

Introduction to Metadata for Decision-Makers Briefing Talking Points

- Identifies points in the script and slides where you may want to customize the briefing to include case studies and examples from your own experiences, and exercises which may be more appropriate for your audiences.

Customization Index

(listed by slide number)

Ack-1
Page 1
Page 3
Page 4
Page 6
Page 15
Page 16
Page 20
Page 21
Page 22
Page 23
Page 25
Page 26
Page 37
Page 38
Appendices

Slide	Talking Points
<p>Briefing Background, Acknowledgements, and Contact Information</p>	<p>Briefing Background, Acknowledgements, and Contact Information</p>
<p>Ack-1 This briefing and all related materials are the direct result of a two-year grant to the State Archives Department of the Minnesota Historical Society (MHS) from the National Historical Publications and Records Commission (NHPRC). Work on the “Educating Archivists and Their Constituencies” project began in January 2001 and was completed in May 2003.</p> <p>The project sought to address a critical responsibility that archives have discovered in their work with electronic records: the persistent need to educate a variety of constituencies about the principles, products, and resources necessary to implement archival considerations in the application of information technology to government functions. Several other goals were also supported:</p> <ul style="list-style-type: none"> • raising the level of knowledge and understanding of essential electronic records skills and tools among archivists, • helping archivists reach the electronic records creators who are their key constituencies, • providing the means to form with those constituencies communities of learning that will support and sustain collaboration, and • raising the profile of archivists in their own organizations and promoting their involvement in the design and analysis of recordkeeping systems. <p>MHS administered the project and worked in collaboration with several partners: the Delaware Public Archives, the Indiana University Archives, the Ohio Historical Society, the San Diego Supercomputer Center, the Smithsonian Institution Archives, and the State of Kentucky. This list represents a variety of institutions, records environments, constituencies, needs, and levels of electronic records expertise. At MHS, Robert Horton served as the Project Director, Shawn Rounds as the Project Manager, and Jennifer Johnson</p>	<p>Ack-1 Explain:</p> <ul style="list-style-type: none"> • slide appearance • space for notes • briefing background in brief • encourage contact with instructors • thank any partners <p><input checked="" type="checkbox"/> <i>Note to instructor: Include your contact information at the bottom of the slide</i></p> <p><input checked="" type="checkbox"/> <i>Note to instructor: It is helpful to handout all the exercises and examples at the beginning of the day, when handing out course books, so that you do not have to interrupt the briefing later. Consider copying them on sheets of different colored paper so that they are easy to distinguish from one another.</i></p>

<p>as the Project Archivist.</p> <p>MHS gratefully acknowledges the contribution of Advanced Strategies, Inc. (ASI) of Atlanta, Georgia, and Saint Paul, Minnesota, which specializes in a user-centric approach to all aspects of information technology planning and implementation. MHS project staff received training and guidance from ASI in adult education strategies and briefing development. The format of this course book is directly based on the design used by ASI in its own classes. For more information about ASI, visit http://www.advstr.com/</p> <p>For more information regarding the briefing, contact MHS staff or visit the briefing web site at http://www.mnhs.org/preserve/records/workshops/edarchivists.html</p>	
<p>Page-1 This briefing includes:</p> <ul style="list-style-type: none">• Briefing objectives• What do we mean by information resources, digital objects, and electronic records?• Definitions of metadata.• Why is metadata useful?• Systems management metadata• Access metadata• Recordkeeping metadata• Preservation metadata• Putting it all together	<p>Page-1 Discuss the list on the slide.</p> <p><input checked="" type="checkbox"/> <i>Note to instructor: Consider substituting local examples in place of the Minnesota examples given for each metadata function.</i></p>

<p>Page-2 Briefing Objectives</p> <p>Upon completion of this briefing, you will be able to:</p> <ul style="list-style-type: none"> • understand what is meant by digital objects and electronic records • understand the definition of metadata • discuss what metadata may be needed for digital objects • describe different functions of metadata • discuss systems management, access, recordkeeping, and preservation metadata functions and some example standards 	<p>Page-2 Discuss the list on the slide.</p>
<p>Page-3 What do we mean by information resources, digital objects, and electronic records?</p> <p>Information resources: The content of your information technology projects (data, information, records, images, digital objects, etc.)</p> <p>Digital object: Information that is inscribed on a tangible medium or that is stored in an electronic or other medium and is retrievable in perceivable form. An object created, generated, sent, communicated, received, or stored by electronic means.</p> <p>An electronic record is a specific type of digital object with unique characteristics described by archivists and records managers.</p> <p>Types of digital objects:</p> <ul style="list-style-type: none"> • e-mail • spreadsheets • PowerPoint presentations • web pages • word processing documents • digital images • databases • Portable Document Format (PDF) files <p>...and many more</p>	<p>Page-3 As we begin to discuss metadata, let's make sure we're all on the same page by defining some of the terms we'll be using today.</p> <p>We'll need to start with a common definition of information resources, digital objects, and electronic records in order to understand how we can use metadata to describe them.</p> <p>Define Information resources as: The content of your information technology projects (data, information, records, images, digital objects, etc.)</p> <p>Our definition of digital objects comes from E-Sign</p> <ul style="list-style-type: none"> • The Electronic Signatures in Global and National Commerce Act passed by Congress in 1999 to create a common legal framework for electronic commerce and electronic government in the nation. • Digital object is defined as "Information that is inscribed on a tangible medium or that is stored in an electronic or other medium and is retrievable in perceivable form. An object created, generated, sent, communicated, received, or stored by electronic means." • Electronic record <ul style="list-style-type: none"> ◦ is a specific type of digital object with unique characteristics described by archivists and records managers. ◦ very broad and generic definition, which is exactly why it was chosen ◦ it's not exclusive to anything – records, digital objects, data, information, knowledge, all words we may use interchangeably. <p><input checked="" type="checkbox"/> <i>Note to instructor: You may want to add your own definitions here as they relate.</i></p> <p>This is a good place to start.</p> <ul style="list-style-type: none"> • In a practical sense, we need to break that definition down right away - it's too broad.

	<ul style="list-style-type: none"> • To determine what we want to manage and how, we need to be much more precise. • We also need to consider that people often think in terms of types or genres, like e-mail, web pages, databases, word processing documents, and the like. <ul style="list-style-type: none"> ◦ But it's not enough to know what application or file format a digital object is linked to if we want them to be accessible for however long we may need them, especially if we need to share them, re-use them, and/or if they are expected to outlast their original systems <p>Where do we start?</p>
<p>Page-4 Digital objects have three components:</p> <p>Content: Informational substance of the object.</p> <p>Structure: Technical characteristics of the objects (e.g., presentation, appearance, display).</p> <p>Context: Information outside the object which provides illumination or understanding about it, or assigns meaning to it.</p>	<p>Page-4 Archivists and records managers came up with a definition that applies to electronic objects, but we think it applies to all information objects, and it is applicable to anything in a digital format. Content, Structure, and Context was first defined by the Pittsburgh Project in the early 1990s which focused on helping archivists deal with Information Technology and Electronic Records. It's just one of the many ways to describe a record or an object.</p> <ul style="list-style-type: none"> • Content: Informational substance of the object. (What it says) • Structure: Technical characteristics of the objects (e.g.; presentation, appearance, display). (How the record looks) • Context: Information outside the object which provides illumination or understanding about it, or assigns meaning to it. (What it is about) <p><input checked="" type="checkbox"/> Illustration: \$20 bill example [hold one up]</p> <ul style="list-style-type: none"> • Content: basic information about the bill - \$20, serial number, image (2 Jacksons), statement about legal tender. informational substance of records • Structure: what says that this object is authentic - hologram/ghost image, hidden strip, color of the ink, feel of the paper, etc. The structure assures us that it is valid, we doubt it it's validity if any of these components is missing. • Context: information outside the object. foreign currency market – how related to other currency, what's it worth <p>Content, Context, and Structure are the necessary components to help us understand what an object is and what it's worth</p>
<p>Page-5 Defining information objects</p> <p><u>Pittsburgh Project</u> <u>Order of Values</u> <u>Information Technology</u></p>	<p>Page-5 Take those three components and compare it to some other ways to define digital objects.</p> <p>In the first column we have the scheme we just talked about.</p>

<u>Definitio n</u>		<u>ogy Architec ture</u>	
Content	Data	Data	<p>Let's move to the second column and read down. Many people break down information objects in terms of the value they represent to an organization.</p> <ul style="list-style-type: none"> • one may say that it's data, information or knowledge. • look at this column in terms of our \$20 bill again. <ul style="list-style-type: none"> ◦ Data here is the lowest common denominator. \$20 is the equivalent of 2000 pennies, but if you wanted to purchase something for \$14.97, would you be welcome anywhere if you pulled out a bag of pennies and said, "Just wait, I have exact change!" Data is accurate, but it's not necessarily useful. It's the least functional value. ◦ Information – here data is structured for more functionality; it's data presented in a practical format. A \$20 bill is more useful than 2000 pennies. It's in a specific format that's designed to be more easy to use. ◦ Knowledge – is data available for a wide variety of uses. Think of \$20 in the bank. You have different ways to access it through a check or debit card. You can automatically withdraw money to pay bills. It can earn interest, or the bank can use it to loan to other people. This is the level where you get most value. <p>Why isn't this enough? Let's look at the third column.</p> <ul style="list-style-type: none"> • To use technology, all of these components and values have to be captured in a specific information technology architecture, a configuration of hardware and software that allows us to use computers to manage our information. • And an architecture has three components: <ul style="list-style-type: none"> ◦ data – the actual stored bits and bytes ◦ format – data in a particular format (e.g., Word file, PDF, TIFF image) (how it looks to us) ◦ application – a program that takes a particular format and puts the data to use, gives it some functionality. • The architecture traditionally represents a limitation. • Applications are subject to rapid obsolescence. And, they often don't do all the things we want them to do <ul style="list-style-type: none"> ◦ we try to take data configured for one application and make it work in another. We lose a lot in the process usually. Because applications and their associated formats are usually proprietary something is almost always lost in the process. Very often, we're lucky to preserve just the data, let alone structure and functionality, given our limitations.
Structure	Information	Format	
Context	Knowledge	Application	
			<ul style="list-style-type: none"> • The \$64,000 question for archivists and for anyone

	<p>trying to preserve an investment in information: How do we preserve the value of our asset, our knowledge, over time, when the tools that help us realize that value (the hardware and software) are so unreliable?</p> <p>Think of the first column as the conceptual framework for your records, the second as your business needs, and the third as your practical IT structure that you have to work within.</p> <p>We have to capture everything that's required to meet to meet our business needs -- all the values and structure of an object or a record that will allow us to manage and access it for as long as we need to. This information is metadata, which is important in and of itself. Taking the extra step to standardization, however, will allow us to consistently and efficiently use, re-use, and share the object or record in order to get the best return possible.</p> <p>That's where metadata and XML come in.</p> <ul style="list-style-type: none"> • Metadata and XML are means of describing and capturing content, structure and context, of preserving data as knowledge, more importantly as executable knowledge, knowledge we can use in computer applications. • Metadata and XML address these three things and allow us to do what we want, preserve digital objects and electronic records.
<p>Page-6 What do you think metadata is? (<i>exercise</i>)</p>	<p>Page-6 Exercise: What do you think metadata is?</p> <p>A brainstorming session,</p> <ul style="list-style-type: none"> • Feel free to throw out ideas. • Don't expect an ideal definition or perfect understanding right now, because we're going to explain metadata to you throughout the day. • We'll revisit this list throughout the day to understand what metadata actually is and is capable of and how that fits with the preconceptions we identify here. <p><input checked="" type="checkbox"/> <i>Note to instructor: Use a flip chart to write down responses. Post the sheets so that you can refer back to them throughout the day.</i></p>
<p>Page-7 Different people and professions have different definitions of metadata</p> <ul style="list-style-type: none"> ○ data about data ○ information about information ○ data about objects ○ descriptive information which facilitates management of, and access to, other information ○ evaluation tool 	<p>Page-7 Different people/professions have different definitions of metadata, but they all boil down to:</p> <ul style="list-style-type: none"> ○ data about data ○ information about information ○ data about objects ○ descriptive information which facilitates management of, and access to, other information ○ evaluation tool that assists us in judging authenticity, reliability, and suitability - there is an enormous amount of information out there, how do we judge

	<p>which is more accurate or reliable? (What is the information we want?)</p> <p>Want to stress that metadata is not something that you can go out and buy, like a software package. It's descriptive information that you assemble using what schemes and tools you find appropriate.</p> <ul style="list-style-type: none"> o Think of a card catalog in the paper world. For a book, a card will tell you such items as author, title, publisher, publication date, shelf location, and so on. That's metadata for the book. If you're a librarian, it's up to you to collect the information (metadata) you need, create the actual card, file it, and maintain it.
<p>Page-8</p> <p>Different people and professions use metadata to fulfill different functions:</p> <p>Description: what is in the object, what the object is about</p> <p>Discovery: the location of the object</p> <p>Evaluation: the value of the object, is this the object I want to use</p> <p>Management: control of the access, storage, preservation, and disposal of an object</p>	<p>Page-8</p> <p>Metadata helps us with several general functions.</p> <p>Go through the list on the slide.</p>
<p>Page-9</p> <p>Why is metadata useful?</p> <p>Everyone needs metadata to help manage and use digital objects. Collaboration with partners and stakeholders is crucial to ensure that everyone's requirements are met and that efforts are coordinated.</p> <p>Metadata helps with:</p> <ul style="list-style-type: none"> • Legal discovery and admissibility issues • Data access requirements • Data management tasks such as: <ul style="list-style-type: none"> o knowing who created, modified, and accessed a file over time (reliability) o determining ownership o finding files o version control o tracking hardware and software requirements o planning for migration and conversion o implementing retention schedules 	<p>Page-9</p> <p>Our partners and stakeholders have their own needs and uses of metadata that must be considered. We live in an increasingly collaborative world, of formal and ad-hoc collaborations.</p> <p>If we each create our own individual types of metadata to help us use and manage our own material, that's fine, as long as no one else needs to use it.</p> <ul style="list-style-type: none"> • But, if we're really interested in data sharing and access, that has implications. • Then, we need to define our use of metadata or else no one else will be able to use it. • We need to agree on and apply standards consistently, in coordination with our partners. <p>Only when you consider an enterprise outlook, when you have standards across the organization, with your partners, can you really start meeting your management functions effectively.</p> <p>For instance, legal mandates require that we use metadata. Our legal mandates haven't changed just because we now keep records in electronic format. We still have to be able to find records and produce them when needed. Metadata will help us in this task.</p> <ul style="list-style-type: none"> • Legal admissibility: Electronic records and e-mail are increasingly recognized as evidence in court and admitted without question. Metadata about current versions, official copies, meeting records retention

schedules, etc. helps us to respond to discovery requests and assists us in establishing authenticity and accountability.

- Data privacy – There are often access requirements. In government, we cannot share certain data with certain parties. How do we block out part of an object (confidential) and share the rest (public)? Metadata can let us know who can see what and under what circumstances.
- Data sharing – Sometimes, in government, we're legally required to share data, for example between agencies. And now in the health care sector, there are HIPAA (Health Insurance Portability and Accountability Act) regulations requiring security components and access documentation. Agencies need to share information, but the consumer or citizen expects that confidential and private material will stay that way. Metadata can tell you what information has been shared and when, and what the circumstances were.

Metadata can help us address a wide range of new questions and issues are raised by information technology, such as:

- Who created the file and who has accessed it? Is the file reliable?
 - Who owns the file?
 - Where is my file?
 - What version of the file is this? Is it the most current file? Is it the official file?
 - Which files are duplicates?
 - Where are the backup files?
 - What storage media are the files saved on? What software/hardware was used to save these digital objects?
 - When do I need to migrate or convert my files?
 - What is the retention period for my records?
- ...and many more

<p>Page-10 Primary and secondary uses of data requires metadata</p> <p>Primary use: Why you create or use data.</p> <p>Secondary use: When anyone else wants to use the data.</p> <p>Metadata makes re-usage possible. Metadata standards allow for more consistent and efficient description, discovery, evaluation, and management.</p>	<p>Page-10 Data, digital objects, and electronic records have primary reasons for existing, each also has secondary uses.</p> <ul style="list-style-type: none">• This is why you may see some overlap in functions of the different types of metadata, and why some of the various elements of the metadata standards can be the same. <p>A primary function of the data or a record is your justification for creating that data.</p> <ul style="list-style-type: none">• It's primary/principal purpose comes from your immediate business need for the data, and that's where you want to get your first return on your investment. <p>Then there are the other things you or someone else are able to do with the data. These become the secondary purposes for the data. For instance, a record that was created to document a shipment, may be taken into a data warehouse and used to track inventory trends.</p> <ul style="list-style-type: none">• As we talked about earlier, in order for there to be secondary uses of data, there needs to be metadata, the foundation of a common understanding of what the data is and how it fits into a larger context.• Metadata makes re-usage possible.• Metadata standards allow for more consistent description, discovery, evaluation, and management of data, and ease re-use and re-purposing of data.
<p>Page-11 Different metadata standards serve different functions</p> <ul style="list-style-type: none">• Data Modeling metadata: a graphic representation of a process or system (metadata). Data models graphically capture and record business decisions, facilitate planning, and offer a means of understanding information relationships, structures, and processes. Models range from conceptual to physical (What is actually needed to implement the system).• Systems management metadata: metadata for structured data like that in a database or data warehouse.• Recordkeeping metadata: information that facilitates both management of, and access to, records.• Access metadata: information that facilitates the search for, access to, and use of digital objects.• Preservation metadata: metadata used for carrying out, documenting, and evaluating the processes that support the long-term retention and	<p>Page-11 As we mentioned earlier, metadata is a tool that assists us with different functions.</p> <ul style="list-style-type: none">• For example, systems management (data modeling and data administration), records management, access, and preservation.• We will talk about four of these today. <p>As we talk about functions and standards</p> <ul style="list-style-type: none">• You will notice that there are some points of commonality between these. That's because there is a basic core of information that you need for all digital objects. <p>While enterprise-wide standards are the best scenario, you may find that your partners are using other ones specifically tailored to their functions. This does not necessarily doom your coordination efforts, but crosswalking may be possible.</p> <ul style="list-style-type: none">• Crosswalking, or mapping, is a generic term for the process of moving between standards that have some point of commonality.• Crosswalks provide the ability to create and maintain one set of metadata and map that metadata to any number of related metadata standards.• Crosswalks allow metadata created by one community to be used by another group that employs a different metadata standards.

<p>accessibility of digital content.</p> <ul style="list-style-type: none">• GIS (Geographic Information System) metadata: combines aspects of data administration, recordkeeping, access, and preservation functions with application to geospatial data. <p>Standards have some points of commonality because there is a basic core of information that is needed for all digital objects. There are also points of difference, since each was created to support a particular function.</p> <p>Crosswalking, or mapping, allows you to move between different metadata standards with points of commonality (see Appendix A for more information).</p>	<ul style="list-style-type: none">• Crosswalks are a communication tool for sharing metadata and data.
---	--

<p>Page-12</p> <p>What is systems management metadata?</p> <ul style="list-style-type: none">○ necessary for day-to-day system functions○ examples include field size, allowable values○ associated with data administration, databases, data warehouses, etc.○ Users include systems analysts, data administrators, business analysts, software developers, planners, and auditors	<p>Page-12</p> <p>Let's start by looking at system operations, which would fall under the function of management, although on the IT side of things.</p> <p>Systems management metadata is necessary for day-to-day system functions and functions of databases, data warehouses, etc.</p> <ul style="list-style-type: none">● information on the rules for your databases, not necessarily how to set up your database.● defines the structure of database elements. A database element is a combination of characters that is one discrete piece of information, like a city name or area code● examples include field size, and allowable values● it's associated with data administration, databases, data warehouses, etc.● used by systems analysts, data administrators, business analysts, software developers, planners, and auditors. <p>Systems management metadata is important because we need to understand what data has been created, we need to understand how it's been structured, and we need to document the systems we are using.</p> <p>As mentioned, there are different users of systems management metadata</p> <ul style="list-style-type: none">● And, we also said different people use metadata in different ways.● So how are these different users using this type of metadata?● To an IT person like a data administrator, metadata is information like field size, allowable values, and file structure. In a data warehouse environment, they will want information about the data's source system, and information about whether the data was cleansed and/or transformed when it was loaded into the warehouse.<ul style="list-style-type: none">○ Metadata is the information that <i>supports</i> the operation of the system.● A business analyst generally takes a broader view<ul style="list-style-type: none">○ They want information that goes beyond the operation to understanding the "why's" and "how's". For instance, why do we collect information on individuals and how is it represented in the system? How current is the information and where did it come from? How do the system and database function? They are interested in related information like data models and business rules, and may make use of decision-support systems like data warehouses.<ul style="list-style-type: none">○ The analyst wants metadata in order to <i>understand</i> the systems and the business.
<p>Page-13</p>	<p>Page-13</p>

<p>Systems management metadata (continued)</p> <ul style="list-style-type: none">➤ <i>Specification and Standardization of Data Elements</i>. ISO/IEC 11179, Final draft international standard.➤ <i>ISO/IEC 11179: Metadata Registries</i> (2001 draft revisions) <p>Part 1: Framework for the Specification and Standardization of Data Elements</p> <p>Part 2: Classification for Data Elements</p> <p>Part 3: Basic Attributes of Data Elements (Registry Metamodel)</p> <p>Part 4: Rules and Guidelines for the Formulation of Data Definitions</p> <p>Part 5: Naming and Identification Principles for Data Elements</p> <p>Part 6: Registration of Data Elements</p> <p>Purpose of standard: “to give concrete guidance on the formulation and maintenance of discrete data element descriptions and semantic content (metadata) that shall be used to formulate data elements in a consistent, standard manner. It also provides guidance for establishing a data element registry.”</p>	<p>There is a standard for systems management metadata: ISO/IEC 11179 <i>Specification and Standardization of Data Elements</i>.</p> <ul style="list-style-type: none">• ISO: International Organization for Standardization. It creates, maintains, and distributes hundreds of different standards in use worldwide.• IEC: International Electrotechnical Commission <p>There are 6 parts to the standard</p> <p>It's designed to define database structures and the accompanying data dictionary structures. It defines the architecture for how the fields, tables, and relations work.</p> <ul style="list-style-type: none">• 11179 describes the standardizing and registering of data elements to make data understandable and shareable.• Purpose is to give concrete guidance on the formulation and maintenance of discrete data element descriptions and semantic content (metadata) that shall be used to formulate data elements in a consistent, standard manner• Geared toward users and managers of data, systems analysts and data stewards <p>The standard is too complex to cover in detail today. The links in your course books take you to drafts of each of the Parts. In order to get the most recent edition you'll have to order it from the ISO, and since they require you to purchase each part separately, it could run you a few hundred dollars</p>
---	---

Page-14
Systems management metadata (continued)

Useful for data warehouses

What is a data warehouse?

“Data warehouses are computer based information systems that are home for "secondhand" data that originated from either other applications and/or from external systems or sources. Warehouses optimize database query and reporting tools because of their ability to analyze data, often from disparate databases and in interesting ways. They are a way for managers and decision makers to extract information quickly and easily in order to answer questions about their business. In other words, data warehouses are read-only, integrated databases designed to answer comparative and "what if" questions. Unlike operational databases that are set up to handle transactions and that are kept up to date as of the last transaction, data warehouses are analytical, subject-oriented and are structured to aggregate transactions as a snapshot in time.”

- This metadata helps you evaluate data and answer the following questions:

What’s the source of the data?

Has the data recently been cleansed, or transformed?

Is this data appropriate for my needs?

Page-14
Examples of systems management metadata can be found in data warehouses.

What’s a data warehouse?

- Data warehouses are computer-based information systems that are home for "secondhand" data that originated from either other applications and/or from external systems or sources.
- Purpose of a data warehouse is to collect data from different agencies or departments for the purpose of analysis of that data, and creating reports from that data.
- A data warehouse is both a repository and an application that allows for standard queries and research. For example, your might want to compare travel expenses from different agencies. And perhaps you want to do that over time.
- Data warehouses store historical data so that you can track information over time, not just work with current data.
- Collecting and comparing information from different sources means you have to understand what’s there and what potential it has (calls for evaluation), as well as how old it is, and what’s changed (collecting information in the warehouse usually means extracting only selected information from the source and transforming it to compatible formats). You are evaluating the potential or appropriateness of the information for your specific purpose.

Systems management metadata allows you to analyze whether or not the information suits your needs.

Page-15-16

NAME	DESCRIPTION	LENGTH	TABLES
ACCT_PAYABLE_PHONE_NBR	Phone number for the accounts payable contact	15	ACCT_PAYABLE_PHONE
ACCT_PAYABLE_PHONE	Phone number for the accounts payable contact	15	ACCT_PAYABLE_PHONE
ACCT_PAYABLE_PHONE_NBR	Phone number for the accounts payable contact	15	ACCT_PAYABLE_PHONE

Page 15-16
Two pages of examples

- Note to instructor: You should use local examples of systems management metadata if possible.*

These two pages are examples of metadata from the Minnesota Department of Finance’s data warehouse.

Example no. 1: What you see here is a list of the elements in the data warehouse, definitions for each, how many characters each is, what type it is (date, character string, etc.), and what tables each appears in.

- For instance, we see that there is an element called “ACCT_Payable_Phone_NBR”. The description says it is the phone number for the accounts payable contact. The element can be up to 15 characters long, and can

TABLE NAME	TABLE DESCRIPTION	TABLE FREQUENCY	NEW LAST UPDATE DATE/TIME	EXTRACT PROCEDURE	LOAD PROCEDURE
T_CUSTOMER	Customer Information	Daily	2003-05-01 10:00:00	Customer	Customer
T_CUSTOMER_HISTORY	Customer History	Weekly	2003-05-01 10:00:00	Customer_History	Customer_History
T_PRODUCT	Product Information	Daily	2003-05-01 10:00:00	Product	Product
T_PRODUCT_HISTORY	Product History	Weekly	2003-05-01 10:00:00	Product_History	Product_History
T_ORDER	Order Information	Daily	2003-05-01 10:00:00	Order	Order
T_ORDER_HISTORY	Order History	Weekly	2003-05-01 10:00:00	Order_History	Order_History
T_INVENTORY	Inventory Information	Daily	2003-05-01 10:00:00	Inventory	Inventory
T_INVENTORY_HISTORY	Inventory History	Weekly	2003-05-01 10:00:00	Inventory_History	Inventory_History

contain numbers and symbols. It appears in the “Customer” and “T_Customer” tables.

Example No. 2:

- Along with the metadata we just looked at, this gives us a fairly complete picture of the information the warehouse contains.
- In this case, we see table names (and we know what’s contained in the tables from the previous example) and descriptions of each table.
- Two other very important pieces of metadata here – Update Frequency – this tells us how often information is brought in from the source systems; and Last Update Date/Time – this tells you exactly when the last update occurred. Combined, this tells you how current the information in any given table is and whether or not it is suitable for your needs.

Page-17
What is access metadata?

Access metadata is metadata which facilitates your search for, access to, and use of records and information. It makes the process of finding objects faster and more precise.

Users include web page creators, search engines, archivists, records managers, librarians, researchers, and record creators.

Page-17
Access metadata

- metadata which facilitates your search for, access to, and use of digital objects and information. It makes the process of finding your records faster and more precise.
- obviously important to archivists and records managers because they’re in the business of providing people access to objects and information.
- used for resource discovery, for example finding resources on the World Wide Web
- used by web page creators, search engines, archivists, records managers, librarians, researchers, and records creators.

Page-17
Dublin Core Metadata Standard

- ISO/NISO Standard: Dublin Core Metadata Element Set (NISO Z39.85-2001, approved July 2001) (ISO 15836, approved February 2003)
- Used for resource discovery for networked resources (e.g., web pages, PDFs)
- Audiences: Web users, page owners, page creators, search engine developers
- Goals of Dublin Core:
 - o Simplicity of creation and maintenance
 - o Commonly understood semantics
 - o International Scope
 - o Extensibility
 - o Flexibility with respect to

Page-18
Dublin Core Metadata Element Set.

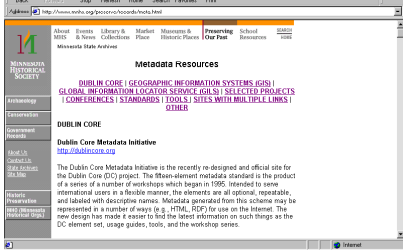
- named after Dublin, Ohio, not Dublin, Ireland, where the development work started several years ago.
- NISO (National Information Standards Organization) standard in October 2001, the Dublin Core Metadata Element Set (Z39.85-2001). ISO standard in February 2003, The Dublin Core Metadata Element Set (15836), so it has international recognition and support.
- The Dublin Core web site serves as a home base for the standard and as a forum for Dublin core users and on-going updates.

Dublin Core Metadata is used for resource discovery primarily for networked resources (e.g. web pages, PDFs, or any type of resource)

- Users include: Web users, page and resource creators, search engine developers
- The development of the Dublin Core was driven by the proliferation of electronic resources and the inability of the library profession to catalog all these resources. The fifteen elements were established through consensus by an international, cross-disciplinary group of librarians, computer science professionals,

<p>implementation</p>	<p>representatives from the museum community, and others.</p> <p>The goal was</p> <ul style="list-style-type: none"> • To define a few elements and some simple rules that could be applied by noncatalogers, by the resource authors themselves. • Developed to be simple to create and maintain, and to be concise • Designed to feature semantic interoperability with other metadata standards and build on international consensus. <p>The Dublin Core</p> <ul style="list-style-type: none"> • Recommends the use of controlled values for fields where they are appropriate (i.e., controlled vocabularies for the subject field), but is not required • Meant to be extensible as needed. It is the core set or the lowest, common denominator of information needed about a resource (extend as needed within your community) • Flexible, Dublin Core offers guidelines to what should be captured, but not how to capture it or store it
<p>Page-19 Dublin Core Metadata Standard (continued)</p> <p><u>3 categories of elements:</u></p> <p>Content:</p> <p>Title: A name given to the resource. Subject: The topic of the content of the resource. Description: An account of the content of the resource. Type: The nature or genre of the content of the resource. Source: A reference to a resource from which the present resource is derived. Relation: A reference to a related resource. Coverage: The extent or scope of the content of the resource.</p> <p>Intellectual property:</p> <p>Creator: An entity primarily responsible for making the content of the resource. Publisher: An entity responsible for making the resource available. Contributor: An entity responsible for making contributions to the content of the resource. Rights: Information about rights held in and over the resource</p>	<p>Page-19 Dublin Core features</p> <ul style="list-style-type: none"> • Elements which are optional and repeatable. <p>There are 3 categories of elements:</p> <ul style="list-style-type: none"> • Content: title, subject, description, type, source, relation, coverage • Intellectual property: creator, publisher, contributor, rights • Instantiation (information about version, time and identification): date, format, identifier, language <ul style="list-style-type: none"> • Title: A name given to the resource. Typically, a Title will be a name by which the resource is formally known. • Creator: An entity primarily responsible for making the content of the resource. <ul style="list-style-type: none"> ◦ Examples of a Creator include a person, an organization, or a service. • Subject: The topic of the content of the resource. Typically, a Subject will be expressed as keywords, key phrases, or classification codes that describe a topic of the resource. Recommended best practice is to select a value from a controlled vocabulary or formal classification scheme. • Description: An account of the content of the resource. Description may include but is not limited to: an abstract, table of contents, or a free-text account of the content. • Publisher: An entity responsible for making the resource available.

<p>and over the resource.</p> <p>Instantiation (version):</p> <p>Date: A date associated with an event in the life cycle of the resource.</p> <p>Format: The physical or digital manifestation of the resource.</p> <p>Identifier: An unambiguous reference to the resource within a given context.</p> <p>Language: A language of the intellectual content of the resource.</p>	<p>Examples of a Publisher include a person, an organization, or a service.</p> <ul style="list-style-type: none">• Contributor: An entity responsible for making contributions to the content of the resource. Examples of a Contributor include a person, an organization, or a service.• Date: A date associated with an event in the life cycle of the resource. Typically, Date will be associated with the creation or availability of the resource.• Type: The nature or genre of the content of the resource. Examples include Text, Image, Sound, Data.• Format: The physical or digital manifestation of the resource. Typically, Format may include the media-type or dimensions of the resource. Format may be used to determine the software, hardware or other equipment needed to display or operate the resource, i.e., HTML, ASCII, JPEG.• Identifier: An unambiguous reference to the resource within a given context. Recommended best practice is to identify the resource by means of a string or number conforming to a formal identification system. Example formal identification systems include the Uniform Resource Identifier (URI) (including the Uniform Resource Locator (URL)), and the International Standard Book Number (ISBN).• Source: A Reference to a resource from which the present resource is derived. The present resource may be derived from the Source resource in whole or in part. Recommended best practice is to reference the resource by means of a string or number conforming to a formal identification system.• Language: The language of the intellectual content of the resource.• Relation: A reference to a related resource. Recommended best practice is to reference the resource by means of a string or number conforming to a formal identification system.• Coverage: The extent or scope of the content of the resource. Coverage will typically include spatial location (a place name or geographic coordinates), temporal period (a period label, date, or date range) or jurisdiction (such as a named administrative entity).• Rights: Information about rights held in and over the resource. Typically, a Rights element will contain a rights management statement for the resource, or reference a service providing such information. Rights information often encompasses Intellectual Property Rights (IPR), Copyright, and various
--	---

	<p>Property Rights. If the Rights element is absent, no assumptions can be made about the status of these and other rights with respect to the resource.</p> <p>You can see that it's called the Dublin Core set because this metadata is very basic information about any type of digital object.</p>
<p>Page-20 Example web page</p> 	<p>Page-20 This first slide is a screen shot of a Minnesota State Archives web page. The page is entitled Metadata Resources.</p> <p><input checked="" type="checkbox"/> <i>Note to Instructor: Consider using a local example.</i></p>
<p>Page-21 Example web page metadata</p> <pre> <html> <head> <title>Metadata Resources</title> <meta name="resource-type" content="document"> <meta name="revisit-after" content="30 days"> <!-- Start Dublin Core - Do Not Modify Tags in This Block --> <!-- Dublin Core Meta Tags generated by TagGen - The Meta Tag Management System --> <meta name="DC.Title" content="Metadata Resources"> <meta name="DC.Description" content="This site provides an annotated list of on-line resources relating to metadata."> <meta name="DC.Creator.CorporateName" scheme="AACR2" content="Minnesota State Archives"> <meta name="DC.Publisher.CorporateName" scheme="AACR2" content="Minnesota State Archives"> <meta name="DC.Contributor.PageDesigner" scheme="AACR2" content="Goertz, Angela"> <meta name="DC.Date.Creation" scheme="ISO 8601" content="1998-12- 11"> <meta name="DC.Date.Modified" </pre>	<p>Page-21 Dublin Core metadata for the web page</p> <p><input checked="" type="checkbox"/> <i>Note to Instructor: Consider using a local example, which should match the web page on the previous slide.</i></p> <p>Go through the elements on the slide.</p> <p>Evaluation: metadata helps us evaluate its value, worthiness to be read by letting us know who did it, how recently its been updated, who published it Description: key words describe the content Discovery: helps us locate the page using search engines</p> <p>This metadata is created using a product called TagGen. Brings up issue of what metadata is contained in a record? What metadata do you capture when you print something out? In this case it's additional metadata. Do you want to capture both?</p> <p>Also brings up the issue of automation. With TagGen, you can set defaults for many of the elements, so that you're not having to create all of the metadata manually. As we go through these standards, keep in mind that much of the metadata can be automatically captured by the system or by software, so it's not as daunting as it might first appear.</p>

<pre> scheme="ISO 8601" content="2003-01-02"> <meta name="DC.Type" content="Text"> <meta name="DC.Format" scheme="HTML" content="text/html"> <meta name="DC.Rights" content="../mhsuse.html"> <meta name="DC.Language" scheme="ISO639-1" content="en"> <LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements"> <!-- End Dublin Core - Do Not Modify This Block --> </pre>	
<p>Page-22</p> <p>Bridges: Minnesota's Gateway to Environmental Information</p> <ul style="list-style-type: none"> • Example of government implementation • Agencies tag own web pages using TagGen • Feeds into state search engine, powered by Inktomi, which has been optimized for Dublin Core <p>Reasons for adopting Dublin Core in Minnesota</p> <ul style="list-style-type: none"> ◦ Dublin Core is easy to create and provides uncomplicated descriptions. ◦ Dublin Core is simple to index and use for describing a resource's location, form, etc. ◦ Dublin Core allows for the use of controlled vocabularies that enable greater searching precision than full-text searches. ◦ Dublin Core is a standard agreed upon by the World Wide Web Consortium (W3C). ◦ Dublin Core offers extensibility and interoperability with other standards. ◦ Dublin Core enhances the quality of resource management. 	<p>Page-22</p> <p>We've talked in some detail now about an access metadata standard, but what we're going to look at now is an example of practical implementation.</p> <p><input checked="" type="checkbox"/> <i>Note to instructor: Consider using a local example.</i></p> <p>The Bridges Project</p> <ul style="list-style-type: none"> • Multi-agency collaboration in Minnesota aimed at improving public access to environmental and natural resources data and information. • Bridges is a metaphor for connecting and spanning Minnesota state agencies • Project started in 1998 and was managed by the Minnesota Department of Natural Resources. • Goal was to add Dublin Core metadata to electronic data and information resources including web pages, PDF documents, tabular data and geographic data • Designed to be intuitive and easy to use for both specialists and non-specialists • Included a cost-benefit analyses <p>Reasons for adopting Dublin Core in Minnesota (from the business case)</p> <ul style="list-style-type: none"> • Dublin Core is easy to create and provides uncomplicated descriptions. • Dublin Core is simple to index and use for describing a resource's location, form, etc. • Dublin Core offers controlled vocabularies that enable greater searching precision than full text searches. • Dublin Core is a standard agreed upon by the World Wide Web Consortium). • Dublin Core offers extensibility and interoperability with other standards. • Dublin Core enhances the quality of resource management. <p>Project staff developed metadata cataloging guidelines, a training manual, and other searching aids.</p> <p>Result: Agencies tag their own web pages. There is a</p>

	<p>strong training component covering what metadata is and how to use the tool. Software product called TagGen (Hiawatha Island Software), which is available free of charge to state agencies in Minnesota.</p> <p>Why would they want to go to the trouble of doing this?</p> <ul style="list-style-type: none"> • The metadata from the Bridges project meshes with the state search engine, which has been optimized to look for Dublin Core metadata tags, to search the actual web page for that metadata. • The search engine then gives higher ranking to those pages in result lists. • Offers an immediate benefit to agencies that participate because attention is drawn to their pages first. • And it's not just customized search engines which utilize metadata. Other general web search engines rely on tags "title" and "description" when indexing and returning search results.
<p>Page-23 Bridges: Minnesota's Gateway to Environmental Information (continued)</p> <p>Bridges is also an example of another key metadata concept: controlled vocabularies</p> <p>Controlled vocabulary: a limited set of consistently used and carefully defined terms.</p> <p>Controlled vocabularies may take many forms, including:</p> <ul style="list-style-type: none"> • Taxonomies • Thesaurus (e.g., the Minnesota Legislative Indexing Vocabulary) • Naming conventions 	<p>Page-23 The Bridges project is also a good example of another key metadata concept: controlled vocabularies.</p> <p><input checked="" type="checkbox"/> <i>Note to instructor: Consider using local examples.</i></p> <p>Content data for some elements may be selected from a controlled vocabulary</p> <ul style="list-style-type: none"> • A limited set of consistently used and carefully defined terms. • Purpose is to bring together like concepts, synonyms, and related terminology. • Requires you to build a consensus among catalogers about which terms to use. • Can dramatically improve search results because computers are good at matching words character by character, but weak at understanding the way people refer to one concept using different words, i.e., synonyms. • Without basic terminology control, inconsistent or incorrect metadata can profoundly degrade the quality of search results. <ul style="list-style-type: none"> ◦ For example, without a controlled vocabulary, "candy" and "sweet" might be used to refer to the same concept. Or, does "pop" refer to a soft drink (slang term/ pop/soda/coke) or my father? • Controlled vocabularies may also reduce the likelihood of spelling errors when recording metadata. <p>Controlled vocabularies may take many forms:</p> <ul style="list-style-type: none"> • Taxonomies: Classification of major subject areas. Think of how Yellow pages are structured or subject trees • Thesaurus: Assisting catalogers in selecting appropriate search terms the searcher may not have envisioned. Defining allowable terms for specific metadata fields (e.g., subject).

	<ul style="list-style-type: none"> o In Minnesota we use the Minnesota Legislative Indexing Vocabulary, which is derived from the Library of Congress Legislative Indexing Vocabulary. • Naming conventions: Brings consistency in recording names of people and organizations. <ul style="list-style-type: none"> o For instance, rather than letting people use the shorthand “Admin” under Publisher, a naming convention could stipulate “Department of Administration.” <p>Controlled vocabularies are also used in other metadata standards, although it’s associated primarily with web metadata.</p> <ul style="list-style-type: none"> • Systems management metadata: 11179 Metadata Registry • Minnesota Recordkeeping Metadata Standard which we’ll talk about next, also advises the use of them.
<p>Page-24 What is recordkeeping metadata?</p> <p>The act or process of creating, managing, and disposing of records.</p> <p>Recordkeeping metadata is information that facilitates that process.</p> <p>Users include archivists and records managers, recordkeeping staff, IT staff, information users and creators, and developers.</p> <p>Used for records and information systems including: file types and record types, word processing documents, e-mail, databases, data warehouses, web pages, spatial data, geographic files, microform, videotapes, audio tapes, correspondence, maps, and many, many more.</p>	<p>Page-24 Let’s move onto a second function which metadata supports – recordkeeping.</p> <p>A simple definition of recordkeeping, or records management, is the act or process of creating, managing, and disposing of records.</p> <ul style="list-style-type: none"> • Recordkeeping metadata is, therefore, information that facilitates that process. <p>Users of recordkeeping metadata include: archivists and records managers, IT staff, information users and creators, and developers</p> <p>Used for records and information systems including: word processing documents, e-mail, databases, data warehouses, web pages, spatial data, geographic files, microform, videotapes, audio tapes, correspondence, maps, and many, many more.</p> <p>Everyone can use recordkeeping metadata, for example, to track records in a system, to know who’s using it when, the date it was last changed, who has authority to make changes or look at a record.</p> <p>Recordkeeping metadata helps ensure that the system is creating and maintaining trustworthy, authentic, and reliable records.</p>
<p>Page-25 Minnesota Recordkeeping Metadata Standard</p> <p>Minnesota Government Business Case for Metadata and Recordkeeping Metadata Guidelines</p> <p>Facilitate compliance with the</p>	<p>Page-25 <input checked="" type="checkbox"/> <i>Note to instructor: Consider using a local example.</i></p> <p>The <i>Minnesota Recordkeeping Metadata Standard</i>. May 2002, is based on the National Archives of Australia’s <i>Recordkeeping Metadata Standard for Commonwealth Agencies</i> (Version 1.0, issued in May 1999).</p> <ul style="list-style-type: none"> • Good example of a well-known standard. • Australian standard is concise and a lowest common

<p>Minnesota Government Data Practices Act (MGDPA). Facilitate accountability to citizens. Facilitate location and retrieval of records for increased proper public access, for use in a government information locator service, and for litigation, for business use, etc. Reduce costs by reducing redundancy, eliminating records kept beyond retention periods, and decreasing development costs within agency. Improve records management with respect to retention periods (short-term, permanent, archival, etc.), storage, preservation, and access. Reduce paperwork (decrease use of hard copies) by increasing agencies' confidence in locating and managing electronic records. Achieve greater consistency of information within and across agencies. Facilitate sharing (when appropriate and allowed by law) within and across agencies by knowing what information is available and what is not, and carrying out cross-agency queries. Reduce the number of ad-hoc, agency-specific, recordkeeping metadata schemes. Provide recordkeeping metadata standards and guidance for consultants and vendors to allow easy reference, consistency, and agency projects to build on what others have done. Provide pointers to other related metadata (data dictionaries, etc. – metadata for metadata). Increase the reliability of recordkeeping metadata; reduce errors.</p>	<p>denominator standard, it's practical and in use right now</p> <ul style="list-style-type: none"> • Based on Dublin Core, which is another standard we cover in the next unit. • In Australia, this standard -feeds into their Government Information Locators System, and is a component of their overall comprehensive records management program, <p>The Minnesota Recordkeeping Metadata Standard is an adaption of the Australian standard for use by Minnesota government entities, and it will fit into the Minnesota statewide enterprise architecture.</p> <p>The Minnesota project began with the Recordkeeping Metadata Study Committee, which met from September-December 2000 to see if the standard was suitable for adaptation and adoption in Minnesota. The Committee recommended to the state's Information Policy Council that the standard was suitable. The group morphed into the Recordkeeping Metadata Development Committee, which wrapped up its work on adapting the standard in April 2002.</p> <ul style="list-style-type: none"> • Members of the committees from state agencies, and local government. <ul style="list-style-type: none"> ◦ made up specifically of people from different agencies who could bring different skills to the table, for example an IT or a records management background. ◦ This helped get wide buy-in for the standards adoption. <p>Committee established a business case:</p> <ul style="list-style-type: none"> • helped to justify the agencies' interest in creating metadata, because this is a system where the records creators are creating the metadata. • had to be a benefit to the agencies or there would be no reason to create a standard in the first place, let alone create the metadata <p>Minnesota Government Business Case for Metadata and Recordkeeping Metadata Guidelines</p> <p>Go over the list on the slide.</p> <p>Business case</p> <ul style="list-style-type: none"> • Fairly generic, so it's really a good business case for the adoption of recordkeeping metadata in general. • Throughout development process, kept referring to it for guidance and used it to justify the expenditure of our time and effort. • For any sort of metadata project, it's very useful to do this sort of analysis.
<p>Page-26-33 <u>Mandatory Elements</u> <u>Optional</u></p>	<p>Page-26-33 Quickly run through the twenty elements in this standard.</p>

	<u>Elements</u>	
Agent	Description	<ul style="list-style-type: none"> • Some of these are very similar to the Dublin Core elements we have already seen, but there are additional ones that are particular to the function of recordkeeping. • Standard is intended to work with just electronic records or with electronic-paper hybrid systems. • Some are mandatory to use and others optional. • Most have sub-elements that further refine the information, which may themselves be mandatory or optional (sub-elements listed in course book). • This is because the standard is not meant to be burdensome <ul style="list-style-type: none"> ◦ Organizations can use just the basic core set if they have minimal requirements ◦ Can adopt any of the optional elements as well if they have a business need to collect that extra information. ◦ And of course, the agency's business needs determine which records warrant this collection of metadata in the first place. <p>Like the Dublin Core, this standard is just a guideline</p> <ul style="list-style-type: none"> • It is completely architecture-independent and does not stipulate how it should be implemented • Suggestions offered in the comments section of each element. • Points out what might be automatically captured. <p>If you want to see the whole standard, it's available online.</p> <p>Start with the mandatory elements.</p> <ul style="list-style-type: none"> • Agent: An agency or organizational unit which is responsible for some action on or usage of a record. An individual who performs some action on a record, or who uses a record in some way. Example: creator, records manager, security officer Analogous to Creator, Publisher, and Contributor in Dublin Core • Rights Management: Policies, legislation, caveats, and/or classifications which govern or restrict access to or use of records. Example: Minnesota Government Data Practices Act Analogous to Dublin Core's Rights element. • Title: The name given to a record. Analogous to Dublin Core's Title element. • Subject: The subject or topic of a record which concisely and accurately describes the record's content. Analogous to Dublin Core's Subject element. • Date: The dates and times at which the fundamental recordkeeping actions of creation, transaction, and registration [into a recordkeeping system] occur. ISO standard for date/time recording. Analogous to Dublin Core's Date element.
Rights Management	Language	
Title	Relation	
Subject	Coverage	
Date	Function	
Aggregation Level	Type	
Record Identifier	Format	
Management History	Use History	
Location	Preservation History	
Disposal	Mandate	

- **Aggregation Level:** The level at which the record(s) is/are being described and controlled. The options are Record or Record Series.
- **Record Identifier:** A unique identifier for the record(s). Analogous to Dublin Core's Identifier element.
- **Management History:** The dates and descriptions of all records management actions performed on a record from its registration into a recordkeeping system to its disposal. This basically functions as an audit trail noting the event type, event date/time, and event description.
Example: disposed, location changed, published, scheduled, disposal on hold.
- **Location:** The current (physical or system) location of the record.
Example: recordkeeping system, records center
- **Disposal:** Information about policies and conditions which pertain to or control the authorized disposal of records, like retention schedule identifier, retention period, disposal action, and disposal due date.

The following metadata elements would be **optional** for use where the guideline is adopted:

- **Description** (strongly recommended): A description, in free text prose, of the content and/or purpose of the record.
Analogous to Dublin Core's Description element.
- **Language:** The language of the content of the record.
Analogous to Dublin Core's language element.
- **Relation:** A link between one record item and another, or between various aggregations of records. A link between a record and another information resource.
You'll notice that this standard does not have a metadata element for Version Number. That's because it was decided since this standard is being used to track records, once a record is changed, the new version is actually a new record. But you can use this element to tie all incarnations of a record together.
Analogous to Dublin Core's Relation element.
- **Coverage:** The jurisdictional, spatial, and/or temporal characteristics of the content of the record.
Example: county, calendar year, state, tribe
Analogous to Dublin Core's Coverage element.
- **Function:** The general or agency-specific business function(s) and activities which are documented by the record. Usually this should be chosen from a controlled vocabulary.
- **Type:** The recognized form a record takes.
Examples: letter, report.
- **Format:** Content Medium (e.g., audio, text, video), Data Format (e.g., ASCII, JPEG, Word), Storage Medium (e.g., hard disk, CDROM), and Extent (e.g., 14 MB).
Analogous to Dublin Core's Format element.

	<ul style="list-style-type: none">• Use History: The dates and descriptions of both legal and illegal attempts to access and use a record, from the time of its registration into a recordkeeping system until its disposal. Functions as an audit trail.• Preservation History: The dates and descriptions of all actions performed on a record after its registration into a recordkeeping system which ensure that the record remains readable and accessible for as long as it has value to the agency or department, and to the community at large. Example: backed-up, compressed, converted, microfilmed.• Mandate: A source of recordkeeping requirements [applicable laws]. A piece of legislation, formal directive, policy, standard, guideline, set of procedures, or community expectation which (explicitly or implicitly) imposes a requirement to create, keep, dispose of, or control access to and use of a record. Examples: legislation, regulations, internal business requirements. <p>Look at sample record on pages 30-33</p> <p>Go through the information on the slides.</p> <p>{This example includes all of the metadata elements and sub-elements, both mandatory and optional. It's probably rare that you would want to gather all of this information.}</p>
<p>Page-34 Department of Defense. <i>Design Criteria Standard for Electronic Records Management Software Applications</i>. DoD 5015.2STD, June 2002.</p> <p>“Provides implementing and procedural guidance on the management of records in the Department of Defense. This Standard sets forth mandatory baseline functional requirements for Records Management Application (RMA) software used by DoD Components in the implementation of their records management programs; defines required system interfaces and search criteria to be supported by the RMAs; and describes the minimum records management requirements that must be met, based on current National Archives and Records Administration (NARA) guidelines.”</p> <p>Records Management Application: “a software program that helps users file and manage electronic records. It provides</p>	<p>Page-34 The Department of Defense. <i>Design Criteria Standard for Electronic Records Management Software Applications</i>. DoD 5015.2STD, first issued in November 1997 and revised in June 2002.</p> <ul style="list-style-type: none">• A different kind of recordkeeping metadata• Moves from looking at elements to examining functions.• The DoD standard is a general overview.• Not so much a standard as it is a guideline for using metadata if you choose to implement a recordkeeping system using records management software. <p>It is specifically aimed at federal agencies, however, many other organizations are also using the standard as a guide when choosing software.</p> <p>What the standard basically does is provide a means of benchmarking and certifying Records Management Applications (RMA).</p> <ul style="list-style-type: none">• Records Management Application: “a software program that helps users file and manage electronic records.”• Records Management application software features:<ul style="list-style-type: none">◦ Centralized filing and storage of electronic records◦ Version control and security

some of the following functions:
centralized filing and storage of electronic records; version control; and search and retrieval."

As of January 2003, sixteen Records Management Application software packages are DoD 5015.2 compliant, including:

Tower Software's TRIM Context v5
Foremost Enterprise 2.5
Filenet Corporation's Filenet IDM Content Services 5.1.1

The standard gives guidelines for capturing record metadata such as unique record identifier, subject or title, media type, format, record dates, date filed, publication date, author or originator, and location.

- o Search and retrieval capabilities
- o Retention and disposal tools
- o Links to retention schedules
- o Disposal dates for disposition and purging from system
- Distinction between RMA software and document management software.
 - o This is key because more and more companies and organizations have one or the other.
 - o Document management applications don't have records management features/capabilities, but instead focus on workflow.
 - o Whereas Records Management Applications have both document management features and records management capabilities. Unfortunately it doesn't go both other way.

Prior to the revision in 2002, over 30 software packages were certified as compliant. After the revision, packages must be re-certified, and certification lasts for two years.

For the RMA to be certified it must meet minimums of performance in several areas, including:

- Managing records regardless of format
- Accommodating four-digit dates
- Supporting backwards compatibility with earlier product versions
- Implementing file plans
- Scheduling records
- Storing records
- Searching for and retrieving records
- Transferring records
- Destroying records
- Access control
- System audits

The standard also gives guidelines for record metadata that includes capturing:

- Unique record identifier
- Subject or title
- Media type
- Format
- Record dates
- Date filed
- Publication date
- Author or originator
- Location

All are metadata elements that are common and that we've seen in the Dublin Core and Minnesota standards.

- If an agency wants to work with the Department of Defense, their Records Management Applications have

	to meet the DoD standards, many software packages are now or are becoming DoD 5015.2 compliant.
--	---

<p>Page-35 What is preservation metadata?</p> <p>The key to the preservation of digital resources over the long-term is to maintain:</p> <p>Viability: requires that the archived digital object's bit stream is intact and readable from the digital media upon which it is stored.</p> <p>Renderability: refers to the translation of the bit stream into a form that can be viewed by human users, or processed by computers.</p> <p>Understandability: involves providing enough information such that the rendered content can be interpreted and understood by its intended users.</p> <p>Used for carrying out, documenting and evaluating the processes that support the long-term retention and accessibility of digital content.</p> <p>Often associated with digitization projects.</p>	<p>Page-35 Now let's talk about the function of preservation.</p> <ul style="list-style-type: none"> • Used for carrying out, documenting, and evaluating the processes that support the long-term retention and accessibility of digital content. You'll recall that the Minnesota Recordkeeping Metadata Standard accommodated preservation-related information. • Associated with digitization projects. <ul style="list-style-type: none"> ◦ Organizations often image materials to make them more widely accessible and to cut down on storage costs. ◦ You don't usually image materials that you only want to keep for a little while. ◦ It's an investment you want to keep long-term, which means gathering information, or metadata up-front that will help you search and use those records, as well as facilitate any media migrations or format conversions that might be necessary down the road. ◦ Anyone doing a digitization project wants to track what's been done to the object and it's version history, for example.
<p>Page-36 Minnesota Historical Society Metadata Capture for Digital Imaging</p> <p>Metadata information captured during the digital imaging process:</p> <p>Full-Size File Name, Date, File Format</p> <p>Thumbnail File Name, Department/Transcriber, Compression</p> <p>Off-Site File Name, Producer, Color</p> <p>Date Sent, Capture Device, Color Space</p> <p>Date Received, Capture Details, Color Management</p> <p>Unique Identifier, Change History, Color Bar/Grayscale Bar</p> <p>Negative Number, Source, Control Targets</p> <p>Creator Last Name, Encryption, Usage rights</p> <p>Creator First Name, Watermark, Notes</p> <p>Title/Object Description, Resolution</p> <ul style="list-style-type: none"> • A combination of elements from the 	<p>Page-36 <input checked="" type="checkbox"/> <i>Note to instructor: Consider using a local example.</i></p> <p>An imaging example. The Minnesota Historical Society has developed what it calls the Visual Resources Database, or VRDB.</p> <ul style="list-style-type: none"> • Online catalog of information about photographs, artwork, and posters from MHS' collection. • Over 180,000 entries in the database, of which about 60% have accompanying images. • Uses a combination of Dublin Core and a preservation metadata standard issued by the Research Libraries Group (RLG) (a not-for-profit membership corporation of over 160 universities, libraries, archives, historical societies, and other institutions). • Metadata is captured during the imaging process. <p>Here are the metadata fields they try to capture for each image. Go over list:</p> <ul style="list-style-type: none"> • Full-Size File Name: MHS file name for the full-size digital version (string of numbers up to eight) • Thumbnail File Name: MHS file name for the thumbnail digital version • Date Sent: Date image was sent to outside vendor for scanning (if applicable) • Date Received: Date image to be scanned was returned to MHS (if applicable) • Unique Identifier: AV, Classification, Accession or other unique identifier assigned to the work (location of photo print)

Dublin Core Metadata Standard and elements created by the RLG Working Group on Preservation Issues of Metadata

- **Negative Number:** Negative number assigned to the work of art when applicable (not every image has a negative)
- **Creator Last Name**
- **Creator First Name**
- **Title/Object Description:** Title or brief description of the original object
- **Date:** Date file was created; format = yyyyddmm
- **Department/Transcriber:** Name of department responsible for transcribing the metadata
- **Producer:** Agency responsible for the physical creation of the file
- **Capture Device:** Indicate make and model of the digital camera or scanner
- **Capture Details:** Name scanner software or lens type (digital camera) & other relevant details
- **Change History:** Record modifications to the file, identify when & by whom they were made (e.g., thumbnail created)
- **Source:** Indicate the characteristics of the source document from which the digital version was created (e.g., negative, photoprint, 35mm slide, etc.)
- **Encryption:** Is the file encrypted or not, if so the type of encryption should be indicated
- **Watermark:** Indicate whether or not a digital watermark has been used, indicate type used
- **Resolution:** Determine number of pixels, pixels per inch or dots per inch
- **File Format:** Indicate type of file format used to create the digital version
- **Compression:** Indicate type of file compression used to create the digital version
- **Color:** Indicate pixel depth
- **Color Space:** Enter the color reference (e.g., Kodak, RGB, CMYK)
- **Color Management:** Identify system used to improve color consistency
- **Color Bar/Grayscale Bar:** Indicate presence or absences of either and, if present, identify type
- **Control Targets:** Indicate information about targets used for quality control, calibration, etc.
- **Usage rights:** Enter any copyright or usage restrictions
- **Notes:** any other pertinent information

Note that there are several elements that overlap with the Dublin Core and also the Minnesota Recordkeeping standard, including, Creator Name, Date, Unique Identifier, Title/Object Description, Source, File Format, and Usage Rights. Additional elements were added to meet the unique requirements of the digitization and access processes.

<p>Page-37 Minnesota Historical Society Preservation Metadata example</p> <p>Full-Size File Name: pf072079 Thumbnail File Name: pt072079 Date Sent: N/A Date Received: N/A Unique Identifier: HD4.6 p24 Negative Number: 77428 Creator Last Name: Creator First Name: Title: Control Data's CDC 160 Computer Department/Transcriber: Processing Department. SLD Producer: Minnesota Historical Society 345 Kellogg Boulevard W., St. Paul, MN 55102-1906 Date: 20000610 Capture Device: Umax Powerlook III Capture Details: Umax Powerlook III scanning software, Photoshop 5.0 Change History: Thumbnail created Source: 8 x 10 photoprint Validation Key: Encryption: None Watermark: None Resolution: 150 dpi File Format: JPEG Compression: JPEG Color: 8-bit Color Space: Grayscale Color Management: None Color Bar/Grayscale Bar: None Control Targets: None Usage rights: None Notes:</p>	<p>Page-37 The Minnesota Historical Society <input checked="" type="checkbox"/> <i>Note to instructor: Consider using a local example.</i></p> <p>Go through the information on the slide.</p>
<p>Page-38 Exercise: E-mail</p> <p>From: Rounds, Shawn Sent: Tuesday, February 19, 2002 9:09 AM To: Craig A Steiner (E-mail); Eileen McCormack (E-mail); Horton, Robert; James Taylor (E-mail); Johnson, Jennifer; Karen Bondy (E-mail); Linda Feist (E- mail); Marsha K Haagenson (E-mail); Patricia Dunlop (E-mail); Steve Retzlaff (E-mail) Cc: Bruce Yurich (E-mail); Ed Potter (E- mail); Janice Zarkin (E-mail); Jay. Achenbach (E-mail) Subject: Metadata Comm: New Draft</p> <p>Hello --</p>	<p>Page-38 <input checked="" type="checkbox"/> <i>Note to instructor: Consider using a local example.</i></p> <p>We've talked now about some different functions that metadata supports, including system administration, access and evaluation, records management, and preservation. We also seen that even though all of these have unique characteristics and concerns, there is much overlap.</p> <p>Everybody worries about what to do with e-mail, since they are often serve as audit trails of how decisions are made, and they are often a weak point in legal actions. So let's take a look at one as an example of how the metadata we talked about today might all come together to meet your business needs.</p> <ul style="list-style-type: none">• What do we see here?• What don't we see here that we might want to worry about?

As promised at our last meeting, here is the latest draft of the standard, now complete with all elements. I've cleaned it up, removing the track-changes marks and doing some editing to provide consistency in phrasing, formatting, etc. It's NOT the final draft, but it's the starting point for our next round of work. It should also be online at the committee web site in the next few days
 (<http://www.mnhs.org/preserve/records/metadata.html>)

Believe it or not, we're nearing the end! Thanks for your continued interest and participation --



Shawn

- o Recordkeeping metadata: who sent it, when it was sent, who it was sent to, and the subject. This information can also be used for access – for instance, you could search on any of these fields to find this e-mail.
 - But was somebody blind CC'd? Someone who doesn't show up on this copy?
 - What are these people's actual e-mail addresses?
- o Preservation: there are attachments to worry about.
 - One looks to be an HTML file, the other a Word document – but what version Word is it in? That's not immediately apparent.
 - And what about metadata for each of the attachments – creation information, etc? How do we know they haven't been altered?
- o System administration: this copy doesn't show it, but behind in the e-mail system, there would have been metadata about the size of the message and its attachments, what type of message it is (plain text, rich text? HTML?), is there a message identification number? Date and time stamps to indicate more precisely when it was sent and received. This last, of course, can also be recordkeeping and access metadata.

So you see, just in a simple e-mail there's a lot of metadata that you might want to capture for different purposes and there's a lot of overlap. For any given environment, your business needs guide your determination of exactly what constitutes a record, and how much and what kind of metadata you need to keep.

Page-39
 Recordkeeping, Access, and Preservation metadata don't exist in a vacuum.

1. Used in conjunction with other types of metadata (e.g., systems management metadata)
2. Stakeholders will use a variety of metadata
3. Assists in evaluation
4. Metadata provides a common language to use when building partnerships and collaborations.
5. Make metadata part of your overall strategic plan.

Page-39
 We've learned a lot about recordkeeping, access and preservation metadata in the last three units

- remember that while recordkeeping, access, and preservation metadata may be the types of metadata archivists and recordkeepers use primarily, they do not exist in a vacuum.
- Often, they are used in addition to other types of metadata, such as systems management metadata
- archivists and recordkeepers should be aware that their partners and stakeholders use a variety of metadata.

Metadata influences our evaluation decisions.

- We're more apt to select and retain objects if we know what they are and how to use them.
- We're more apt to decide we'll take the objects from our partners and stakeholders if we understand where they fit into our stakeholders business and how they were used

We all (archivists, IT, etc.)

- need to have understanding of information systems, of

the whole process and flow of information.

- need to build partnerships because we cannot do it alone.
- Archivists need to promote active and mutually beneficial partnerships with records creators, records managers, and the information technology communities. Many of our needs overlap, and metadata provides us with a common language.

The more documentation there is, the greater the basis for understanding for whatever activity you're aiming at, whether it's day-to-day operations, recordkeeping, access, etc. Metadata should play a role in every stage in the lifecycle of digital information.

The bottom line is, of course, metadata is a tool. It is important for paper records, and absolutely necessary for digital objects.

The best way to make use of metadata is to make it part of an overall plan, which analyzes what digital objects you have, what you want to do with them, and why you need metadata.

<p>Page-40 How to fit metadata into your overall plan</p> <p>Step 1: Identify why you need metadata. What purpose(s) are you trying to serve? What do your partners need?</p> <p>Step 2: Identify and define all the pertinent variables (tools, resources, partners, etc.)</p> <p>Step 3: Assign values to the variables to determine the appropriate strategy for your organization. Start by asking the following questions:</p> <p>What resources are available inside your organization? Outside of it?</p> <p>What skills do you have? Or, how can you obtain the skills you do not have?</p> <p>What standards are available and applicable?</p> <p>How will you train and educate people, and encourage them to become willingly involved?</p> <p>What mandates, regulations, industry standards, business directive, etc. apply?</p> <p>Where are the intersections among your organization, partners, and stakeholders?</p> <p>Is there already metadata that's being created that you could build on?</p>	<p>Page-40 How do you fit metadata into your overall plan?</p> <p>Here is a process:</p> <ul style="list-style-type: none"> • Even before step one, consider your goals and create a business case for why you need metadata: <ul style="list-style-type: none"> ○ What are your business goals, or what is your business function? ○ Are you meeting those goals or functions right now? ○ What's the gap? ○ Are you willing to stop doing something you are currently doing in order to do something else? ○ How can you prioritize your goals? ○ What's practical and reasonable? ○ What do your partners need? Where can you both benefit? • Step 1: <ul style="list-style-type: none"> ○ Identify and define all the pertinent variables (resources, tools, standards/methodologies, education, technology, partners, etc.) to meet your goals. ○ Take the time to evaluate your resources. • Step 2: <ul style="list-style-type: none"> ○ Assign values to the variables to determine the appropriate strategy for your organization. ○ Start by asking the following questions. Go over list on slide, adding others that may be appropriate to your audience <p>Keep in mind that one size does not fit all</p> <ul style="list-style-type: none"> • Can't take one model and set it up, for example, what has worked in Minnesota may not work for you • You need to make decisions about what's appropriate for you and your partners and environment. <p>This does not have to be a formal, rigid program.</p> <ul style="list-style-type: none"> • Whether you realize it or not, you are probably doing this sort of analysis whenever you think about adopting a new tool or technology, starting a new activity, or revamping what you do right now. • This is just one framework to use. • Will assist you in clarifying and documenting your decisions for yourself and others with an interest in your organization, both now and in the future. • Need to re-visit your decisions and actions on a regular basis to make sure that everything is still in line with your goals. • Always use your goals as your touchstone for evaluating your success <p><i>{Address quote on the slide.}</i></p>
<p>Thank you!</p>	<p>Thank you!</p>
	<p><input checked="" type="checkbox"/> <i>Note to instructor: You may choose to use the current appendices or change them for your audience</i></p>

{Appendix A} Metadata Crosswalks	{Appendix A} Metadata Crosswalks
{Appendix B} List of Acronyms	{Appendix B} List of Acronyms
{Appendix C} Bibliography	{Appendix C} Bibliography.