



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



3

COMMON METADATA AMBIGUITIES

Common Metadata Ambiguities

Metadata is often casually used, with ambiguous semantics

And so many problems ensue, especially for robots.

It may not be clear (to a robot) what an attribute/value pair is *about*

Though it is obvious enough to humans

Attributes themselves can be, to robots, problematically ambiguous

Even when perfectly clear to a human

Yes, humans can figure things out

Ok, but that's not safe, and *it is not fair to robots!*

To get your intuitions going . . .

In a collection of digital images of paintings; it is not uncommon to see, attached to the same image:

...

ID: BNF4242a

...

Artist: Leonardo da Vinci

...

Format: JPEG

...

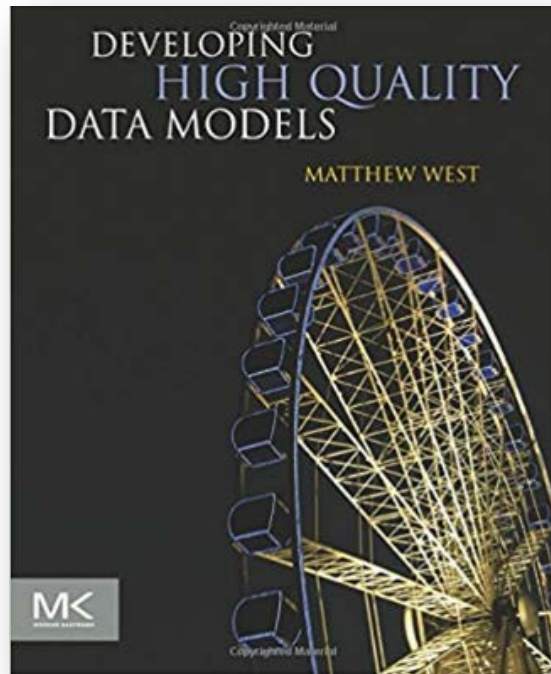
Format: oil on panel

What's wrong with this?

Adapted from work by Richard Urban

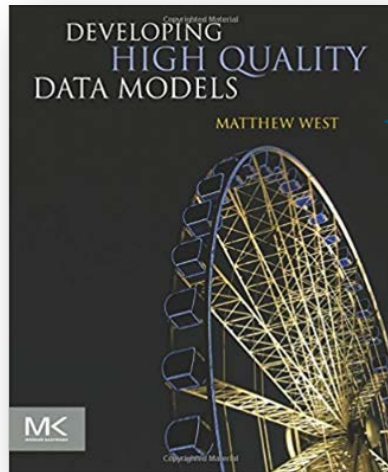


Recall: Bibliographic metadata

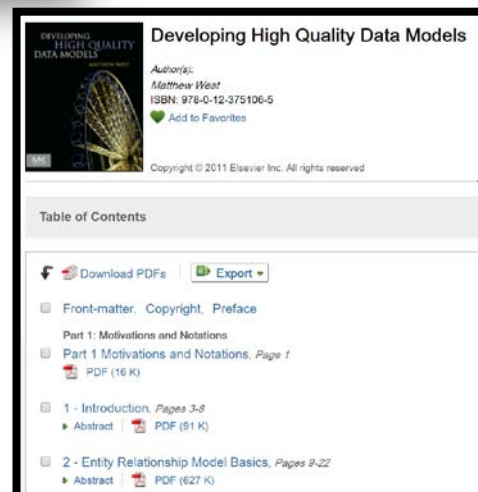


identifier (ISBN):	978-0123751065
creator:	Matthew West
title:	Developing High Quality Data Models
date:	2011
publisher:	Morgan Kaufmann
subject:	database design
subject:	data structures (computer science)
pages:	408

“What is being described?” Problems (think FRBR)



title: Developing High Quality Data Models
date: 2011
publisher: Morgan Kaufmann
subject: database design
subject: data structures
(computer science)



title: Developing High Quality Data Models
date: 2011
publisher: Morgan Kaufmann
subject: database design
subject: data structures
(computer science)
language: English
pages: 389

A converse problem: two meanings for same attribute

For a while this was common practice

DC:Identifier=Book2424a DC:lang=EN to indicate a book in English

and

DC:Identifier/Coll99LB DC:lang/EN to the books in a collection are in English

What's wrong with this?

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 Ludæscher, Ruth Duerr and Simone Sacchi.**

Comments and corrections to: renear@illinois.edu.