Entity-Relationship Data Model

Spring 2025

Thinking about data

How do we represent data?

To know how to represent data, we need to know what the data represents

- books collection for a library à la goodreads
- messages à la Slack
- customer survey responses à la Qualtrics

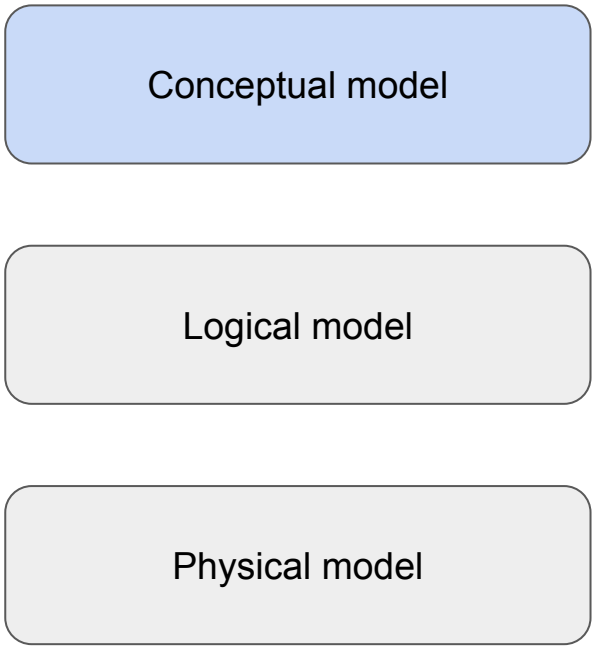
You need to understand the data to know what it represents

- we model the data to understand it

Thinking about data

Conceptual model	What is the data about?
Logical model	How does the data fit within a database model?
Physical model	How is the data represented in memory or on disk?

Thinking about data



What is the data about?

How does the data fit within a database model?

How is the data represented in memory or on disk?

The entity-relationship data model

The entity-relationship (ER) data model

- Chen (1976)
- human-centric model to help understand the data

The ER model is never implemented directly

- it is a thinking (and communicating) guide

We'll use the Chen notation, but notation is not what is important

- it's the exercise of working through the model that is important

Entities

An entity is a "thing" that can be distinguished from other "things"

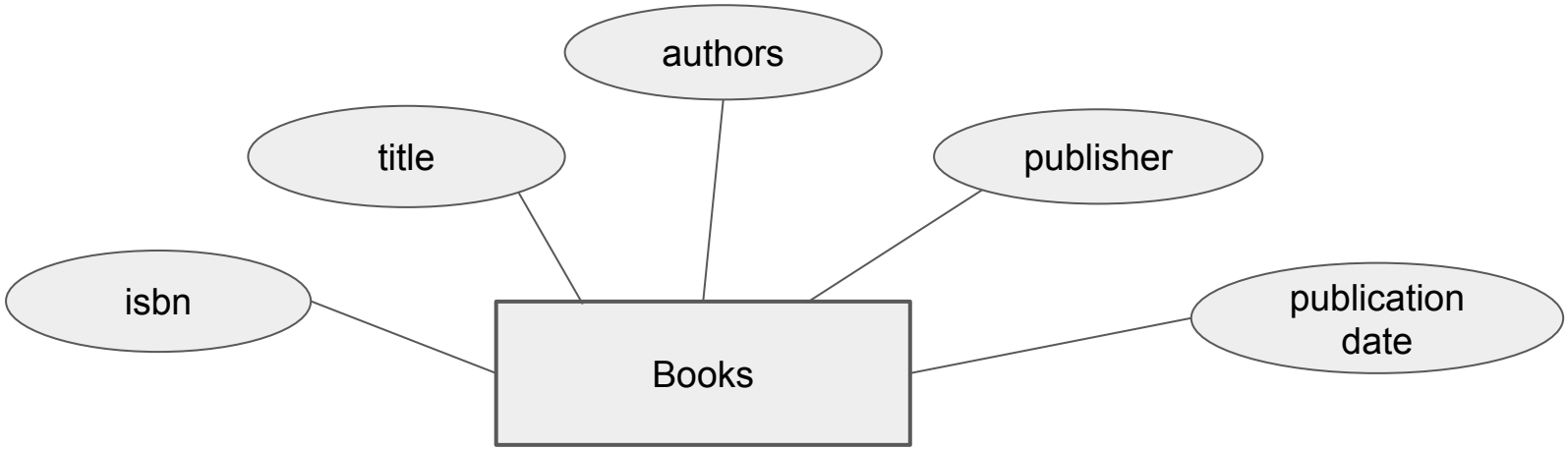
- an entity is described by attributes
- an attribute takes a value out of some domain of values

An entity set is a collection of similar entities (same set of attributes)

Entity set **Books:**

- title
- isbn
- page count
- publication date
- authors
- …

Example



Keys

Sets of attributes used to uniquely identify an entity in an entity set

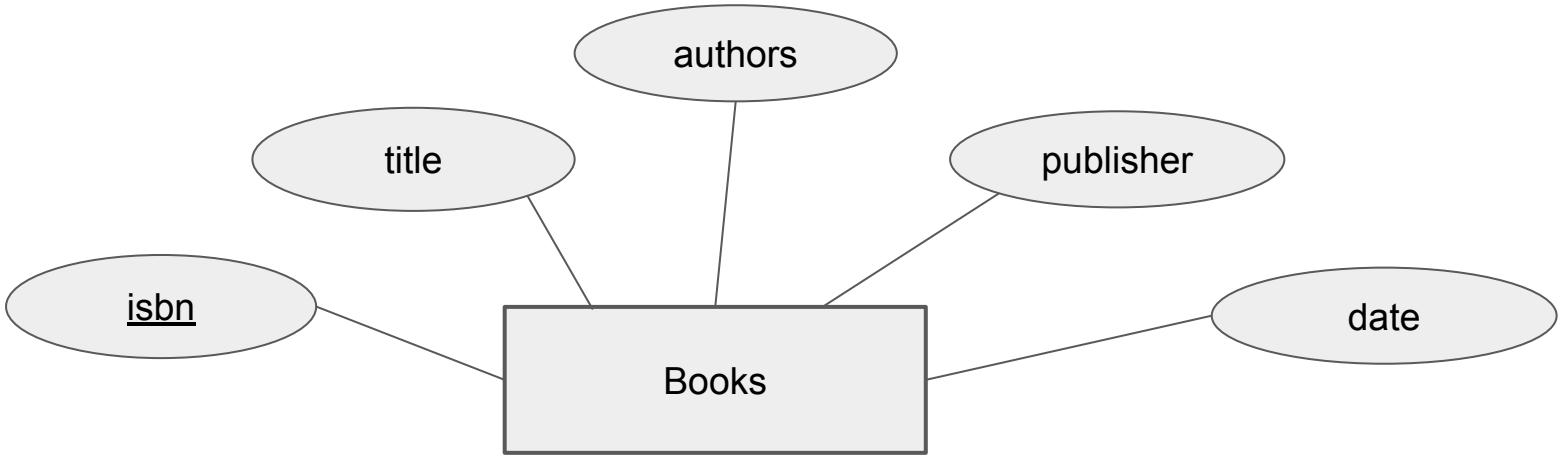
Superkey: one or more attributes that together uniquely identify an entity

- all attributes together should be a superkey
- if not, you can't distinguish repetitions

Candidate key: a minimal superkey

Primary key: a candidate key **chosen** to identify entities

Example



Relationships

A relationship is an association between two or more entities

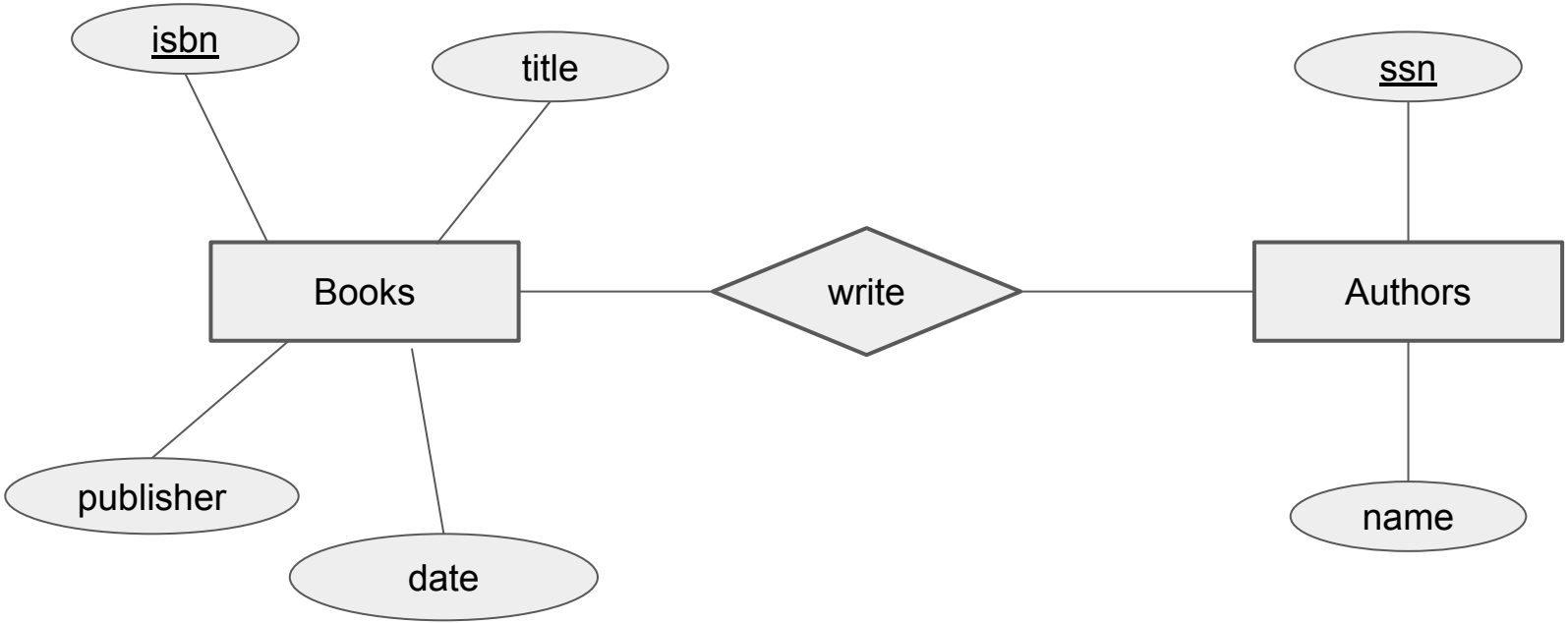
- an **author** writes a **book**

A relationship set is a collection of similar relationships

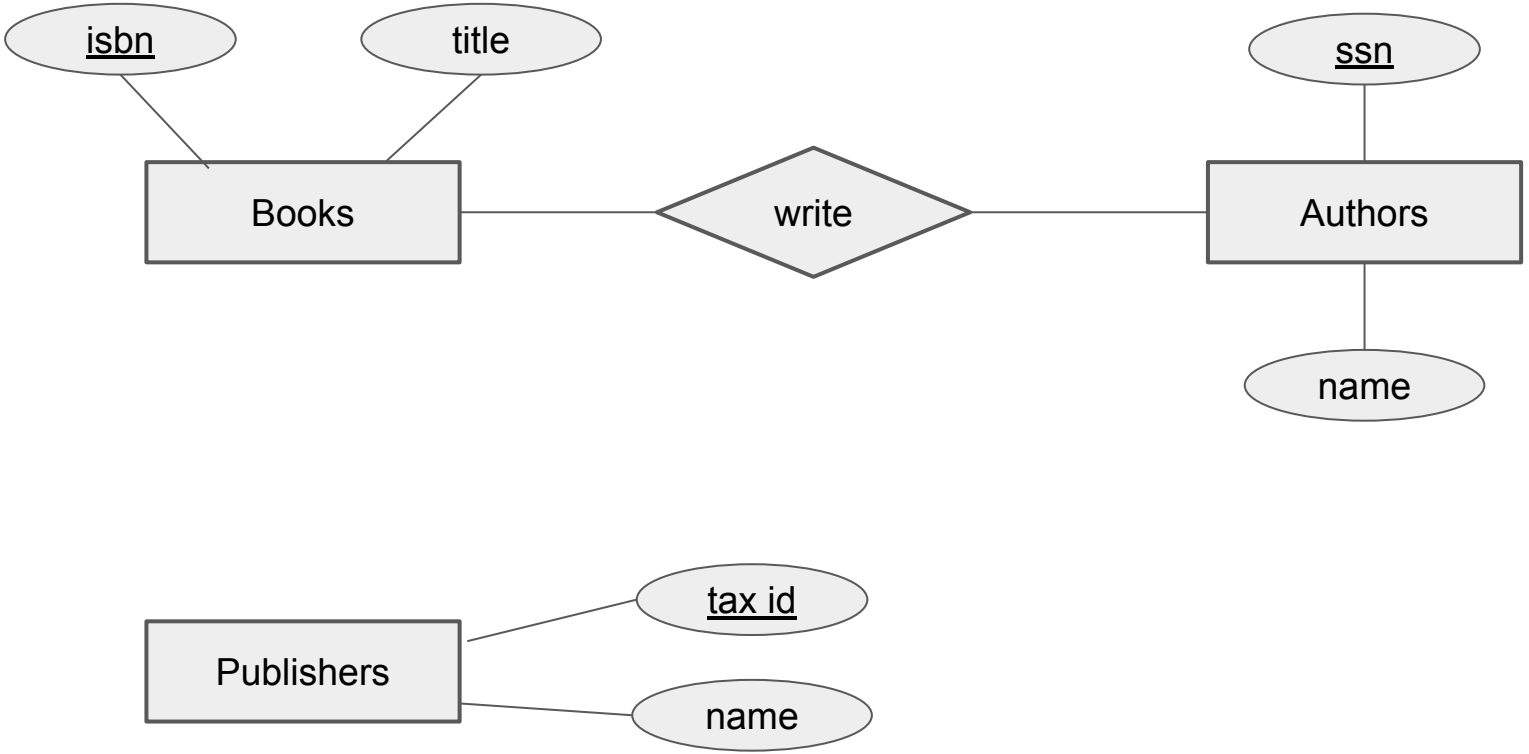
Write is a relationship between **Books** and an **Authors**

- relationships can have attributes
- if the same entity set appears multiple times in a relationship, can distinguish using *roles*

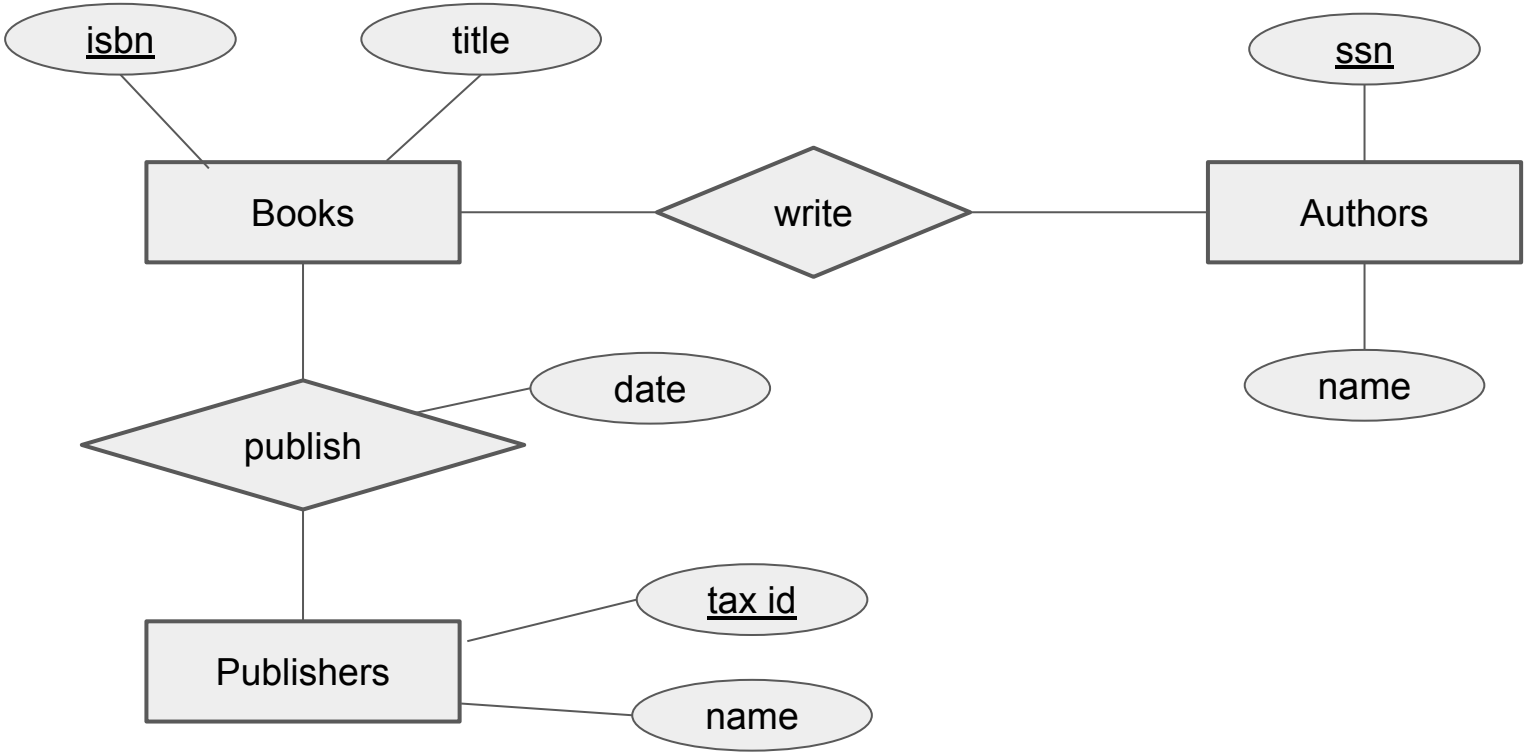
Example



Example


isbn
title
ssn
Books
write
Authors
name
Publishers
tax id
name

Example


isbn
title
ssn
Books
write
Authors
date
publish
name
tax id
Publishers
name

Key (cardinality) constraints

How many relationships can an entity participate in?

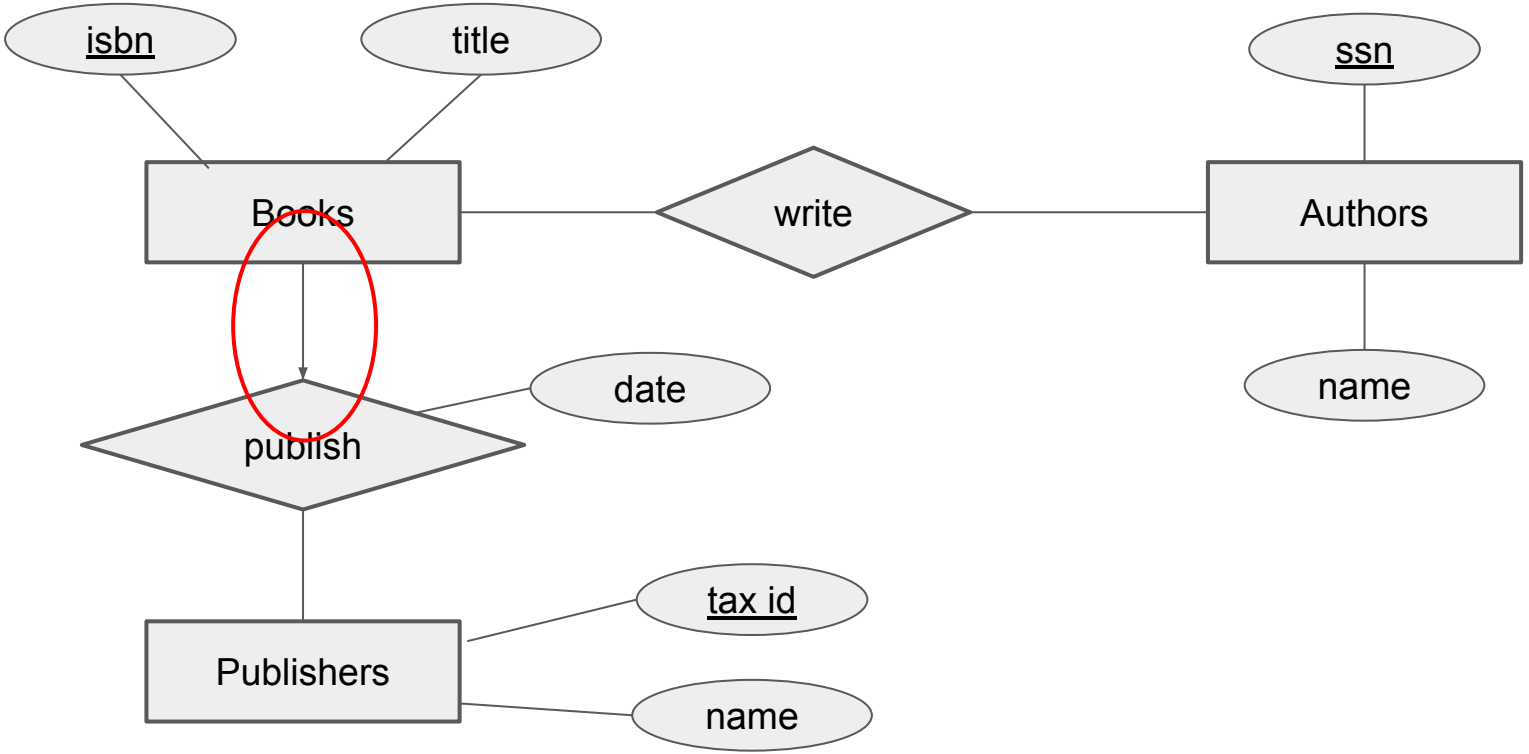
Examples:

- a book may have multiple authors, but a single publisher
- a book review applies to one book and is written by one reviewer

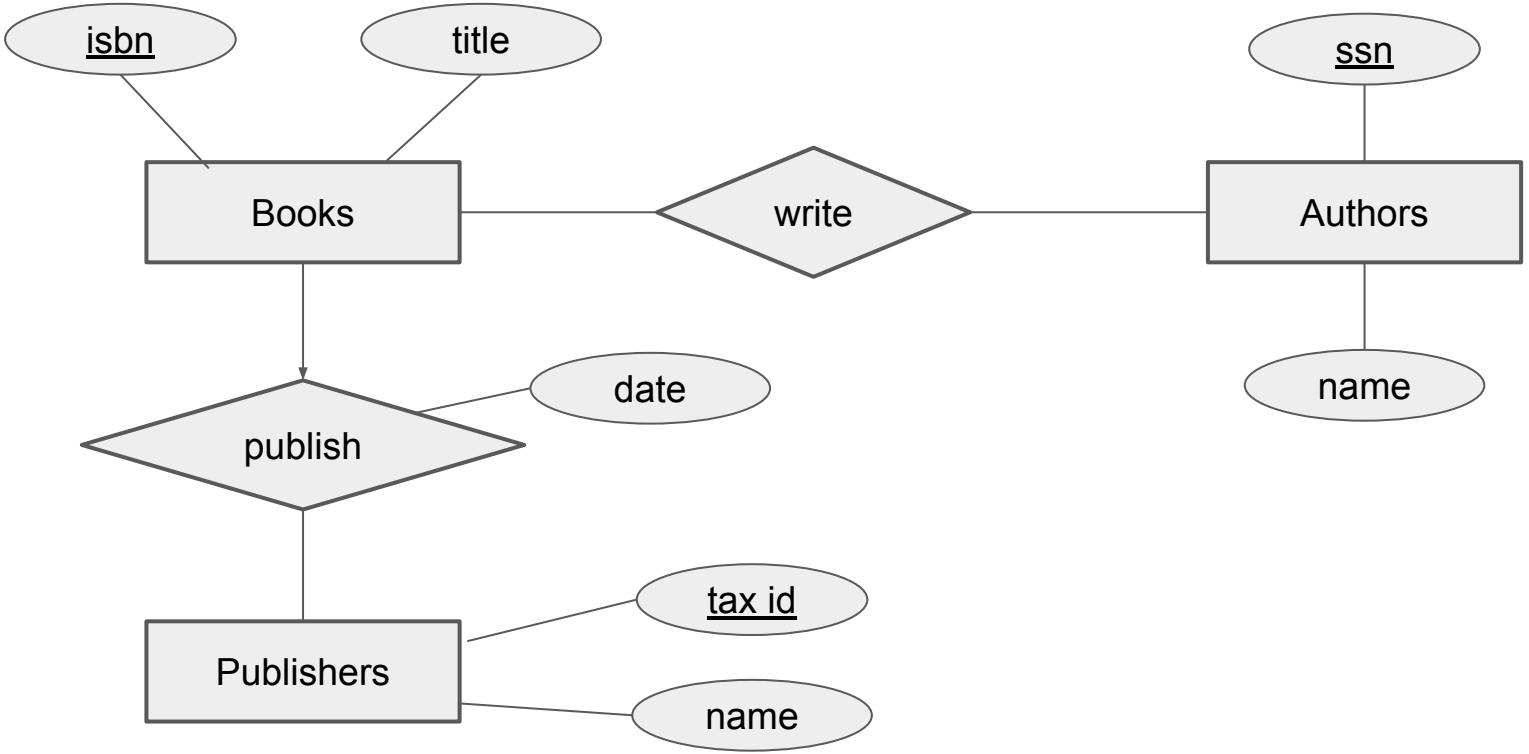
Annotate the relationship with cardinality constraints

- **1** or **N** ("many") on an entity line says how many relationships it participates in
- alternatively, use ← instead of **1** and — instead of **N**
- we talk of **1**:**1** or **1**:**N** or **N**:**1** or **N**:**N** relationship sets
- most DBs can enforce cardinality constraints

Example


isbn
title
ssn
Books
write
Authors
date
publish
name
tax id
Publishers
name

Example


isbn
title
ssn
Books
write
Authors
date
publish
name
tax id
Publishers
name

Participation constraints

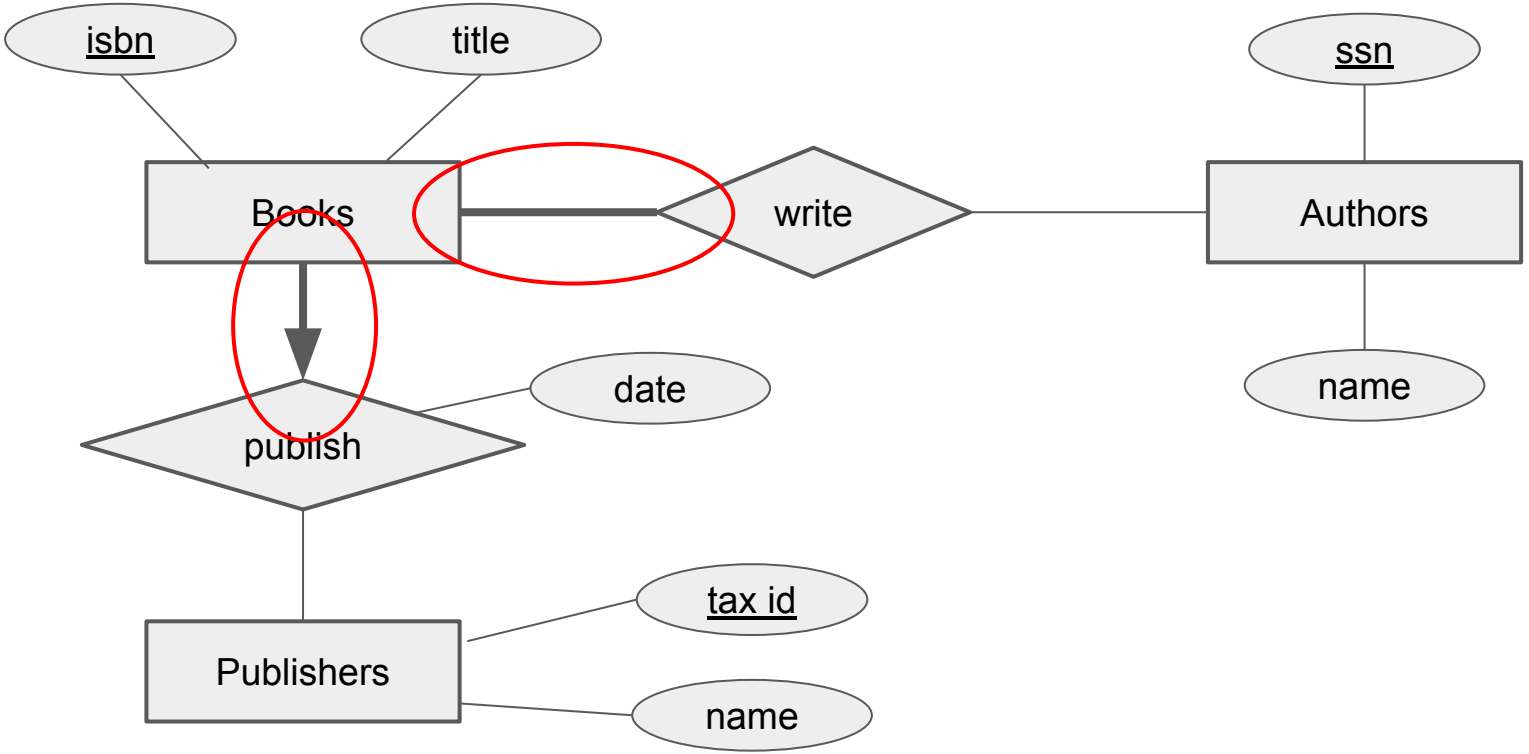
Must an entity participate in a relationship?

- a book must have at least one author
- an author need to not have written any book (yet)

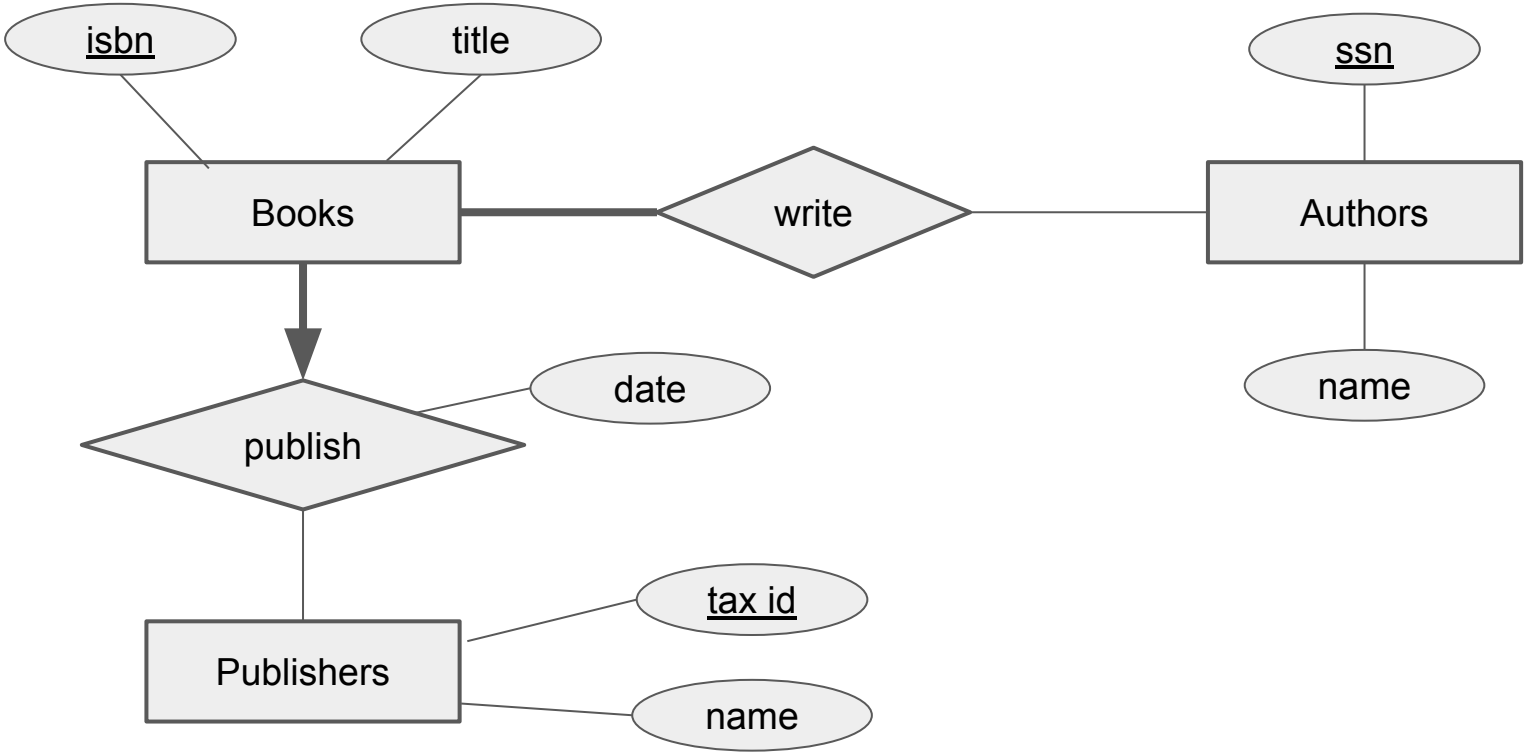
Annotate the relationship with participation constraints

- Use **0** on the entity line to indicate optional participation
- Compatible with **1** (exactly 1) and **N** (one or more)
- alternatively, use — instead of **0** and **—** instead of **1** or **N**
- most DBs can enforce participation constraints (useful, errors if they fail)

Example


isbn
title
ssn
Books
write
Authors
date
name
publish
tax id
Publishers
name

Example


isbn
title
ssn
Books
write
Authors
date
publish
name
tax id
Publishers
name

Ternary relationships

Examples of relationships have been binary — relating two entities

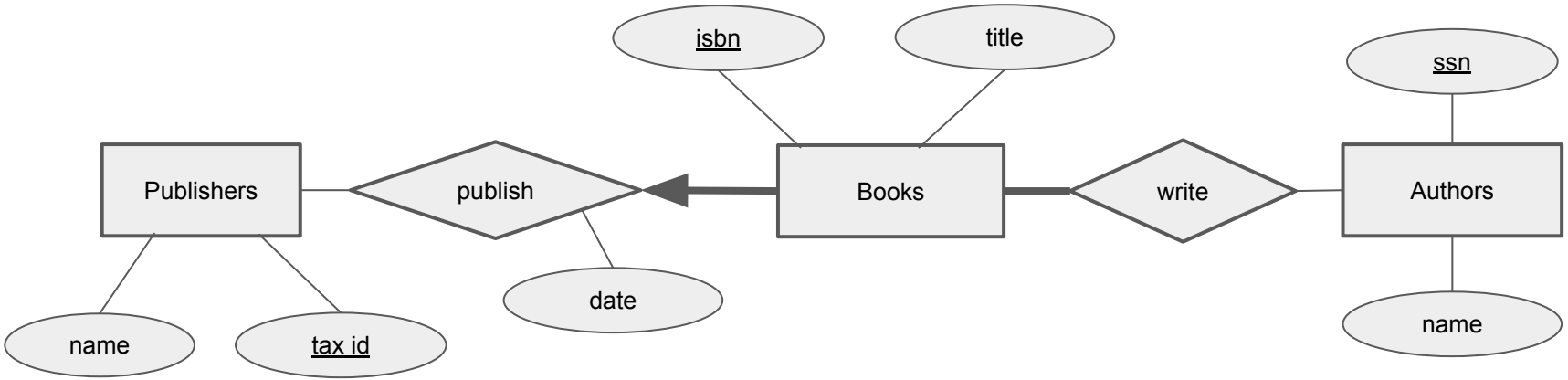
This obviously generalizes:

A **distributor** distributes a **book** in a **region**

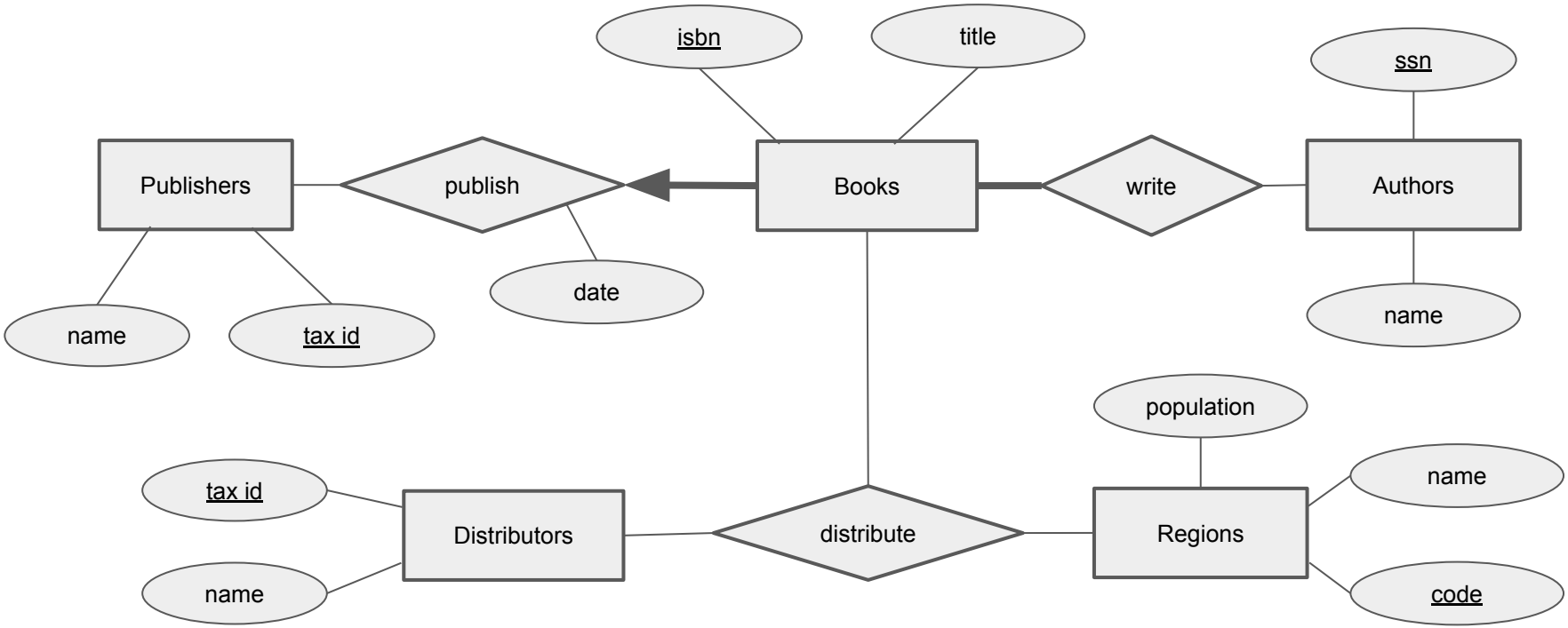
- A distributor may distribute different books in different regions
- A region may have different distributors distributing different books
- A book may be distributed by different distributors in different regions

This is a ternary relationship

Example


isbn
title
ssn
Publishers
publish
Books
write
Authors
date
name
tax id
name

Example


isbn
title
ssn
Publishers
publish
Books
write
Authors
name
tax id
date
name
population
tax id
Distributors
distribute
Regions
name
name
code

Other examples

Qualtrics-like surveys

Slack-like messaging

Appendix: relationship keys

We can identify a relationship by the primary key of the entities in the relationship

You can define the primary key of a relationship set A — R — B if you have cardinality constraints on R

- **N**:**N** primary(R) = primary(A) ∪ primary(B)
- **1**:**N** primary(R) = primary(B)
- **N**:**1** primary(R) = primary(A)
- **1**:**1** primary(R) = primary(A) or primary(B)