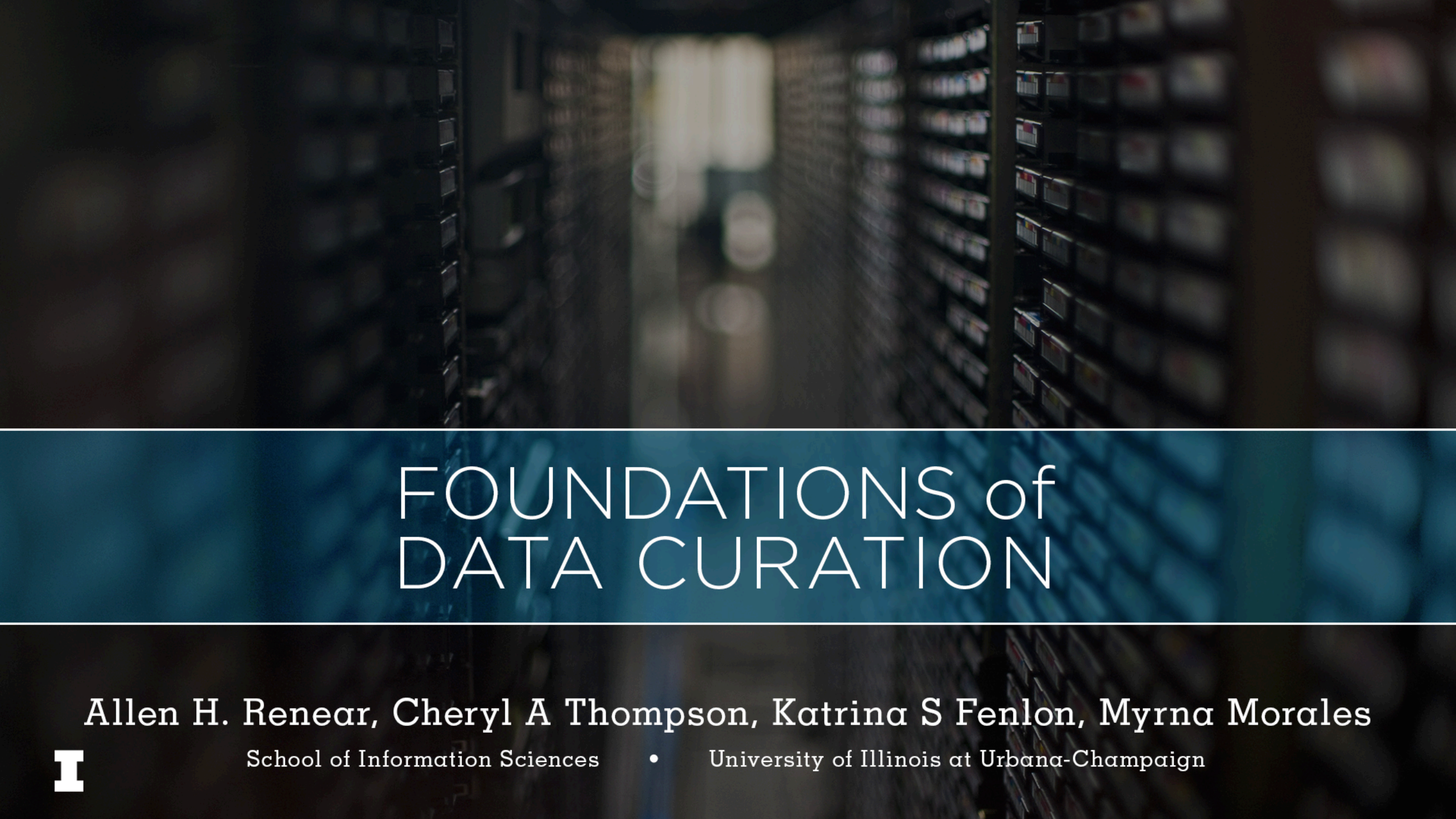

FOUNDATIONS of DATA CURATION

Allen H. Renear, Cheryl A Thompson, Katrina S Fenlon, Myrna Morales

School of Information Sciences • University of Illinois at Urbana-Champaign


METADATA


2
METADATA SCHEMAS

Metadata Schemas

Metadata schemas
Vocabulary independence
Syntax/serialization independence
Mixing and matching metadata schemas
Examples of schemas

Metadata Schemas

A set of data elements, with specified meanings

for supporting metadata statements in particular contexts

Sometimes vocabulary independent*

Often syntax/serialization independent**

Pure metadata schemas are *conceptual*

*Sometimes vocabulary independent

Metadata elements in a metadata schema can be purely conceptual, without a controlled vocabulary.

By allowing different vocabulary terms for the same concept the schema can more gracefully support different languages, as well as avoid, or take advantage of, common meanings for common words.

However advantages of having a level of indirection between concept and term may be outweighed by the advantages of providing a controlled term vocabulary, and so avoiding unnecessary variation.

So many metadata schemas specify a controlled vocabulary as well as conceptual elements; users who wish to use an alternative vocabulary for the same concepts can provide a mapping.

Pure metadata schemas are *conceptual*

**Often syntax/serialization independent

Similarly metadata schemas need not specify any particular syntax for applying specified concepts.

In this case most metadata schemas do strictly separate the conceptual schema from the variety of options for applying metadata to objects.

Given the variety of contexts (data models, file formats etc.) in which metadata will be applied allowing metadata statements to implemented in different serialization syntaxes is profoundly useful, and essential for wide adoption.

Recommended serialization syntaxes can be developed and standardized independently.

e.g., Dublin Core

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

- 15 elements for describing resources on the web
- With defined semantics and recommended vocabularies for elements

Contributor	Format	Rights
Coverage	Identifier	Source
Creator	Language	Subject
Date	Publisher	Title
Description	Relation	Type

Schemas vs. their serialization

Dublin Core metadata element set (select terms)

Creator: William Blake

Title: “A Sick Rose”

Date: 1794

Serialized as RDF/XML

<xml> <?namespace href = "http://www.w3.org/schemas/rdf-schema" as = "RDF">
<?namespace href = "http://www.purl.org/RDF/DC/" as = "DC"> <RDF:RDF>
<RDF:Description RDF:HREF="http://purl.org/metadata/dublin_core_elements"

DC:Title = **"The Sick Rose"** DC:Creator = **“William Blake"** DC:Date = **“1794"** />

</RDF:RDF> </xml>

Serialized with HTML meta elements.

<meta name="DC.Title" content=”The Sick Rose">
<meta name="DC.Creator" content=“William Blake">
<meta name="DC.Date" content=”1794">

Combining, specializing, and extending metadata schemas

Metadata schemas can be combined, specialized, and extended in various ways.

For example, a defined application of Dubin Core,

- may limit the values of the Dublin Core *dc:format* element to the controlled vocabulary for IANA media types (MIME types).

or

- specify the refinements *spatial* and *temporal* for dc:coverage

To ensure graceful processing of adaptations techniques for schema modification or other notifications may also be specified.

Take a look at these standard metadata schemas

Lists of Metadata Schemas:

Biodiversity: http://www.dcc.ac.uk/resources/metadata-standards/list

Libraries and museums: http://jennriley.com/metadatamap/seeingstandards_glossary_pamphlet.pdf

Other areas: https://en.wikipedia.org/wiki/Metadata_standard

Individual research groups, businesses, and other organizations may develop custom metadata schemas for private or limited use; usually these are specializations of existing schemas.

FOUNDATIONS OF DATA CURATION (IS531)

Allen H. Renear, Cheryl A Thompson, Katrina S Fenlon, Myrna Morales
School of Information Sciences
University of Illinois at Urbana-Champaign

Includes material adapted from work by Carole Palmer, Melissa Cragin, David Dubin, Karen Wickett, Bertram Ludaescher, Ruth Duerr and Simone Sacchi.

Comments and corrections to: renear@illinois.edu.