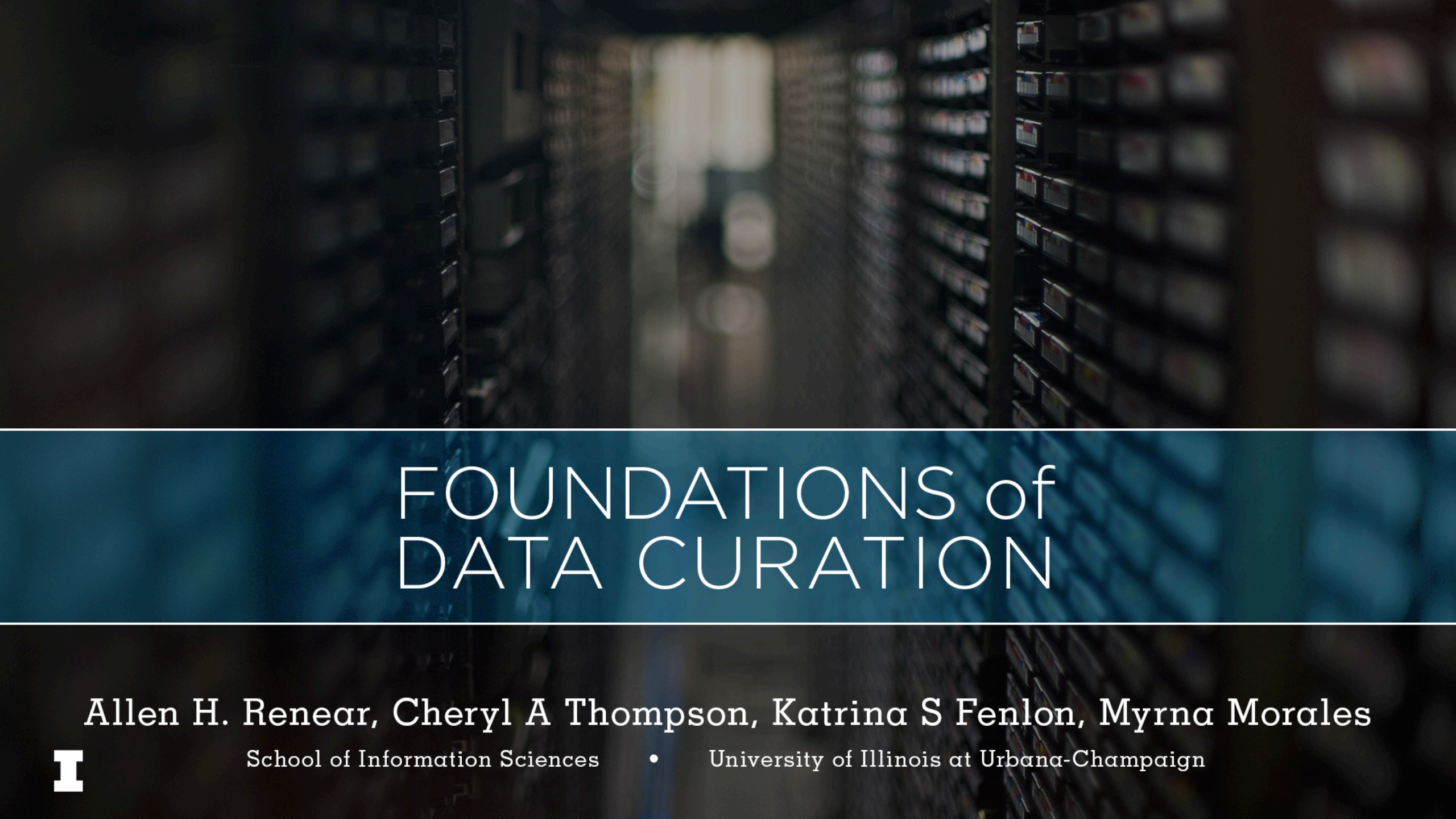

FOUNDATIONS of DATA CURATION

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

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

DATA MODELS: TREES


5
WHY THE SOLUTION WORKS

Clarification: these models have two parts

The heart of the relational model can be understood as the combination of the relational data structure with attributes for data values.

Similarly the tree model, as it is typically implemented, combines the tree data structure with descriptive markup node labels (or in XML terminology “generic identifiers”) such as “`stanza`”.

We will now refer to the tree model as the ”tree/DM” model.
The phrase “relational model” does not need the clarification as the role of attributes is commonly understood.

Drum roll

Can we say in general terms why the tree/DM model works so well?

Why it succeeds so well in meeting data management challenges?

We can.

It works the same way the relational model works: *Abstraction* and *Indirection*

Abstraction

Both models focus on the data *itself*, separate from storage and processing.

This explicit identification of *data attributes* in one case, and *logical text objects* in the other, brings enormous new functionality and efficiency

Once again: *Indirection*

Both models support an indirect relationship to storage and processing,
but in practice the emphasis is often different:

For the relational model abstracting away from storage is dominant
For the tree/DM model abstracting away from processing is dominant

In both case the separation is mediated by a mapping:
logical schema to physical schema in the case of the relational model
text component (type) to processing instructions in the case of trees/DM

Formal vs colloquial understanding of these data models

Although in the relational model we commonly think of attributes as representing dyadic properties or relationships in the world, technically they are names for domains of values,

.

Similarly in the tree/DM model we think of these node labels as indicating the kind of enclosed text object (stanza, formula, etc), but in the model they are really simply names.

It is our colloquial understanding of these models that enables us to actually use them to secure useful abstraction and data management..

This may seem a small point now, but it motivates a further advance in abstraction, as we well see when we discuss ontologies.

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.