

## **10 -TAXMET-Planning-a-content-model**

### **1. How to Start Planning a Content Model**

#### ***1.1 Welcome***



#### **Notes:**

Welcome to this course in how to start planning a content model.

## **1.2 Learning Objectives**

# Learning Objectives

*click each tab below for more details*

- Define
- Distinguish
- Articulate Value

©AIIM 2018

 aiim

### **Notes:**

We're now going to explore the topic of content modeling. And the learning objectives for this particular session include defining a few key terms. That includes content model, content type, and content element. We've touched on these a little bit in previous sessions, but here, we'll get into exactly how you define them and explore each one in-depth.

We'll also explore how to distinguish between structured and unstructured content, and we'll also articulate the value as well as the difficulty in exploiting structured content.

**Define (Slide Layer)**

The slide has a teal header bar with the title "Learning Objectives" and a sub-instruction "click each tab below for more details". Below the header are three tabs: "Define", "Distinguish", and "Articulate Value". The "Define" tab is currently active, showing a white content area with the text "Define: Content Model, Content Type, and Content Element". The other tabs are shown in smaller, semi-transparent boxes. At the bottom left is the copyright notice "©AIIM 2018" and at the bottom right is the AIIM logo.

**Learning Objectives**

*click each tab below for more details*

Define

Distinguish

Articulate Value

Define: Content Model, Content Type, and Content Element

©AIIM 2018

**aiim**

## Distinguish (Slide Layer)

The slide features a teal header bar with the title "Learning Objectives" and a sub-instruction "click each tab below for more details". Below the header are three teal tabs: "Define", "Distinguish", and "Articulate Value". A callout box with a white arrow points from the "Distinguish" tab to the text "Distinguish between structured and unstructured content". At the bottom left is the copyright notice "©AIIM 2018" and at the bottom right is the AIIM logo.

**Learning Objectives**

*click each tab below for more details*

Define

Distinguish

Articulate Value

Distinguish between structured and unstructured content

©AIIM 2018

**aiim**

## Articulate (Slide Layer)

**Learning Objectives**

*click each tab below for more details*

Define

Distinguish

Articulate Value

Articulate the value and difficulty in exploiting unstructured content

©AIIM 2018

**aiim**

### **1.3 Agenda**

The slide features a teal header with the word "Agenda". Below the header is a white horizontal bar containing the title "Introduction to Content Modeling". Underneath this bar is a large, semi-transparent background image showing a diverse group of people looking upwards. Overlaid on this background are three more teal bars, each containing one of the four agenda items: "Modeling a Repository", "Understanding Content Types", and "Structured vs Unstructured Content".

# Agenda

- Introduction to Content Modeling
- Modeling a Repository
- Understanding Content Types
- Structured vs Unstructured Content

#### **Notes:**

Let's start by defining what we mean by content modeling.

## 1.4 What is a content model?

# What is a Content Model?

**Components or “elements” that make up a body of content**

- The folder or “meta”-structure of a repository or enterprise information set
- The document specs
- Associated metadata
- Elements within a (structured) document

**A framework applied to content to create relevant information**

- Making those related pieces useful to the people who need it

This is how you need to see and think about content

**aiim**

### Notes:

A content model is the representation of components that make up a body or a corpus of content. And I used this golf hole as an example simply because a picture of a golf hole is made up of many separate components. If you took those components separately, it wouldn't necessarily carry a lot of meaning. The number 250, which is on the left of that picture, if you take it by itself, it doesn't necessarily mean something specific. But when you take that content and you put it in the larger context of this picture, it actually represents the distance between two points on this particular golf hole.

So that's really about the metastructure of this particular representation. And a content model is really the metastructure of an entire repository as well, or an entire information set. It's also about the document types, and about those elements within a structured document or any piece of content. In this case of the picture of the golf hole, those elements include the picture itself, the numbers, and the different ways that those numbers and pictures are associated with one another.

A content model can also be defined as a framework applied to content to create

information. It's making those related pieces useful to the people who need it. If you're embarking on an information management project, this is really how you need to see and think about information. What are the components? How do they relate to one another?

You might be wondering a little bit, well, what exactly is the difference between a content model and an ontology? Well, a content model also occupies itself with the granularity of content, how content breaks down, what are all the pieces of each instance of content. So in this case, this golf hole, as I mentioned, each individual number makes up the content model. Each part of the picture is a different part of the content model. The content model, as opposed to the ontology, is more occupied with how a machine or a repository will store the particular information and how it will be able to process it.

### **1.5 What makes up a content model?**

## What Makes Up a Content Model?

- Taxonomy**
  - Law for categorizing information
- Metadata and/or Content Attributes**
  - Information about content: "data about the data"
- Content Architecture**
  - How information is "chunked" and organized in a repository or across an enterprise
- Content Types**
  - How documents and other content can be grouped and managed functionally
- Content Elements**
  - Basic chunks or building blocks of structured content



©AIIM 2018

**aiim**

#### **Notes:**

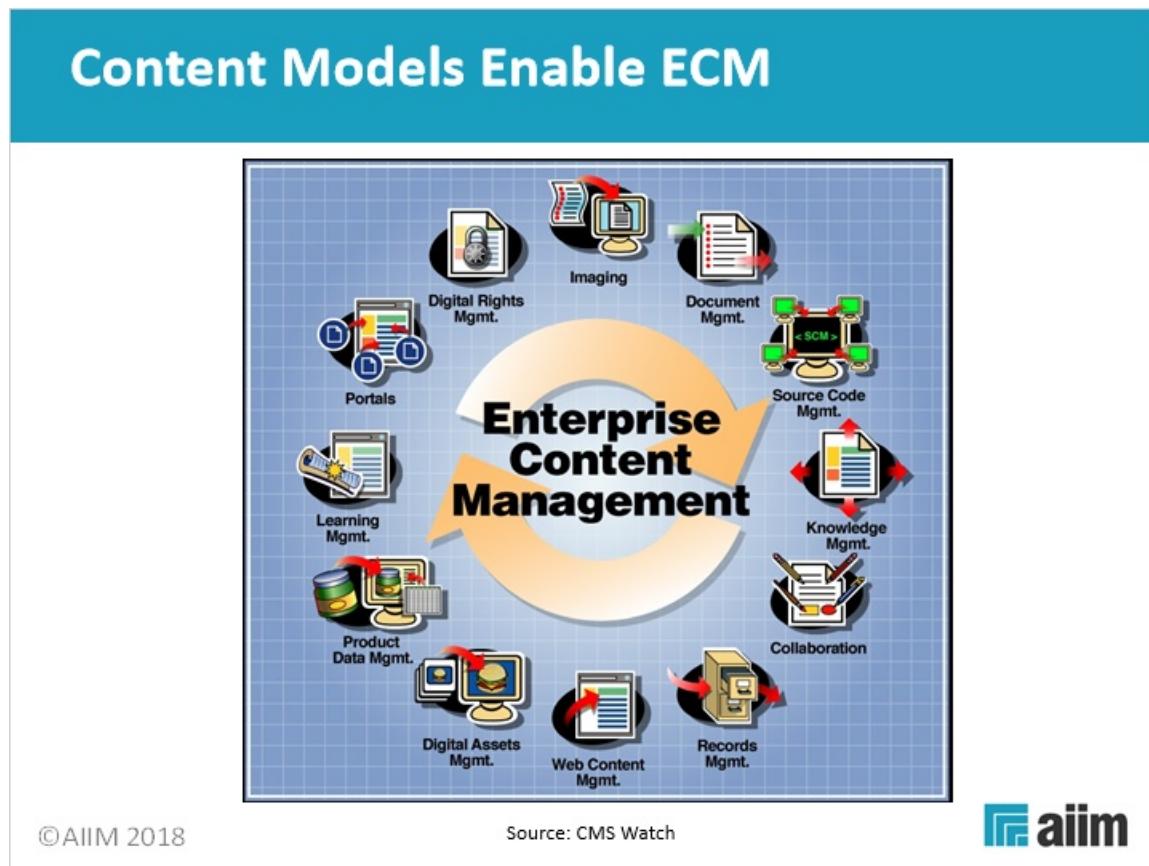
There are a number of things that play into the full scale content model. First of all, you

have the taxonomy, which, as you know, is the law for categorizing information. You have the metadata or the information about the content. You have the content architecture. So that's how the information is broken down, and at what level or granularity you wish to store it. So in my golf hole example, you have to decide, are you going to store that piece of information as a whole unit, so the picture with all of the numbers associated with it? Or are you going to break it down at a more granular level and perhaps store the image separate from the numbers that are associated with it?

Then you have content types. So this is, as we spoke about a little earlier, how documents and other content can be grouped and managed functionally. So a press release is a software content type. And you would put those together in a group and probably manage them all in the same way. And then finally content elements. These are the basic building blocks of structured documents. What are the pieces within a document? How specifically would they break down? So those fit into the content architecture.

Content models are just as important if not more important than the technology you use. Many people get hung up on technology as the solution to findability problems or information access problems. In reality, if the content model isn't consistent, those pieces of technology can't access content in a way that's really effective. So remember, the content model is just as important if not more important than the process of buying technology, implementing technology. None of that's going to work if your content model isn't solid.

## 1.6 Content models enable ECM



#### Notes:

Remember that these content models enable ECM. I talked earlier about taxonomies and categorization enabling enterprise content management and content technologies as a whole. Content models in the larger sense, of which taxonomies are a part, do the same. So how content is broken down, the metadata around the content, and the structure is all very significant to these different technologies. They all use structure and categorization and content architecture in different ways.

So as you look at these technologies, you also need to consider that some of these technologies can handle very granular content models, such as rows and tables in a database, while others can only consider document level content models and won't actually manage content that's broken down very granularly. So as you look at these technologies and you design your content model. You need to know what's possible within the technology you already have. Or, if you're considering a new technology, be sure to look at how your content is already structured, how you might want to structure it in the future, and make sure that that strategy is aligned with the type of technology that you're choosing.

## ***1.7 Agenda***

The slide features a teal header with the word "Agenda". Below it is a white horizontal bar containing the title "Introduction to Content Modeling". Underneath this is another white bar with the title "Modeling a Repository". Below that is a third white bar with the title "Understanding Content Types". At the bottom is a fourth white bar with the title "Structured vs Unstructured Content". The background of the slide is a grayscale photograph of a diverse group of people looking upwards. A faint blue overlay with various icons (globe, magnifying glass, document, etc.) is visible across the top half of the image.

# Agenda

- Introduction to Content Modeling
- Modeling a Repository
- Understanding Content Types
- Structured vs Unstructured Content

### **Notes:**

Let's start by defining what we mean by content modeling.

## 1.8 Repositories

Click each tab for audio and more details

Modeling

Hierarchy

Example

Properties

©AIIM 2018

**aiim**

### Notes:

Modelling content for a repository and figuring out how the content should break down, into what sorts of pieces and how that should be categorized is fundamentally an exercise in taxonomy development. It's really taking your taxonomy to the next level and looking at the content itself, and figuring out, well, based on what I need to do with this content, does it make sense to store it as a whole document? Or does it make sense to break it into pieces? Based on how I need to retrieve this in the future, do I need to categorize each section of a document? Or can I just categorize the entire document and achieve the same result?

Remember that the more granular you get with a repository and the more the content breaks down, the more work it is to manage. So you need to strike a balance there based on your needs.

Folders and virtual folders do tend to be hierarchical in general, by nature. However, there's a lot of opportunities for faceted organization as well.

You might be thinking about modelling the entire enterprise information set. This is really challenging to do at once, because you're going to have different repositories. You're going to have different opportunities to organize content logically rather than physically. And in some cases, those different repositories are already different repositories for that reason, because the information within them lends itself to be managed in different ways. So maybe in one repository, you do need a very granular content model, whereas in another, it might be sufficient to just manage, tag, and categorize at a much higher document level.

Figuring this out is really a matter of looking at what you need to do with the information and how it needs to be accessed in the future. You should determine a set of use cases for future access so that you're not overdoing the modelling process. You don't want to store at too granular of a level. It's just far too much work. So know how that information needs to be searched and accessed in the future, and model your repository to accommodate that.

Here's a possible repository hierarchy which could also be seen as a taxonomy or one-part of our larger content model. You could have cabinets within your repository with specific work spaces. And maybe those work spaces break down into different groups or folders, which then can break into subfolders. And then you have a set of documents within those folders. And this is essentially putting pieces of content into a hierarchical taxonomy or categorisation scheme.

Looking at something that's a little bit more granular, here's an example of a repository that is a database record. So on the left, we have a navigation tree of folders very similar to the last example I showed. But it's organised by various departments. And then within accounts payable, we're looking at dates.

And then we get a list of different accounts that were used and that invoices were issued for. And it gives essentially rows and table information that's displayed to the end user. And you see the difficult metadata that is then populated with various values. So this is a little bit more granular of a content model, but it follows a very similar paradigm. Here, you're using both a taxonomy as well as metadata and it's managed at quite a granular level. Vendor is stored separately from the invoice number from the authorisation person from the invoice date, et cetera.

This level of granularity is particularly useful in this scenario because it allows the user to sort by each of these individual criteria. So if they wanted to look at the vendors in alphabetical order, or if they wanted to sort by invoice date, they also have the option to do that. If this information was stored at a document level and not this level of granular, they wouldn't be able to sort by these various criteria. So again, you want to look at your content usage before you determine the level at which you want to break it down and store it within the repository.

Some software allows you to set properties on something like a cabinet, whereby if you put a document into that cabinet, it then inherits specific metadata or attributes, or, on a document-by-document basis, you can assign specific attributes. In this particular example, you can see that there's a name associated with a department as well as an owner which is chosen from a drop-down menu. You can add a description if you like. And all of this contributes to the future findability of the information. Each of these specific fields would be searchable in the future, so that if you wanted to look for all of the documents that were owned by Cameron, you would be able to do that, simply by picking his name from the drop-down, and then retrieving based on that criteria.

So as you think about assigning attributes to documents, as you check them into some sort of system, be sure to think about, again, what is the future access need for this particular document? And those are the sorts of attributes you should be putting into your content model, and allowing people to assign as they add documents to a system.

## Modeling (Slide Layer)

# Repositories

*Click each tab for audio and more details*

- Modeling**
- Hierarchy**
- Example**
- Properties**

### Modeling a Repository

Fundamentally an exercise in taxonomy development

- Folders and “virtual” folders tend to be hierarchical
- But there are opportunities for faceted organization as well

Modeling the enterprise information set

- Huge challenge to do at once
- Model different repositories, but look for opportunities to organize content logically, rather than physically

©AIIM 2018



## Hierarchy (Slide Layer)

# Repositories

*Click each tab for audio and more details*

- Modeling
- Hierarchy
- Example
- Properties

A Possible Repository Hierarchy

```
graph TD; Root[Repository] --> Cabinets[Cabinets]; Root --> Workspaces[Workspaces]; Root --> FG[Folder Groups]; Root --> DV[Document Versions and Renditions]; Cabinets --> Folders[Folders]; Folders --> Subfolders[Subfolders]; Subfolders --> Documents1[Documents]; Subfolders --> Documents2[Documents]; Subfolders --> Documents3[Documents]; Subfolders --> Documents4[Documents]
```

The diagram illustrates a possible repository hierarchy. At the top level is the **Repository**, which contains **Cabinets**, **Workspaces**, **Folder Groups**, and **Document Versions and Renditions**. The **Cabinets** node contains **Folders**, which in turn contain **Subfolders**. Each **Subfolder** contains four **Documents**.

©AIIM 2018

**aiim**

## Example (Slide Layer)

# Repositories

*Click each tab for audio and more details*

Modeling

Hierarchy

Example

Properties

### What a Repository Looks Like

The screenshot shows the DocuXplorer - Enterprise software interface. On the left, there is a 'Library Tree' pane with the following structure:

- Library
  - DX In Box
  - Recycle Bin
  - Work Folder
  - Accounts Payable
    - AP 2002
      - 01 Jan
      - AP 2003
        - 01 Jan
        - 02 Feb
        - 03 Mar
    - Clients
    - Human Resources
    - Marketing

On the right, there is a grid view titled 'Index Set: AP' with the following data:

Vendor	Inv. #	Authorized by	Inv. Date	Inv. Am
1 Staples	47225	Preston Sturges	1/9/2003	\$13.
2 Gross Foods	485	Howard Hawks	1/24/2003	\$8.
3 Fred Astaire Dan	230	Preston Sturges	1/15/2003	\$25.
2 PC World	2544	Ella Kazan	1/18/2003	\$2.

Source: ScanPaper

©AIIM 2018



## Properties (Slide Layer)

# Repositories

*Click each tab for audio and more details*

- Modeling
- Heirarchy
- Example
- Properties

### Setting Properties in a Cabinet

Edit Accounting / Finance Cabinet

General Membership

Attributes

Name	Accounting / Finance
Owner	Cameron Siguenza
Description	(empty)
Default Schema	Accounting

This Cabinet Contains

Drawers	2
Folder Groups	3
Folders	3
Files (latest version)	37

Source: FileHold

©AIIM 2018

**aiim**

## **1.9 Agenda**

The slide features a teal header with the word "Agenda". Below the header is a background image of a diverse group of people looking upwards. Overlaid on the image are four white rectangular boxes, each containing a topic title. From top to bottom, the titles are: "Introduction to Content Modeling", "Modeling a Repository", "Understanding Content Types", and "Structured vs Unstructured Content".

- Introduction to Content Modeling**
- Modeling a Repository**
- Understanding Content Types**
- Structured vs Unstructured Content**

### **Notes:**

Let's start by defining what we mean by content modeling.

## 1.10 Intro to Content Types

# Introduction to Content Types

*click each number below for more details*

©AIIM 2018

**aiim**

### Notes:

Wrapping up, let's summarize what we learned in this course. A taxonomy is a controlled vocabulary standard and a law for categorizing information. The categories of a taxonomy should be derived based on user needs, enterprise requirements and the nature of the content that you're dealing with.

When possible, start with an industry standard. There's also a lot of standard taxonomies out there that you can use and modify for your needs. Content technologies use taxonomies and other sorts of controlled vocabularies constantly. They need them to operate, to access content, to display content, and deliver things to the right user at the right time.

And finally, folksonomies are consumer-driven classification schemes that are particularly beneficial with social networking applications, but you should question the value of them for the enterprise. While they might seem like the easy way to classify your content, it won't make your information more interoperable and it won't necessarily make your publication process, your content management process any simpler. You

need to establish a standard for that really to happen.

### Arc 1 (Slide Layer)

## Introduction to Content Types

*click each number below for more details*

Core building block for managing a corpus of content - often known as "document type"

©AIIM 2018

aiim

Arc 2 (Slide Layer)

## Introduction to Content Types

*click each number below for more details*

1

2

3

4

5

Typically a functional distinction  
(case study, research finding,  
interoffice memo, claim form,  
press release, etc.)

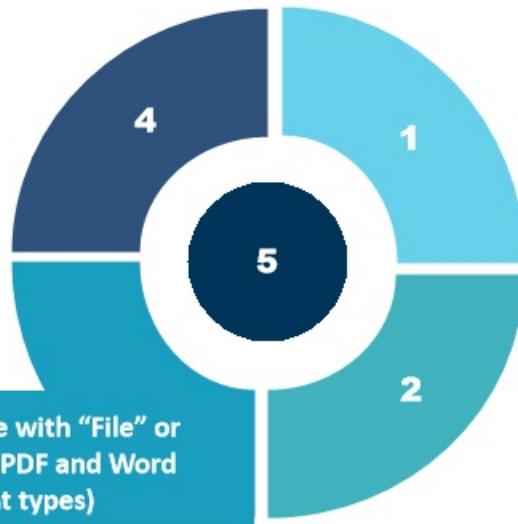
©AIIM 2018

**aiim**

Arc 3 (Slide Layer)

## Introduction to Content Types

*click each number below for more details*



©AIIM 2018



## Arc 4 (Slide Layer)

# Introduction to Content Types

*click each number below for more details*

Sometimes have “sub-types”  
*(applied research finding, monthly budget reconciliation)*

1  
2  
3  
4  
5

©AIIM 2018

**aiim**

Oval - 5 (Slide Layer)

# Introduction to Content Types

*click each number below for more details*

A donut chart divided into five segments, each containing a number from 1 to 5. Segment 5 is highlighted with a callout box containing a note about "one-offs".

Segment	Value
1	Light Blue
2	Medium Blue
3	Dark Blue
4	Very Dark Blue
5	Very Dark Blue (highlighted)

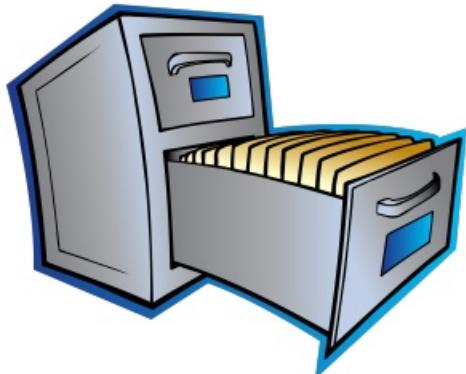
Some documents may seem to be  
“one-offs” (or belong to a  
“generic” type – be careful here)

©AIIM 2018

**aiim**

## **1.11 When will you use content types?**

### **When Will You Use Content Components and Types?**



- Whenever you create or ingest new content
- Whenever you store content
- Whenever you process, route, or publish content
- Whenever you archive or delete content
- Whenever you gather statistics

#### **“Content Item”**

- Instance of a content type (the actual content itself)

©AIIM 2018



#### **Notes:**

You might ask the question, well, when will I actually use content types? And when will I use the components that content types break into? Well, that's all part of the process of creating your content model. And your content model is used whenever you create or ingest new content, whenever you store content, process it, route it through something or publish it, whenever you archive or delete it, and whenever you gather statistics. All of these things happen based on the content types and the components within that content type.

A content item is a single instance of a content type. It's the actual content itself. So when you think of a press release as a content type, you might have 600 press releases that you've published over the last few years. Each of those 600 instances is a content item and is an example of that particular content type. That content type in turn breaks down into content components. Those content components, in a press release, would typically be a title, a city or a location where the press release was put out, body text, perhaps some footer information about the company that's issuing the information. It's those components that make up the content type and what helps to dictate what exactly happens to that content.

Oftentimes if a component changes, based on the type of content it is, it then gets routed to a certain person. So as an example, if a press release is updated, that might send the notification in a workflow process to an editor, a specific editor whose job it is to edit and then publish press releases.

So all of these items are tied together. And it's that larger content model and how content is placed into that content model that triggers a lot of these different actions.

### **1.12 Importance of Content Types**

## Importance of Content Types

**Automated content management, business processing, and search pivots off them**

**Allows you to generalize, analyze, and create processes**

- Processing rules (e.g., routing and retention schedules)
- Publishing rules (e.g., authoring norms and approval processes)
- Security rules (e.g., access schedules)
- Other rules ...

**• Helps distinguish structured from unstructured or semi-structured documents**

©AIIM 2018



**aiim**

#### **Notes:**

As I just eluded to, the importance of content types, first and foremost, is that it really allows you to automate content management, business process, and search tools. In fact, these sorts of tools really pivot off of content types. You specify how certain content types should be published, should be routed through a workflow, should be archived or deleted. In most cases, that's based on a content type.

It allows you to generalize and analyze content based on those particular types, setting those processing rules, those publishing rules, and of course security rules as well. As an example, you might not want most of your organization to have access to financial documents or other sorts of confidential materials. Based on that content type, financial documents, you can map that to a specific group of users, perhaps the executive team or the financial department, and only they will have access to that particular type. That's exactly the sort of information that content technology hinges off of.

It also helps us distinguish structured from unstructured or semi-structured content. You can have specific content types for each of these different formats of content, and then manage each appropriately or route them to the right repository so that the structure can be accommodated.

### **1.13 Governance: Sponsor**

The slide has a teal header with the title "Governance: Sponsor" and a sub-instruction "Click each tab for more details". On the left, there is a vertical stack of six teal rectangular tabs labeled "Ensure", "Help", "Harness", "Get", "Provide", and "Validate". To the right of these tabs is a white callout box containing the text "Click each box on the left to learn more". In the background, there is a faint image of a diverse group of people looking up at a woman in a blue blouse and black pants who is standing with her arms crossed, smiling.

#### **Notes:**

A taxonomy is a controlled resource, which means that there are defined roles, processes and structures for controlling it. It is also a living resource which means there has to be a mechanism for updating it as its information environment changes. This is why Governance is such an important topic.

Ensure you know the taxonomy sponsor. THE role of a taxonomy project sponsor is to:

- ENSURE senior management's objectives are met
- HELP you prioritize
- HELP you identify stakeholders
- HARNESS support from senior colleagues
- GET you access to the people and resources you need
- PROVIDE political cover when you run into problems
- VALIDATE your business case, plan and evaluation process
-

## Ensure (Slide Layer)

# Governance: Sponsor

*Click each tab for more details*

Ensure senior management's objectives are met

**Harness**

**Get**

**Provide**

**Validate**

## Help (Slide Layer)

# Governance: Sponsor

*Click each tab for more details*

The slide features a central image of a woman standing with her arms crossed, smiling, with a group of diverse people in the background looking up at her. To the left, there is a vertical stack of five teal-colored tabs with white text: "Help", "Harness", "Get", "Provide", and "Validate". A large, semi-transparent white speech bubble is positioned over the tabs, containing the text "Help you:" followed by a bulleted list: "• Prioritize" and "• Identify stakeholders".

Help

Harness

Get

Provide

Validate

Help you:

- Prioritize
- Identify stakeholders

## Harness (Slide Layer)

# Governance: Sponsor

*Click each tab for more details*

The slide has a teal header with the title 'Governance: Sponsor' and a sub-instruction 'Click each tab for more details'. Below the header is a large white speech bubble containing the text 'Harness support from senior colleagues'. To the right of the speech bubble is a photograph of a woman with short dark hair, wearing a blue blouse and black trousers, standing with her arms crossed. The background of the slide features a teal sidebar with four tabs: 'Harness', 'Get', 'Provide', and 'Validate'. Below the sidebar is a faint background image of a diverse group of people looking up at the woman.

Harness support from senior colleagues

**Harness**

**Get**

**Provide**

**Validate**

## Get (Slide Layer)

# Governance: Sponsor

*Click each tab for more details*

The slide has a teal header with the title 'Governance: Sponsor' and a sub-instruction 'Click each tab for more details'. Below the header is a large white speech bubble containing the text 'Get you access to the people and resources you need'. To the right of the speech bubble is a photograph of a woman with short dark hair, wearing a blue blouse and black pants, standing with her arms crossed. In the background, there is a faint image of a diverse group of people looking up at the woman. On the left side of the slide, there are four teal-colored tabs with white text: 'Harness', 'Get', 'Provide', and 'Validate'. The 'Get' tab is highlighted with a white border.

- Harness
- Get
- Provide
- Validate

## Provide (Slide Layer)

# Governance: Sponsor

*Click each tab for more details*

Provide political cover when you run into problems

**Harness**

**Get**

**Provide**

**Validate**

## Validate (Slide Layer)

# Governance: Sponsor

*Click each tab for more details*

Validate you business case, plan and evaluation process

**Harness**

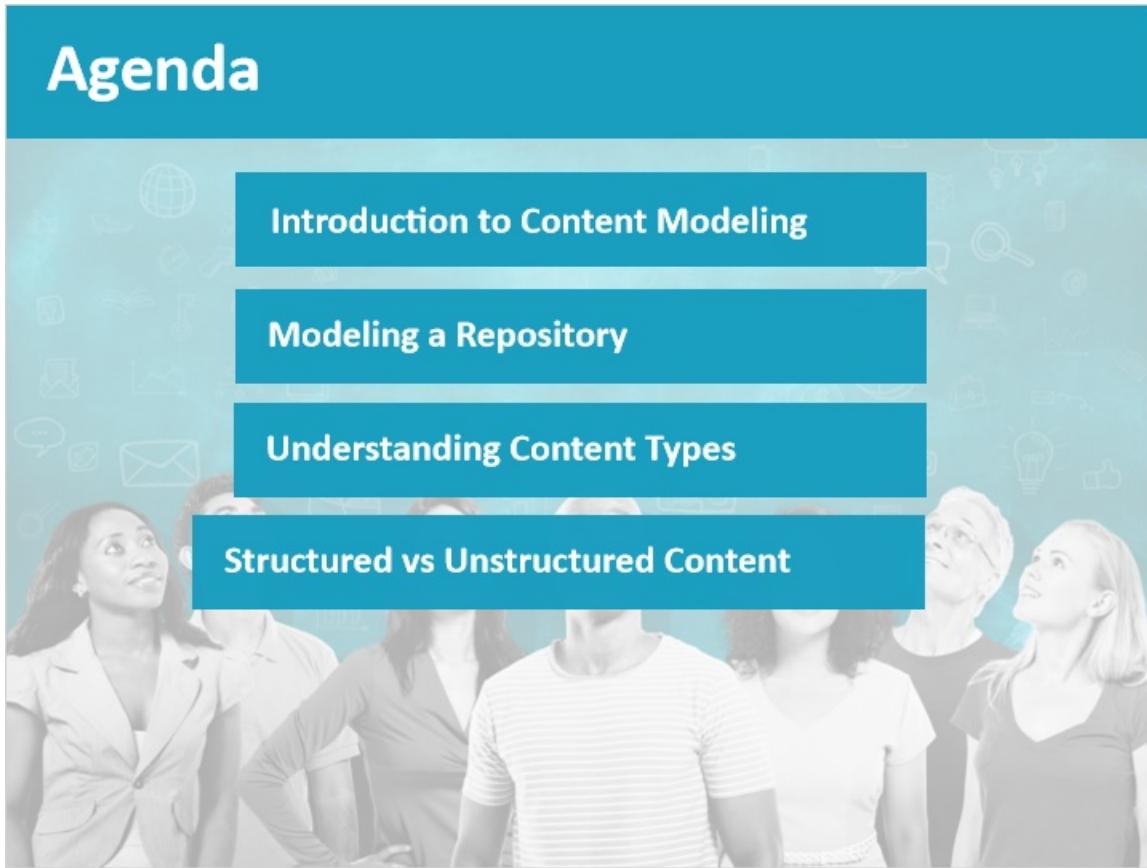
**Get**

**Provide**

**Validate**

## **1.14 Agenda**

# **Agenda**



The background of the slide features a photograph of a diverse group of people of various ages and ethnicities, all looking upwards and to the right with expressions of interest and engagement. Overlaid on this image are four horizontal blue bars, each containing white text. From top to bottom, the bars read: "Introduction to Content Modeling", "Modeling a Repository", "Understanding Content Types", and "Structured vs Unstructured Content".

- Introduction to Content Modeling**
- Modeling a Repository**
- Understanding Content Types**
- Structured vs Unstructured Content**

### **Notes:**

Let's start by defining what we mean by content modeling.

## 1.15 (Un)Structured Content

# (Un)Structured Content

 Structured Content	 Unstructured Content
<p>Follows a consistent model</p> <p>Is comprised of discrete, "chunkable" elements</p> <ul style="list-style-type: none"><li>And therefore often, though not always, stored in a database</li></ul> <p>Not all content types are structured</p> <ul style="list-style-type: none"><li>In fact most enterprise content is unstructured</li></ul>	<p>Does not follow an internal model with consistent elements</p> <ul style="list-style-type: none"><li>Most imaged documents are unstructured</li><li>Therefore stored most readily in a file-system repository</li></ul> <p>Unstructured content can still be indexed</p> <ul style="list-style-type: none"><li>Structure does not equal metadata or classification</li><li>Indexing content (even with full-text OCR) is not decomposing it</li></ul>

©AIIM 2018



### Notes:

Structured content follows a consistent model. It's very commonly in rows and databases, and it's comprised of discrete and decomposable elements meaning they break down and they're very granular.

Not all content types are structured. As we've said a few times, most enterprises have about 80% of their content unstructured. That unstructured content is more what you're commonly used to dealing with every day - email, Word documents, image documents, or sound documents, audio files, whatever you'd like to call them. These are very freeform. They don't necessarily break down into any fixed structure.

Unstructured content is a little bit more challenging to manage. It can still be indexed, and often that index information is structured, but note that that doesn't necessarily equal metadata or classification. Let's get into some more depth on this topic and the challenges around structured versus unstructured content.

## ***1.16 Press Release as structured content***

# Press Release As Structured Content

Structured content types have elements

Elements are:

- Named parts of a content type
- Individually stored and accessible units within a content type
- A basic unit of content

The diagram illustrates the structure of a press release as structured content. It features a central box labeled "Press Release" with various sections and their corresponding labels:

- LOGO**: Points to the "idev" logo.
- PRESS RELEASE**: Points to the main title area.
- HEADLINE**: Points to the headline section.
- SUBHEAD**: Points to the subhead section.
- FIRST PARAGRAPH**: Points to the first paragraph of the body text.
- DATE**: Points to the date element.
- CONTACT**: Points to the contact information element.
- FOR IMMEDIATE RELEASE**: Points to the "FOR IMMEDIATE RELEASE" text.
- DATE**: Points to the "May 15, 2001" date.
- CONTACT**: Points to the "CONTACT: Sarah Tricha, Associate Producer" section.
- NAME**: Points to the "Production Director" role.
- PHONE**: Points to the "1-800-445-7345 x322" phone number.
- EMAIL**: Points to the "Sarah.Tricha@idev.com" email address.
- HEADLINE**: Points to the "IDEV Redesigns Website for Health Industry Group Purchasing Association" headline.
- SUBHEAD**: Points to the "Redesigned Website Intended to Expand the Resources of the Association Staff" subhead.
- FIRST PARAGRAPH**: Points to the "Silver Spring, MD – May 15, 2001 – IDEV, a full-service Web development and consulting agency in metro-D.C., launched the redesign of the Health Industry Group Purchasing Association (HIGPA) website today."

Below the main box, two quotes provide context:

HIGPA, a broad-based trade association that represents group purchasing organizations (GPOs) and their trading partners, had maintained a web presence for the past two years, but their current website was no longer accurately communicating their mission. "Initially we were very pleased with the web site," Sandy Gold, Operation Manager at HIGPA stated of their former web presence, "but it was no longer meeting our goals."

IDEV's initial goals for the project were to improve site organization and user interface. "We did that and more without losing respect for the constraints of their budget," said Sarah Tricha, Associate Producer on the project.

"IDEV was great to work with throughout the entire process. They helped us to prioritize our needs and focus on what was important. We could really focus on the features that would be of the greatest benefit to our membership and staff. Consequently, we were able to afford much more functionality than originally anticipated," said Gold.

©AIIM 2018

aiim

## Notes:

You can bring structure to unstructured content. So you might have a press release, for example, that's in a Word document that isn't formatted particularly consistently. However if you need to put it on a website, as in this example, you can break down the pieces of it into a little bit more of a discrete format so that you can then present perhaps pieces of it in different areas of the page, or even different areas of a website.

Structured content types have elements. So that word ‘element’ has come up quite a few times. And in this particular case, the elements are the logo, the date, the contact information, the headline for the article, the subhead, the first paragraph, and then it goes into the rest of the text. So these elements are those named parts of the content type. So if I was to say to you, “What are the elements of a press release?” you would name these particular items.

When you move content like this into a more structured repository, you can then work with it in a more flexible manner. It's a basic unit of content, essentially. Part of the power of adding this structure is that if elsewhere on this website you simply wanted to

list headlines, you could list those headlines without having to retype or repeat that same piece of content. The headline is stored once and can be reused numerous times. If you were storing this entire press release as one single chunk of content, you wouldn't be able to pull up just the headline individually.

So this is how you need to think when you're looking at your content model, thinking about your content types and how they will break down. Again, you need to think about, well, what's the intended use of this content? What are the access needs going to be? And then structure and breakdown the content appropriately to accommodate those needs.

### 1.17 Chunking the press release

## Chunking the Press Release

### Elements ...

Are stored separately, so they can “travel” separately

- Often stored in database, but not essential
- Indeed, not all database content is “structured”
- XML is a common “persistence,” representation, or management format

Enable parts of a document to be re-used in other contexts

- Essential to lightweight syndication

```
- <heading>
- <headline>
  IDEV Redesigns Website for British Health Indus
</headline>
- <subhead>
  Redesigned Website Intended to Expand the Res
</subhead>
<heading>
- <Contact>
  <Name>Lynn Cheryan</Name>
  <Title>Production Director</Title>
  <Phone>+44-72-435-1234</Phone>
  <Email>lcheryan@idev.co.uk</Email>
</Contact>
- <Body>
- <Para>
  London, UK, 15 October, 2007 – IDEV, a full-s
</Para>
- <mainBody>
  HIGPA , a broad-based trade association that rej
  their mission. "Initially we were very pleased with
  organization and user interface. "We did that and
</mainBody>
- <closingPara>
  Founded in 1993 and headquartered in London, I
  technical as well as creative talent under a single r
</closingPara>
</Body>
```

The logo for AIIM (Articulate Instructional Media) features a stylized blue square icon followed by the lowercase letters "aiim".

©AIIM 2018

### Notes:

Let's decompose that press release a little bit. So the elements - as I mentioned, those elements are stored separately so that they can travel separately. By travel of course I mean publish them as discrete pieces on various parts of the website, multiple pages, and yet only have one instance of that content.

Published by Articulate® Storyline www.articulate.com

They're often stored in a database, but that's not essential. And let's remember though that database content is not necessarily all structured. You can store a big blob such as an image within a database. That's not necessarily a structured piece of information. The XML that's depicted on the right here is essentially a representation of how that content depicted on the last slide for the press release might be stored if it was in XML. So you can see each of these elements are represented by an XML tag. We have the headline, the subhead, the contact information, which then breaks down even to more discrete elements such as name, title, phone, email, et cetera. And then the paragraphs of the press release itself are also stored individually.

And this is what enables parts of a document to be reused in other contexts. It's essential for lightweight syndication. It's a very simple way of modelling your content that adds a lot of power and a lot of portability.

### **1.18 Syndicating the press release**

## **Syndicating the Press Release**

Elements can be “recomposed” for different channels and formats

Example: lightweight headline syndication using RSS

- Employs an XML format
- Each node is an element (*Headline* becomes *Title*; *First paragraph* becomes *description*)

Content Management system can automatically generate from components in repository

```
- <item>
  - <title>
    IDEV Redesigns Website for British Health Industry Group Purchasing Association
  </title>
  - <description>
    London, UK, 15 October, 2007 – IDEV, a full-service Web development and consulting agency in metro London
    launched the redesign of the British Health Industry Group Purchasing Association (BHIGPA) website today....
  </description>
  <link>http://www.idev.co.uk Releases/2007/BHIGPA.html</link>
</item>
```

©AIIM 2018



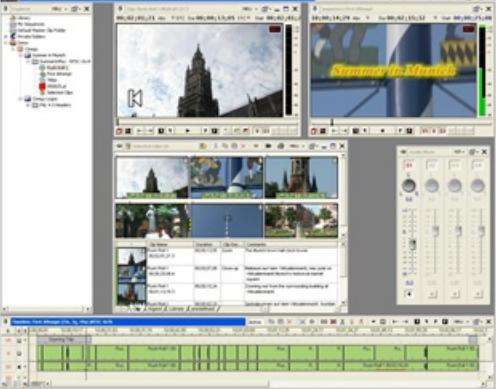
#### **Notes:**

So let's pretend that we're going to syndicate that particular press release. Your elements can be recomposed for different channels and formats. So depending upon who's receiving it and how they might like to receive that particular content, you can map particular fields to how that person wishes to receive that information.

So here's a lightweight headline syndication using RSS that uses the XML format. And each node is a different element. So in this case what this particular company is storing is headline, maps to the title in the RSS feed. And then the first paragraph maps to what is the description in the RSS feed. And a content management system can automatically generate this from components in the repository. So if it's creating this RSS feed, you can simply say, "When you generate this RSS feed," you being the content management system, "...make the headline the title, the first paragraph the description." And then when the feed is ready to go out, it has the correct tags that it can understand and use in order to syndicate it.

### **1.19 Media can be unstructured too**

## Media Can Be Unstructured Too



Tools and Methods for decomposition and subsequent re-use are arguably more complex than in the text world

But the concepts remain the same!

©AIIM 2018

**aiim**

#### **Notes:**

While I've been using audio and video frequently as examples of unstructured content, it is possible to structure media files. The methods for the decomposition of those files and the subsequent reuse are arguably a bit more complex than in the text world, but the concepts remain the same. You need to determine if it makes sense to break down your media files frame by frame or into components, and whether or not you're going to take on that labor because you might need to just use little segments of a video or little segments of audio. All the same concept, but the technology is a little bit less advanced to enable you to do that.

### **1.20 A Final Word - Elements**

## A Final Word About Elements

*click each number below for more details*

The chart consists of four segments arranged in a circle. Segment 1 is light blue, segment 2 is medium blue, segment 3 is dark blue, and segment 4 is very dark blue. Each segment contains a white number corresponding to its position: 1 at the top right, 2 at the bottom right, 3 at the bottom left, and 4 at the top left.

©AIIM 2018

**aiim**

#### **Notes:**

Wrapping up, let's summarize what we learned in this course. A taxonomy is a controlled vocabulary standard and a law for categorizing information. The categories of a taxonomy should be derived based on user needs, enterprise requirements and the nature of the content that you're dealing with.

When possible, start with an industry standard. There's also a lot of standard taxonomies out there that you can use and modify for your needs. Content technologies use taxonomies and other sorts of controlled vocabularies constantly. They need them to operate, to access content, to display content, and deliver things to the right user at the right time.

And finally, folksonomies are consumer-driven classification schemes that are particularly beneficial with social networking applications, but you should question the value of them for the enterprise. While they might seem like the easy way to classify your content, it won't make your information more interoperable and it won't necessarily make your publication process, your content management process any simpler. You need to establish a standard for that really to happen.

#### Arc 1 (Slide Layer)

## A Final Word About Elements

*click each number below for more details*

Element labels and relationships can be: SEMANTIC - rooted in meaning (e.g., Introduction, title, subject, contact, etc.).

4

3

2

©AIIM 2018

**aiim**

Arc 2 (Slide Layer)

## A Final Word About Elements

*click each number below for more details*

Element labels and relationships can be: SYNTACTIC – having to do with hierarchy and order (e.g., Paragraph one, footer, etc.).

3

3

4

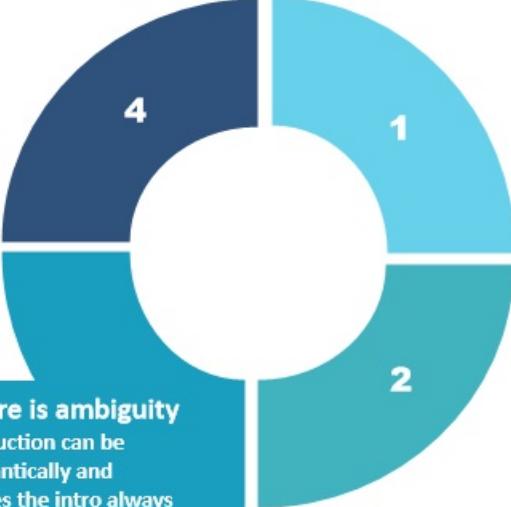
©AIIM 2018

**aiim**

### Arc 3 (Slide Layer)

## A Final Word About Elements

*click each number below for more details*



**Sometimes there is ambiguity**

- terms like Introduction can be overloaded semantically and syntactically (does the intro always come first?)
- People are used to representing structure through formatting

©AIIM 2018

**aiim**

Arc 4 (Slide Layer)

## A Final Word About Elements

*click each number below for more details*

Consistency is important for machines .... AND people

1

2

3

©AIIM 2018

**aiim**

## 1.21 The value of structured content

# The Value of Structured Content



**Structured Content:**

- Can reduce the cost of translation and customization
- Can improve information sharing and reuse
- Can support a “single source of the truth” across multiple content channels
- But: structured content can take more effort to manage
  - Major change management challenges: authors and managers are most familiar with documents, not elements
  - Systems to manage chunks tend to be more complex and costly to design and run

©AIIM 2018

 aiim

### Notes:

I've mentioned that a lot of enterprises are trying to make more of their content structured and taking a lot of their unstructured content and attempting to structure it. And that's because there is a lot of value to structured content. Content technologies work better in general with structured information. It tends to reduce the cost of things like translation and customisation if you can work with discrete elements rather than large documents. It can improve information sharing and reuse as I showed in that very particular press release example.

It can also support the idea of a single source of truth. If you have one source of things, one document, one version, it's a lot better than having potential duplication across the enterprise when you're not sure exactly which version of something you should be looking at.

Adding structure tends to be in tandem with adding a single source of that particular information. It eliminates a lot of duplication. And as I've mentioned structured content is also a little bit more difficult to manage. It takes a little bit more effort. You have change

management challenges. Authors and managers are used to dealing with documents. They're not used to dealing with chunks. And if they look at one of those chunks out of context, they might get confused and not really understand what it relates to. So you have to think about your content management user interface or whatever application you're using to manage your information, and make sure that when someone is looking at a chunk, that they're looking at it in the larger context of what relates to it.

Systems that manage these discrete chunks tend to be a lot more complex and costly, both to design and to run. They also require more education on your part to the content managers, to the people that will be managing these chunks. They need to have a larger understanding of how the systems work. So the more structured your content, the more granular you get, the more education and work that tends to be required. However, that does usually correlate to more payoff and better information access in the long-run.

## 1.22 A Quick Review

### A Quick Review

 **A Content Model is a representation of components that make up a body or corpus of content**

- The folder or “meta”-structure
- Content Types
- Content Elements within structured documents
- The relationship among those components

 **Content types are an essential building block in any content management, business process improvement, or search strategy**

 **Internally, a Content Type may exhibit some degree of structure**

- A consistent set of labeled internal elements that allow for greater reuse of content

©AIIM 2018



**Notes:**

Let's review what we covered in this particular session on content modeling. A content model is a representation of components that make up a body or a corpus of content. Essentially it's a metastructure. We also talked about content types like a press release or a financial document that describe the nature of content. And then we talked about content elements, so those are the pieces of content types such as the title or the headline of a press release.

We talked about how content types are an essential building block in any content management, business process improvement, or search strategy. Really any type of project that has to do with information management, you need to think about content types and you need to standardize around them.

Internally, a content type tends to exhibit some degree of structure. And that's a good thing. A consistent set of labeled internal elements allows for the greater reuse of content. That reuse and that single versioning of content will cut down on the very laborious process of sifting through duplication, not just for the searchers of content, but also for the software that's managing it.

## 1.23 Pitfalls and Caveats

# Pitfalls and Caveats

*click each number below for more details*

### Challenges in Content Modeling include:

©AIIM 2018

**aiim**

#### Notes:

There's a number of pitfalls and caveats around content modeling. And we'll get into some of those during the specialist track when we talk about methodologies for content modeling. But let's just list a few here.

First, you need to set a reasonable scope for the analytical exercise. You can't consider all enterprise information at once. It's simply too overwhelming. Dealing with variances and exception is also very difficult. Trying to standardize across the board is a challenge. The more you can do that, the more successful you'll be.

Organizing those large masses of unstructured documents into a reasonable and manageable set of content types which fits into that dealing with variances and exceptions is equally challenging. Trying to standardize is the name of the game. If you can do that, you'll be a lot more successful with your information management project.

Representing documents in a multifaceted environment is also a little challenging. If

you're planning to categorize in multiple ways and have faceted navigation, and yet people have the concept of a physical location of a document in mind, that can be a little bit confusing. A lot of document-based people are accustomed to the idea that a document lives somewhere. So it's a little tough to break out of that whole model, and not so much think of it as living somewhere as having characteristics and being able to find it based on those characteristics.

Also decomposing to just the right level of structure from documents is a little bit of a challenge. Some people will break down their content too much or they will over-chunk their content. It's a common mistake. And then they realize what a burden it is to manage all those chunks of content. Then they have to go through an even more painful process of what we often call up-chunking. I know that doesn't sound pleasant, but it's the idea of taking all those very granular chunks and having to put them back together into a larger less granular corpus. So be careful. You really need to strike the balance there. Don't store something at an extremely discrete level if it doesn't need to be used that way.

Arc 1 (Slide Layer)

# Pitfalls and Caveats

*click each number below for more details*

## Challenges in Content Modeling include:

Setting a reasonable scope for the analytical exercise

5

4

3

2

©AIIM 2018

**aiim**

Arc 2 (Slide Layer)

## Pitfalls and Caveats

*click each number below for more details*

### Challenges in Content Modeling include:

The diagram consists of a circle divided into five equal sectors by radial lines. The sectors are colored: one is light blue, and the others are dark blue. Each sector contains a white number: 1, 2, 3, 4, and 5. A callout box with a light blue background and white text points to the sector containing the number 1. The text in the callout box reads: "Dealing with variances and exceptions".

1  
2  
3  
4  
5

Dealing with variances and exceptions

©AIIM 2018

**aiim**

Arc 3 (Slide Layer)

## Pitfalls and Caveats

*click each number below for more details*

### Challenges in Content Modeling include:

Organizing large masses of unstructured documents into a reasonably manageable set of content types

©AIIM 2018

**aiim**

## Arc 4 (Slide Layer)

# Pitfalls and Caveats

*click each number below for more details*

## Challenges in Content Modeling include:

Representing document locale in multifaceted environment

- People are used to a document "living" somewhere

©AIIM 2018

**aiim**

Oval - 5 (Slide Layer)

# Pitfalls and Caveats

*click each number below for more details*

## Challenges in Content Modeling include:

1  
Chunking “just enough” structure from documents  
• “Over-chunking” is a common mistake

5

4

3

2

©AIIM 2018

**aiim**

## **1.24 Summary**

# Summary

*click each tab below for more details*

**Define**

**Distinguish**

**Articulate  
Value**

©AIIM 2018

 aiim

### **Notes:**

We have now reached the end of this course. In this course, you learned to:

- Define Content Model, Content Type, and Content Elements,
- Distinguish between structured and unstructured content, and
- Articulate the value and difficulty in exploiting unstructured content

## Define (Slide Layer)

**Summary**  
*click each tab below for more details*

Define

Distinguish

Articulate  
Value

**Define: Content Model,  
Content Type, and Content  
Element**

©AIIM 2018

**aiim**

## Distinguish (Slide Layer)

The screenshot shows a slide titled "Summary" with a teal header bar containing the text "click each tab below for more details". Below the header are three tabs: "Define", "Distinguish", and "Articulate Value". A callout box points from the "Distinguish" tab to a text box containing the instruction "Distinguish between structured and unstructured content". At the bottom left is the copyright notice "©AIIM 2018" and at the bottom right is the AIIM logo.

**Summary**

*click each tab below for more details*

Define

Distinguish

Articulate Value

Distinguish between structured and unstructured content

©AIIM 2018

**aiim**

## Articulate (Slide Layer)

**Summary**

*click each tab below for more details*

Define

Distinguish

Articulate Value

Articulate the value and difficulty in exploiting unstructured content

©AIIM 2018

**aiim**