

# Lecture

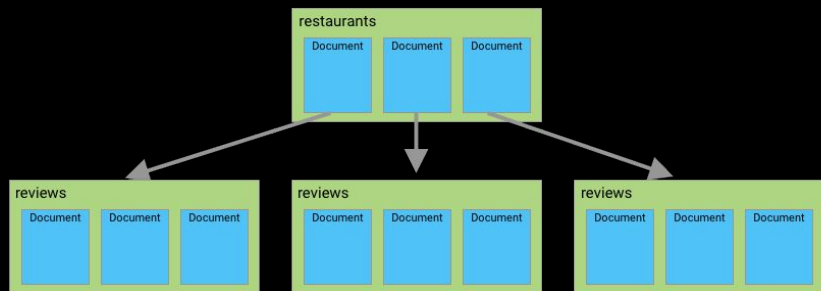
## Introduction to Relational Databases



# Topics

- **Structuring Data**
- **Tabular Data Example**
- **Designing Databases**
- **Joins Between Tables**

# Structuring Data



- We all structure the information we work with
    - so we can find what we need, when we need it
    - to facilitate evaluation, comparison and analysis
  - Choosing the right structure is important
-

# Structuring Data



## The structure you select influences...

- The kinds of information you collect
  - How it's possible to interrogate your data
  - The extent to which you can take advantage of your computer's data-handling abilities
  - How easy it is to share data with other
-

# Options for structuring and analysing data

## Tabular data

- Spreadsheets
- Google Sheets

## Relational databases

- Microsoft Access
- Filemaker Pro
- MySQL
- PostgreSQL

## Non-tabular data

- Document-orientated database
  - includes XML databases
- RDF triplestores
  - linked data on the web
- Qualitative data analysis packages
  - NVivo
  - ATLAS.ti

# When do use a relational database

- Your data can be organised in tabular form
  - e.g information about things that share common properties
- You are interested in multiple types of entity
  - and the relationships between them
  - entities may be concrete or more abstract
- You want to identify instances of things that meet certain criteria
- You want to be able to present one dataset in multiple different ways
  - query results can be exported and used elsewhere

# Benefits of relational database

- More accurate representation of complex data
  - and helps avoid duplication of information
- Permits flexible querying
  - wider range of questions possible vs spreadsheet
- Suitable for collaborative use
  - multiple users can access and use same database
  - can encourage (or enforce) consistency in data entry
- Technology has been around for several decades
  - widely supported and well understood

# Tabular data example



# Video: SQL vs No SQL

# Table of bibliographic data

Author	Title	Publisher	Year	Type
C.S. Lewis	The Problem of Pain	Centenary Press	1940	Non-fiction
C. S. Lewis	Miracles: A Preliminary Study	Geoffrey Bles	1947	Non-fiction
Clive Staples Lewis	Miracles: A Preliminary Study	Fontana	1960	Non-fiction
Lewis, C. S.	Till We Have Faces	Geoffrey Bles	1956	Novel
David Lewis	On the Plurality of Worlds	Blackwell	1986	Non-fiction
David Mitchell	Cloud Atlas	Sceptre	2004	Novel
David Mitchell	Back Story: A Memoir	HarperCollins	2012	Autobiography

# Table of bibliographic data

Author	Title	Publisher	Year	Type
C.S. Lewis	The Problem of Pain	Center		tion
C. S. Lewis	Miracles: A Preliminary Study	Geor		tion
Clive Staples Lewis	Miracles: A Preliminary Study	Fontan		tion
Lewis, C. S.	Till We Have Faces	Geoffr		
David Lewis	On the Plurality of Worlds	Blackwell	1986	Non-fiction
David Mitchell	Cloud Atlas	Seep		
David Mitchell	Back Story: A Memoir	Harper		graphy

One author, four different name formats

One name, two authors

# We can refactor and simplify

Author	Birth year	Death year	Notes on author	Title	Publisher	Year	Type
C.S. Lewis	1898	1963		The Problem of Pain	Centenary Press	1940	Novel
C. S. Lewis	1898	1963		Miracles: A Preliminary Study	Geoffrey Bles	1947	Novel
Clive Staples Lewis	1898	1963		Miracles: A Preliminary Study	Fontana	1960	Novel
Lewis, C. S.	1898	1963		Till We Have Faces	Geoffrey Bles	1956	Novel
David Lewis	1941	2001		On the Plurality of Worlds	Blackwell	1986	Novel
David Mitchell	1974	-	Novelist	Cloud Atlas	Sceptre	2004	Novel
David Mitchell	1969	-	Comedian and actor	Back Story: A Memoir	HarperCollins	2012	Autobiography

# We can refactor and simplify

Author	Birth year	Death year	Notes on author	Title	Publisher	Year	Type
C.S. Lewis	1898	1963		The Problem of Pain	Centenary Press	1940	Non-fiction
C. S. Lewis	1898	1963			Geoffrey Bles	1947	Non-fiction
Clive Staples Lewis	1898	1963			Fontana	1960	Non-fiction
Lewis, C. S.	1898	1963			Geoffrey Bles	1956	Non-fiction
David Lewis	1941	2001			Blackwell	1986	Non-fiction
David Mitchell	1974	-	Novelist	Cloud Atlas	Sceptre	2004	Non-fiction
David Mitchell	1969	-	Comedian and actor	Back Story: A Memoir	HarperCollins	2012	Autobiography

But this involves  
lots of repetition

# We can refactor and simplify

Author	Birth year	Death year	Notes on author	Title	Publisher	Year	Type
C.S. Lewis	1898	1963		The Problem of Pain	Centenary Press	1940	Novel
C. S. Lewis	1898	1963			Geoffrey Bles	1947	Novel
Clive Staples Lewis	1898	1963			Montana	1960	Novel
Lewis, C. S.	1898	1963			Geoffrey Bles	1956	Novel
David Lewis	1941	2001			Blackwell	1986	Novel
David Mitchell	1974	-	Novelist		Jeptre	2004	Novel
David Mitchell	1969	-	Comedian and actor	Back Story: A Memoir	HarperCollins	2012	Autobiography

And may get confusing and unwieldy

# An alternative approach - separate table for author details

ID No	Name	Birth year	Death year	Notes
1	C. S. Lewis	1898	1963	
2	David Lewis	1941	2001	
3	David Mitchell	1974	-	Comedian and actor
4	David Mitchell	1969	-	Novelist

# An alternative approach - separate table for author details

Author	Title	Publisher		Year	Type																			
C. S. Lewis	The Problem of	<table><tr><th>ID No</th><th>Name</th><th>Birth year</th><th>Death year</th></tr><tr><td>1</td><td>C. S. Lewis</td><td>1898</td><td>1963</td></tr><tr><td>2</td><td>David Lewis</td><td>1941</td><td>2001</td></tr><tr><td>3</td><td>David Mitchell</td><td>1974</td><td>-</td></tr><tr><td>4</td><td>David Mitchell</td><td>1969</td><td>-</td></tr></table>	ID No	Name	Birth year	Death year	1	C. S. Lewis	1898	1963	2	David Lewis	1941	2001	3	David Mitchell	1974	-	4	David Mitchell	1969	-		
ID No	Name		Birth year	Death year																				
1	C. S. Lewis		1898	1963																				
2	David Lewis		1941	2001																				
3	David Mitchell		1974	-																				
4	David Mitchell	1969	-																					
C. S. Lewis	Miracles: A Preli																							
C. S. Lewis	Miracles: A Preli																							
C. S. Lewis	Till We Have Fac																							
David Lewis	On the Plurality																							
David Mitchell	Cloud Atlas																							
David Mitchell	Back Story: A Memoir	HarperCollins	2012	Autobiography																				



# An alternative approach - separate table for author details

Author	Title	Publisher	Year	Type
C. S. Lewis	The Problem of			
C. S. Lewis	Miracles: A Prel			
C. S. Lewis	Miracles: A Prel			
C. S. Lewis	Till We Have Fac			
David Lewis	On the Plurality			
David Mitchell	Cloud Atlas			
David Mitchell	Back Story: A Memoir	HarperCollins	2012	Autobiography

ID No	Name	Birth year	Death year
1	C. S. Lewis	1898	1963
2	David Lewis	1941	2001
3	David Mitchell	1974	-
4	David Mitchell	1969	-

# Further possible refinements

Author	Title	Publisher	Year	Type
C. S. Lewis	Publishers could also be split out into a separate table	Centenary Press	1940	Non-fiction
C. S. Lewis		Geoffrey Bles	1947	Non-fiction
C. S. Lewis		Fontana	1960	Non-fiction
C. S. Lewis		Geoffrey Bles	1956	Novel
David Lewis		Blackwell	1986	Non-fiction
David Mitchell	Cloud Atlas	Sceptre	2004	Novel
David Mitchell	Back Story: A Memoir	HarperCollins	2012	Autobiography

# Further possible refinements

Author	Title	Publisher	Year	Type
C. S. Lewis	The Problem of Pain	Centenary Press	1940	Non-fiction
C. S. Lewis	Miracles: A Preliminary Study	Centenary Press	1947	Non-fiction
C. S. Lewis	Miracles: A Preliminary Study	Centenary Press	1960	Non-fiction
C. S. Lewis	Till We Have Faces	Centenary Press	1956	Novel
David Lewis	On the Plurality of Worlds	Blackwell	1986	Non-fiction
David Mitchell	Cloud Atlas	Sceptre	2004	Novel
David Mitchell	Back Story: A Memoir	HarperCollins	2012	Autobiography

We could create a standardised list of types

# Further possible refinements

Author	Title	Publisher	Year	Type
C. S. Lewis	The Problem of Pain	Cent		
C. S. Lewis	Miracles: A Preliminary Study	G		
C. S. Lewis	Miracles: A Preliminary Study	Fonta		
C. S. Lewis	Till We Have Faces	Geoff		
David Lewis	On the Plurality of Worlds	Blackv		
David Mitchell	Cloud Atlas	Scept		
David Mitchell	Back Story: A Memoir	Harpe		

We could distinguish different editions of the same title

The right relational database structure lets us do all this

# Video: Database & Entity Diagrams

<https://youtu.be/wR0jg0eQsZA>

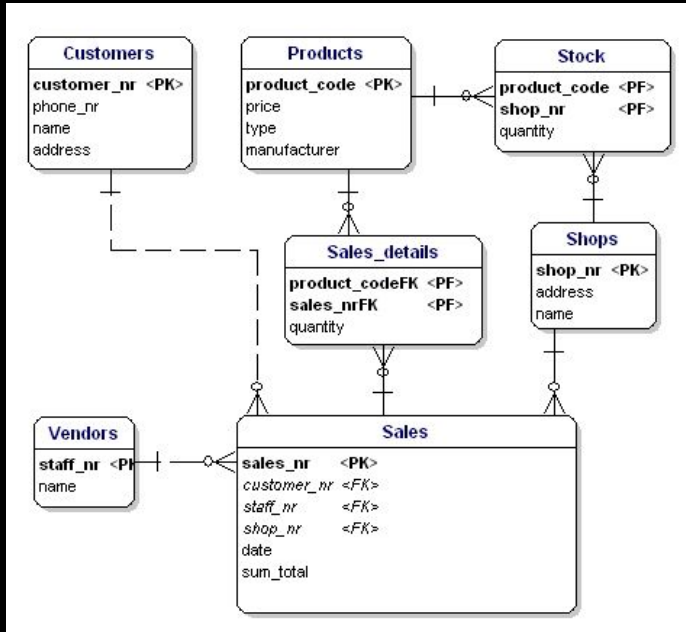
# Designing a Database

# Database Terms

	<b>Name</b>	<b>Role</b>	<b>Town</b>
<i>record 1</i>	Peter	farmer	Oxford
<i>record 2</i>	Mary	weaver	Winchester
<i>record 3</i>	Seth	drover	Bristol

- A database is a collection of data
- Data is organized into one or more **tables**
- Each **row** is a record
- Each **column** is a field

# Designing the database



- Decide on the fields
  - Think of all the facts that will be collected
- Designing the tables
  - Choose the fields, then group them in tables



# Types of data

People	
Surname	text
First name	text
Middle initial(s)	text
Date of birth	date
...	
Notes	memo

Books	
Title	text
Author	text
DatePub	date
...	
Place	text
ISBN	text
...	
...	

Set a data type for each fields:

- Text
- Number
- Date/Time
- Currency
- Boolean

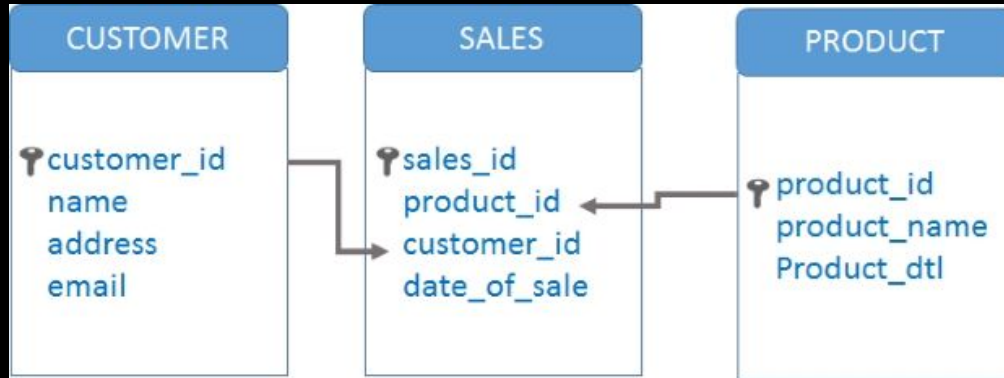
# Joins Between Tables

# Primary Key

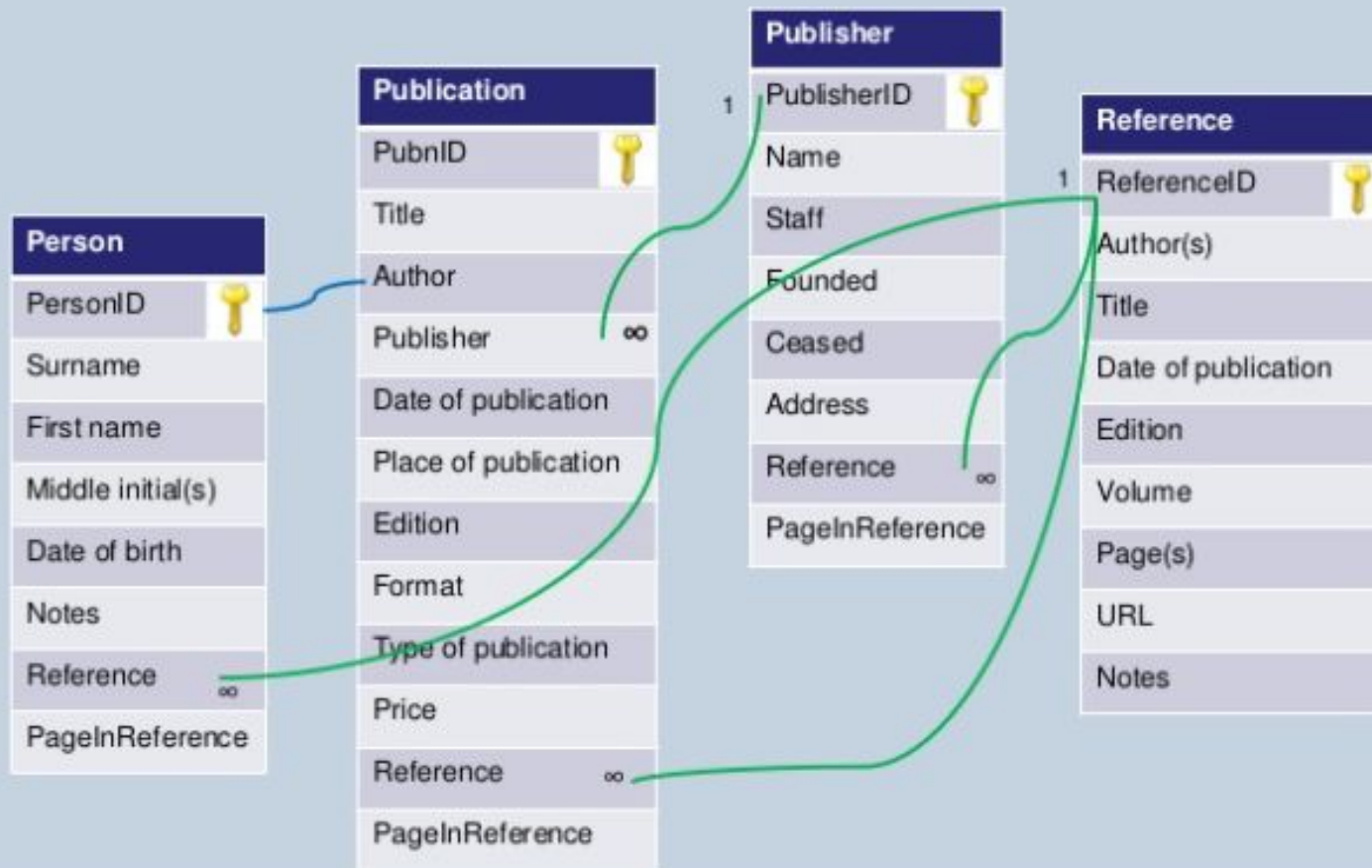
Person	
PersonID	
Surname	
First name	
Middle initial(s)	
Date of birth	
Notes	

- Each table needs a primary key
- Choose (at least) one field that only contains unique values
  - commonly an auto-incrementing whole (integer) number

# Relating two tables - joins



- Mark the field that links this table to that table
- Draw join lines
- Convenient to have same or similar field names



# Types of Relationships

In a **one-to-one relationship**, one record in a table is associated with one and only one record in another table.

In a **one-to-many relationship**, one record in a table can be associated with **one** or more records in another table

**Many-to-many relationships.** A **many-to-many relationship** occurs when multiple records in a table are associated with multiple records in another table.

