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
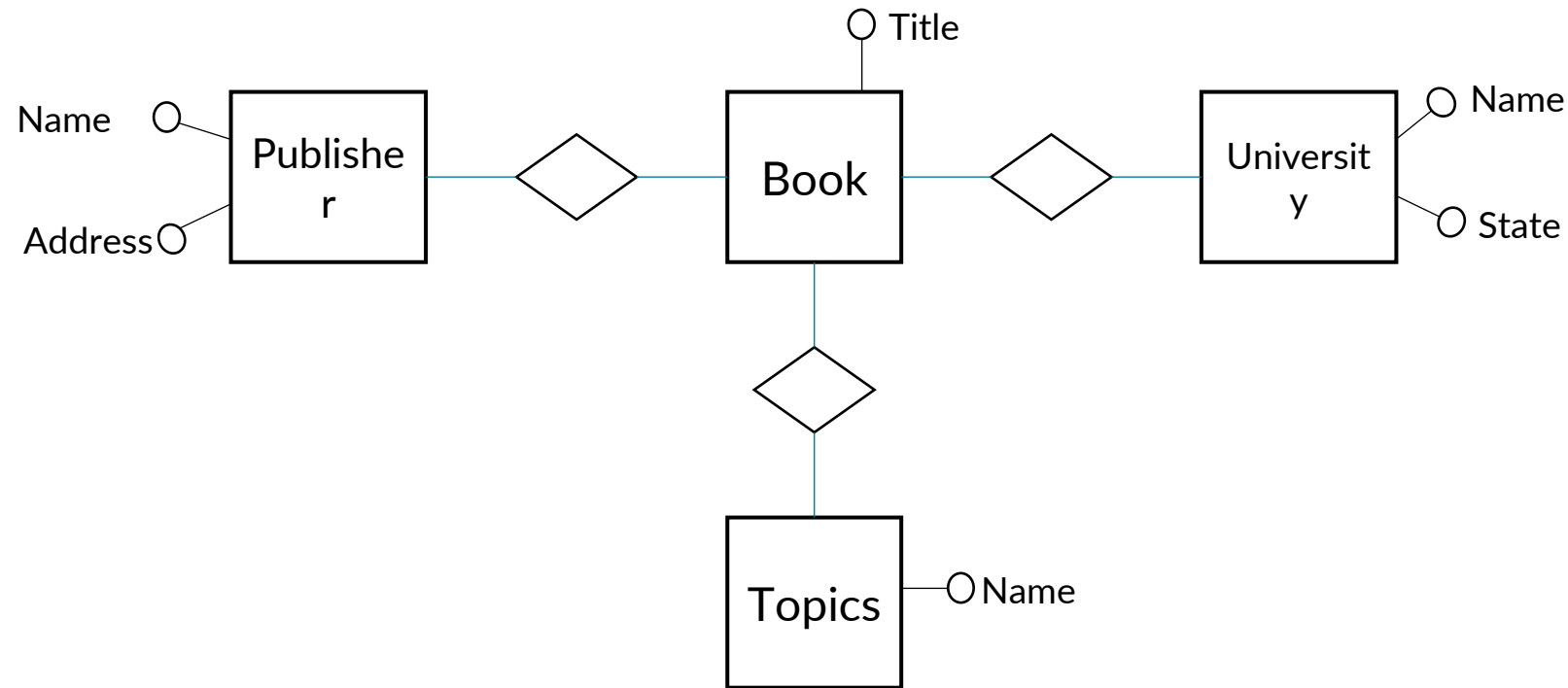
DATA INTEGRATION

④

SCHEMA INTEGRATION: EXAMPLE

Schema integration: example

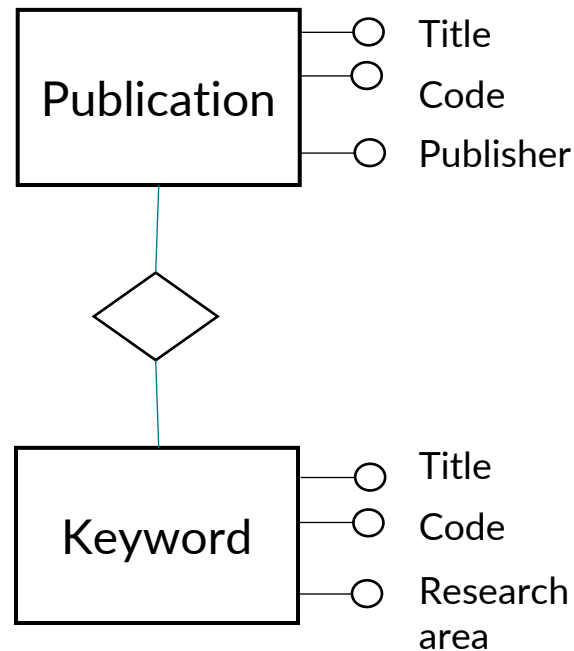Integrating two conceptual schemas.

Books, again

In one model...

Books are published by publishers. They are adopted by universities. They refer to topics.
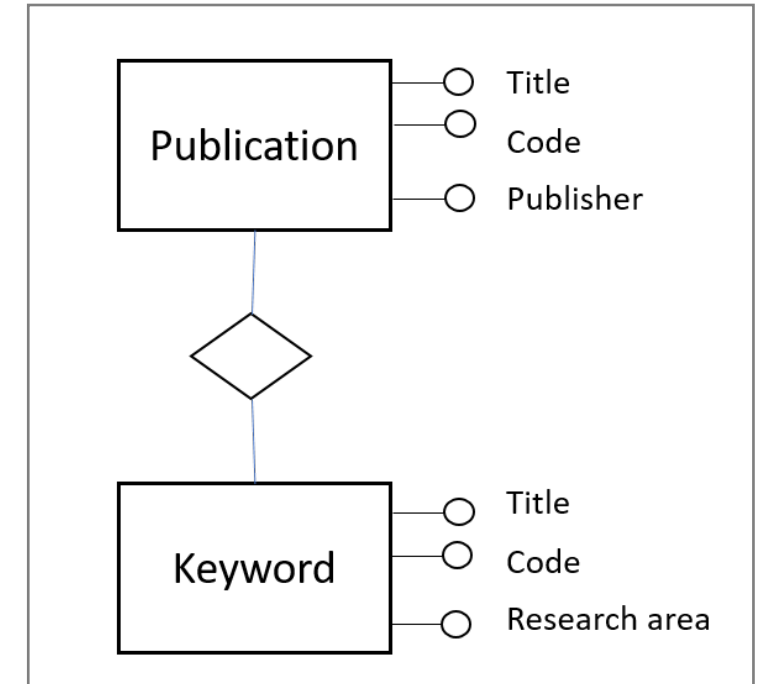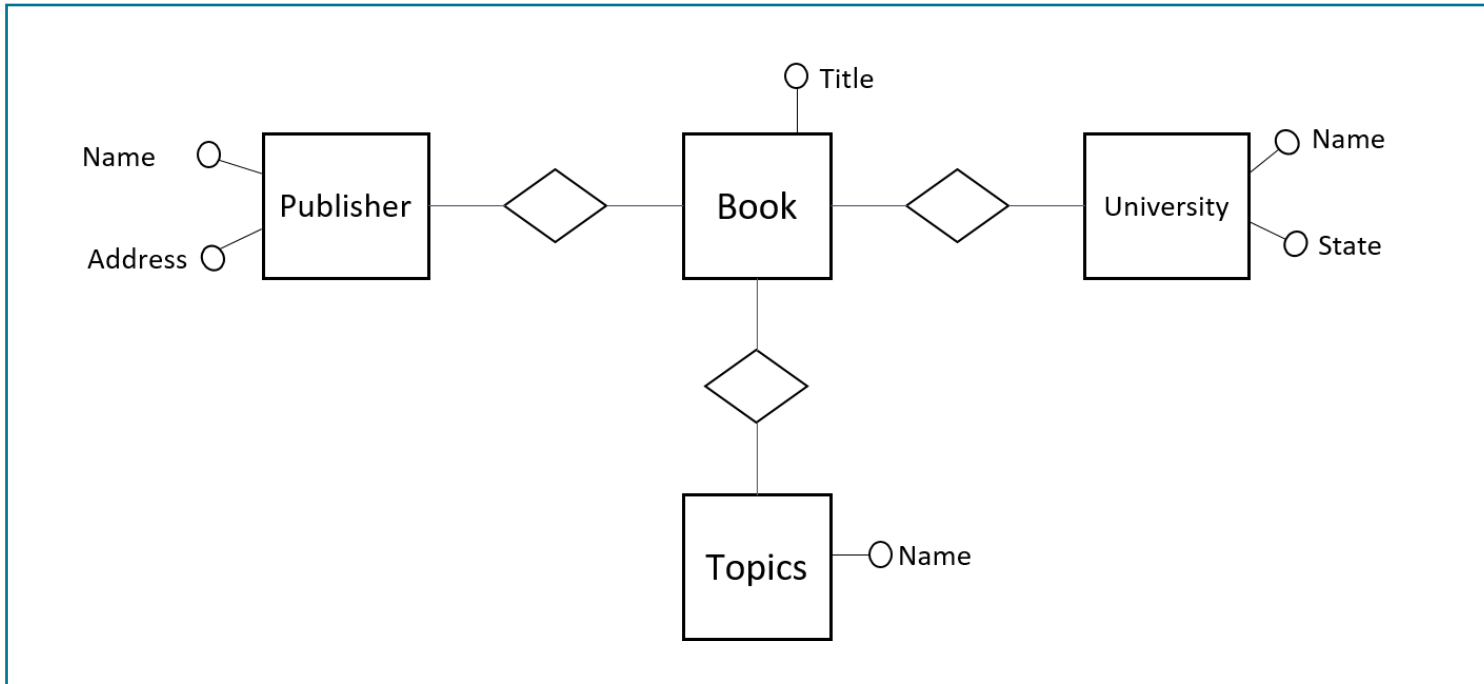
Batini, C., Lenzerini, M., and Navathe, S. B. (1986). A comparative analysis of methodologies for database schema integration. ACM Computing Surveys 18(4). doi>10.1145/27633.27634

Alternative model of same (?) domain

In another model…

There are publications of different types. Each publication has a list of keywords.

Batini, C., Lenzerini, M., and Navathe, S. B. (1986). A comparative analysis of methodologies for database schema integration. ACM Computing Surveys 18(4). doi>10.1145/27633.27634

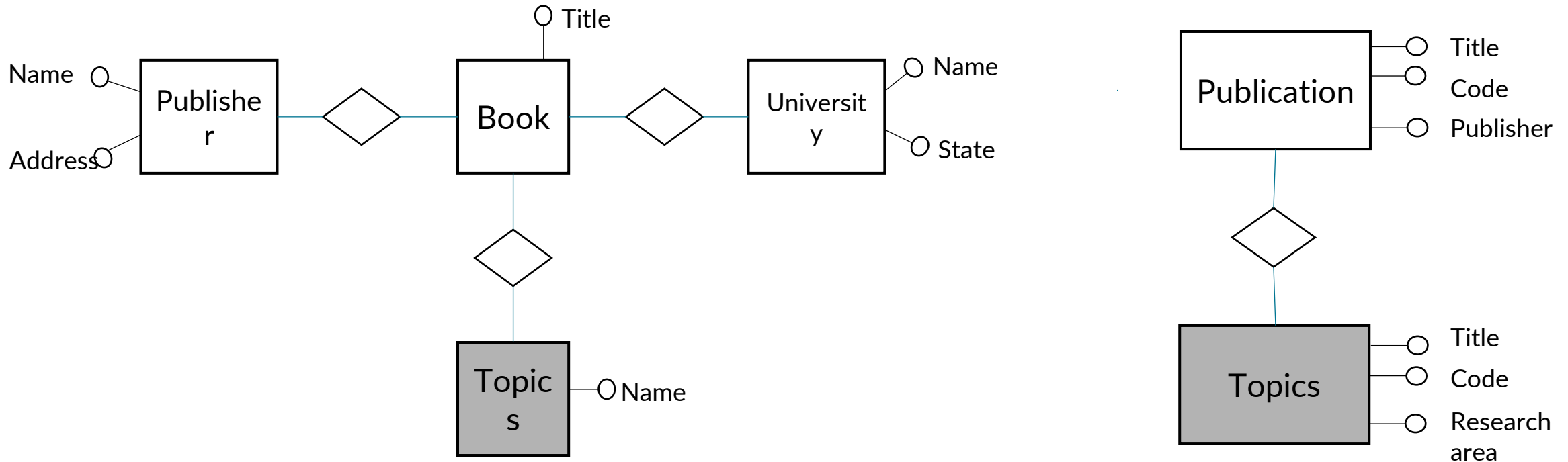# How to integrate these two conceptual schemas?

We observe such thing as:

      Topics and *Keyword* appear to refer to the same thing

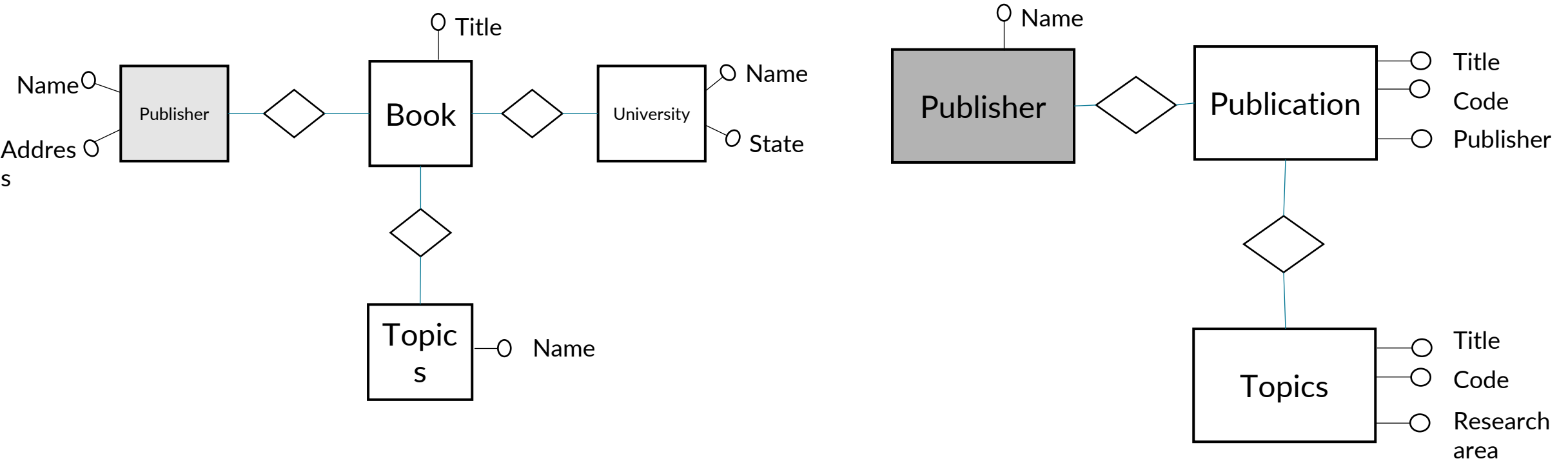   A *Book* /is kind of/ *Publication* (BOOK ⊂ PUBLICATION)

      Publisher is an entity in one schema; an attribute in the other

1. Use "Topics" for "Keyword" as well
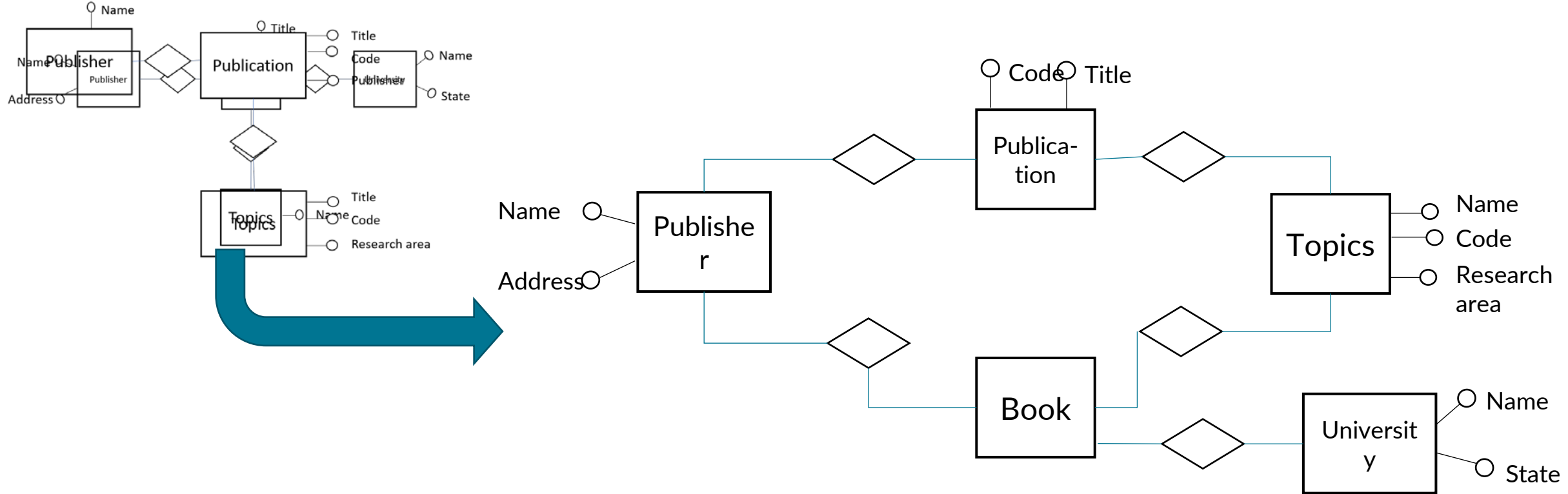

"Topics" and "Keyword" appear to be synonyms

Batini, C., Lenzerini, M., and Navathe, S. B. (1986). A comparative analysis of methodologies for database schema integration. ACM Computing Surveys 18(4). doi>10.1145/27633.27634

2. Make "Publisher" an entity

In one model, "Publisher" is an entity; in the other, an attribute

Batini, C., Lenzerini, M., and Navathe, S. B. (1986). A comparative analysis of methodologies for database schema integration. ACM Computing Surveys 18(4). doi>10.1145/27633.27634

3. Combine schemas

Batini, C., Lenzerini, M., and Navathe, S. B. (1986). A comparative analysis of methodologies for database schema integration. ACM Computing Surveys 18(4). doi>10.1145/27633.27634

4. Create subset relationship
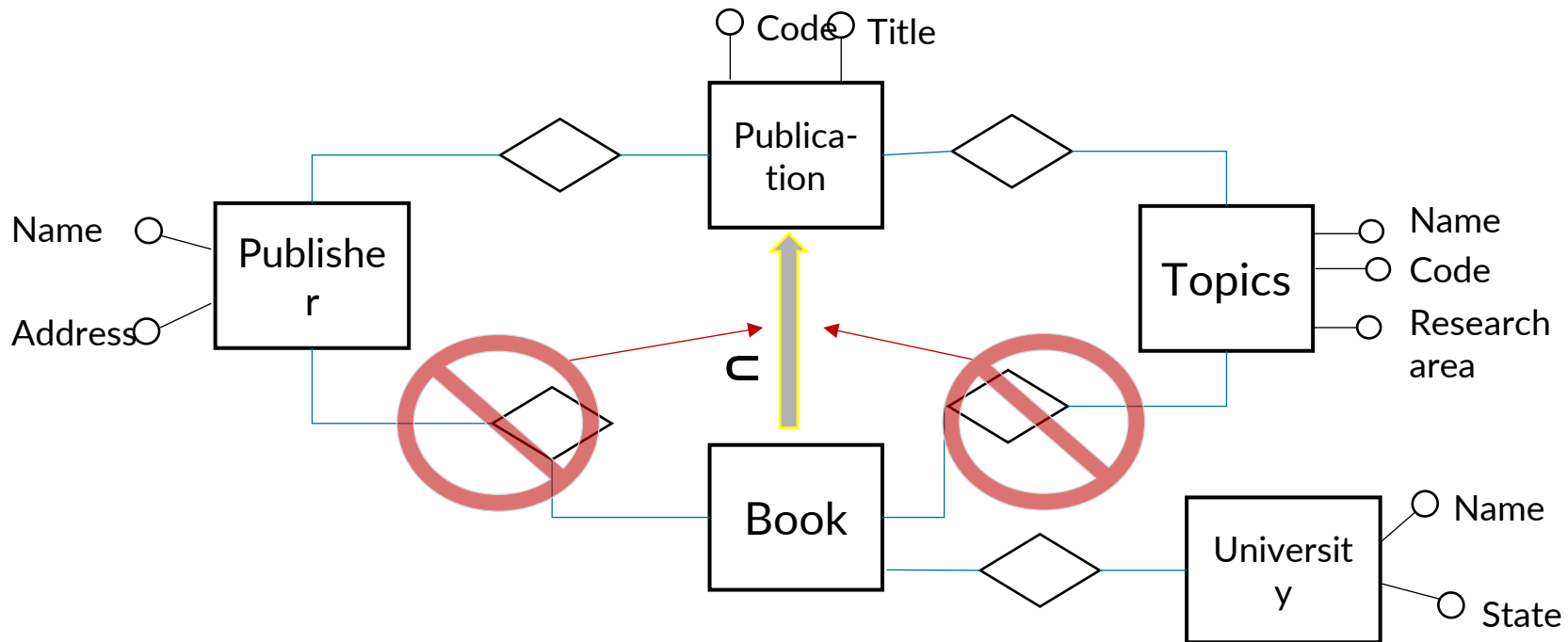
"Publisher" is an entity; an attribute in the other

Batini, C., Lenzerini, M., and Navathe, S. B. (1986). A comparative analysis of methodologies for database schema integration. ACM Computing Surveys 18(4). doi>10.1145/27633.27634
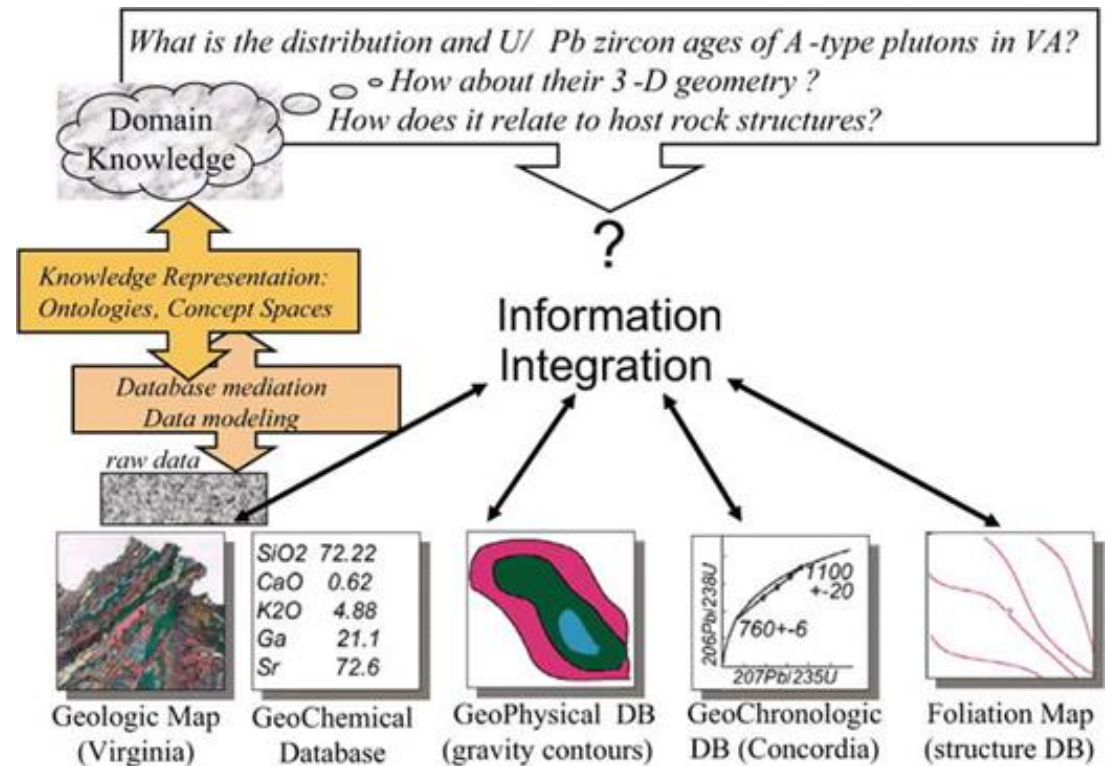
5. Integrated schema
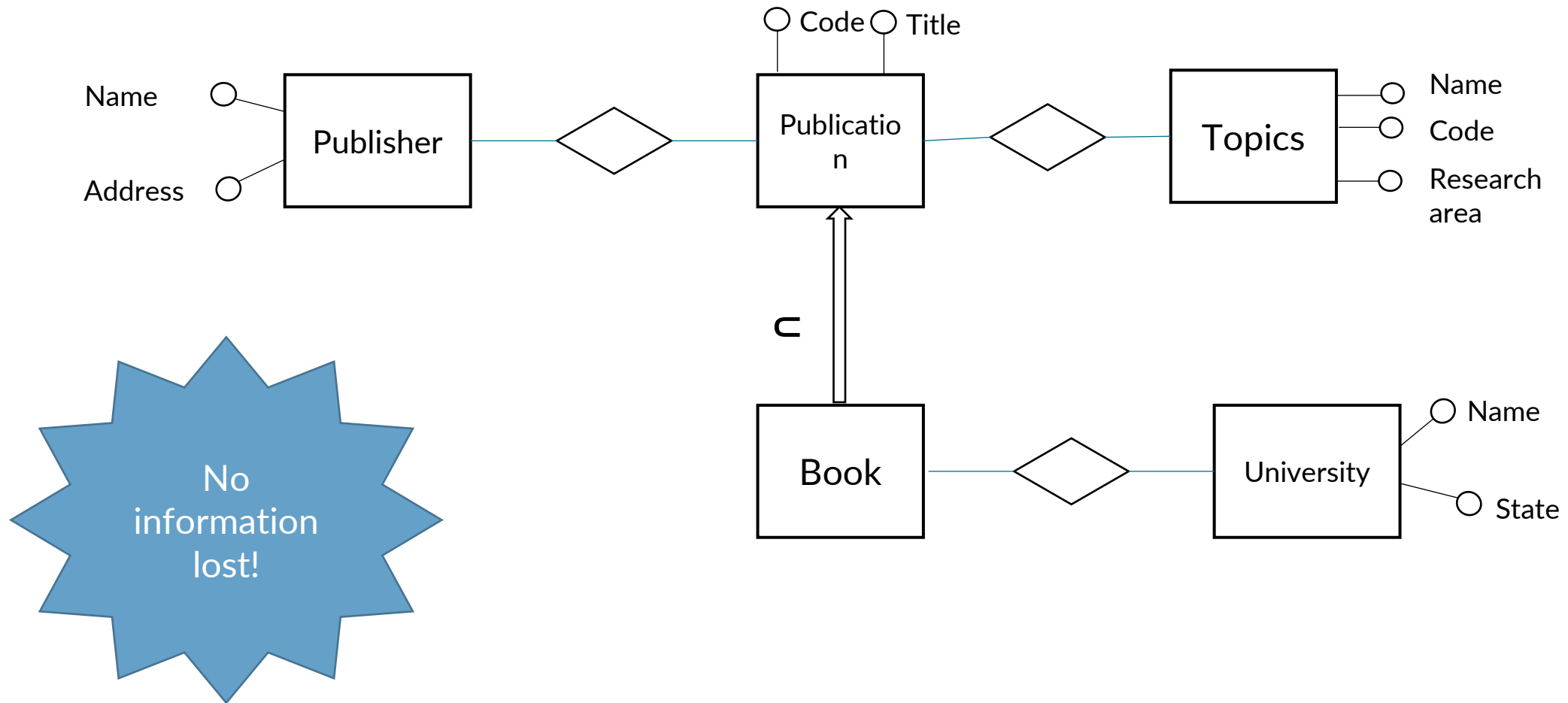
Ok, it isn't really so easy

One-world scenario
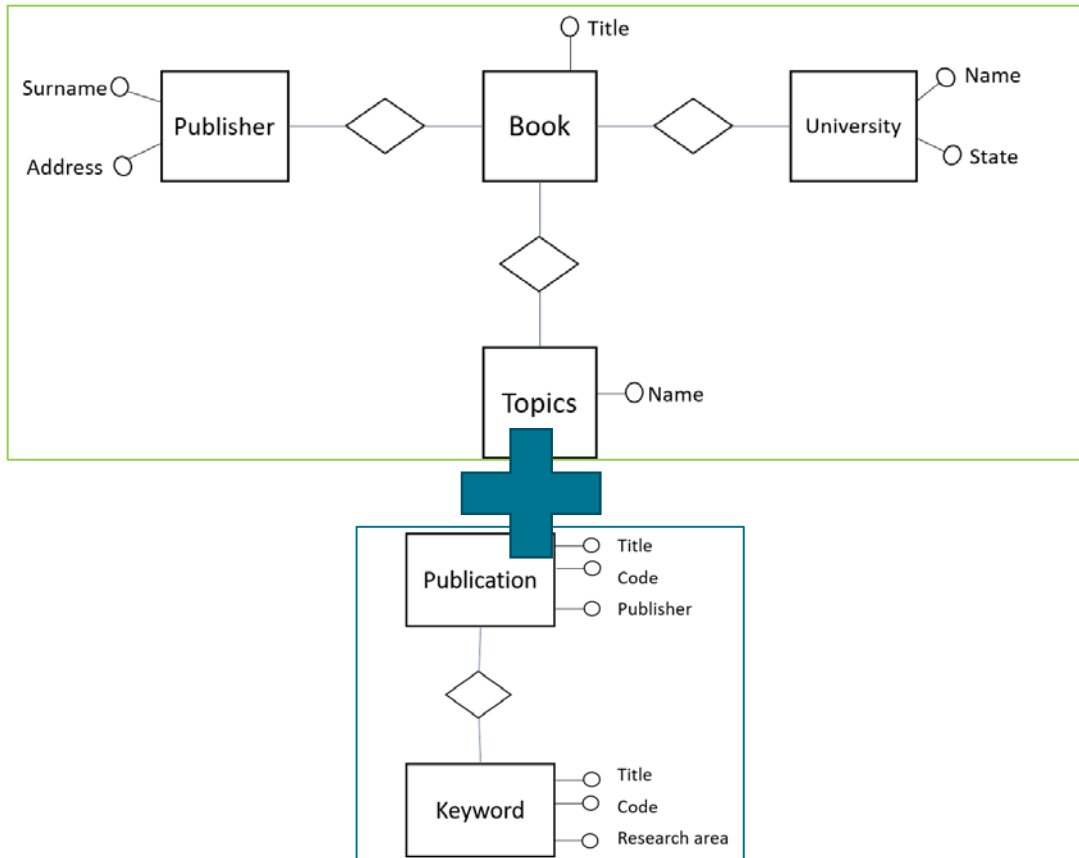
Batini, C., Lenzerini, M., and Navathe, S. B. (1986). A comparative analysis of methodologies for database schema integration. ACM Computing Surveys 18(4). doi>10.1145/27633.27634

Complex, multiple-worlds scenario

Bertram Ludäscher, Kai Lin, Shawn Bowers, Efrat Jaeger-Frank, Boyan Brodaric, Chaitan Baru, 2006. "Managing scientific data: From data integration to scientific workflows* ", Geoinformatics: Data to Knowledge, A. Krishna Sinha

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.