC Core

- **Creator**
- Creator's Job Title
- Contact Information Details: —————>▪ Address
- Headline
- **Title**
- **Description**
- Description Writer
- Keywords
- IPTC Subject Code
- Date Created
- Intellectual Genre
- IPTC Scene Code>▪ City
- Location
- City
- State/Province
- Country
- ISO Country Code
- Job ID
- Instructions
- **Source**
- Copyright Notice
- Copyright Status
- Rights Usage Terms

IPTC Scene Codes



Type Hierarchy

❖ DataType

❖ Thing

- CreativeWork:
- Event:
- Intangible
- MedicalEntity:
- Organization:
- Person:
- Place:
- Product:

- Article:
- Blog:
- Book:
- Comment
- Diet:
- ExercisePlan:
- ItemList:
- Map
- MediaObject:
 - AudioObject:
 - ImageObject:
 - MusicVideoObject
 - VideoObject:
- Movie:
- MusicPlaylist:
- MusicRecording:
- Painting
- Photograph
- Recipe:
- Review:
- Sculpture
- SoftwareApplication:
- TVEpisode:
- TVSeason:
- TVSeries:
- WebPage:
- WebPageElement

Schema.org

itemprop

name
encoding format
content URL
duration
description

Original HTML:

```
<b>12oclock_girona.mp3</b>  
Total Time: 0m:15s - Recorded on a terrace of Girona a sunday morning  
composed by Roger  
  
<script type="text/javascript">  
var fo = new FlashObject("http://google.com/flash/preview-player.swf", "flashPlayer_719", "358", "16", "6",  
"#FFFFFF");fo.addVariable("url",  
"http://media.freesound.org/data/0/previews/719__elmomo__12oclock_girona_preview.mp3");fo.addVariable("autostart",  
"0");fo.write("flashcontent_719");  
</script>
```

With Schema.org:

```
<div itemscope itemtype="http://schema.org/AudioObject">  
  <span itemprop="name"><b>12oclock_girona.mp3</b></span>  
  
  <script type="text/javascript">  
    var fo = new FlashObject("http://google.com/flash/preview-player.swf",  
      "flashPlayer_719", "358", "16", "6",  
      "#FFFFFF");fo.addVariable("url", "http://media.freesound.org/data/0/previews/719__elmomo__12oclock_girona_preview.mp3");fo.a  
ddVariable("autostart", "0");fo.write("flashcontent_719");  
  </script>  
  
  <meta itemprop="encodingFormat" content="mp3" />  
  <meta itemprop="contentURL"  
content="http://media.freesound.org/data/0/previews/719__elmomo__12oclock_girona_preview.mp3" />  
  
  <span class="description">  
    <meta itemprop="duration" content="T0M15S" />  
    <span itemprop="description">Recorded on a terrace of Girona a sunday morning</span>  
  </span>  
</div>
```

Agenda

- ❖ Metadata and interoperability
- ❖ Metadata tools
- ❖ Metadata standards
- ❖ Planning for metadata

Using metadata

- ❖ How do you use metadata associated with digital assets?
- ❖ What metadata do you need to do what you need to do?
- ❖ What questions do you need to answer?

How do you figure out what metadata you need?

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

Identify themes: Retail example

- ❖ A number of stakeholders noted that the company has not given the category a lot of attention.
- ❖ Stakeholders repeatedly underscored the need for clear and timely communication on upcoming changes to the taxonomy and attribution framework. This will help teams plan for remediation efforts and will highlight new selling opportunities.
- ❖ Our stakeholder interviews revealed that there is not currently a mature quality assurance process in place to ensure that metadata is both valid and accurate.
- ❖ A few stakeholders wondered how competitors handle taxonomy and attribution for the category.
- ❖ Stakeholders underscored the need to capture the right amount of detail for each item type and its associated attributes.

Identify use cases: Retail example

- ❖ Category suggestions in search dropdown.
- ❖ Better faceted filtering.
- ❖ In-Store mobile applications that direct customers to products and make purchase suggestions.
- ❖ Dynamically generate seasonal collections.
- ❖ Multichannel seasonal merchandising.
- ❖ Shopping lists.
- ❖ Product research.

Identify key performance indicators (KPI's)

- ❖ Number of total assets
- ❖ Number of assets added during the period
- ❖ Number of assets used and re-used during the period
- ❖ Revenue from assets during the period
- ❖ ... Others?

Metadata Deep Dive

- ❖ How do the mechanics and tools of metadata impact your work? Metadata underpins DAM with a vast array of tools and processes, and the more that you understand it, the greater opportunity for DAM success. The seminar will be a deep dive into metadata (and taxonomy)—how metadata works and what business problems are solved. We'll talk about the metadata standards, how to generate cheap and easy metadata, and the key things you can do with that metadata. Finally we'll look at some examples of metadata (and taxonomy) for different types of assets. This Add On Seminar will:
 - Explore the mechanics of metadata and interoperability
 - Take an analytical look at the tools, some of which are free and some are not
 - Review metadata standards
 - Show how to manipulate metadata
 - Look under the hood of metadata in different file types and formats