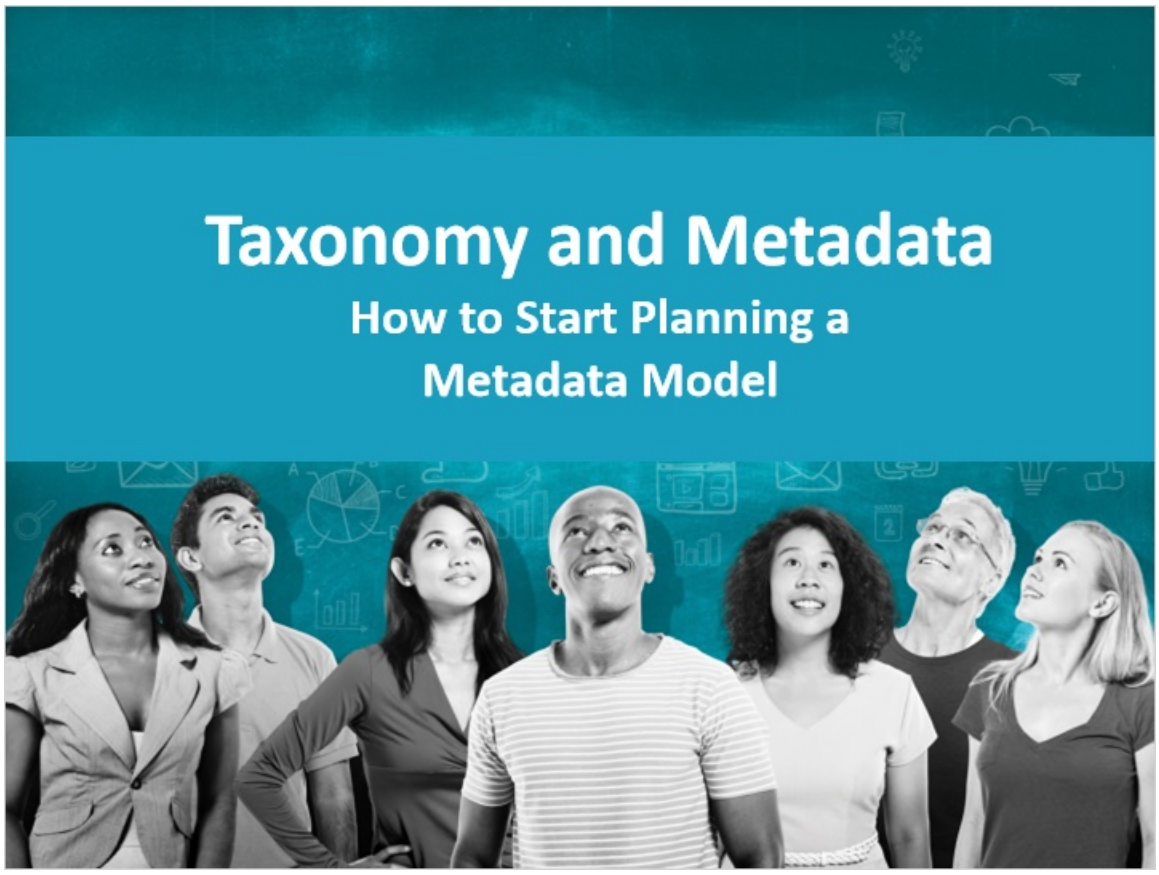
11 -TAXMET-Planning-a-metadata-model

1. How to Start Planning a Metadata Model

1.1 Welcome



Notes:

Welcome to this course addressing how to start planning a metadata model.

1.2 Learning Objectives

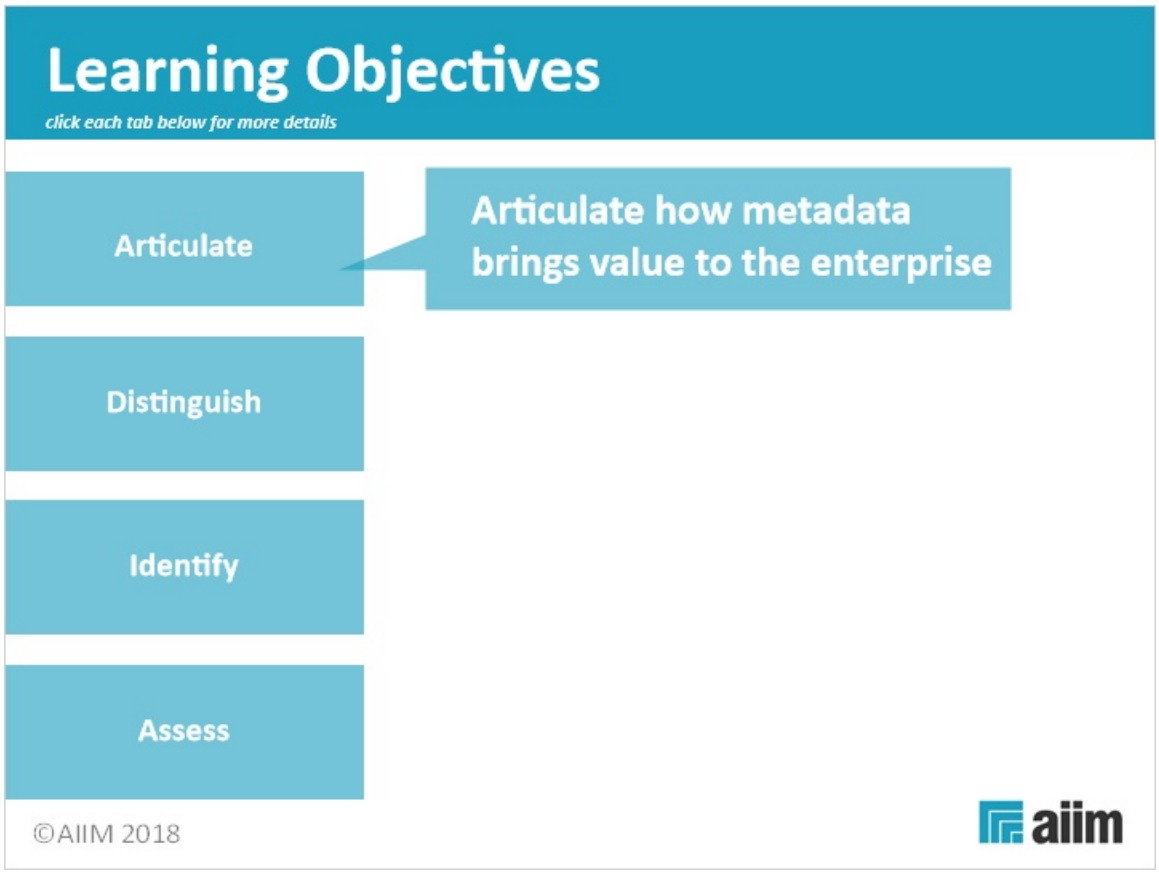
Notes:

At the end of this course, you should be able to, first and foremost, define metadata, and articulate how metadata brings value to the enterprise. We'll also talk about the difference between implicit and explicit metadata, as well as controlled and uncontrolled vocabularies.

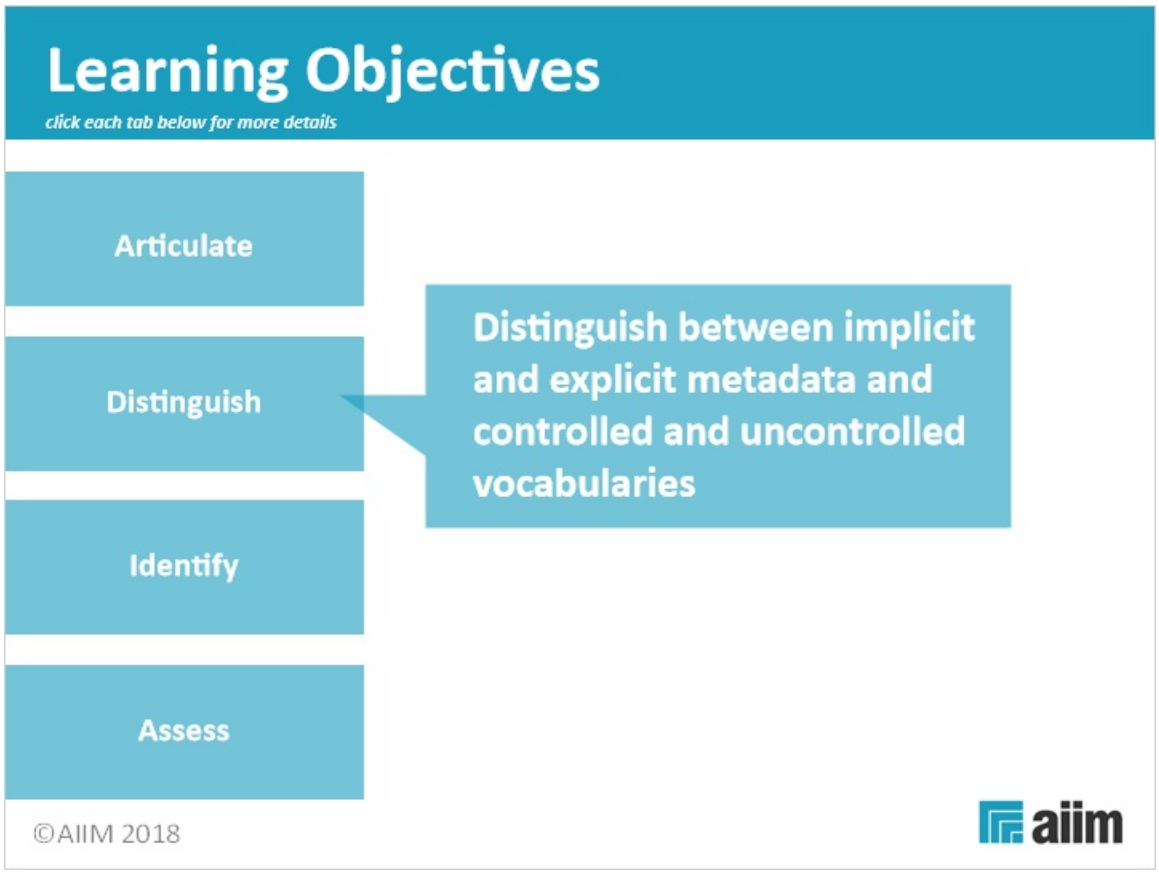
Then we'll talk about the purpose and value of the Dublin core standard elements.

Finally, you'll be able to identify the main dimensions of a metadata strategy and assess the role and value of automated classification.

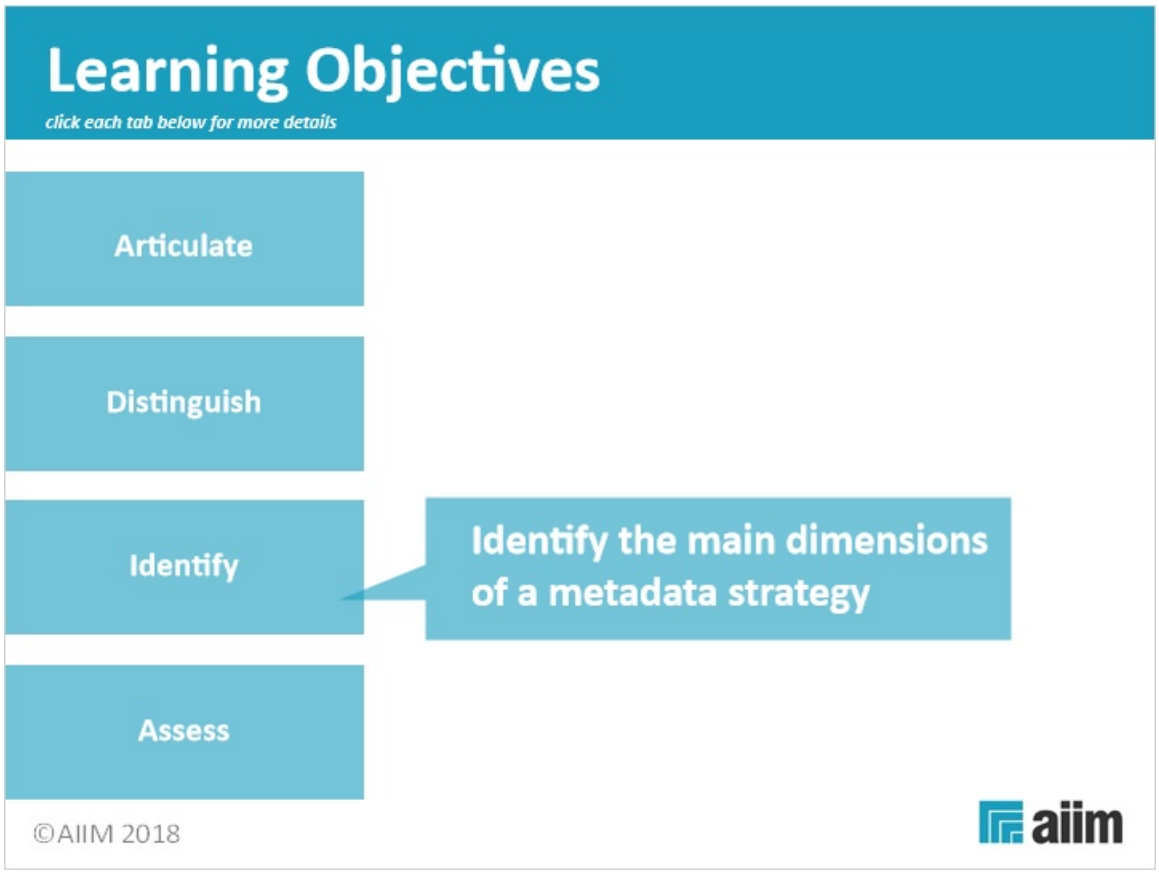
Articulate (Slide Layer)



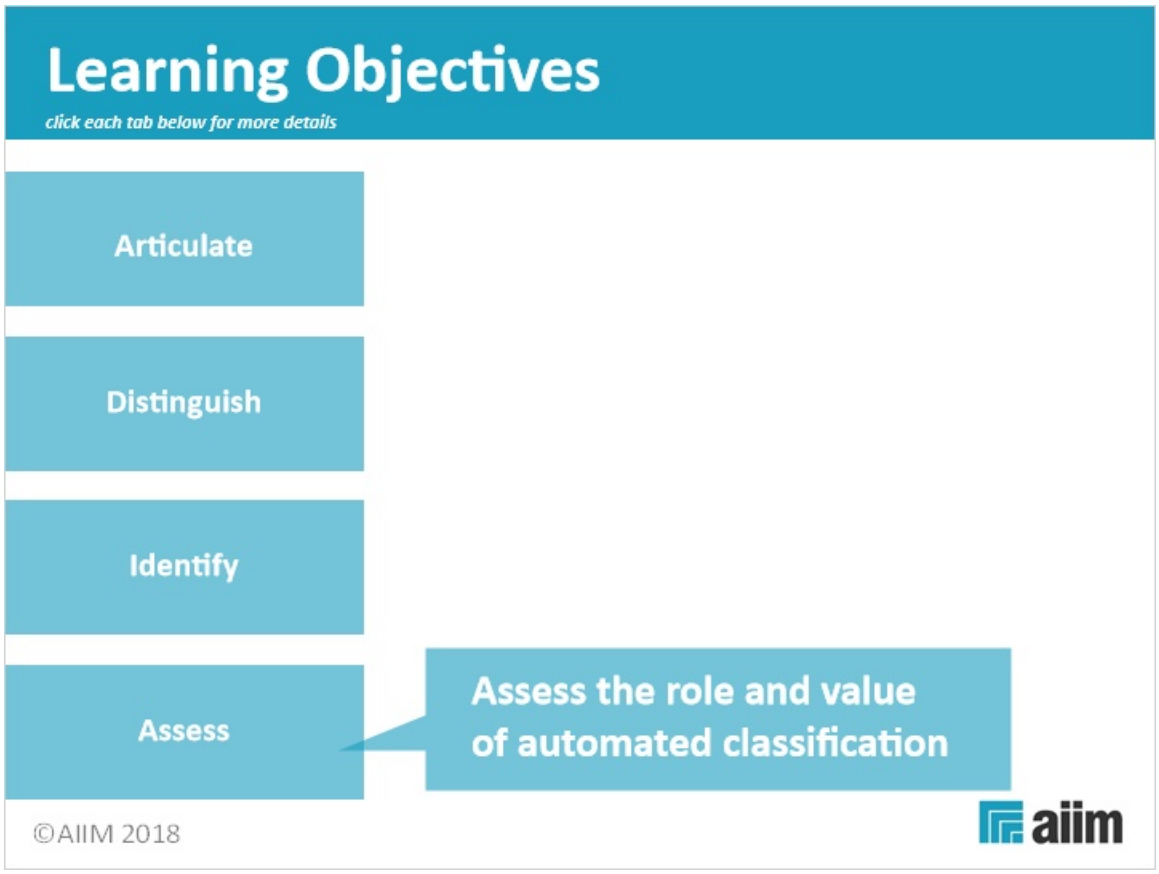
Distinguish (Slide Layer)



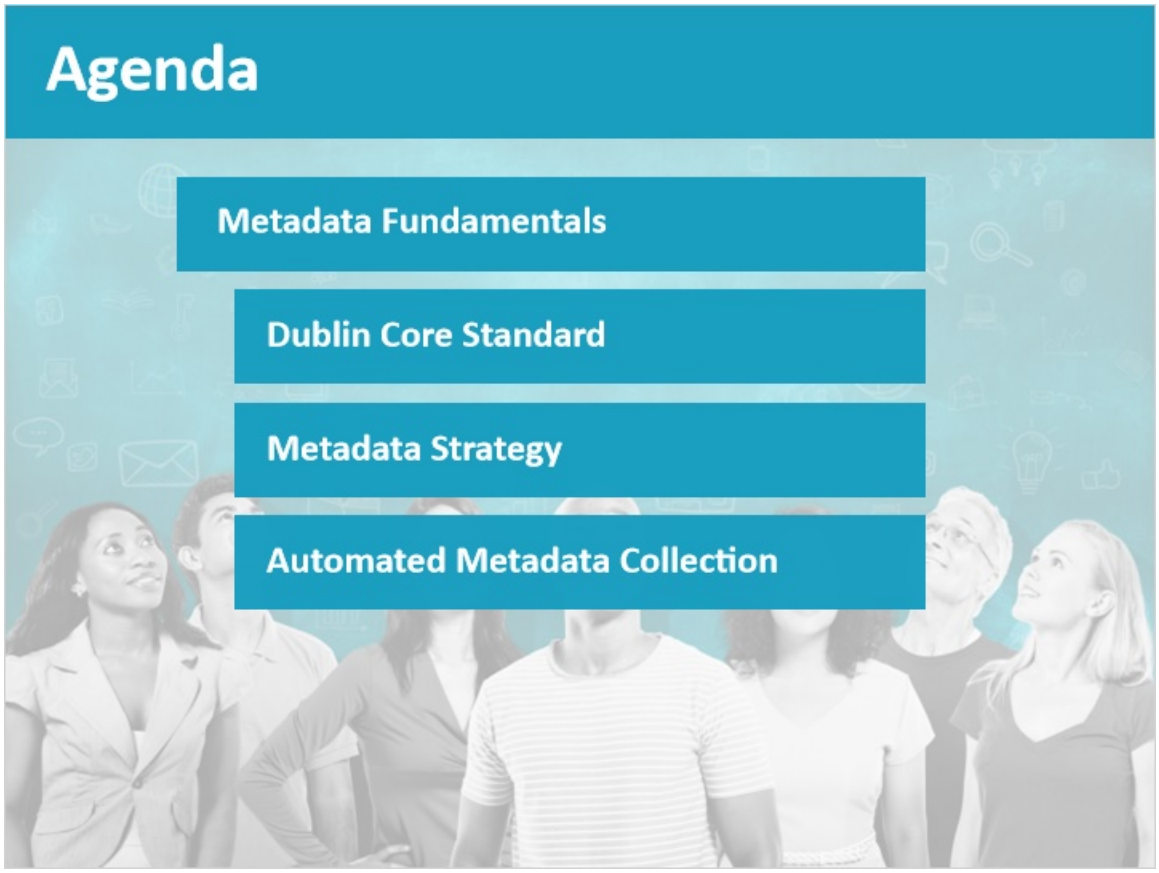
Identity (Slide Layer)



Assess (Slide Layer)



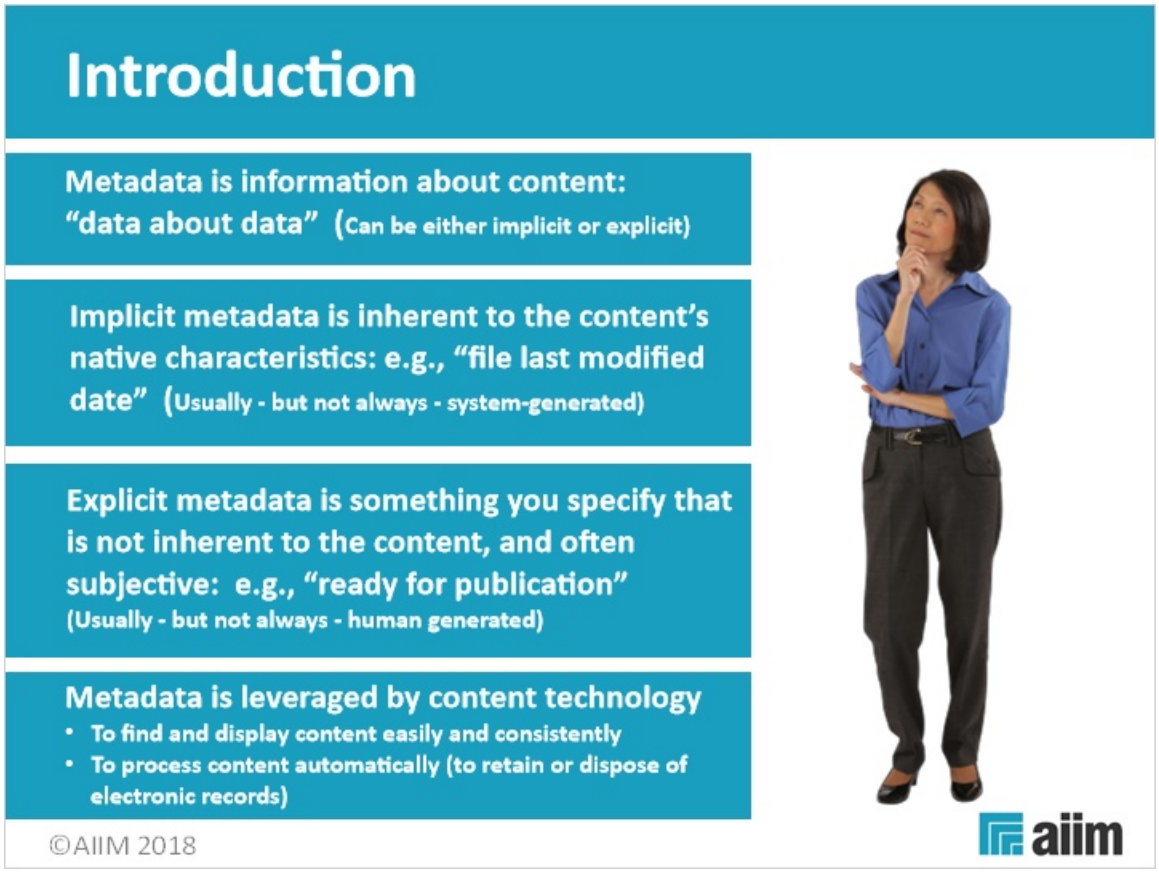
1.3 Agenda



Notes:

Let’s start by defining metadata and the role it plays in information management.

1.4 Metadata Introduction



Notes:

The most basic definition of metadata is data about the data. You could also say information about content. Either way, they both mean the same thing. Metadata is extremely important in the area of information organization and access.

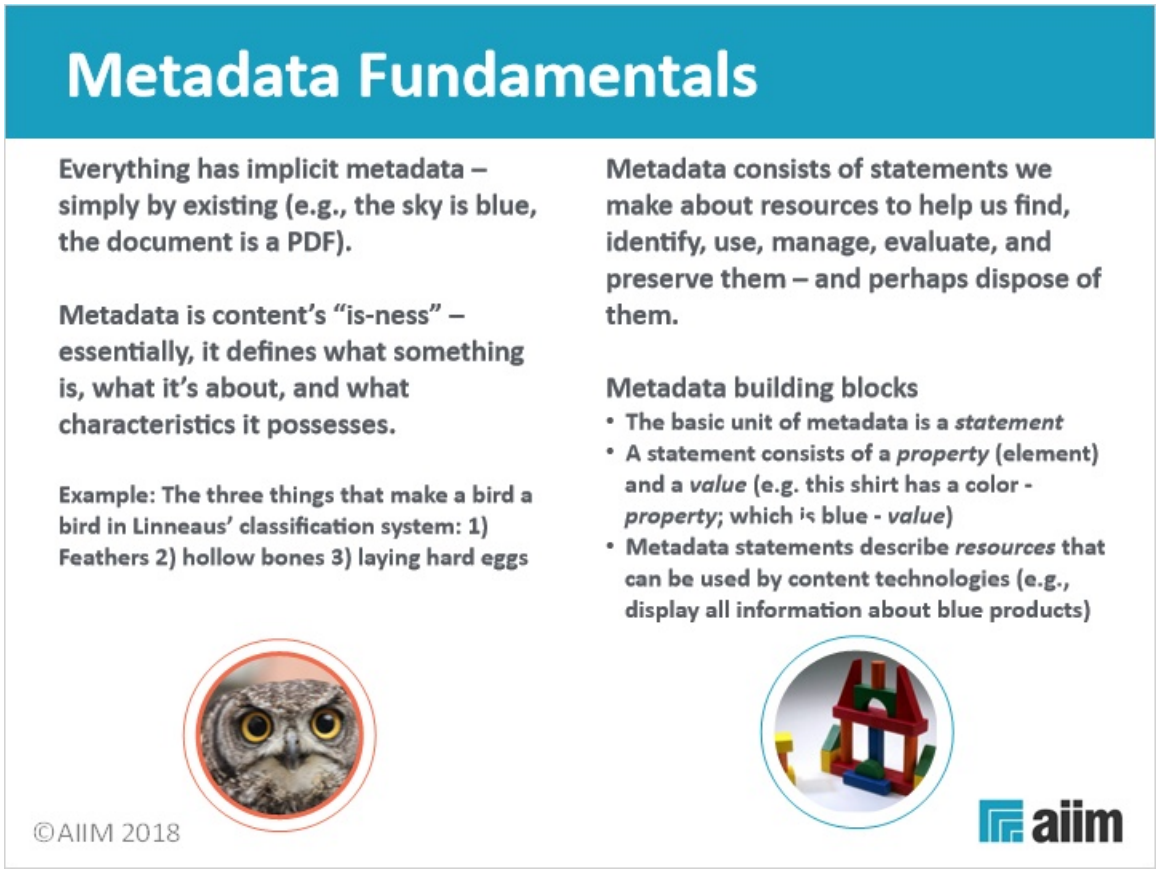
Now metadata can be one of two sorts, either implicit or explicit. Implicit metadata is something that's really inherent to the contents' native characteristics. So as an example, if you work on a file and you save it, it inherently has a file last modified date. You don't actually have to fill in that information. The system takes care of that for you.

However, there might be some details that you need to specify. This is explicit metadata. So this is something that's not inherent to the content. It's very often subjective. So maybe if you're an editor and you decide that something is ready for publication, that's something that you have to make that human judgment call as to whether that piece of metadata is appropriate at a given point in the content life cycle.

As I alluded to, implicit metadata is usually but not always system generated, while explicit metadata is usually but not always human generated. Now why is this important? Well, metadata is extremely important to content technology, whether you're dealing with a Web content management system, a records management system, or an enterprise search system. Anything that is looking for information, retrieving information, and delivering information needs metadata to really work, to find and display content easily and consistently, and also to process content automatically.

So if you wish to retain or dispose of electronic records based on certain characteristics of those records, that metadata is extremely important because that metadata is what the content technology uses in order to take action and do something with that content.

1.5 Metadata Fundamentals



Notes:

To understand metadata, you need to go way back and you need to think about the fundamentals of existence. The Ancient Greeks spent a lot of time occupying themselves with metadata and what exactly were the things that made things what they

are.

So everything that exists in the world around us in fact has very implicit metadata simply by existing. So we look at the sky and you say the sky is blue. On a more mundane level, you can look at a document and say, “Well, this is a PDF.”

Metadata is really content’s is-ness, as we say, so essentially what defines what something is, what it’s about, and what kinds of characteristics it possesses. So there’s nothing in the world around you that you can look at that doesn’t have metadata.

Linnaeus, who is considered one of the fathers of taxonomy-- and we’ll be talking a lot about taxonomy during this course-- in his classification system, he classified all living creatures. And one of the most interesting creatures to think about is birds. What exactly makes a bird a bird? A lot of people will immediately answer, “Well, they fly.” But of course there’s a number of birds such as ostriches or emus, for example, that don’t fly.

What do they have in common? All birds have feathers. They all have hollow bones, which is what of course enables them to be so light and able to fly, and they also lay hard eggs. Anything that doesn’t have these three characteristics is not, in fact, a bird.

So what’s really important to understand with this example is that metadata is not just about what something is, it’s also about what something is not. So when you’re trying to define something, anything, a piece of content, if you’re trying to define it as X or Y, you also need to define, what is it that makes it X or Y, just like Linnaeus did with birds. So when you think about your own enterprise and your own metadata, you can’t be arbitrary about it. You can’t leave it open to interpretation. You need to be as scientific as possible before you assign that metadata as to what makes something what it is.

Metadata that’s used by content technologies is made up of statements. And these are statements that help us find, identify, use, manage, evaluate, and preserve the content within the enterprise, perhaps even dispose of it. There are a few building blocks involved here.

The basic unit of metadata is called a statement. And that statement consists of two things - a property, which can also be called an element-- we’re going to use element a lot over the course of this particular module-- and then a value.

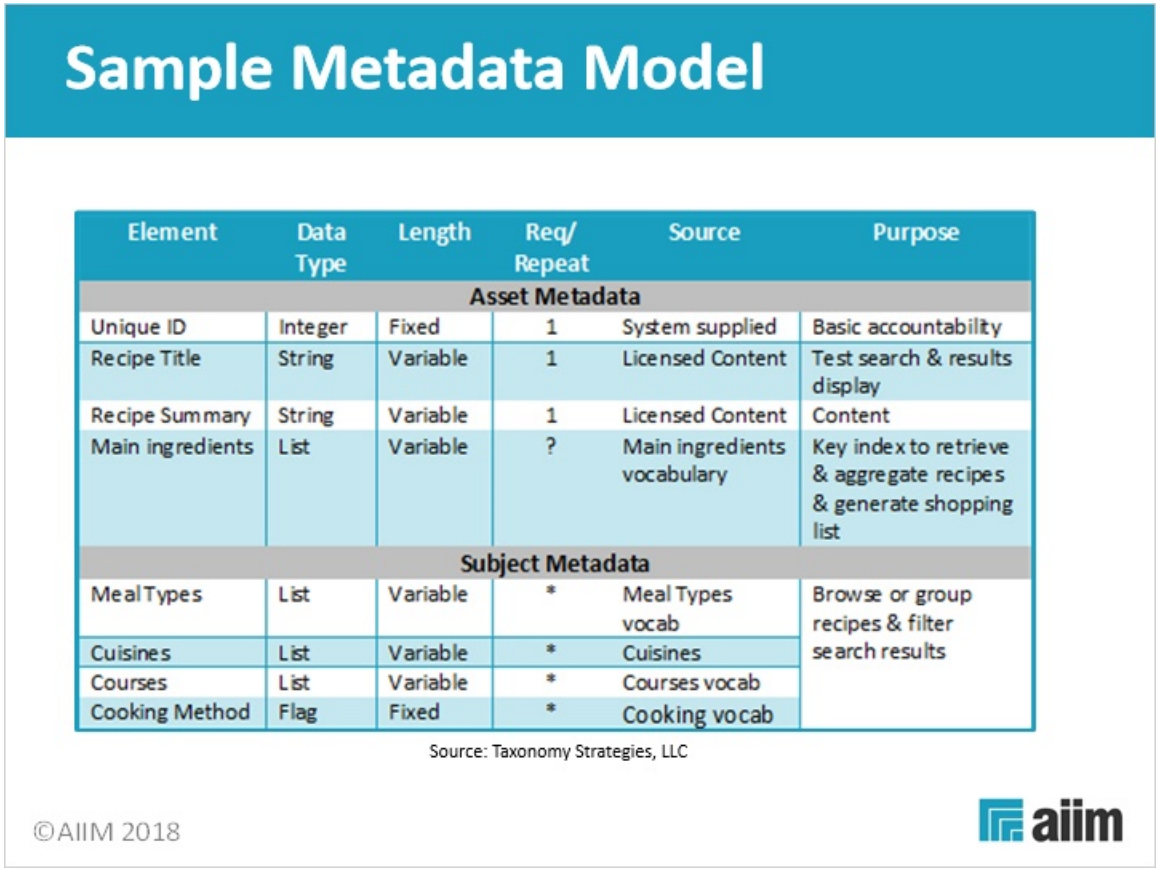
For example, you could take a shirt. A shirt has a property that is called a color. You can’t actually have a shirt without a color. Every shirt inherently has some sort of color. Then you decide based on the color of that shirt what the value of the property is. In this case, it’s blue. The property color has a whole list of values that you could choose from,

which would be a list of colors. And in this particular case, the shirt is blue.

Now the metadata statements describe resources that can be used by content technologies. Content technologies are constantly looking to metadata as a resource to really figure out what to do with it, or to respond to some sort of query. If a search tool, for example, is looking for blue products, the content technology, in this case a search engine, looks for the value that is blue, and then returns all the information about blue products.

This is really fundamental to how content technologies use metadata. They're looking to the metadata to figure out exactly what to do with the content and what sort of action to take with it.

1.6 Metadata Model



Notes:

Let's take those metadata building blocks and put them into a sample metadata model. And this is a model that was put together by two of the world's leading taxonomy experts, Joseph Busch and Ron Daniel of Taxonomy Strategies. And in this case, this particular example is a recipe metadata model. So the idea here is if you had a repository full of recipe information, what is the type of metadata that you would have about recipes?

In the first column, we have element. That was one of the metadata building blocks that I talked about on the last slide. And we've got a few inherent characteristics of recipes, both implicit and explicit, so to speak. For instance, the recipe would have a title. It would have a summary, and it would have a list of main ingredients.

Now, for each of those elements within a metadata model, you tend to specify what type of data it is. The recipe title, for example, would be a string. A string is a set of characters or numbers, for example, that might make up a recipe title. And the same thing with a recipe summary, would usually be a paragraph or a set of instructions that's usually in paragraph form. Whereas main ingredients for a particular recipe would be a list.

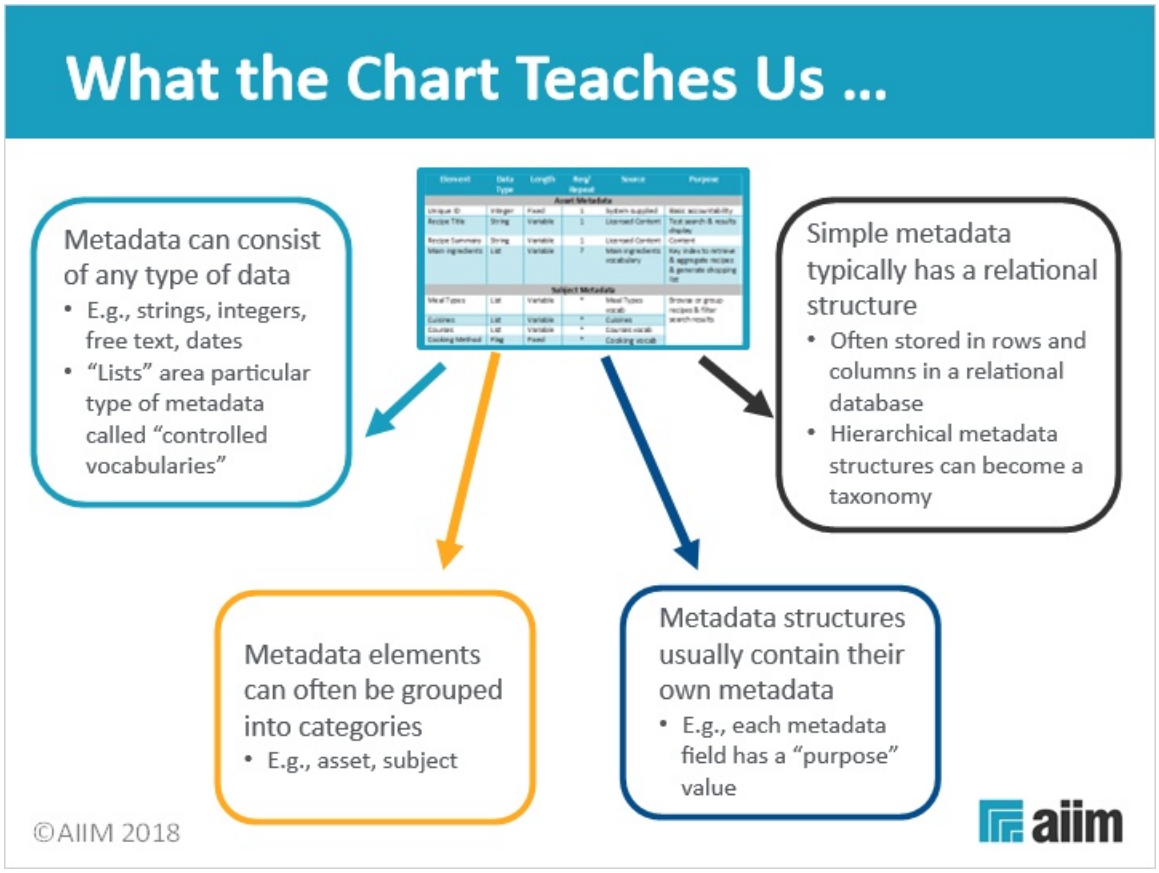
Specifying the data type is very important because it allows you to determine how you should be storing the information within a repository. You also need to specify whether you want the length of that particular metadata to be variable or fixed. So if you're giving a recipe a unique ID, normally what would be a number, in this case, an integer, that would be of fixed length. On the other hand, you might not want to restrict people with the length of the title or the recipe summary, or even the list of the main ingredients.

Then you also need to specify, well, where is this particular metadata coming from? What are the values that I'm going to populate these particular elements with? What's the source of these values?

For main ingredients, you can see here, they've specified a main ingredients vocabulary. This could consist of hundreds of different food items that content managers might specifically select to make up the main ingredients variable length list of this particular element.

And then finally we have the purpose of each element, why exactly are we gathering this metadata, why is it useful, why is it important. I won't go through every single item on this metadata model, but it gives you an idea of how you need to look at the metadata for your own enterprise information. We'll be talking about metadata models more as we continue with the course.

1.7 What the chart teaches us



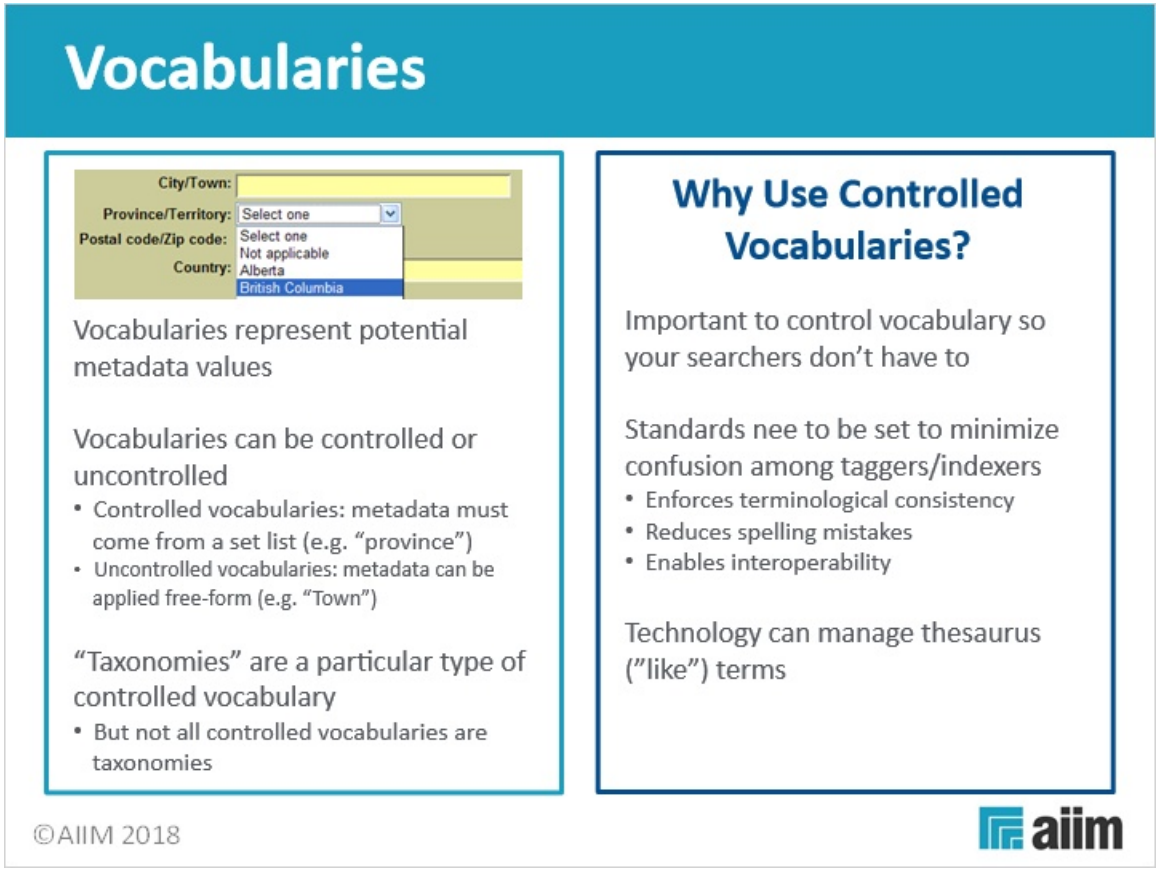
Notes:

What does that chart teach us overall? Well, metadata can consist of any type of data. In the previous slide I mentioned strings, integers, free text. It can be dates. You decide what the format is for each particular type of metadata.

Lists are a particular type of metadata called controlled vocabularies. Metadata elements can often be grouped into categories or lists of terms. So you saw on that particular chart there was asset metadata and subject metadata. And you can decide what makes the most sense for you in particular, in your enterprise in particular.

Metadata structures inherently and often contain their own metadata. Each metadata field has a purpose and value that in and of itself is metadata. Simple metadata has an inherently relational structure. It's therefore often stored in rows and columns in a relational database. However, metadata can also have hierarchical structures. And that's what eventually becomes a taxonomy, which we address in other AIIM Training Program courses.

1.8 Vocabularies



Notes:

We all know the word vocabularies from our everyday conversation, vocabularies being words that make up a language. In this particular case and in the realm of information management, vocabularies represent potential metadata values. In the recipe example, we had a main ingredient vocabulary, so what were the words that populate that particular type of metadata. What are the values that might make up that particular element?

Vocabularies can be controlled or uncontrolled. Controlled vocabularies come from a set list. For example, if you had a list of the provinces of Canada, that is a finite list of places. Uncontrolled vocabularies can be applied more freeform. To further the example, towns, you might have thousands of towns that you don't necessarily want to put into a fixed list, but that you want people to be able to fill in freeform into a field. You can see in the example here, this allows a user to actually type in the name of their town, whereas the province or the territory is selectable from a list.

Now taxonomies are a particular type of controlled vocabulary. But it's important to know that not all controlled vocabularies are taxonomies.

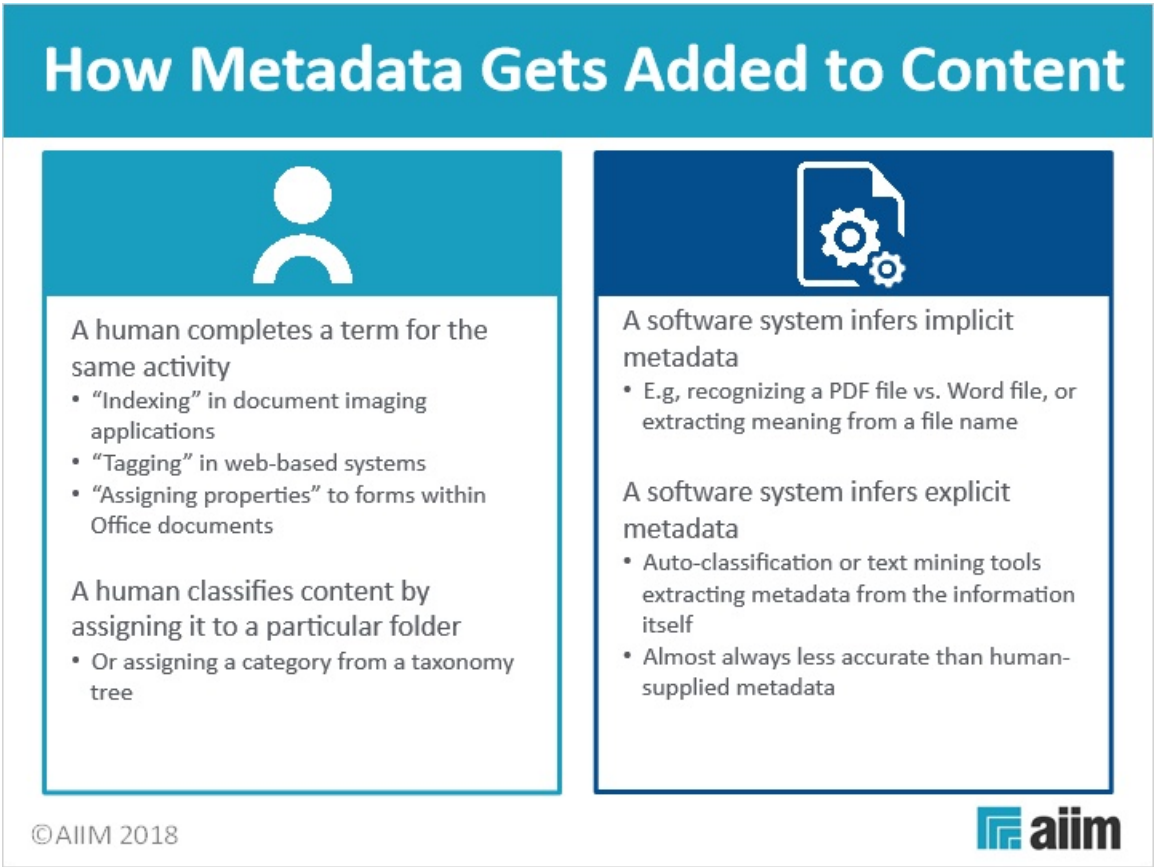
Why should you use controlled vocabularies? Well, it's important to control vocabulary within your enterprise so that those who are searching for information don't have to. They should be able to use a search tool or any sort of information access tool and use their own vocabulary in order to find what they need. Your system needs to know what vocabulary they're using and consistently apply standards across your enterprise. This helps to minimize confusion among taggers and indexers, so they can consistently term things or tag things across the enterprise in the same way.

Controlled vocabulary also helps reduce spelling mistakes. If you allow freeform entry of tags or of metadata, you might have people spelling things incorrectly. And that would make for difficulty in finding that information later.

Controlled vocabularies also enable interoperability among and between different systems. If those systems are using the same vocabulary, they can exchange information more easily. Just like in a regular language, if you speak the same language, if you use the same words, you understand each other better. It's the same thing with your information access tools. They need to use the same words in order to understand each other and what types of information they're trying to communicate to one another.

Technology can also manage thesaurus-like terms. This is where you might have numerous words that mean the same thing. An example of that might be if you have bottle of sparkling water and somebody calls it seltzer or somebody else calls it club soda. Those are all different sorts of sparkling water or water with bubbles. People might look for that same item with those different terms. And the technology can manage those different terms, and, in fact, point people towards the same content object regardless of which term they're using to look for it.

1.9 How metadata gets added to content



Notes:

We all know the word vocabularies from our everyday conversation, vocabularies being words that make up a language. In this particular case and in the realm of information management, vocabularies represent potential metadata values. In the recipe example, we had a main ingredient vocabulary, so what were the words that populate that particular type of metadata. What are the values that might make up that particular element?

Vocabularies can be controlled or uncontrolled. Controlled vocabularies come from a set list. For example, if you had a list of the provinces of Canada, that is a finite list of places. Uncontrolled vocabularies can be applied more freeform. To further the example, towns, you might have thousands of towns that you don't necessarily want to put into a fixed list, but that you want people to be able to fill in freeform into a field. You can see in the example here, this allows a user to actually type in the name of their town, whereas the province or the territory is selectable from a list.

Now taxonomies are a particular type of controlled vocabulary. But it's important to know

that not all controlled vocabularies are taxonomies.

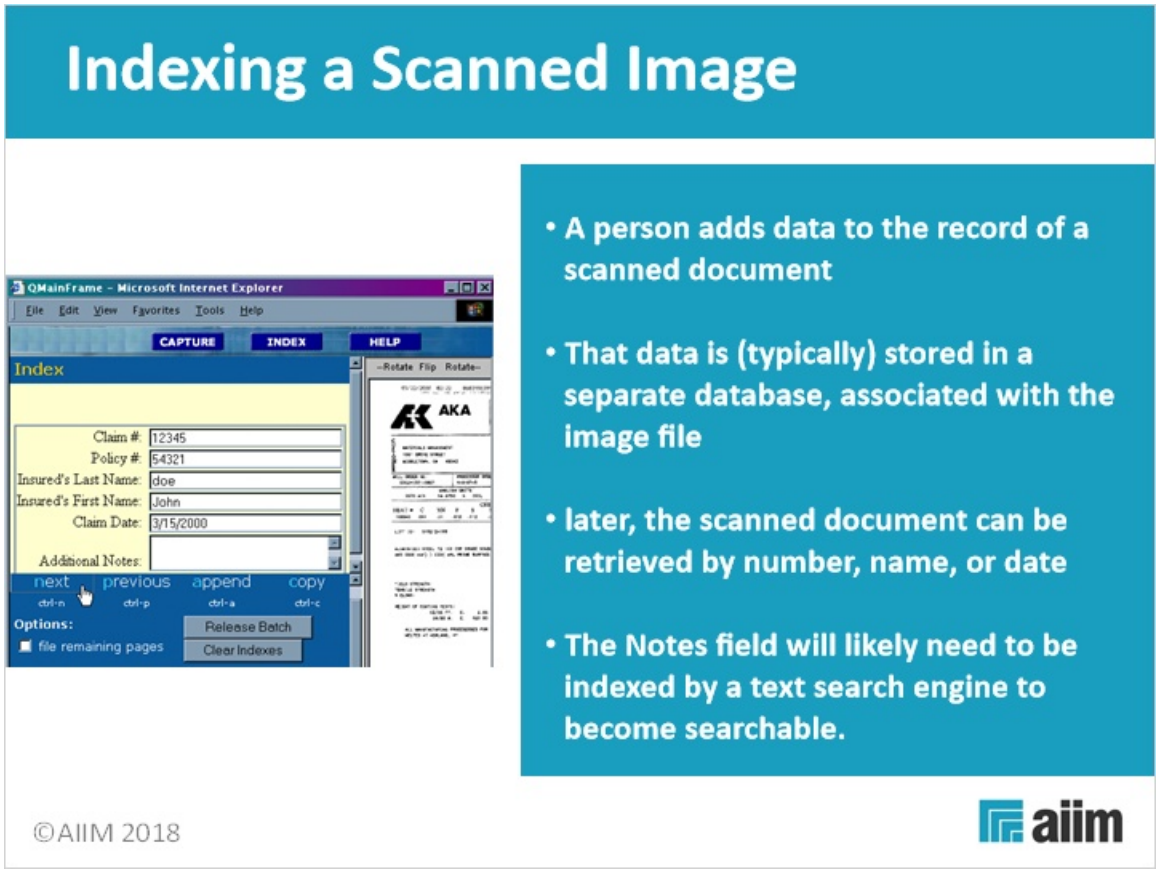
Why should you use controlled vocabularies? Well, it's important to control vocabulary within your enterprise so that those who are searching for information don't have to. They should be able to use a search tool or any sort of information access tool and use their own vocabulary in order to find what they need. Your system needs to know what vocabulary they're using and consistently apply standards across your enterprise. This helps to minimize confusion among taggers and indexers, so they can consistently term things or tag things across the enterprise in the same way.

Controlled vocabulary also helps reduce spelling mistakes. If you allow freeform entry of tags or of metadata, you might have people spelling things incorrectly. And that would make for difficulty in finding that information later.

Controlled vocabularies also enable interoperability among and between different systems. If those systems are using the same vocabulary, they can exchange information more easily. Just like in a regular language, if you speak the same language, if you use the same words, you understand each other better. It's the same thing with your information access tools. They need to use the same words in order to understand each other and what types of information they're trying to communicate to one another.

Technology can also manage thesaurus-like terms. This is where you might have numerous words that mean the same thing. An example of that might be if you have bottle of sparkling water and somebody calls it seltzer or somebody else calls it club soda. Those are all different sorts of sparkling water or water with bubbles. People might look for that same item with those different terms. And the technology can manage those different terms, and, in fact, point people towards the same content object regardless of which term they're using to look for it.

1.10 Indexing a scanned image



Notes:

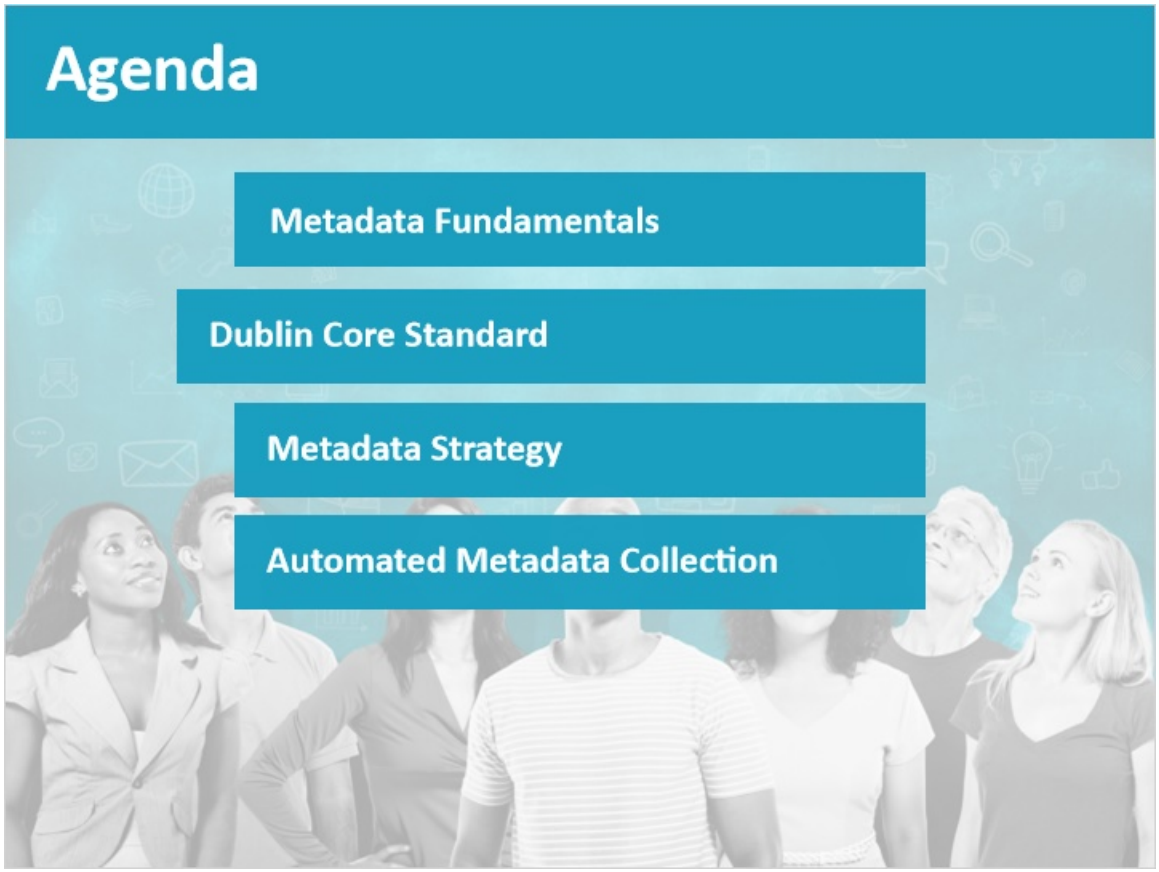
Here's an example of how someone might go about indexing a scanned image. Let's say you have a picture here in this example on the right, where you see a little thumbnail on the far right of a document that's been scanned by some sort of imaging technology. And an indexer might have to fill in information about this particular document. In this case, we have a claim number. We have a policy number. We have the last name of the person who's being insured. So obviously this is some kind of insurance document. There's also a claim date and a random notes field.

This data, when it gets entered into the system, is usually stored in a separate database, and it's associated with the image file so that if someone wants to pull it up, they can type in any of these items, whether they're looking for a particular claim number or a claim date or based on someone's name. And then the system would pull up the document based on this particular metadata that's within the index.

The notes field is a little bit more complex, simply because the information isn't as precise or discrete as the other fields are. So the notes field might actually have to be

indexed by a text-based search engine to become more easily searchable. We're going to talk a lot about search technology a little bit later, and how it might actually deal with a free form field such as this notes area.

1.11 Agenda



Notes:

Let's now look at one of the standards that can help you develop a metadata model.

1.12 Dublin Core - Explained

Notes:

Now that we've been through metadata fundamentals and you have a good idea of why metadata is important and also how you might think about your own information in terms of metadata, we're going to explore the core metadata standard that exists in the information management industry, which is Dublin Core.

You may have heard Dublin Core before and wondered, "Well, what does it have to do with Dublin?" Well, it originated during a meeting that was going on in Dublin, Ohio, USA. And the Dublin Core is really the most important core global metadata standard. And what do we mean by core metadata standard?

Well, it's commonly believed that all enterprise information will have the metadata characteristics that are present in the Dublin Core. And when we get into what those elements are, I think you'll agree that your enterprise information generally carries many of those particular elements.

The Dublin Core came to be in 1995. There was really a dramatic increase going on of the number of document-like resources on the Internet. It was commonly believed that a standard needed to be put in place so that those document-like resources could more easily be exchanged and reused.

Slow improvement in indexing services were making resources very hard to discover. The technology that was looking for information was having a particularly hard time because of the lack of standards and the lack of consistency. There was a common belief that descriptive metadata would improve the discovery and that the technologies that were looking for information would thus perform better if there was a standard. And so there was a perceived need and the Dublin Core was essentially the response and the result of that perceived need.

Dublin Core is a response to a series of different issues. Most content, it's believed, has a set of common properties that are consistently important to both people and machines. As an example, you might have a creator of the document, the date that it was created or last modified, the type of document that it is. The issue here is that there needs to be some kind of agreement as to what to call these different elements consistently. For example, if you have a creator, an author, or a content owner, do those all mean the same thing? To some organizations, they do; to others those are three different people.

The goal was therefore to set this standard of agreed upon terms in order to disambiguate the meaning of those terms. This in turn would allow applications and developers to develop common solutions that would allow systems to exchange information, improving the interoperability of systems and content.

And that's simply because, even if a creator and an author are the same thing, you can't expect two systems to know that inherently. This works very much, again, just like regular spoken languages. If two different words are used, people will not be able to understand each other, despite the fact that they're talking about the same object. Here's the list of the Dublin Core element set. The Dublin Core is considered to be the list of the fundamental or the foundational elements for metadata. And many enterprises have adopted the Dublin Core simply because it does facilitate information exchange. If two enterprises are exchanging information and they both use the Dublin Core element set, they can then very consistently and easily add, for example, a list of creators or a list of contributors to each other's metadata, simply because it's called the same thing. Thus, the two systems can look at the information, add, exchange, et cetera, because of that consistent naming of elements.

Back when I showed you the example of the recipe metadata model, we noted that the metadata was separated into two categories. Let's explore the idea of categorizing metadata a little bit more in-depth.

Sometimes it is a good idea to categorize your metadata because it makes it a little bit easier to digest. Here are four ways to do that. First, you can think about subject

metadata. That's the what, where, and why. These are things like the subject, the title, the description of the document.

Then you might have asset metadata. This is really the who, who's involved with this particular piece of information or piece of content. There might be a creator, a publisher, a contributor, a type, a format, and a unique identifier.

Then we might have use metadata. This is the when and how you can use it. Is there a date associated with when it might be released? Is it in a particular language that limits the audience? What kind of rights do you have to republish it?

Finally, we can think about relational metadata. How does this particular piece of content relate to other content within my body of information? This is particularly important as you get into more complex content applications that might need to display related content or target information based on user preferences or searcher preferences.

These different categories of metadata deliver different value, not only to the people who are searching for information, but also the people who have to manage it on a day-to-day basis. When we think about the value of subject metadata and use metadata, those facilitate better navigation and discovery of the information. People tend to navigate through information and search for information based on subject, whereas if they are someone managing the content, they often navigate or search for information based on use or where it is in the life cycle.

Whereas when we look at asset and relational metadata, that really helps facilitate content processing. Thinking about content management systems, records management systems, they need that sort of metadata in order to decide how to take action on the content. The metadata might determine who needs that content next, perhaps what system that content needs to be passed to, or when that content might be released, distributed, archived, or perhaps disposed of.

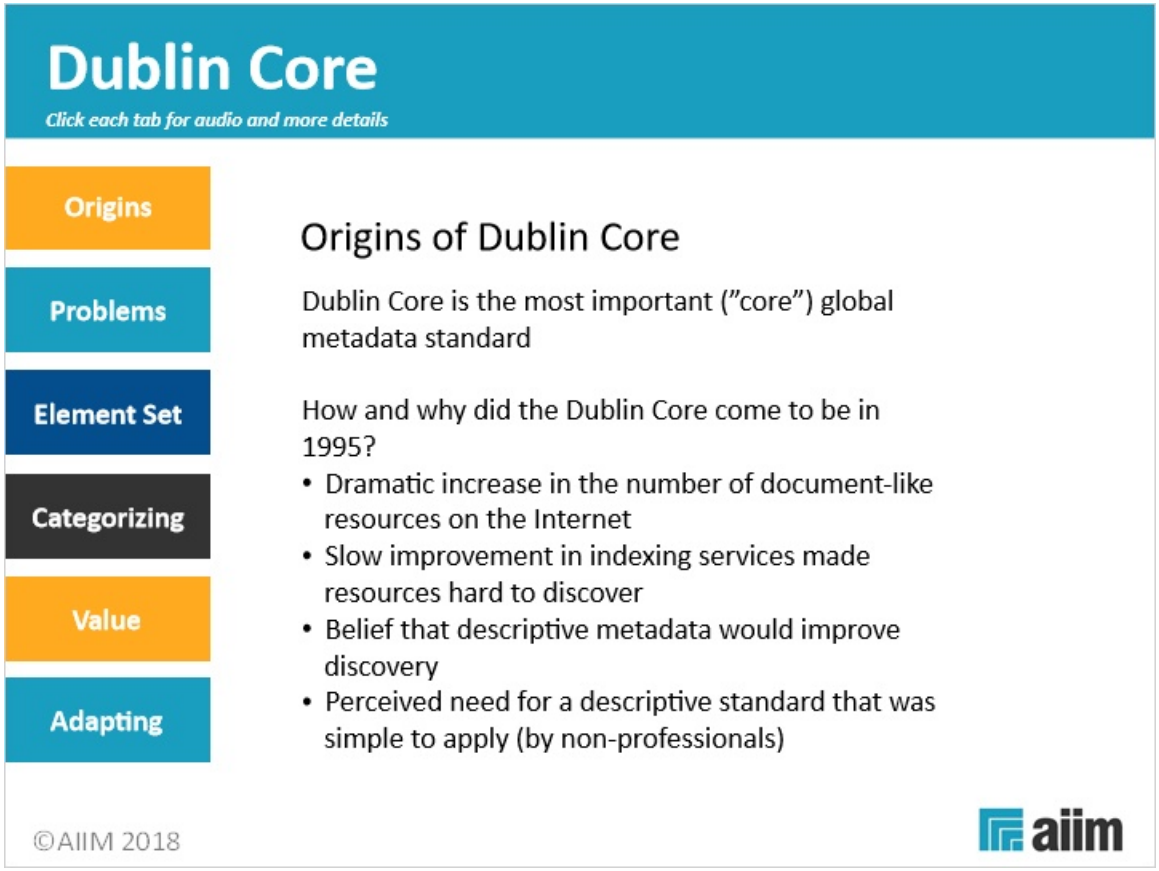
Some would argue that adopting a standard makes things inflexible for your enterprise. In reality, the Dublin Core is a very flexible standard.

First, it's a very flat element structure. All the elements are optional. So that list I showed you, you don't have to use all of them. You can pick and choose which ones are relevant for you. All the elements are repeatable so you don't have to necessarily just use them once. They can be displayed in any order, so you're not required to use them in the order that I listed them. They're also extensible, so if you wish to add elements or qualifiers based on the specific nature of your information, you also have the flexibility to do that.

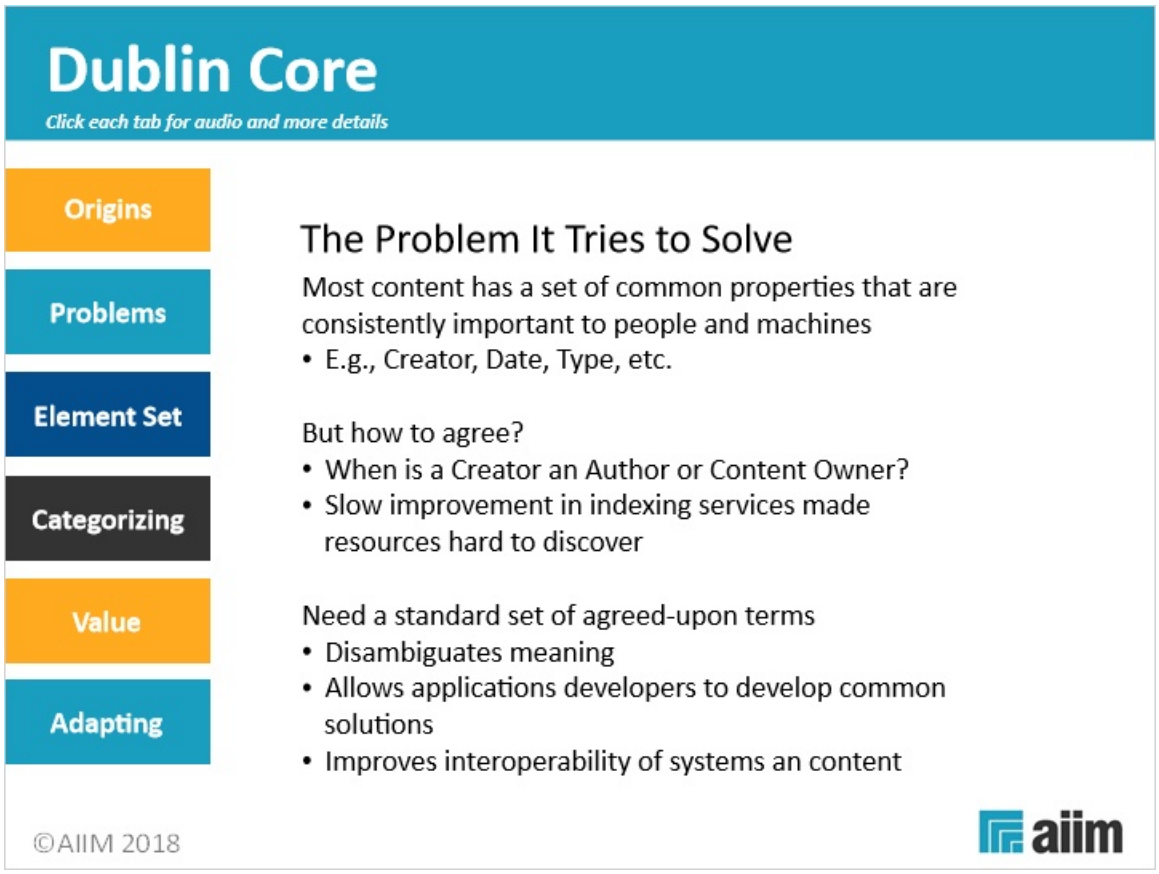
They're syntax-independent. This means that you don't have to worry about what the values are necessarily that are populating those different elements. You can phrase them and arrange them in any way that you wish. It's very international. It can be applied to any language. And it's subject independent. So whether you are in a particular type of business, government, education, in any domain, you can use the Dublin Core metadata

elements regardless of the nature of what you do.

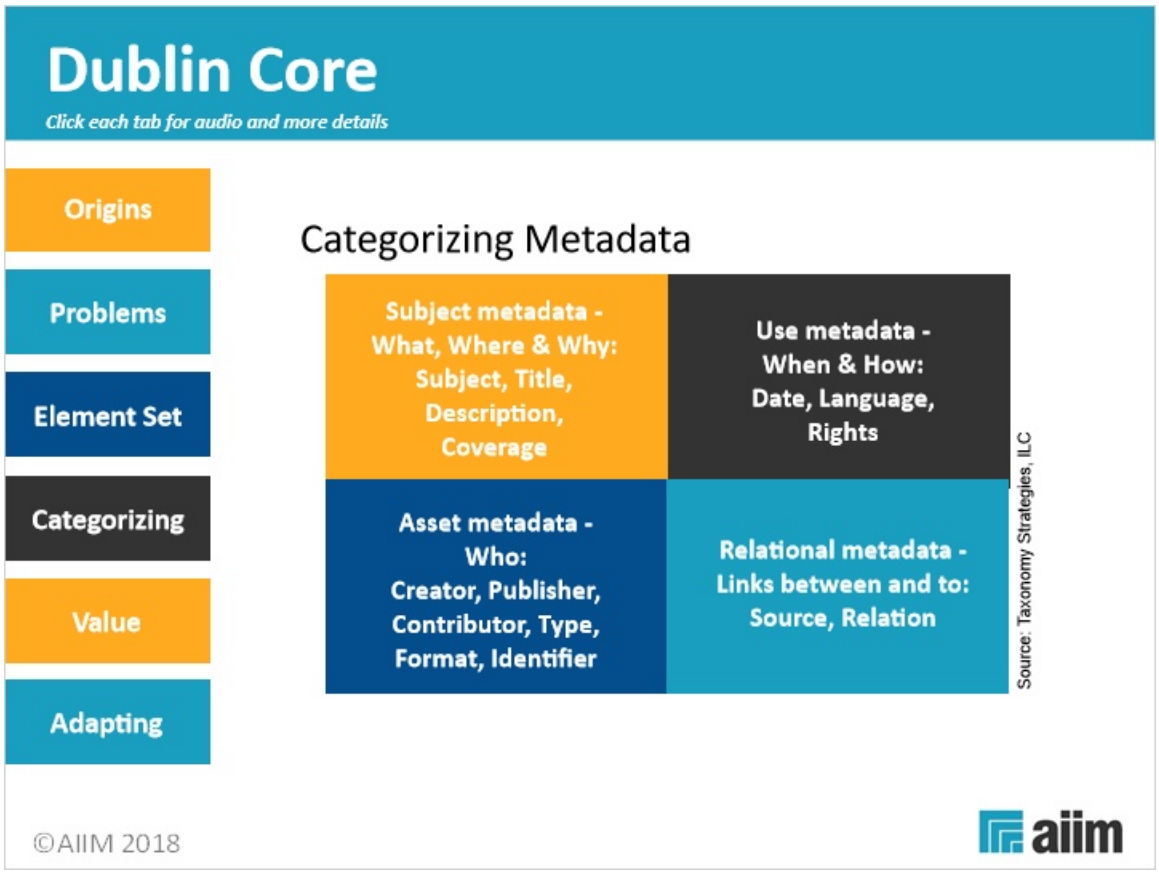
Origins (Slide Layer)



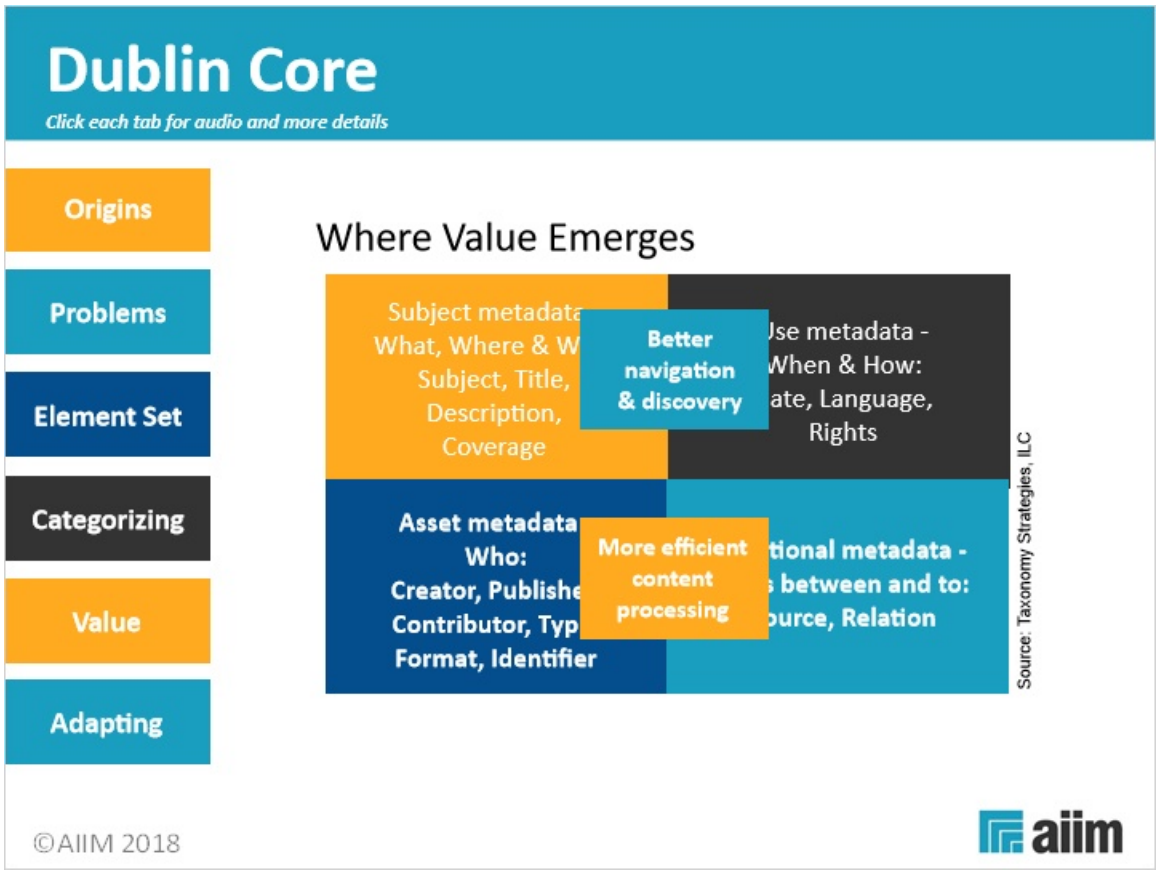
Problems (Slide Layer)



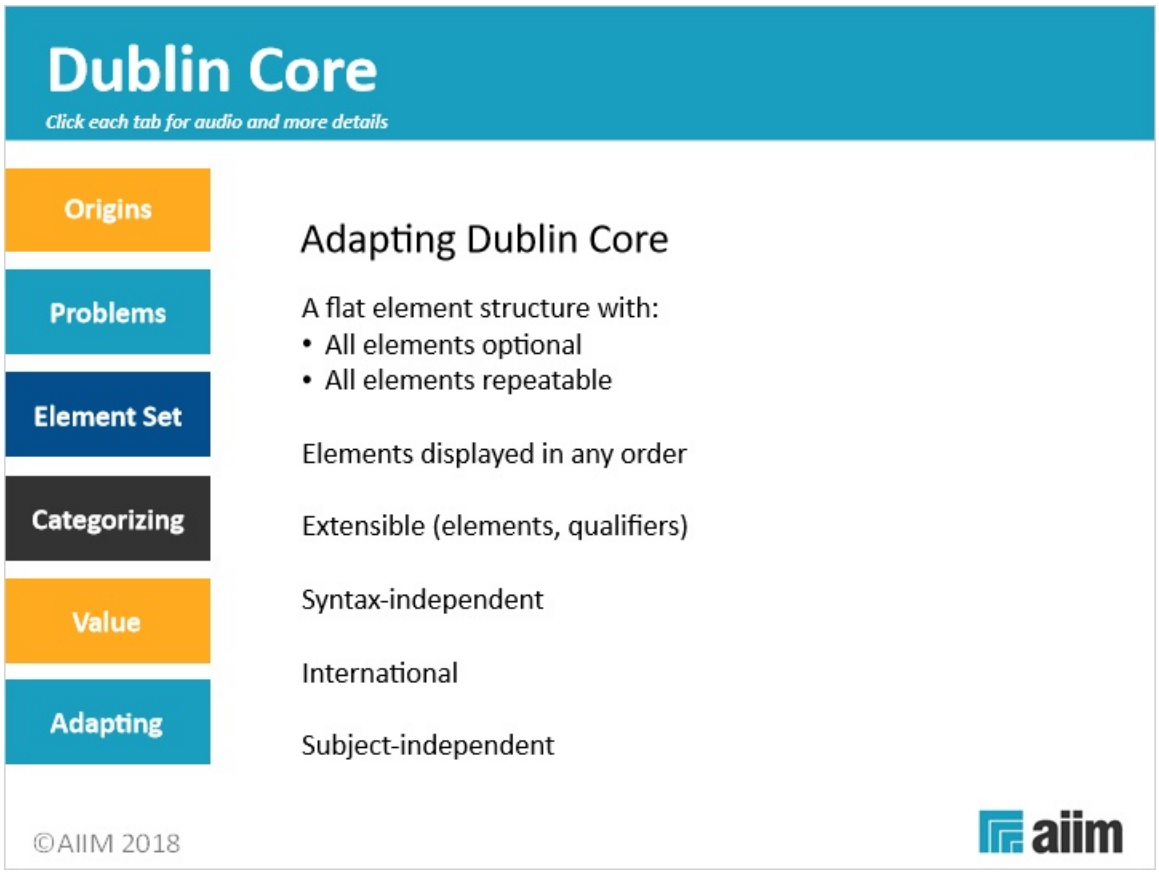
Categorizing (Slide Layer)



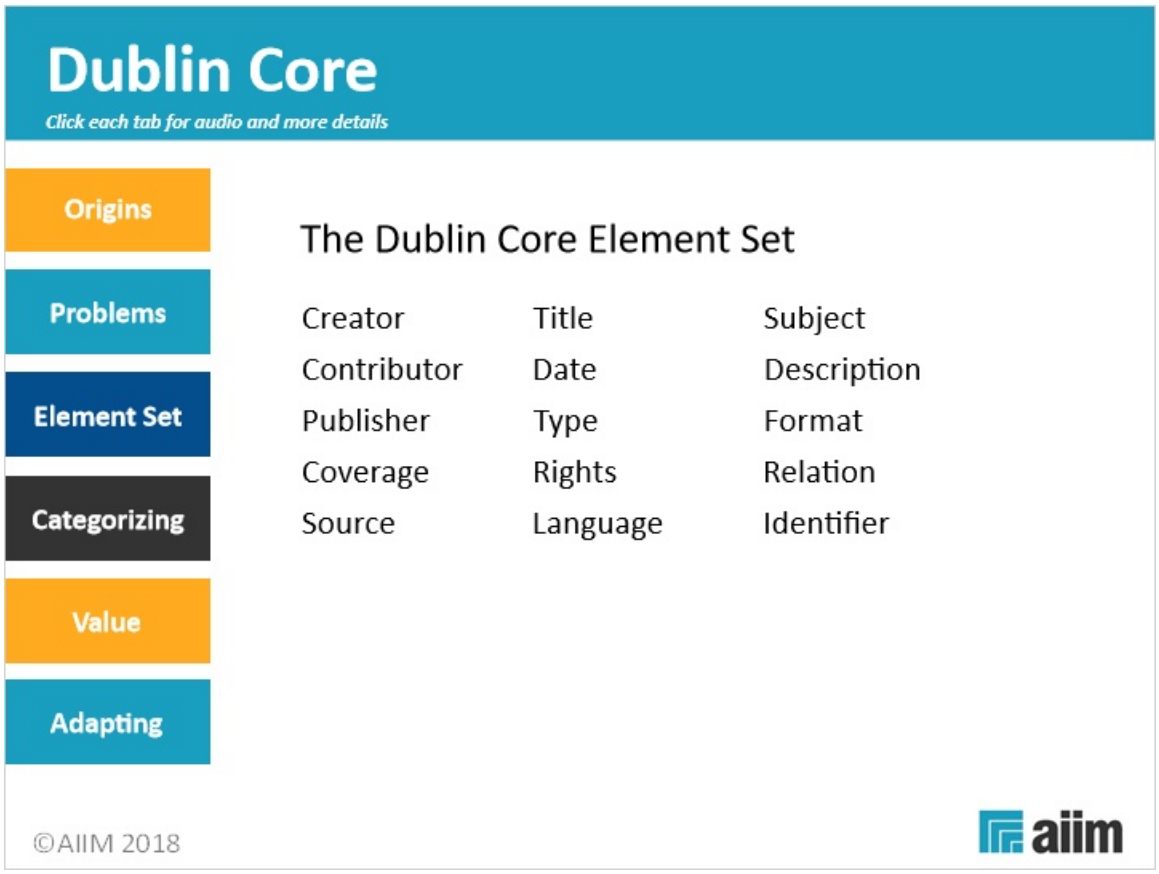
Value (Slide Layer)



Adapting (Slide Layer)



Element Set (Slide Layer)



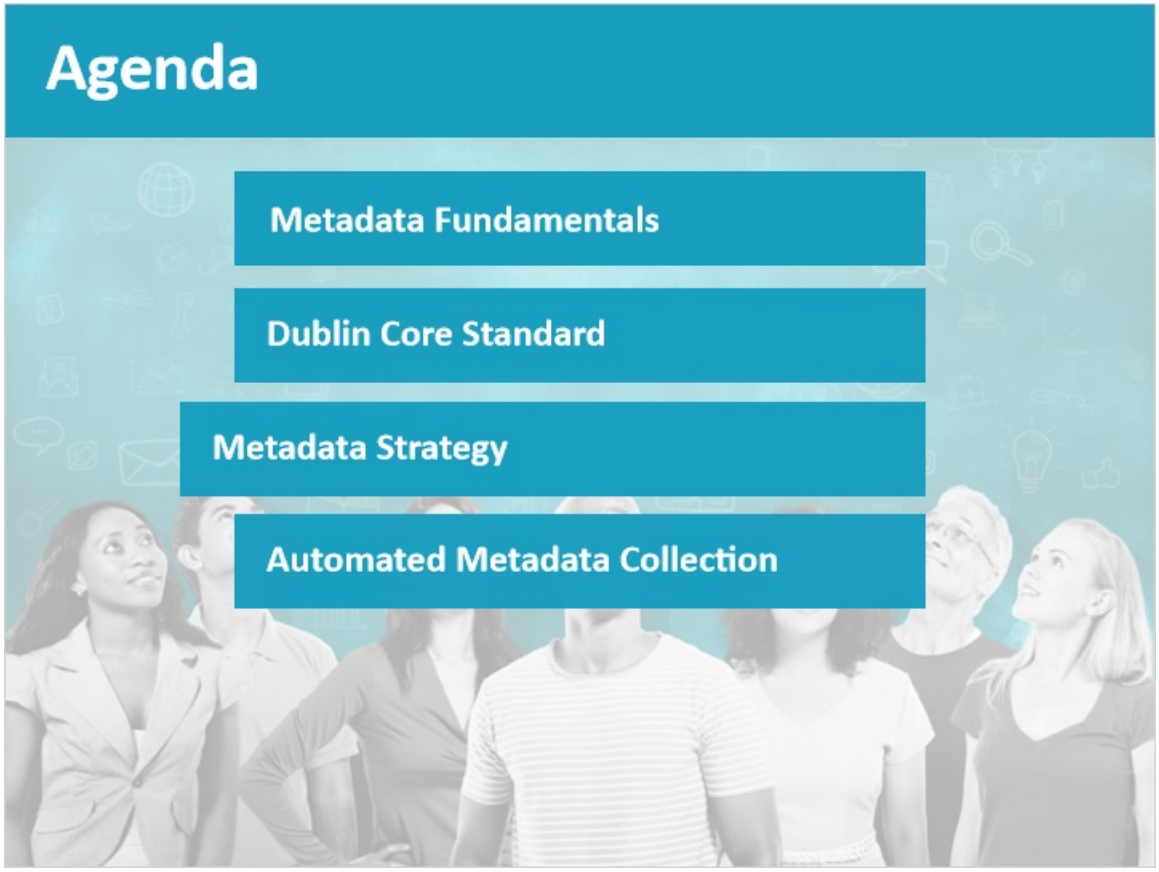
1.13 Why Dublin Core?

Notes:

As more organizations adopt the Dublin Core standards, information is more easily exchanged among repositories. A lot of the content technologies that you might be considering for your enterprise are now offering Dublin Core standard repositories and formats out of the box. So if you decide to license one of these technologies, or perhaps host your content with one of these sorts of technologies, it's good to have adopted the Dublin Core standards, simply because it will be easier to move your content into one of these standard repositories. There's really no better time to standardize your content and adopt one of these standards than it is when you are adopting one of these technologies.

You can also start with just a subset. You don't have to use all of it at once. You can use as much or as little as you need, and gradually, as you move into using more and more of it, you'll find that your information becomes easier to exchange among systems and more easily findable because you are adopting a consistent way of storing metadata.

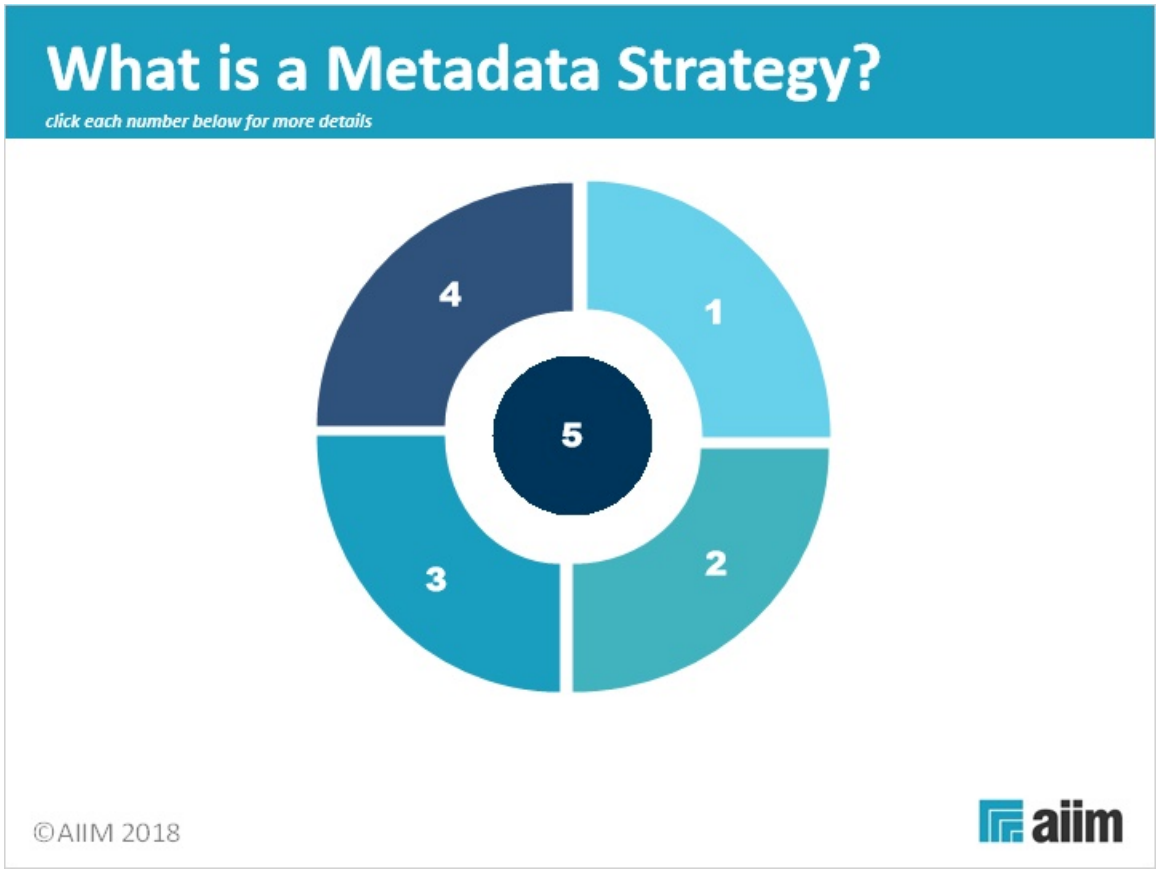
1.14 Agenda



Notes:

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

1.15 What is Metadata Strategy



Notes:

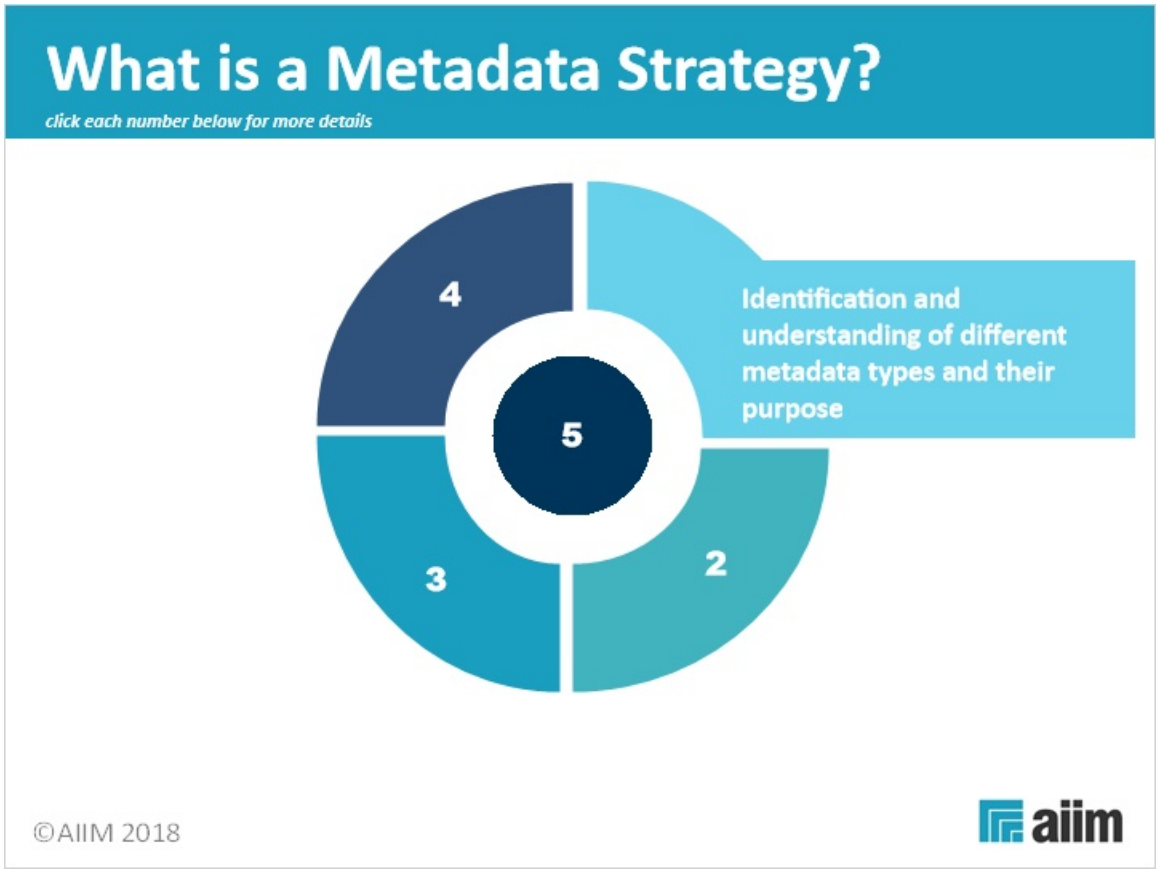
Wrapping up, let's summarize what we learned in this course. A taxonomy is a controlled vocabulary standard and a law for categorizing information. The categories of a taxonomy should be derived based on user needs, enterprise requirements and the nature of the content that you're dealing with.

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.

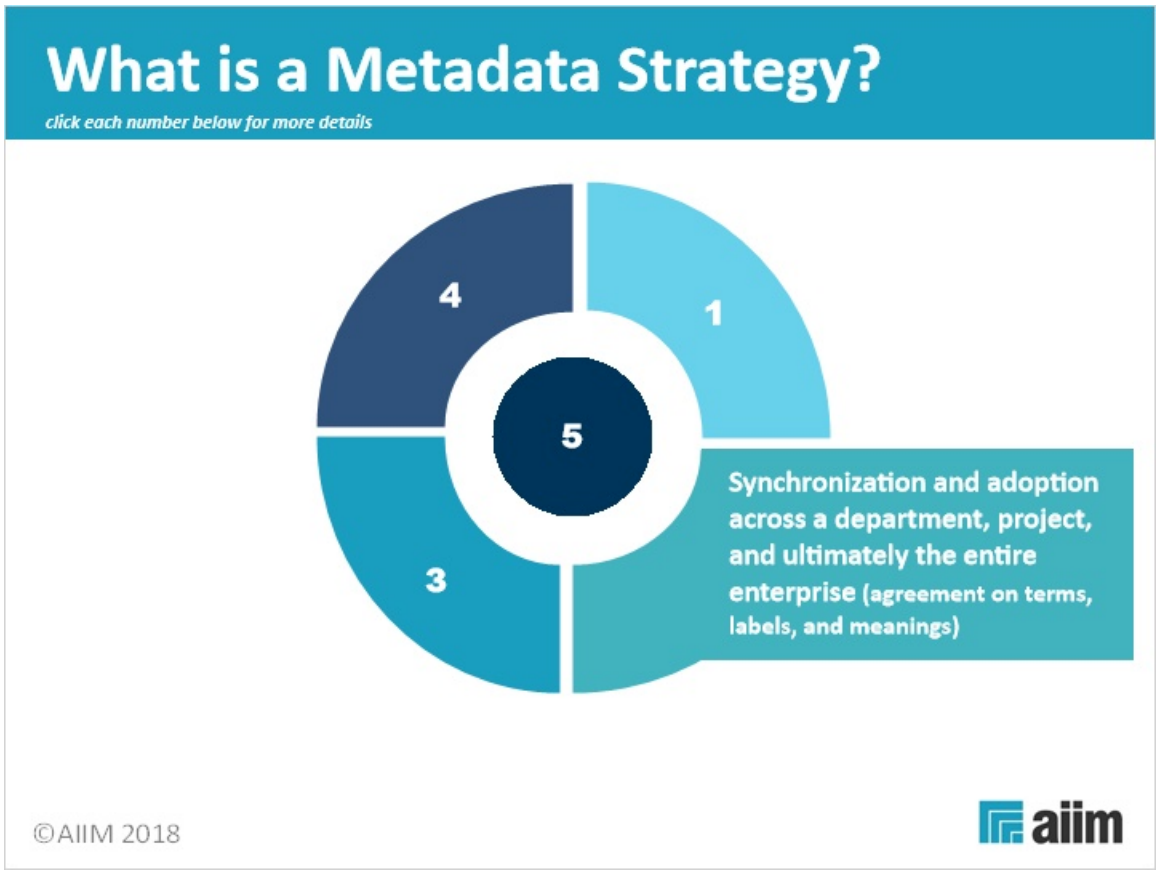
And finally, folksonomies are consumer-driven classification schemes that are particularly beneficial with social networking applications, but you should question the value of them for the enterprise. While they might seem like the easy way to classify your content, it won't make your information more

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.

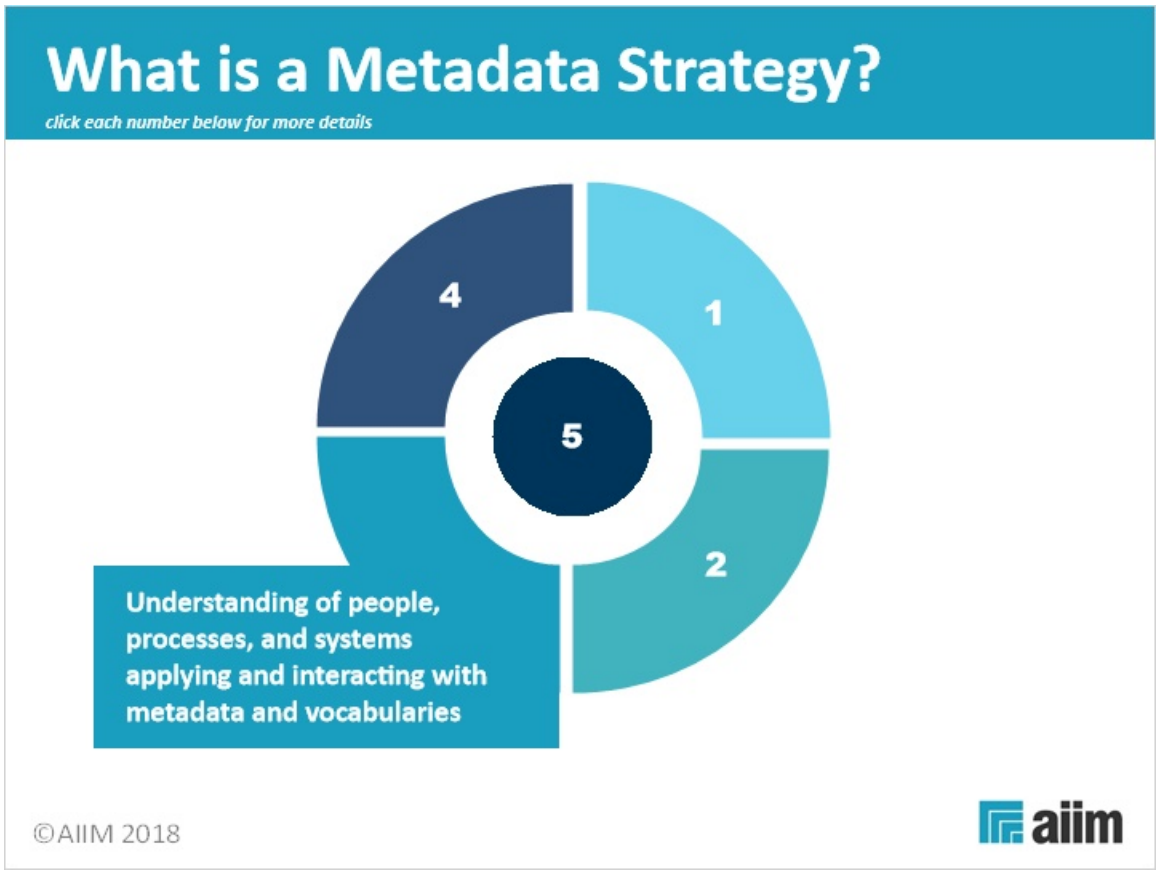
Arc 1 (Slide Layer)



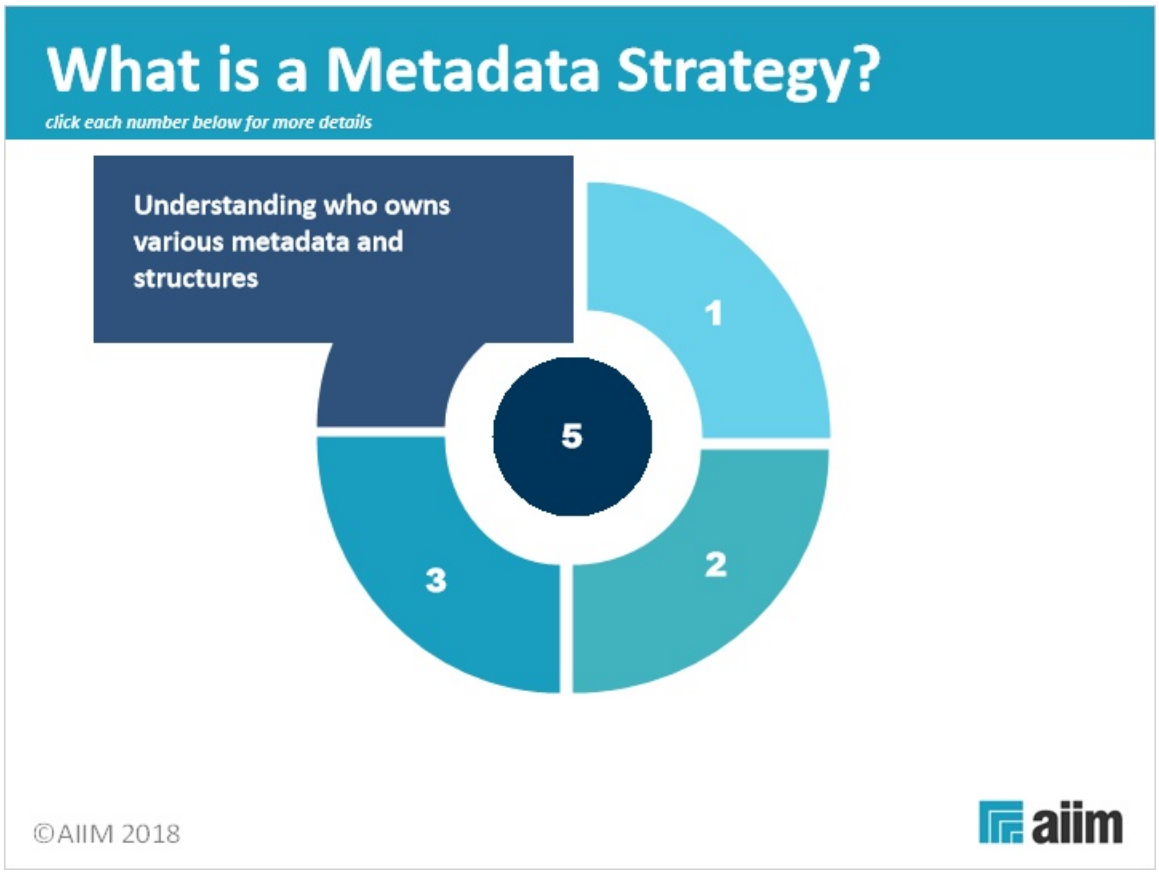
Arc 2 (Slide Layer)



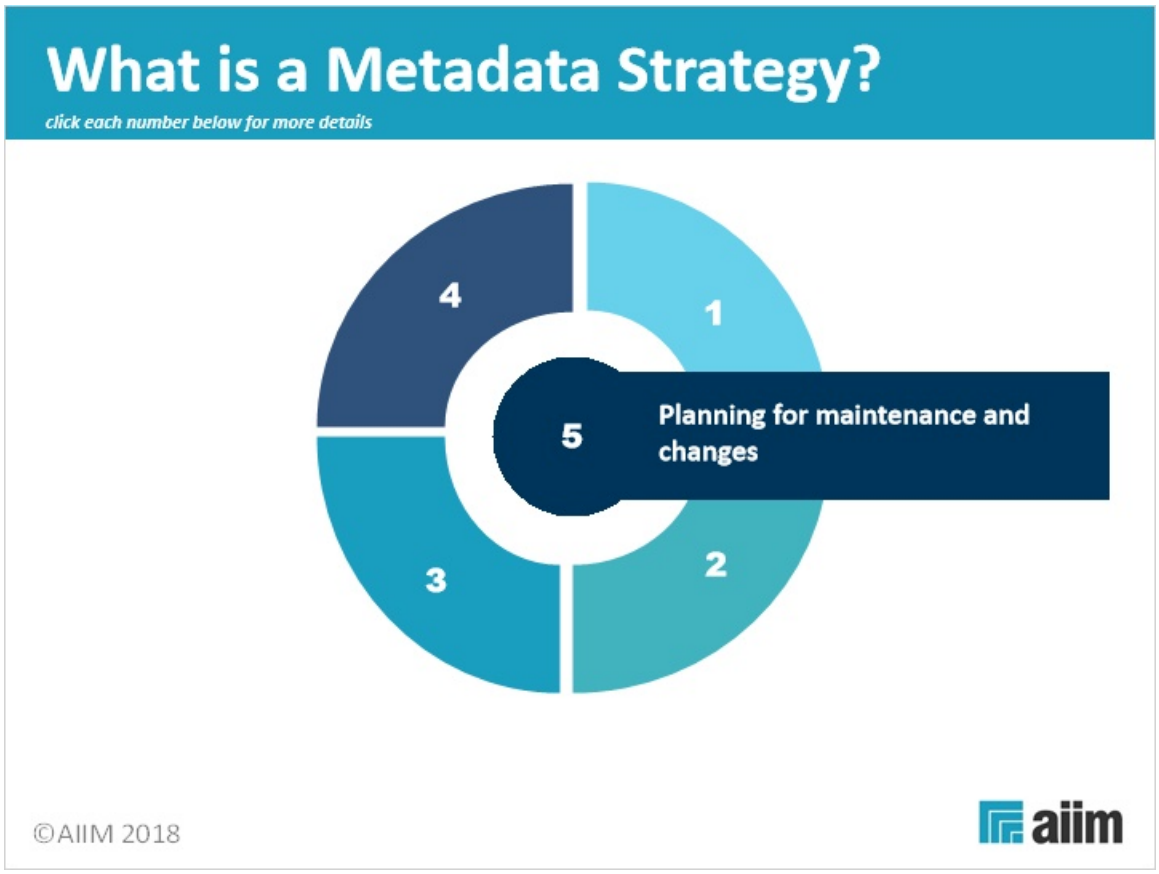
Arc 3 (Slide Layer)



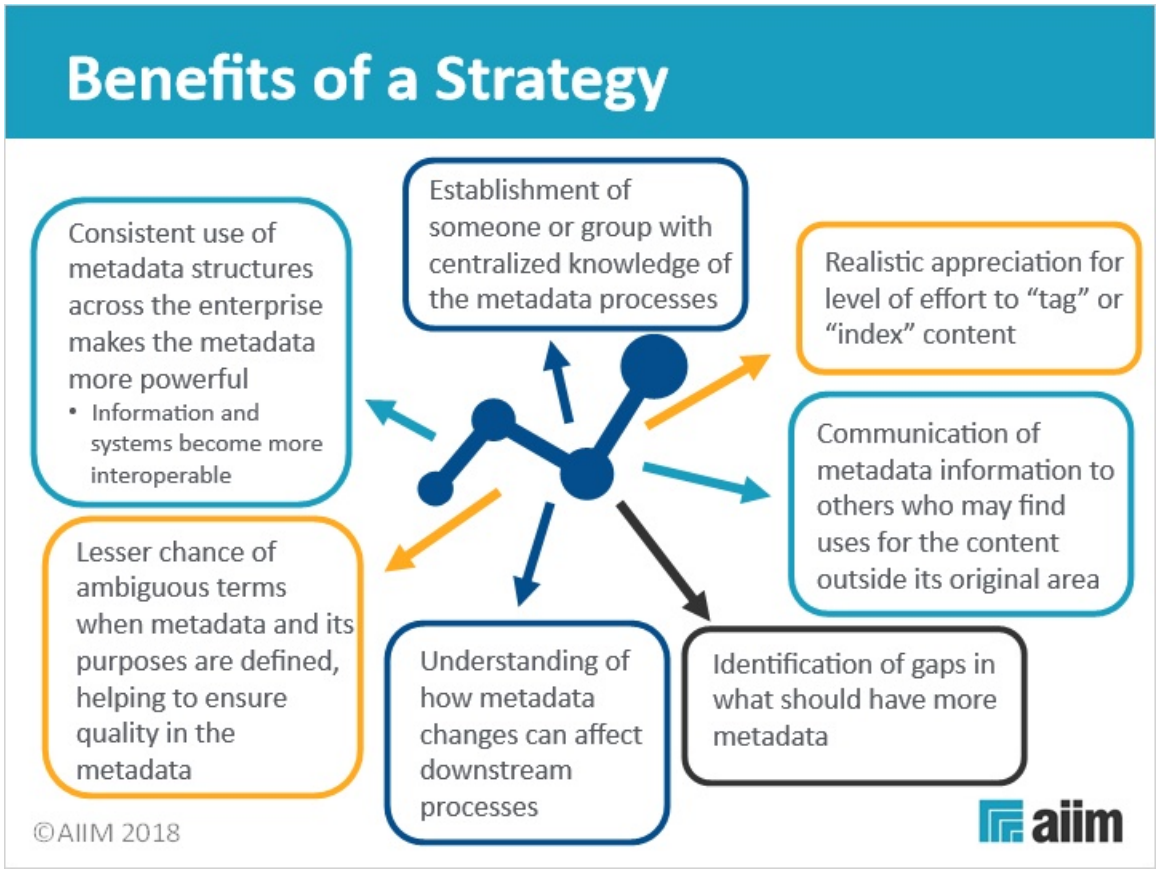
Arc 4 (Slide Layer)



Oval - 5 (Slide Layer)



1.16 Benefits of a strategy



Notes:

There are a number of benefits to establishing this strategy early on. Consistent use of the metadata structures across the enterprise makes the metadata more powerful. If the metadata is not consistent, the effectiveness really gets diluted. You won't be able to consistently find information. And the information and systems will be more interoperable and be able to exchange content more easily if those structures are consistent.

The disambiguation I mentioned earlier is also very important. Understanding how the metadata changes will affect downstream processes is also very important. If someone changes metadata, that might kick off a different process or it might change where the content ends up being displayed. It's very important that your content managers understand the ramifications of tagging content in certain ways.

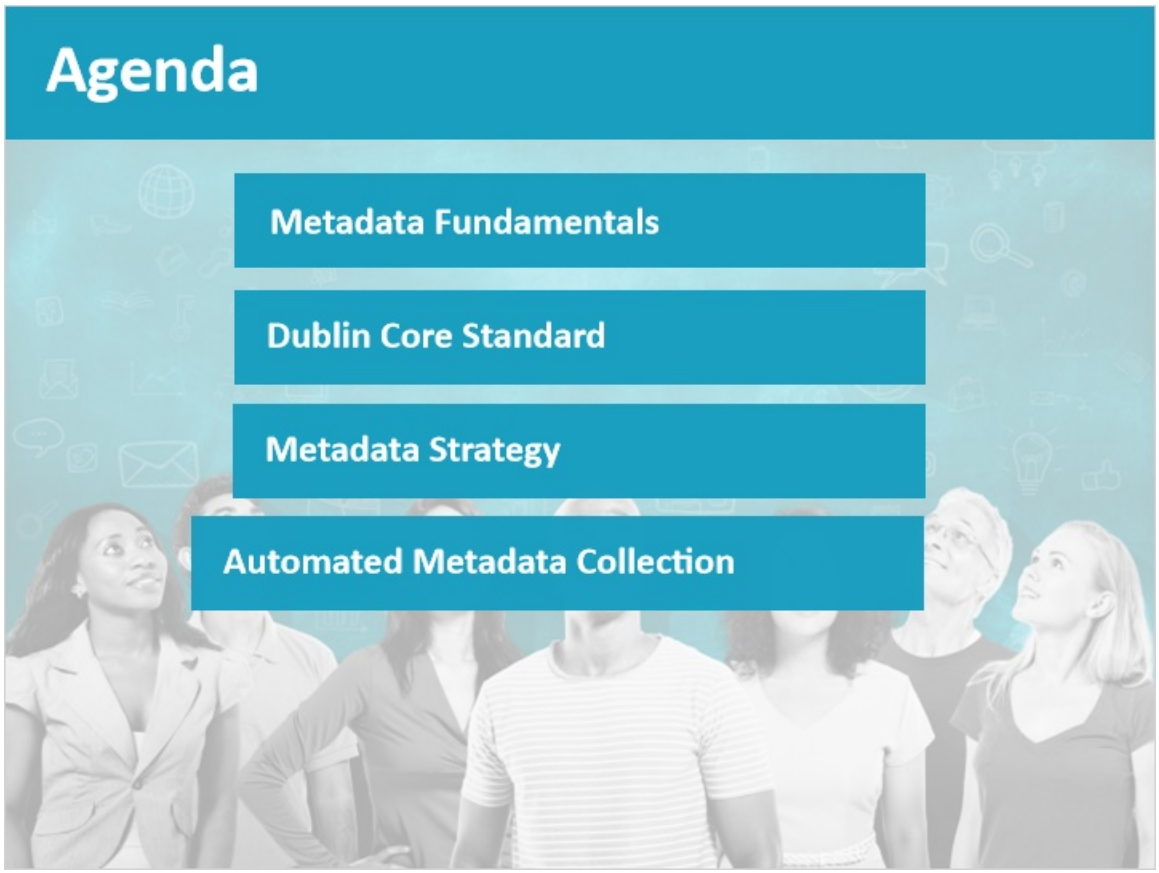
Identifying gaps are also very important because, odds are, you have some implicit metadata already in your organization. These gaps often surface by understanding the end use of the content, how people are searching for it, how that content might need to be used within a system or on a website. That's often where you realize you need

certain types of metadata in order to accomplish that.

Part of the key of establishing a strategy is making sure that everyone's on the same page and that you communicate it out to the organization. With metadata, people outside of that initial circle that's establishing a strategy might actually find a lot of use for that metadata that they hadn't thought about before there actually was a metadata strategy. Understanding the level of effort to tag or index content is also something that tends to come out while establishing a strategy. You need to be prepared with the adequate level of resources and motivation to really tag an index content.

The strategy should also result in the establishment of some sort of centralized knowledge of the metadata process. This could be a single individual or it could be the whole group that was involved with establishing the strategy. Either way, it's important to have that central source that people can go to if they have a question about how to apply metadata, how it's being used across the enterprise.

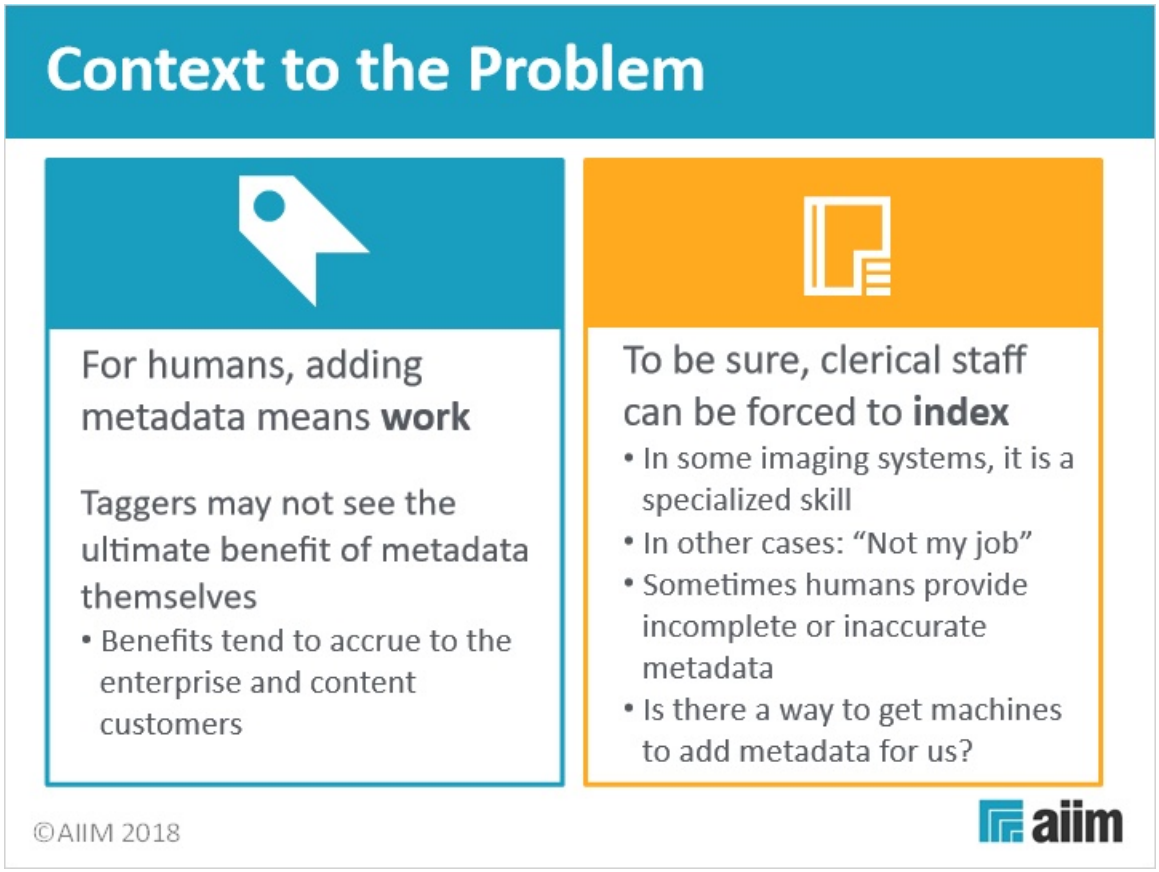
1.17 Agenda



Notes:

We conclude this course with some strategies for how to automate the collection of metadata.

1.18 Context to the Problem



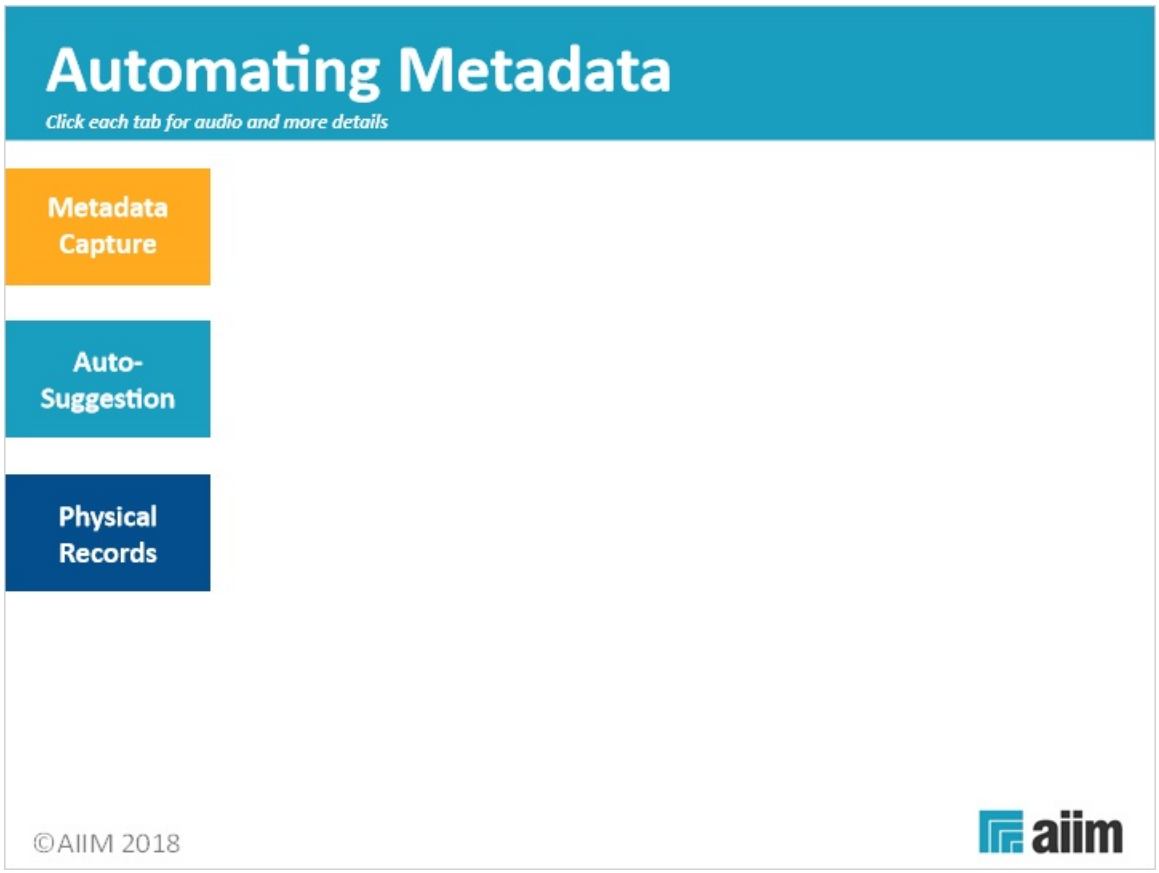
Notes:

In our last section before wrapping things up, let's talk about automated collection of metadata or auto-tagging. For human beings, adding metadata often means quite a bit of work. The people who are assigned to tag may not really see the benefit of getting involved. It becomes to them a chore or laborious process. They might even say, "Well, it's not my job, so why should I do it?"

In many cases it's not just anybody who should be tagging. In some systems, it's a very specialized skill, not only the technical expertise required to use the system and to tag, but also to have the subject matter expertise to be able to do so accurately and effectively.

Commonly, human beings provide incomplete or inaccurate metadata, especially if you have a large number of people tagging. It's very difficult to not leave it open to interpretation. People will tag things differently based on what department they're in, et cetera. Given the challenges in, not only incenting people to tag, but also in keeping what they do complete and accurate, the question arises, well, is there a way to get machines to add metadata for us, consistently, accurately, and to alleviate some of the burden from people within enterprises? Let's explore that question a little bit now.

1.19 Automating Metadata



Notes:

Here is a brief overview of the different ways to automate metadata capture. These are generally arranged in order from least to most automated.

- Auto-suggestion - this approach generally relies on manual metadata entry by users, but offers suggestions based on the first few letters types.

- Bulk import and naming - some tools will allow users to import documents and records in bulk and associate metadata values through the use of scripts
- User-derived - metadata is applied based on the user's credentials
- System-generated - metadata generated automatically such as the date added to the system or the file size
- Default metadata - metadata that is inserted by default but which can often be updated or overridden - for example the date a document was created or edited that defaults to today's date
- Inherited metadata - metadata inherited by a record based on its location - for example a record could inherit a retention period by being placed into a particular folder or classification node
- Workflow - metadata could be assigned and captured as part of a workflow process

Regardless of which approaches are used, another helpful mechanism to make metadata capture more automated and efficient is to use validation. This can range from as simple as verifying that the contents of a numeric field contain only numerals, to adding up individual fields and comparing with a master total using formulas, to actually generating metadata values based on the contents of other fields and/or formulas.

The graphic on this slide shows you an example of auto-suggestion. In this example, the system suggests particular keywords, sectors and locations to associate with this document or record. The user of the system can click yes or no, modify the selections, approve, and save.

This type of integration can help increase metadata efficiency and, otherwise, reduce user resistance to having to manually tag each piece of content. It's a good way to speed up the capture process without necessarily having to read an entire document and manually classify it.

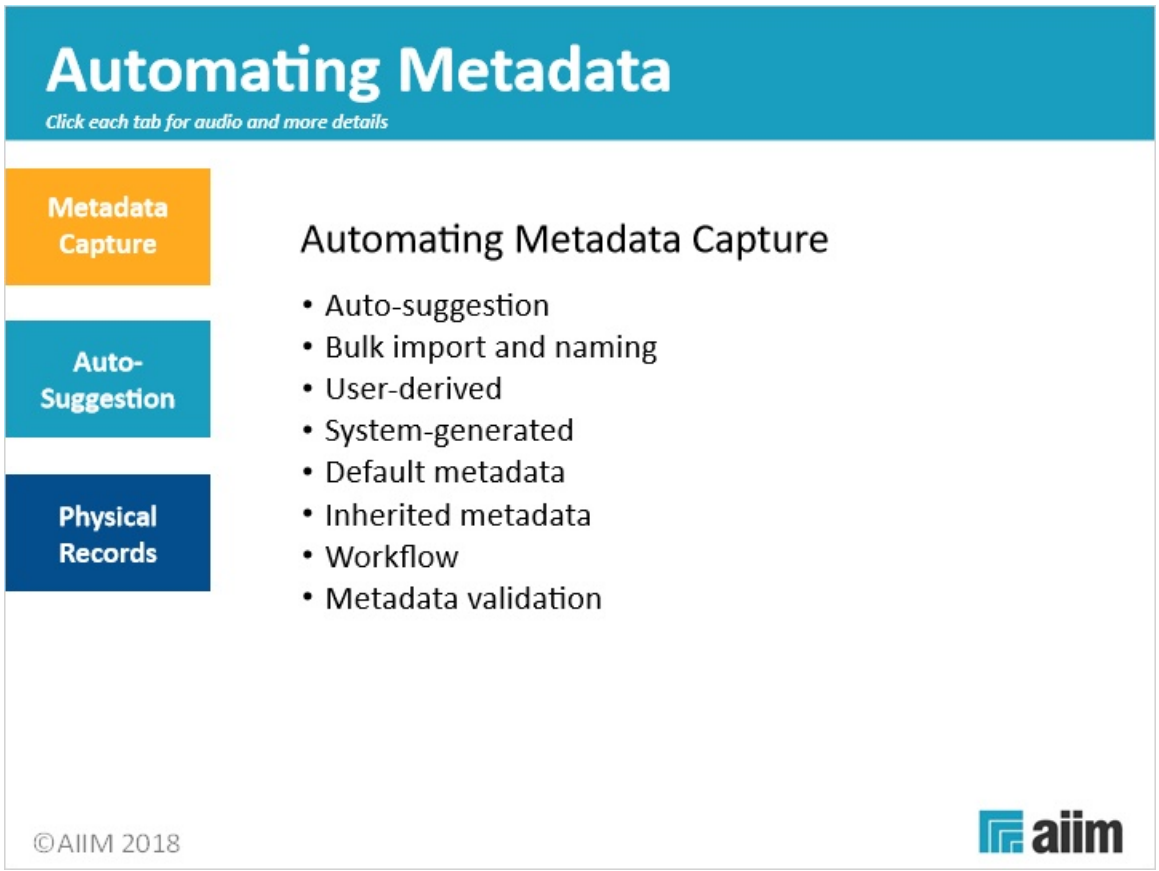
Even a simple approach that captures document-specific metadata (for example the title, and addressee(s) if relevant) can make its capture not only more efficient but more reliable and consistent.

Before we leave this section we should note that it is possible to automate the capture of metadata for physical documents or items as well. There are a couple of ways to do this. The first is to convert them to digital documents, for example by scanning them. Once the documents are digitized, they, and their metadata, can be captured and classified using business rules - for example, a template could be set up to recognize invoices. Rules are defined about how to process invoices based on customer ID, amount, or some other field and then the invoices could be captured and classified appropriately.

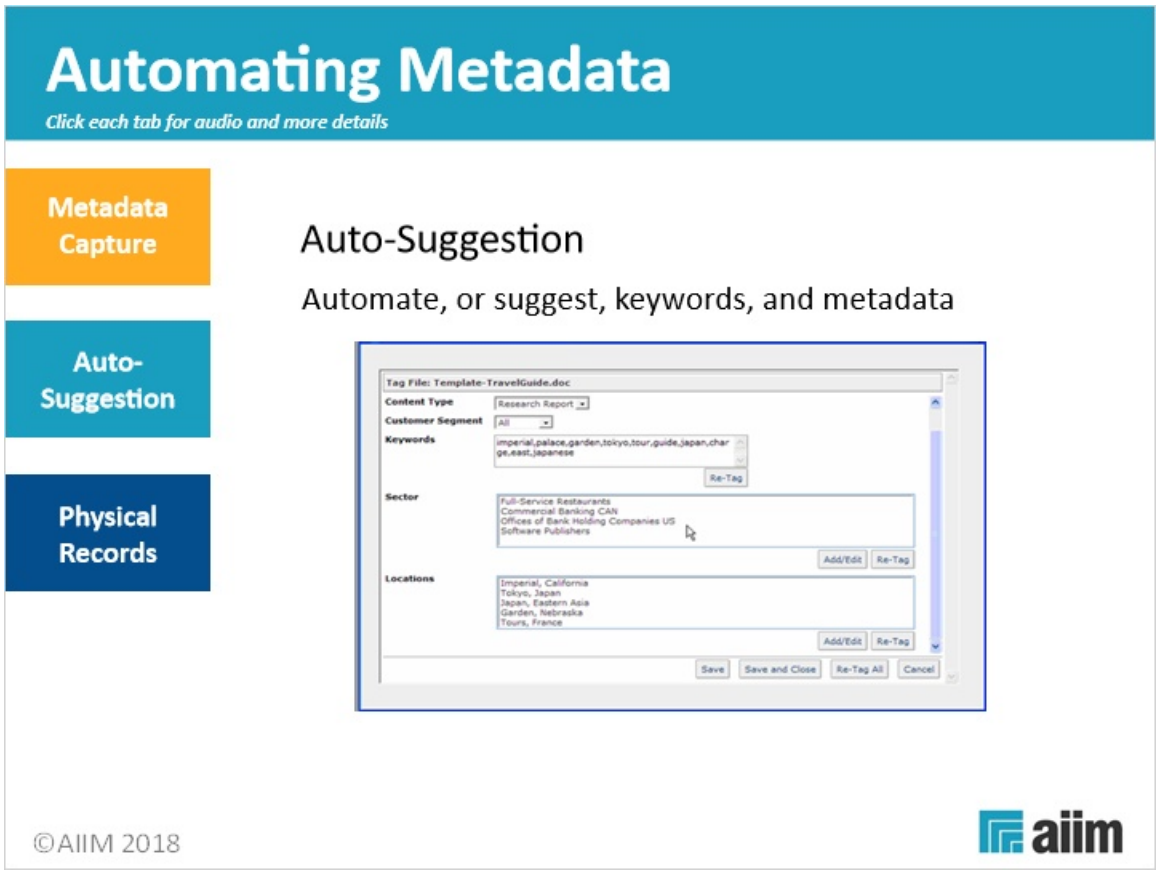
Another mechanism is to use barcodes. This is still a bit more manual process in that someone needs to affix the barcode to the physical document in question and users will need to scan the barcode to access, track, and manage it. But a barcode does make capture and registration of physical documents more efficient and accurate. Some systems will even generate barcodes automatically prior to printing forms or documents or as part of the capture and registration process.

Radio frequency identification, or RFID, is similar to barcodes in that physical documents can be registered and managed in a more automated way. But there are more advanced RFID capabilities available that allow records to be passively tracked as they are moved around in the organization.

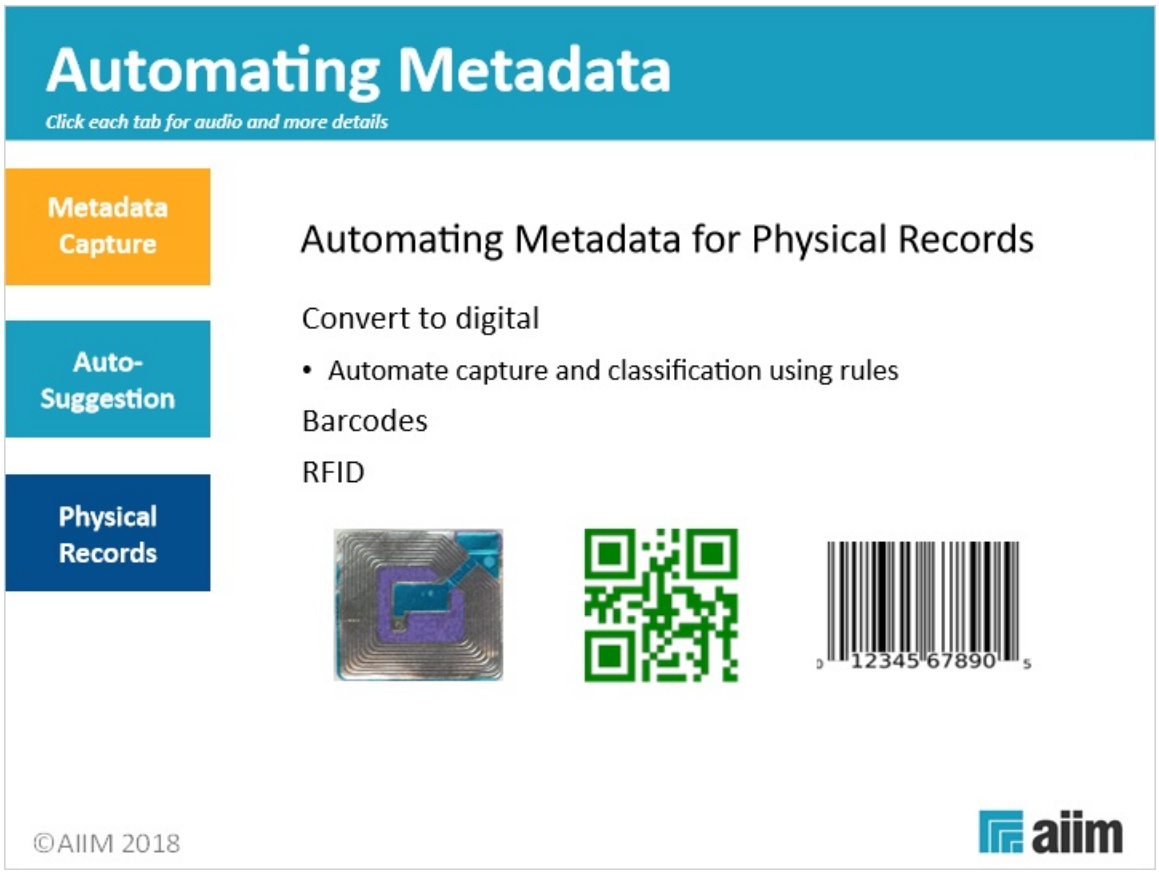
Capture (Slide Layer)



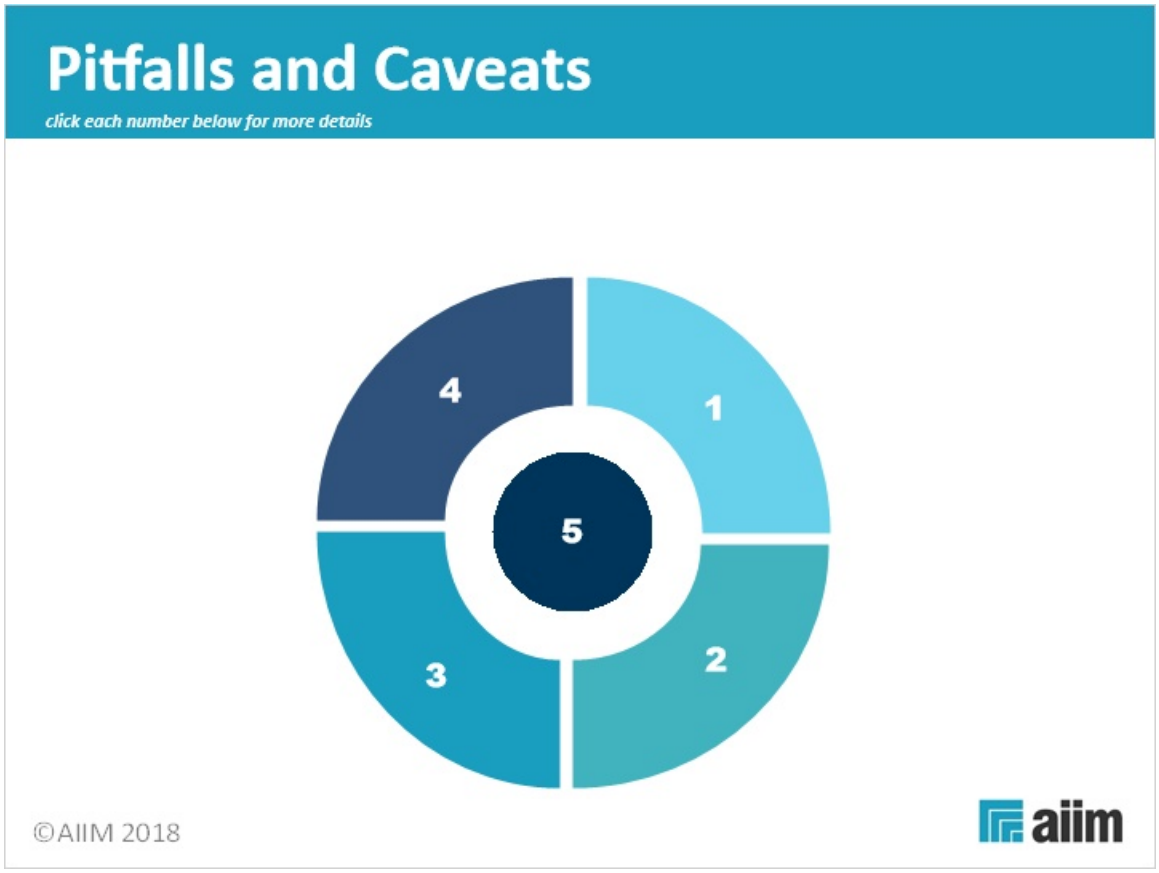
Auto-Suggestion (Slide Layer)



Physical Records (Slide Layer)



1.20 Pitfalls and Caveats



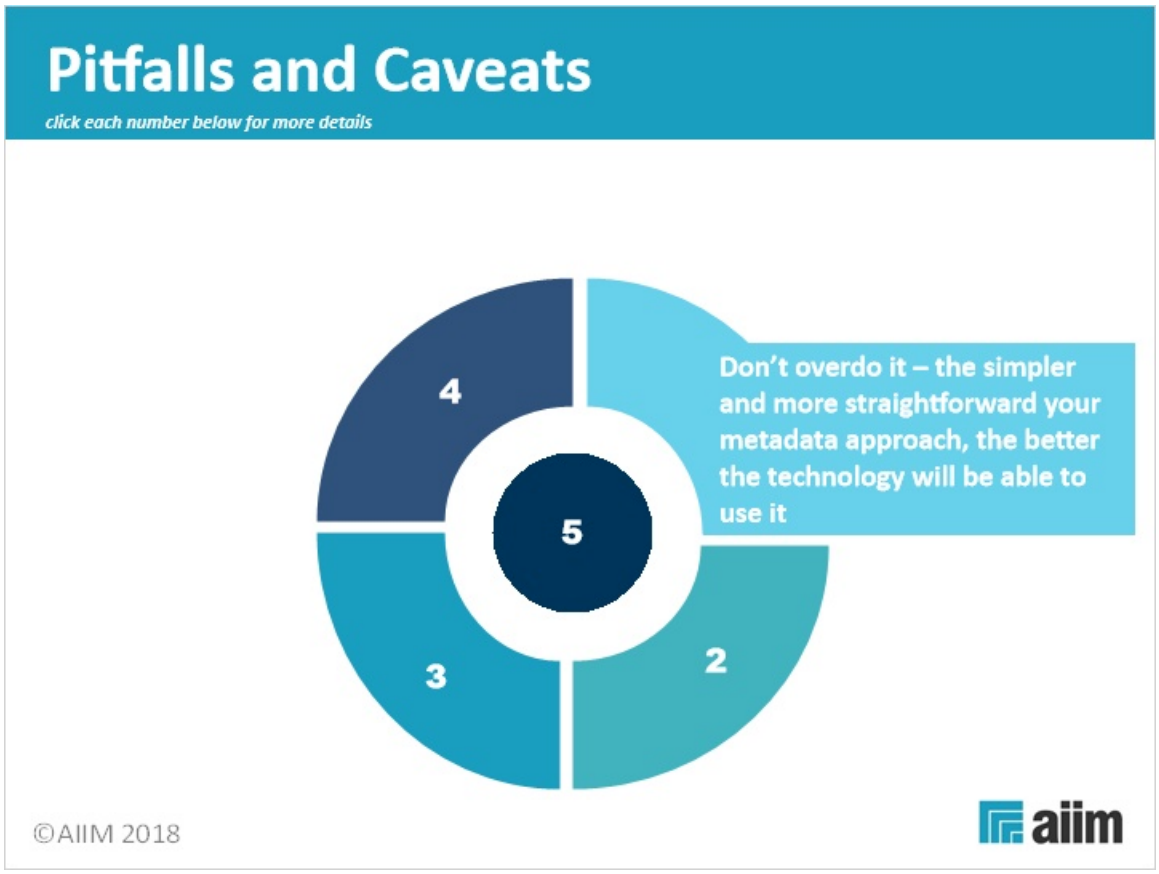
Notes:

A few pitfalls and caveats. Be sure you don't overdo it. The simplest and most straightforward metadata approach tends to be the most usable. It also tends to be the approach that technology can work with the most easily.

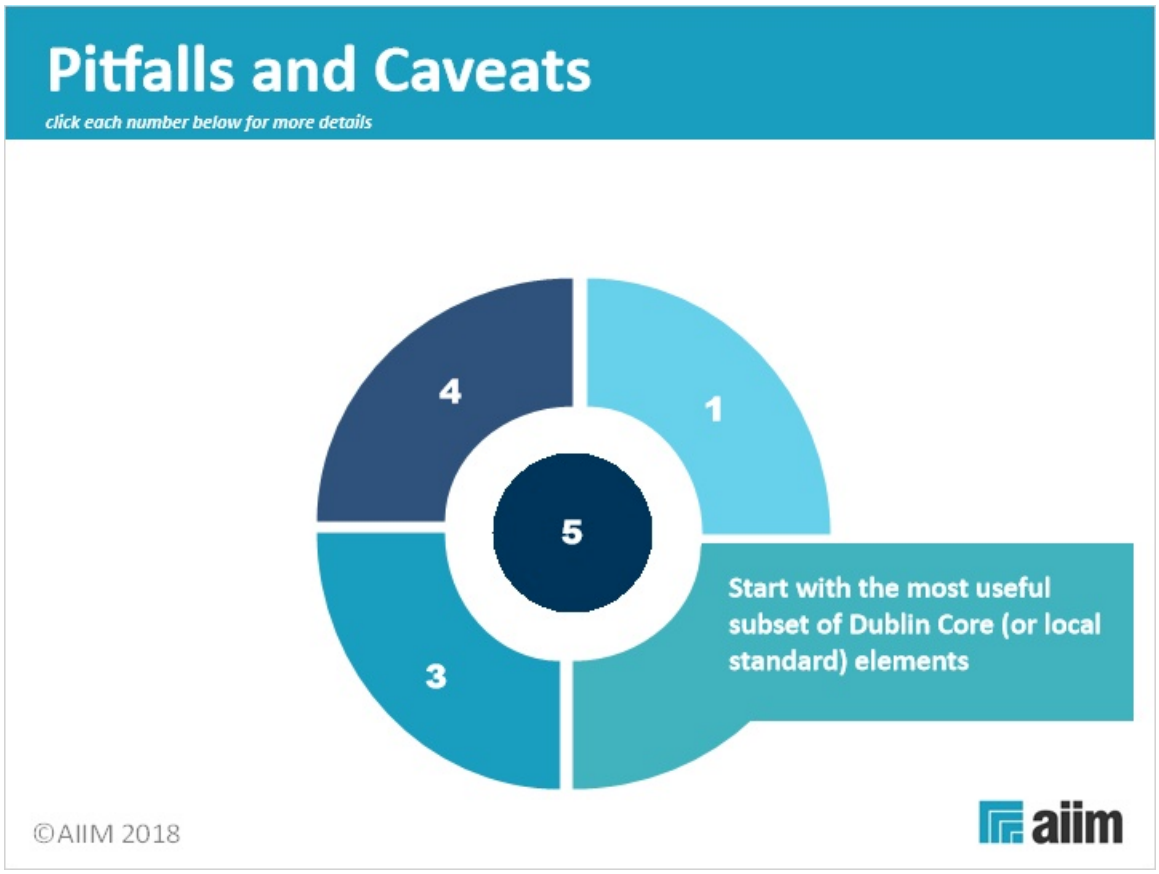
Make sure you start with a foundation, a standard, such as the Dublin Core standard elements. Use what makes sense for your organization and then build on that.

Plowing forward without a metadata strategy results in a complete mess, resulting in confusion and inconsistency. You need to make sure you establish that strategy and inform your organization as to what it is.

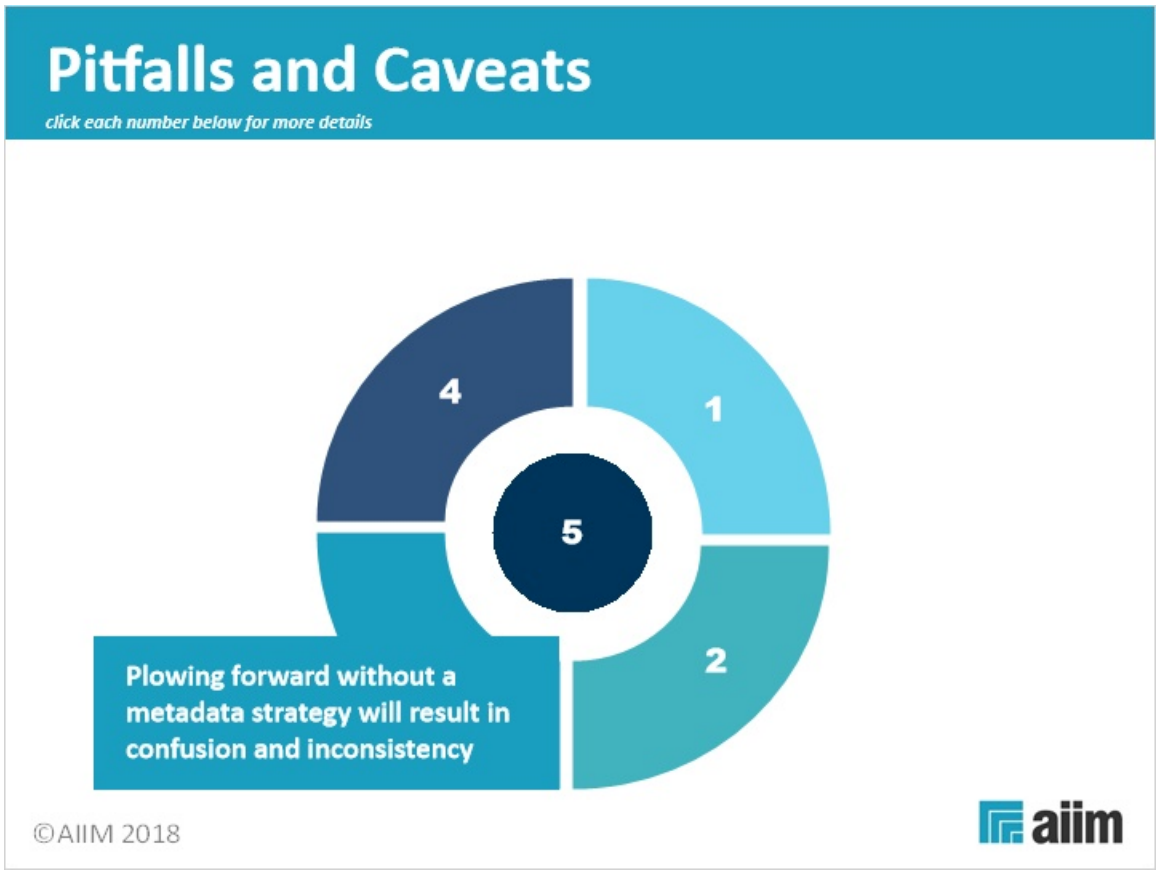
Arc 1 (Slide Layer)



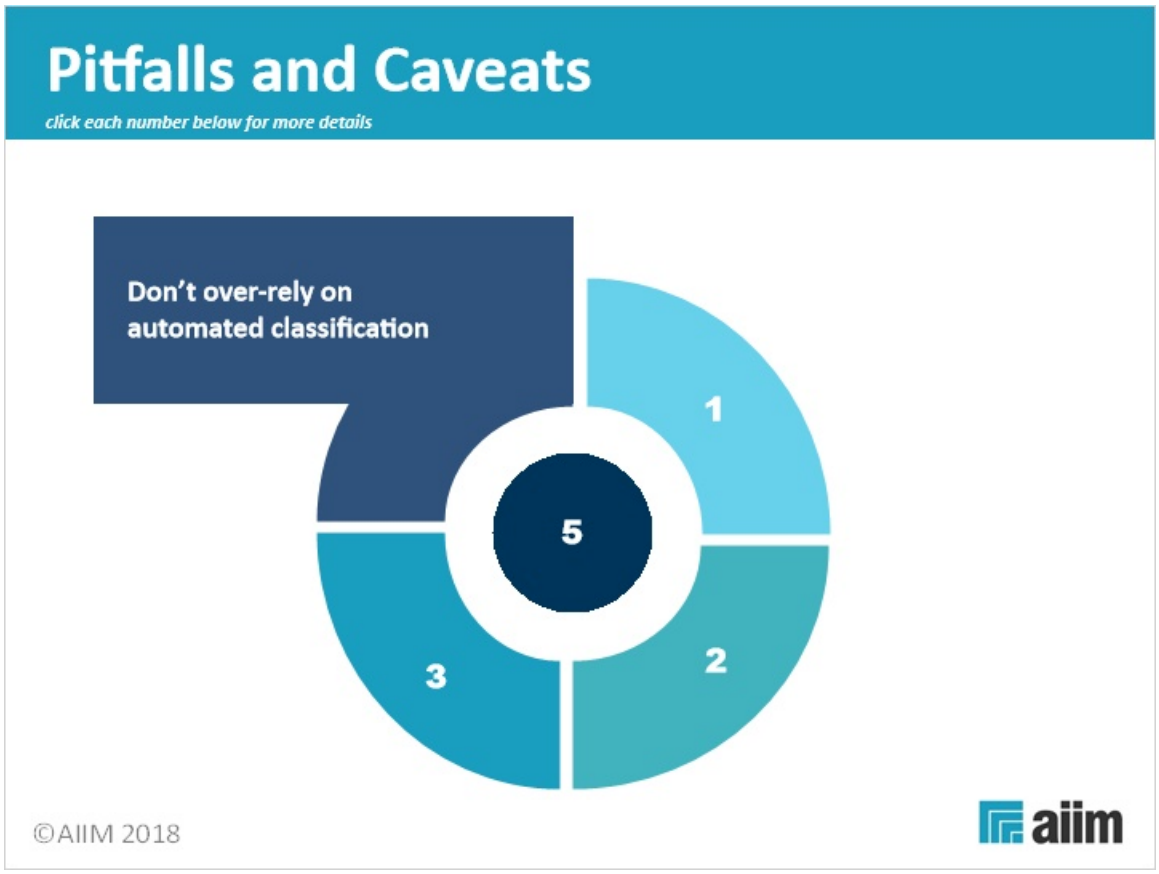
Arc 2 (Slide Layer)



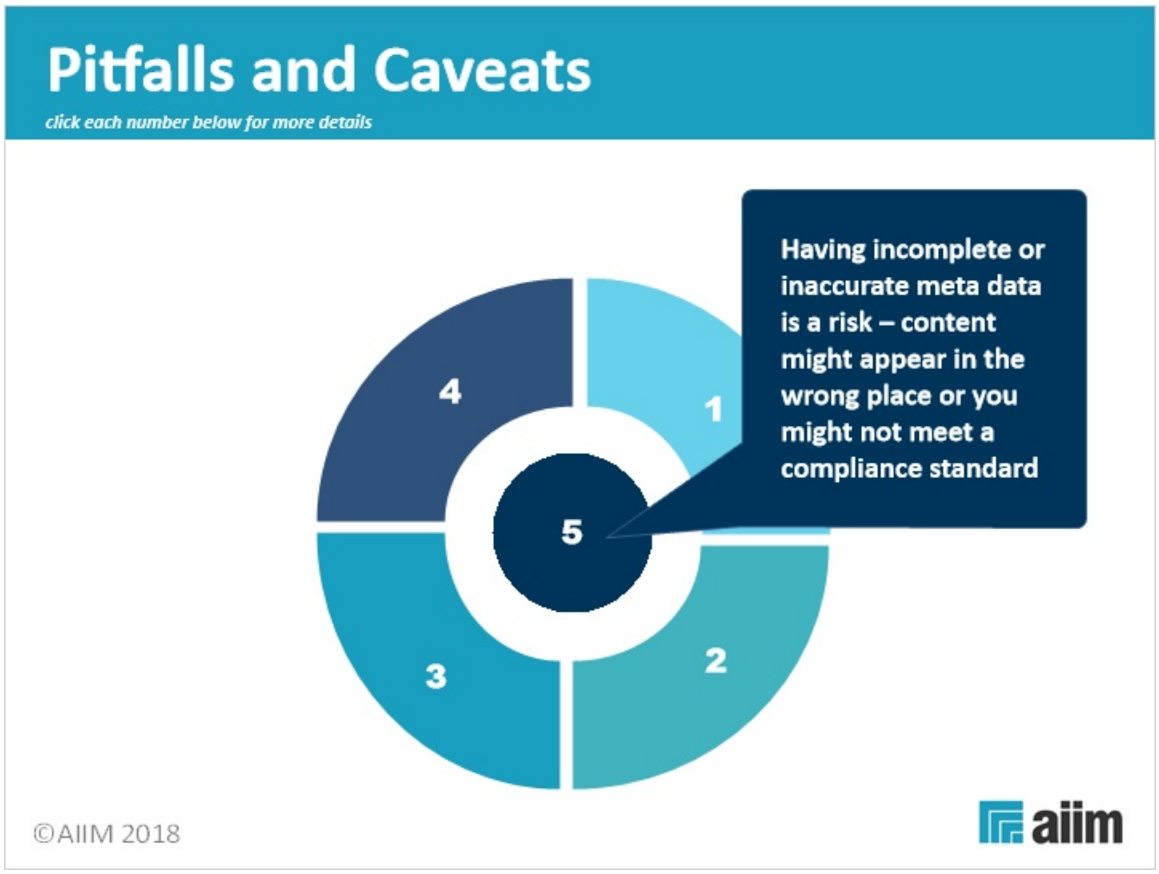
Arc 3 (Slide Layer)



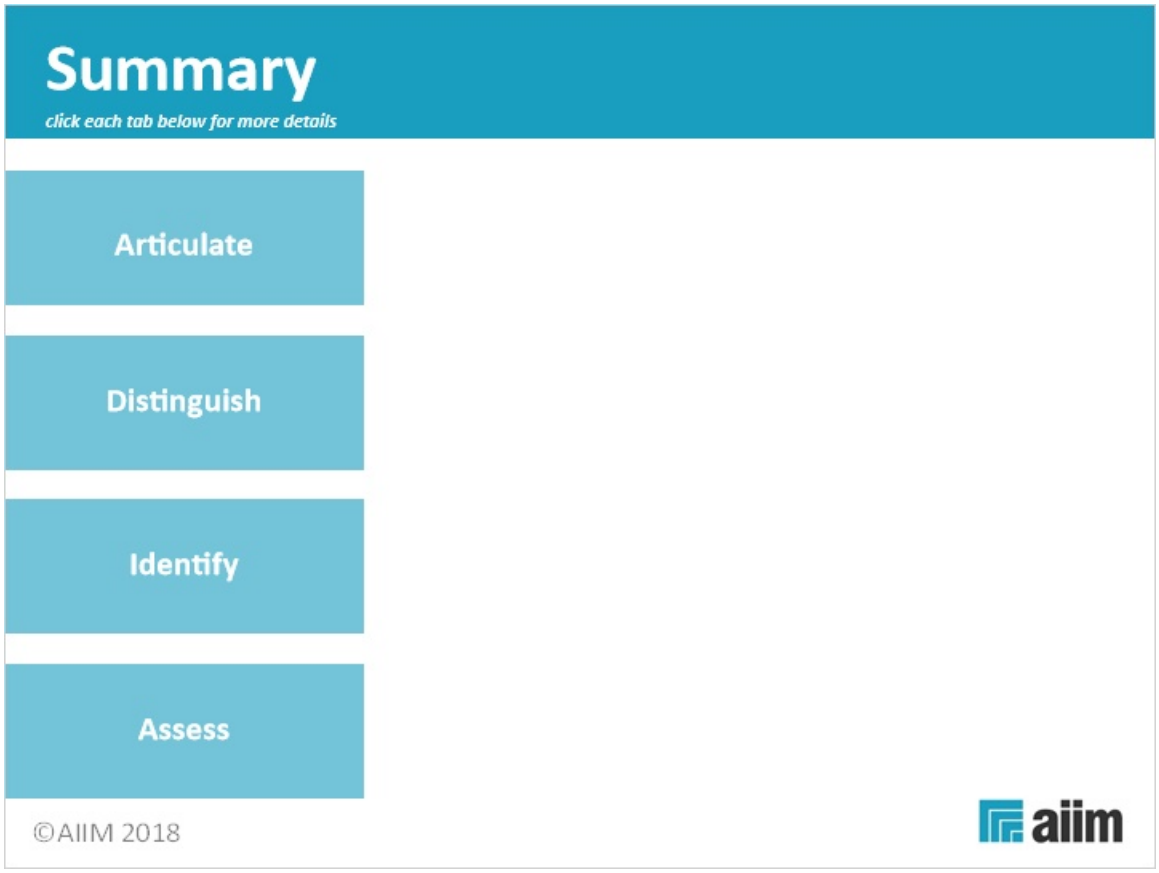
Arc 4 (Slide Layer)



Oval - 5 (Slide Layer)



1.21 Summary

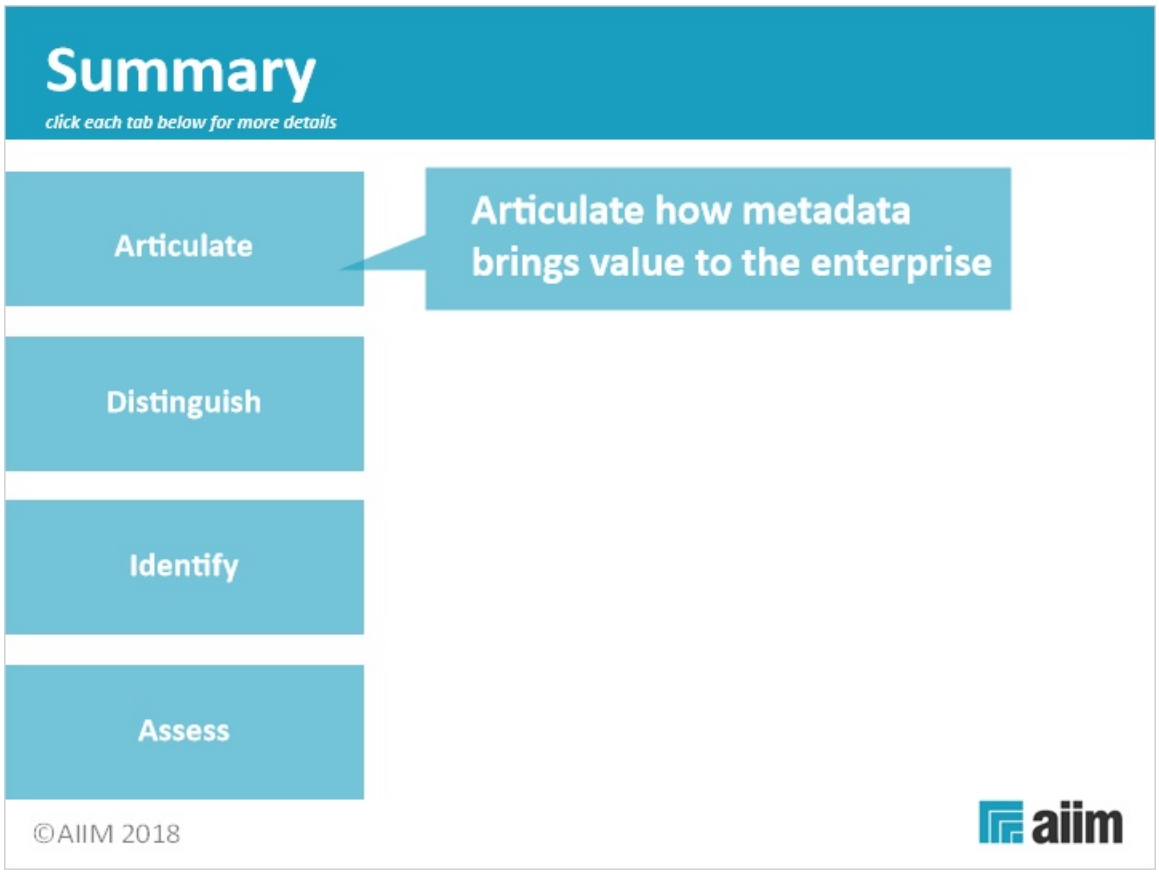


Notes:

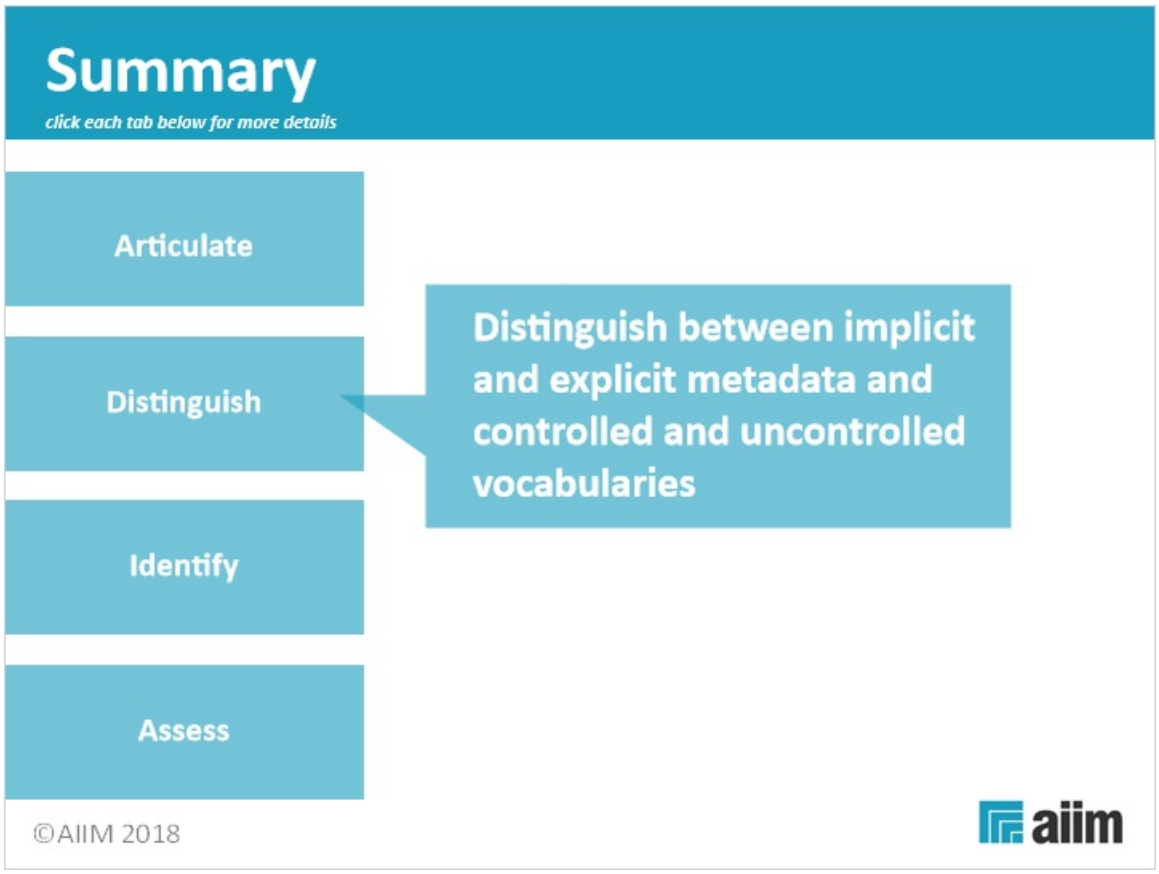
In this course, you learned to:

- Define metadata
- Articulate how metadata brings value to the enterprise
- Distinguish between implicit and explicit metadata and controlled and uncontrolled vocabularies
- Describe the purpose and value of the Dublin Core standard elements
- Identify the main dimensions of a metadata strategy
- Assess the role and value of automated classification

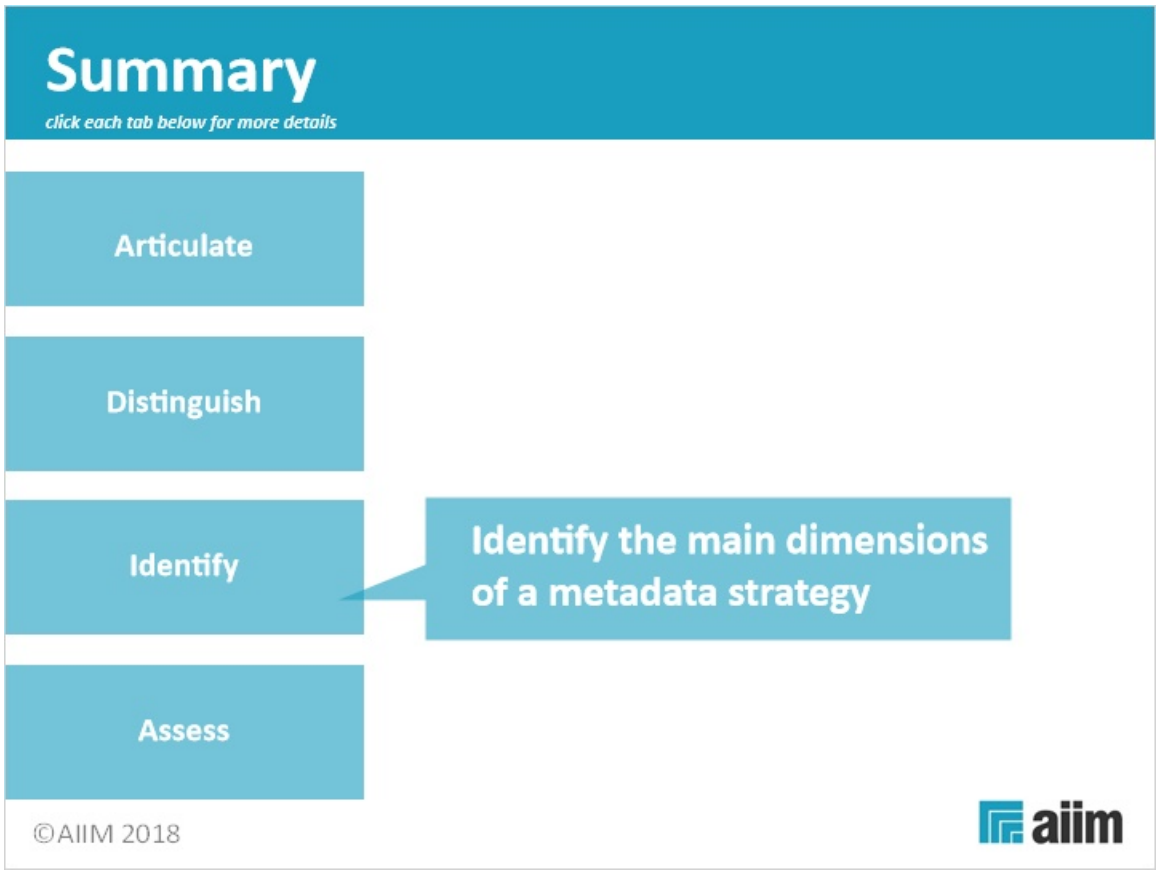
Articulate (Slide Layer)



Distinguish (Slide Layer)



Identity (Slide Layer)



Assess (Slide Layer)