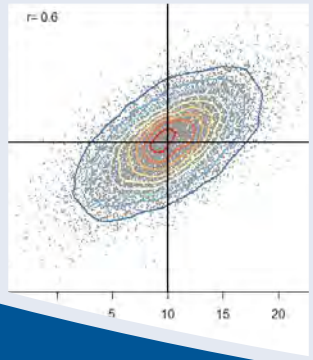
RESEARCH DATA SERVICES

Offering expert data assistance at every stage of the research process.



1: PLANNING

We can assist you with developing a data management plan and designing your planned data analysis, including:

- Implementing plans, using tools, and creating workflows for managing research data
- Advising on study design, power analysis, and choice of statistical methods
- Helping to meet increasingly stringent criteria from funding agencies

2: FINDING & COLLECTING

We have access to thousands of sources of data and experts who will help you:

- Locate, evaluate and format data
- Create metadata and data documentation protocols for new data collection
- Capture data using best practices and appropriate technology

3: ANALYZING

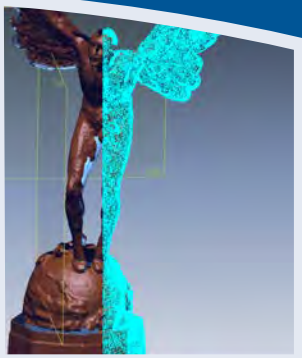
Get expert assistance from statistical, spatial, or media specialists to analyze your data and present your research:

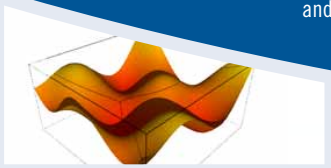
- Learn to use cutting-edge tools and methods
- Experiment with high-resolution visualization technologies
- Develop graphical representations that bring impact to your analysis

4: SHARING & ARCHIVING

We can consult with you on strategies to help others discover or access your research by:

- Adhering to data sharing policies and norms
- Selecting a data-sharing repository
- Making your data easier to discover and reuse



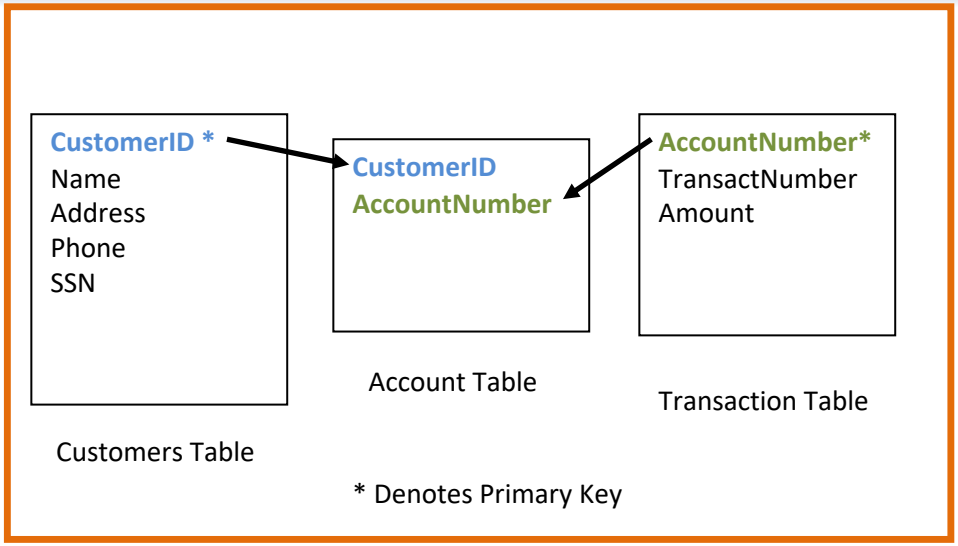


UNIVERSITY of VIRGINIA LIBRARY

Workshops • 1:1 Consultations • Class Presentations

Contact us at researchdataservices@virginia.edu

Introduction to Designing and Building Databases



Sherry Lake
Data Management Consultant
University of Virginia Library
shLake@virginia.edu

Nancy Kechner
Research & Computing Support Specialist
University of Virginia Library
nek8n@virginia.edu

Goals for the Workshop

- Understand differences between Spreadsheets and Databases
- Learn the features of databases
- Learn the 3 steps of database design
- Design a database in MS Access

Some Vocabulary

- **Database** – set of (related) tables
- **Table** – set of rows and columns
- **Column** – field, attribute
- **Row** – Tuple, observation, case
- **RDBMS** – Relational Database Management System

Spreadsheets vs. Database

Spreadsheets are best if:

- Data can be stored in a single datasheet without lots of redundancy
- You are doing calculations or making charts
- Don't need to link several spreadsheets together to get the results you want

Databases are best if:

- Data are readily stored in multiple related tables
- You need multiple user access
- You want to be able to do complex manipulations with the data
- You want to develop data entry tools

Big Issue with Spreadsheets

- Data integrity
  - Internal record consistency is not maintained
  - Updating more than one record
  - Removing information
  - Creating incomplete cells

ID	DeptName	DeptAddress	ContactName	ContactTitle	ContactPhone
1	Finance	110 5th Street	Michael Jones	Manager	555-1111
2	Finance	110 5th Street	Ted Smith	Senior Analyst	555-1112
3	Benefits	118 5th Street	Brian Williams	Manager	555-3333

Database Features

- Collection of data organized into tables
- Each table contains records
- Each record identifies the same set of fields
- Explicit control over data (column) types in tables

Date	Site	Height	Count
<dates only>	<text only>	< real numbers only>	< integers only>

Database Features

- Relationships are defined between tables
- Tools help you manage the table relationships

Date	Site	Species	Height	Diameter
	A			
	B			
	A			
	C			

Site	Latitude	Longitude
A		
B		
C		
D		

3 Steps to Database Design

1. Split Data Into Tables
   - Normalization
   - Each field must contain only one value
   - Each field must have a unique name
2. Determine Data Type for Each Column
3. Identify Relationships Between the Tables
   - No two records can be identical

Normalization

- Process of efficiently organizing data in a database

- Goals of the normalization process:
  - eliminating redundant data (for example, storing the same data in more than one table)
  - ensuring data dependencies make sense (only storing related data in a table)

Normalization Process

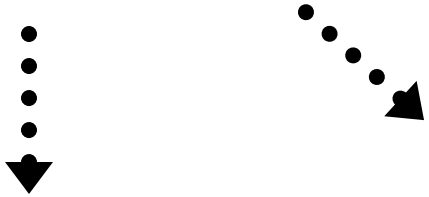
1st Normal Form

- Eliminate duplicate columns from the same table
- Create separate tables for each group of related data
- Identify each row with a unique column or set of columns (the primary key)

First Normal: Eliminate Duplicate Columns and Assign Keys

Books

TITLE	AUTHOR 1	AUTHOR 2	PUBLISHER	ISBN	QTY.
Ecology 101	Smith, A.B.	Gordon, D.A.	Univ. Press	4873895759	4324
Ecology for Dummies	Doe, J.		Wiley & Sons	0493802020	8998
Ecology and Politics	Kim, J.B.		McGraw-Hill	7482929292	900
Ecology and Modern Cinema	Kim, C.B.		Univ. Press	2234849302	1



Books

TITLE	PUBLISHER	ISBN	QTY.
Ecology 101	Univ. Press	4873895759	4324
Ecology for Dummies	Wiley & Sons	0493802020	8998
Ecology and Politics	McGraw-Hill	7482929292	900
Ecology and Modern Cinema	Univ. Press	2234849302	1

Authors

Author
Smith, A.B.
Gordon, D.A.
Doe, J.
Kim, J.B.
Kim, C.B.

First Normal: Eliminate Duplicate Columns and Assign Keys

Books

TITLE	AUTHOR 1	AUTHOR 2	PUBLISHER	ISBN	QTY.
Ecology 101	Smith, A.B.	Gordon, D.A.	Univ. Press	4873895759	4324
Ecology for Dummies	Doe, J.		Wiley & Sons	0493802020	8998
Ecology and Politics	Kim, J.B.		McGraw-Hill	7482929292	900
Ecology and Modern Cinema	Kim, C.B.		Univ. Press	2234849302	1

Primary Key

Books

TITLE	PUBLISHER	ISBN	QTY.
Ecology 101	Univ. Press	4873895759	4324
Ecology for Dummies	Wiley & Sons	0493802020	8998
Ecology and Politics	McGraw-Hill	7482929292	900
Ecology and Modern Cinema	Univ. Press	2234849302	1

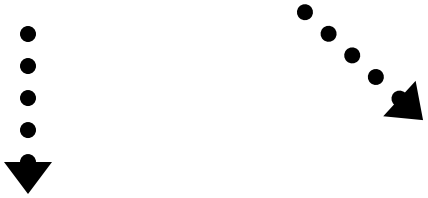
Authors

Id	Author
1	Smith, A.B.
2	Gordon, D.A.
3	Doe, J.
4	Kim, J.B.
5	Kim, C.B.

First Normal: Eliminate Duplicate Columns and Assign Keys

Books

TITLE	AUTHOR 1	AUTHOR 2	PUBLISHER	ISBN	QTY.
Ecology 101	Smith, A.B.	Gordon, D.A.	Univ. Press	4873895759	4324
Ecology for Dummies	Doe, J.		Wiley & Sons	0493802020	8998
Ecology and Politics	Kim, J.B.		McGraw-Hill	7482929292	900
Ecology and Modern Cinema	Kim, C.B.		Univ. Press	2234849302	1



Books

TITLE	PUBLISHER	ISBN	QTY.
Ecology 101	Univ. Press	4873895759	4324
Ecology for Dummies	Wiley & Sons	0493802020	8998
Ecology and Politics	McGraw-Hill	7482929292	900
Ecology and Modern Cinema	Univ. Press	2234849302	1

Authors

Foreign Key

Id	ISBN	Author
1	4873895759	Smith, A.B.
2	4873895759	Gordon, D.A.
3	0493802020	Doe, J.
4	7482929292	Kim, J.B.
5	2234849302	Kim, C.B.

First Normal Exercise

Pause for exercise

Personnel

Last	First	M.I.	Institution	Sector	Position 1	Position 2
Smith	Ann	A	SDSU	Academic	Professor	Community Liaison
Smith	Ann	Z	Acme Inc.	Private	Administrator	Field Technician
Kim	John	B	SDSU	Academic	P.I.	Data Manager

First Normal

Eliminate duplicate columns

Last	First	M.I.	Institution	Sector	Position 1	Position 2
Smith	Ann	A	SDSU	Academic	Professor	Community Liaison
Smith	Ann	Z	Acme Inc.	Private	Administrator	Field Technician
Kim	John	B	SDSU	Academic	P.I.	Data Manager

personnel

Pers_id	Last	First	M.I.	Institution	Sector
1	Smith	Ann	A	SDSU	Academic
2	Smith	Ann	Z	Acme Inc.	Private
3	Kim	John	B	SDSU	Academic

positions

Foreign Key

Pos_id	Pers_id	Position
1	1	Professor
2	1	Community Liaison
3	2	Administrator
4	2	Field Technician
5	3	P.I.
6	3	Data Manager

Normalization Process

2nd Normal Form

- Meet all the requirements of the first normal form
- Remove subsets of data that apply to multiple rows of a table and place them in separate tables
- Create relationships between these new tables and their predecessors through the use of foreign keys

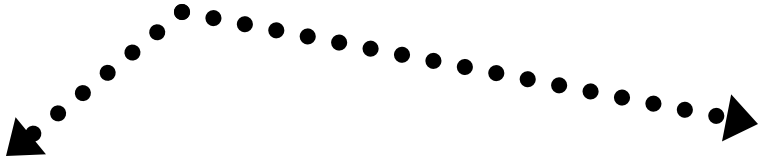
Second Normal: Eliminate Duplicate Rows and Assign Keys

Books

TITLE	PUBLISHER	ISBN	QTY.
Ecology 101	Univ. Press	4873895759	4324
Ecology for Dummies	Wiley & Sons	0493802020	8998
Ecology and Politics	McGraw-Hill	7482929292	900
Ecology and Modern Cinema	Univ. Press	2234849302	1

Authors

Id	ISBN	Author
1	4873895759	Smith, A.B.
2	4873895759	Gordon, D.A.
3	0493802020	Doe, J.
4	7482929292	Kim, J.B.
5	2234849302	Kim, C.B.



Books

TITLE	ISBN	QTY.
Ecology 101	4873895759	4324
Ecology for Dummies	0493802020	8998
Ecology and Politics	7482929292	900
Ecology and Modern Cinema	2234849302	1

Publishers

PUBLISHER
Univ. Press
Wiley & Sons
McGraw-Hill

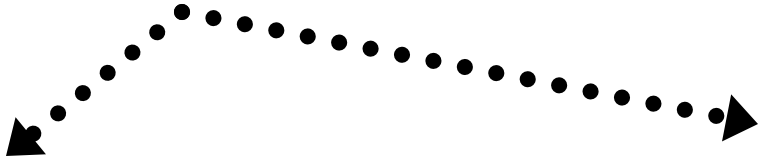
Second Normal: Eliminate Duplicate Rows and Assign Keys

Books

TITLE	PUBLISHER	ISBN	QTY.
Ecology 101	Univ. Press	4873895759	4324
Ecology for Dummies	Wiley & Sons	0493802020	8998
Ecology and Politics	McGraw-Hill	7482929292	900
Ecology and Modern Cinema	Univ. Press	2234849302	1

Authors

Id	ISBN	Author
1	4873895759	Smith, A.B.
2	4873895759	Gordon, D.A.
3	0493802020	Doe, J.
4	7482929292	Kim, J.B.
5	2234849302	Kim, C.B.



Books

TITLE	ISBN	QTY.
Ecology 101	4873895759	4324
Ecology for Dummies	0493802020	8998
Ecology and Politics	7482929292	900
Ecology and Modern Cinema	2234849302	1

Publishers

Publisher_id	PUBLISHER
1	Univ. Press
2	Wiley & Sons
3	McGraw-Hill

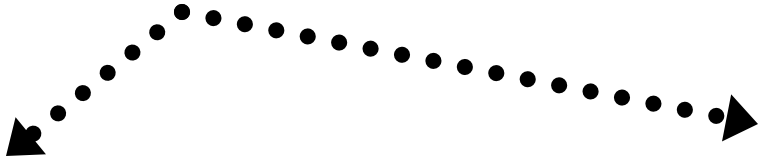
Second Normal: Eliminate Duplicate Rows and Assign Keys

Books

TITLE	PUBLISHER	ISBN	QTY.
Ecology 101	Univ. Press	4873895759	4324
Ecology for Dummies	Wiley & Sons	0493802020	8998
Ecology and Politics	McGraw-Hill	7482929292	900
Ecology and Modern Cinema	Univ. Press	2234849302	1

Authors

Id	ISBN	Author
1	4873895759	Smith, A.B.
2	4873895759	Gordon, D.A.
3	0493802020	Doe, J.
4	7482929292	Kim, J.B.
5	2234849302	Kim, C.B.



Books

TITLE	PUBLISHER_id	ISBN	QTY.
Ecology 101	1	4873895759	4324
Ecology for Dummies	2	0493802020	8998
Ecology and Politics	3	7482929292	900
Ecology and Modern Cinema	1	2234849302	1

Publishers

Publisher_id	PUBLISHER
1	Univ. Press
2	Wiley & Sons
3	McGraw-Hill

Final Tables with Primary and Foreign Keys

Books

TITLE	PUBLISHER_id	ISBN	QTY.
Ecology 101	1	4873895759	4324
Ecology for Dummies	2	0493802020	8998
Ecology and Politics	3	7482929292	900
Ecology and Modern Cinema	1	2234849302	1

Authors

Id	ISBN	Author
1	4873895759	Smith, A.B.
2	4873895759	Gordon, D.A.
3	0493802020	Doe, J.
4	7482929292	Kim, J.B.
5	2234849302	Kim, C.B.

Publishers

Publisher_id	PUBLISHER
1	Univ. Press
2	Wiley & Sons
3	McGraw-Hill

Second Normal Exercise

Pause for exercise

personnel

Pers_id	Last	First	M.I.	Institution	Sector
1	Smith	Ann	A	SDSU	Academic
2	Smith	Ann	Z	Acme Inc.	Private
3	Kim	John	B	SDSU	Academic

positions

Pos_id	Pers_id	Position
1	1	Professor
2	1	Community Liaison
3	2	Administrator
4	2	Field Technician
5	3	P.I.
6	3	Data Manager

Second Normal

Eliminate duplicate rows

personnel

Pers_id	Last	First	M.I.	Institution	Sector
1	Smith	Ann	A	SDSU	Academic
2	Smith	Ann	Z	Acme Inc.	Private
3	Kim	John	B	SDSU	Academic

positions

Pos_id	Pers_id	Position
1	1	Professor
2	1	Community Liaison
3	2	Administrator
4	2	Field Technician
5	3	P.I.
6	3	Data Manager

personnel

Pers_id	Last	First	M.I.	Institution_id
1	Smith	Ann	A	1
2	Smith	Ann	Z	2
3	Kim	John	B	1

institutions

Institution_id	Institution	Sector
1	SDSU	Academic
2	Acme Inc.	Private

Final Tables with Primary and Foreign Keys

personnel

Pers_id	Last	First	M.I.	Institution_id
1	Smith	Ann	A	1
2	Smith	Ann	Z	2
3	Kim	John	B	1

positions

Pos_id	Pers_id	Position
1	1	Professor
2	1	Community Liaison
3	2	Administrator
4	2	Field Technician
5	3	P.I.
6	3	Data Manager

institutions

Institution_id	Institution	Sector
1	SDSU	Academic
2	Acme Inc.	Private

Determine Data Types

Data Type	Definition
Text	0-255 characters
Memo	0-64000 characters
Number	Integer, long integer, single, double
Date/Time	Dates, times, or both at once
AutoNumber	Automatically incremented as records are added
OLE object	Image, sound files
Hyperlink	Link to an internet resource

Determine Data Types

Books

TITLE	Text (255)
PUBLISHER_id	Number (integer)
ISBN	Text(10)
QTY.	Number(integer)

Authors

Id	Number(integer)
ISBN	Text(10)
Author	text(255)

Publishers

PUBLISHER_id	integer
PUBLISHER	text(255)

Determine Data Types Exercise

Group exercise

Personnel

Pers_id	
Last	
First	
M.I.	
Institution_id	

Institutions

Institution_id	
Institution	
Sector	

Positions

Pos_id	
Pers_id	
Position	

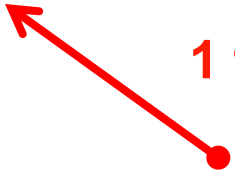
Identify the Relationships

- How the tables are “related” to each other
  - One-to-one
  - One-to-many
  - Many-to-many
- Foreign Keys define the relationships

Identify the Relationships

Books

TITLE	PUBLISHER_id	ISBN	QTY.
Ecology 101	1	4873895759	4324
Ecology for Dummies	2	0493802020	8998
Ecology and Politics	3	7482929292	900
Ecology and Modern Cinema	1	2234849302	1



Authors

Id	ISBN	Author
1	4873895759	Smith, A.B.
2	4873895759	Gordon, D.A.
3	0493802020	Doe, J.
4	7482929292	Kim, J.B.
5	2234849302	Kim, C.B.

Publishers

PUBLISHER_id	PUBLISHER
1	Harcourt Brace
2	Wiley & Sons
3	McGraw-Hill

Identify the Relationships

Group Exercise

personnel

Pers_id	Last	First	M.I.	Institution_id
1	Smith	Ann	A	1
2	Smith	Ann	Z	2
3	Kim	John	B	1

positions

Pos_id	Pers_id	Position
1	**1**	Professor
2	**1**	Community Liaison
3	**2**	Administrator
4	**2**	Field Technician
5	**3**	P.I.
6	**3**	Data Manager

1 to many

1 to many

1 to 1

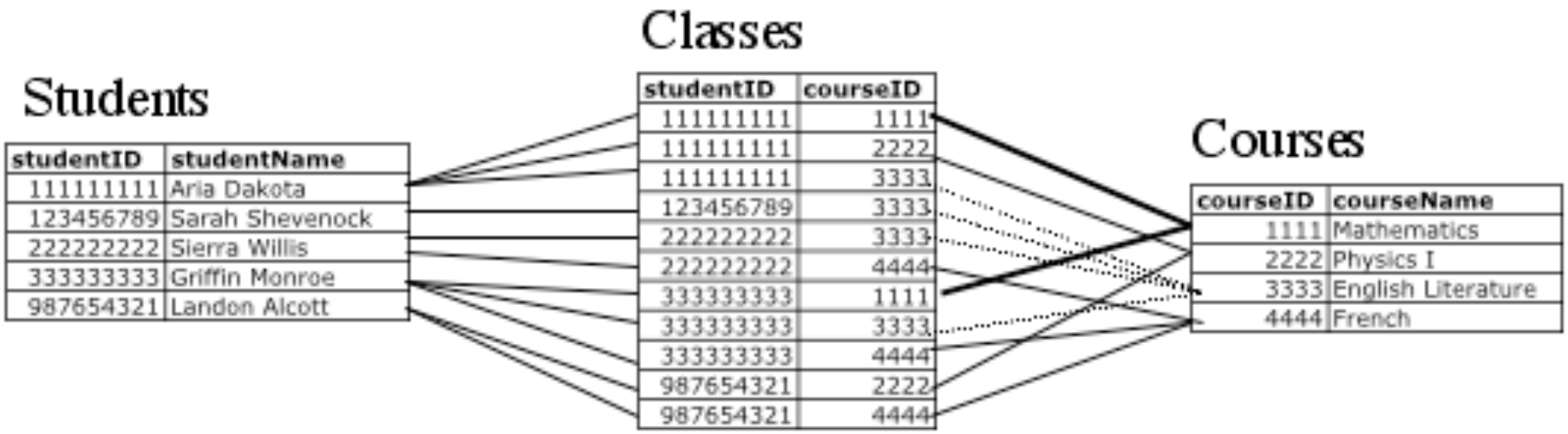
institutions

Institution_id	Institution	Sector
1	SDSU	Academic
2	Acme Inc.	Private

contact_address

Pers_id	Street	City	State
1	523 Main St.	Amherst	MA
2	1010 Sea St.	San Diego	CA
3	99 Ridge Way	Portland	ME

Identify the Relationships



Many-to-many: need to create a 3rd table to relate the two

Relational Database Functions

- Organize data – reduce or eliminate redundancy
- Improve data quality – reject “bad” data
  - Wrong type of data
  - Only good “codes” allowed
- Retrieve data – query/search/select
- Sort data
- Update data
- Output – link to other software with statistical and graphical functionality

Database Systems

Access

- Access workstation based, single user application
- Platform dependent
- Cannot be accessed concurrently
- No security other than workstation
- Part of MS Office suite, not free

MySQL

- MySQL is cross platform, multi user access
- Accessible to more users thru the web, client program or other admin tools to access database (via authentication)
- Can be integrated with Web Server (web programming languages)
- Data available remotely
- Free, open-source

Designing a Database in MS Access

Closer Look at Access

- Access workstation based, single user application
- Platform dependent
- Cannot be accessed concurrently
- No security other than workstation
- Part of MS Office suite, not free

Personnel

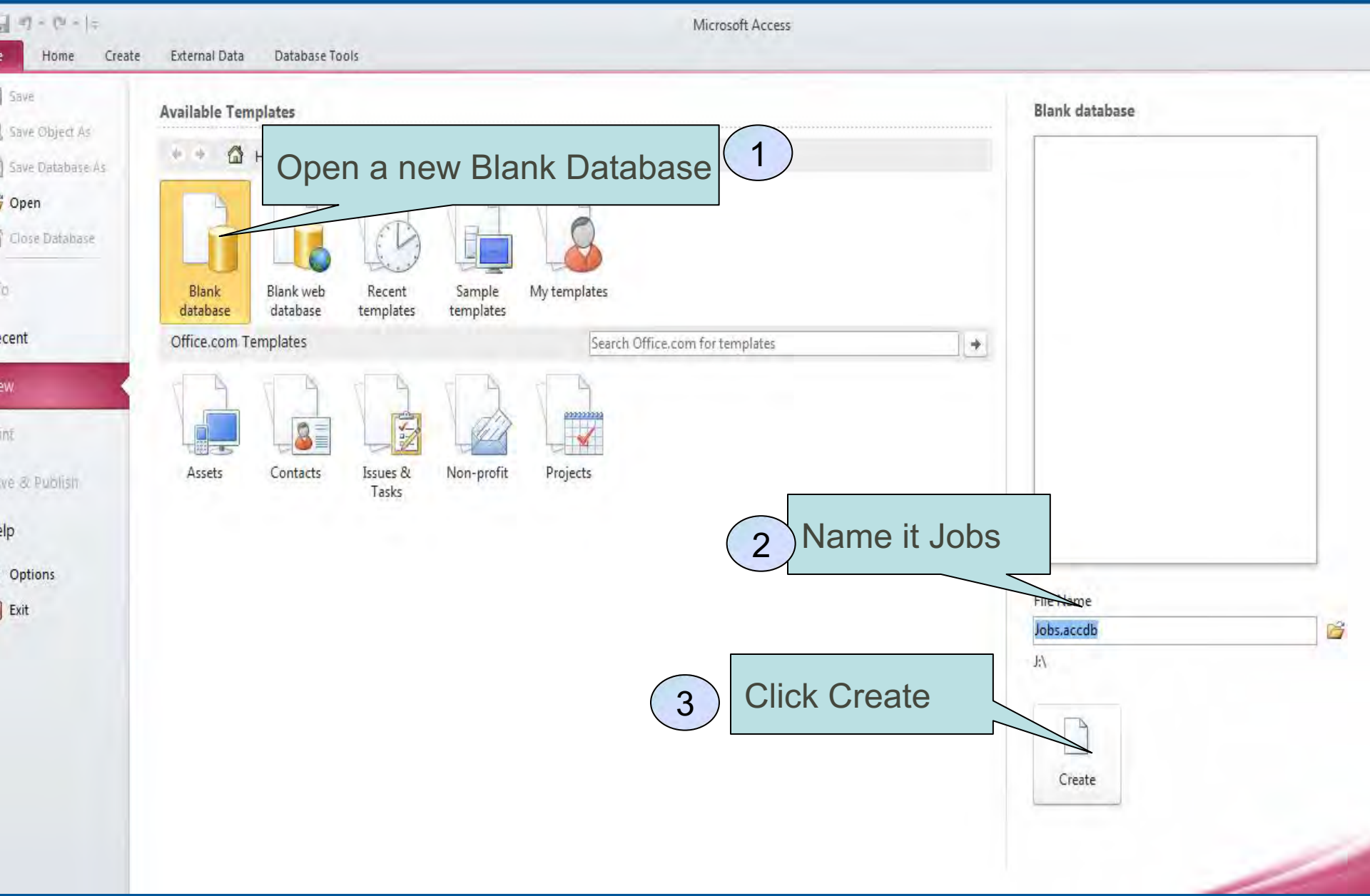
Pers_id	Last	First	M.I.	Institution_id
0	Smith	Ann	A	0
1	Smith	Ann	Z	1
2	Kim	John	B	0

Positions

Pos_id	Pers_id	Position
0	0	Professor
1	0	Community Liaison
2	1	Administrator
3	1	Field Technician
4	2	P.I.
5	2	Data Manager

Institutions

Institution_id	Institution	Sector
0	SDSU	Academic
1	Acme Inc.	Private


Microsoft Access
Home
Create
External Data
Database Tools
Save
Save Object As
Save Database As
Open
Close Database
Options
Exit
Available Templates
Blank database
Blank web database
Recent templates
Sample templates
My templates
Office.com Templates
Search Office.com for templates
Assets
Contacts
Issues & Tasks
Non-profit
Projects
Blank database
Jobs.accdb
J:\
Create
Open a new Blank Database
1
2
Name it Jobs
3
Click Create

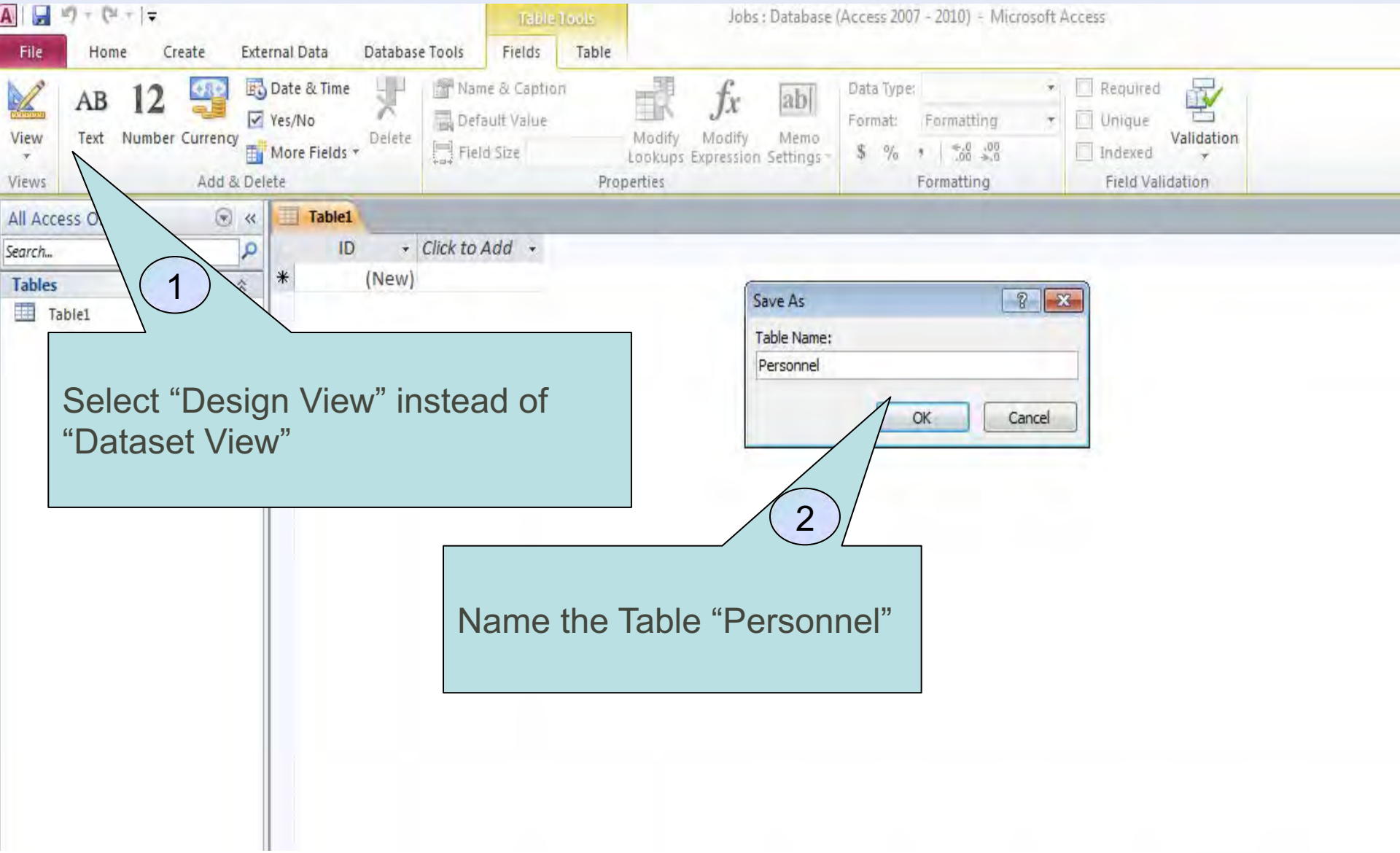
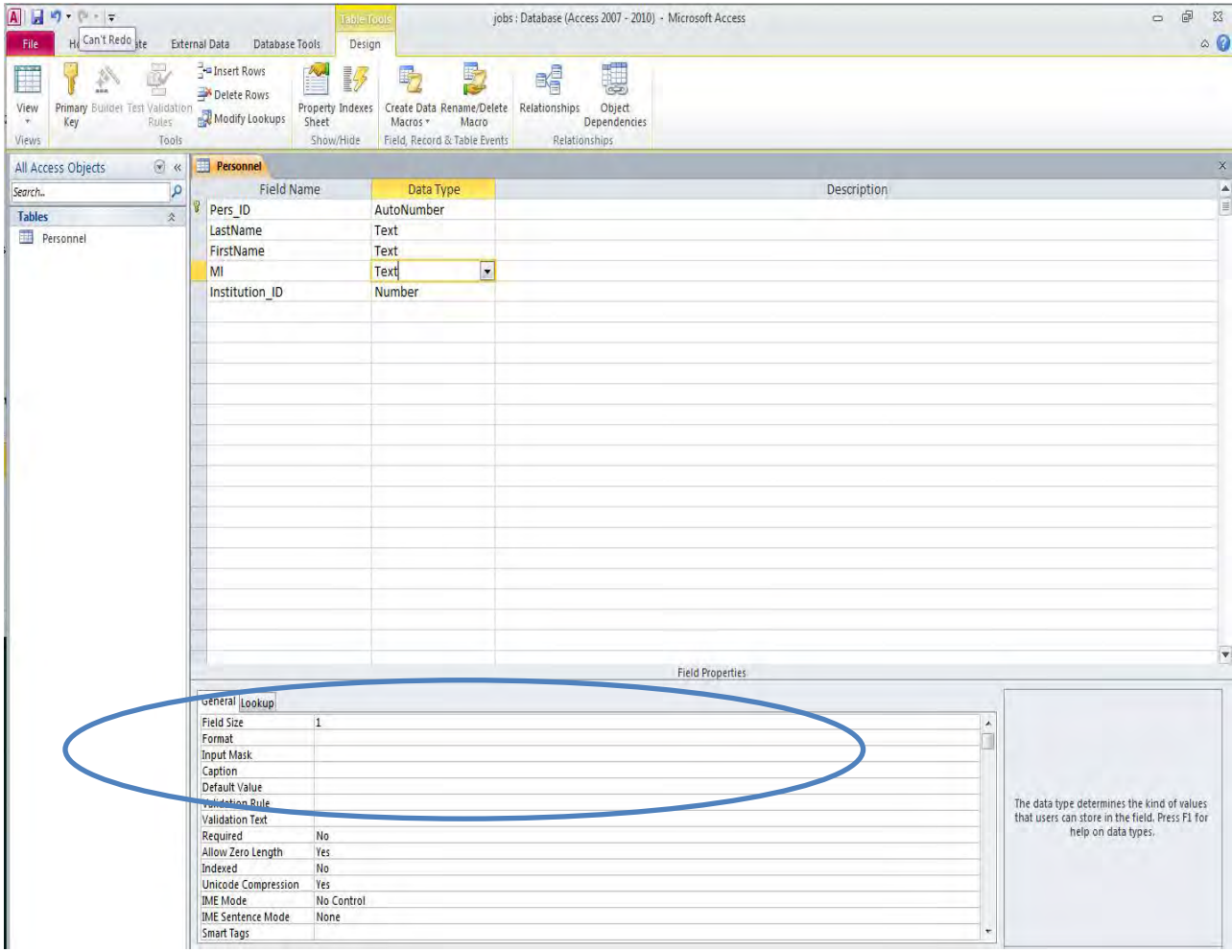


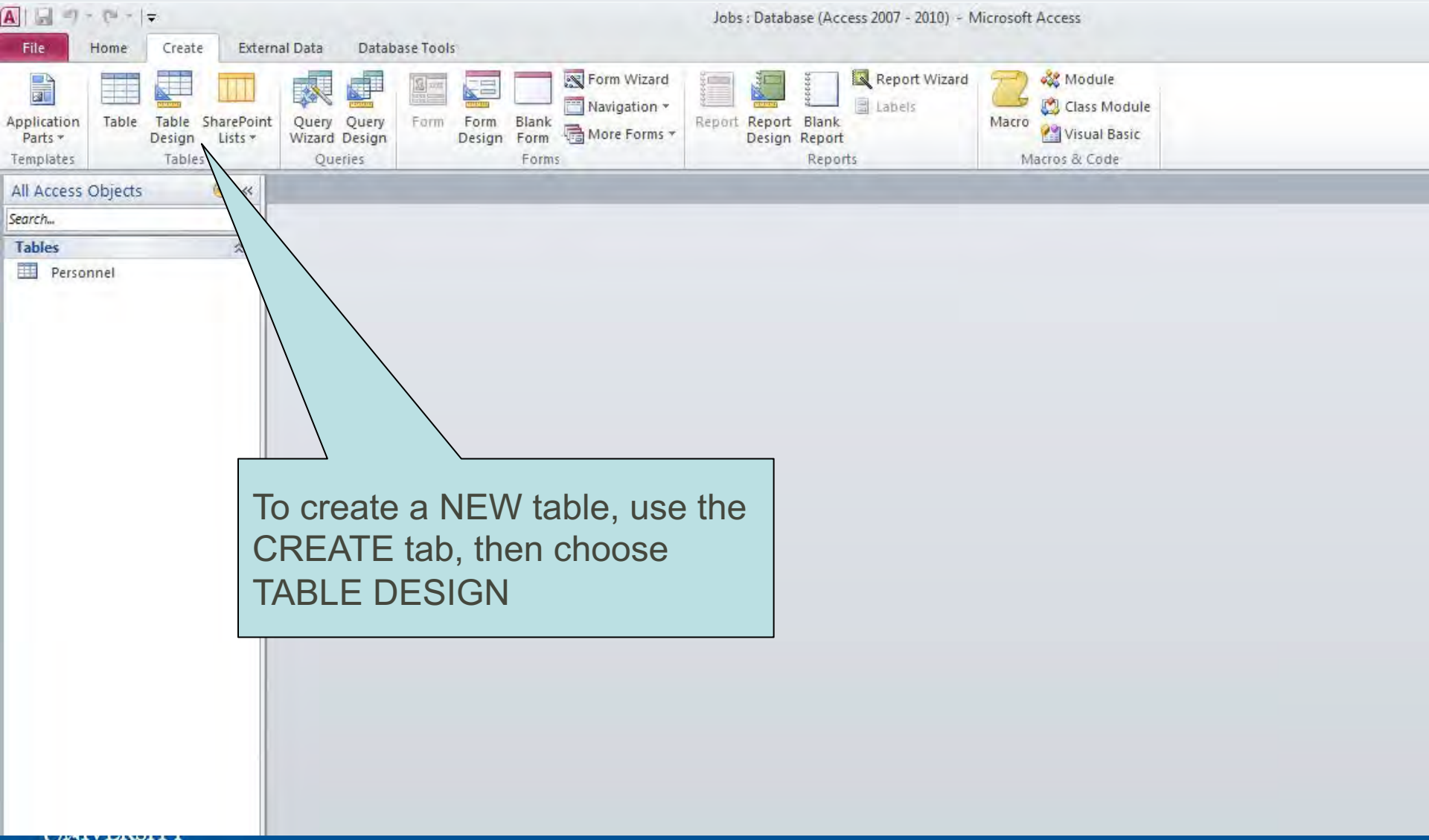
Table Tools
Jobs : Database (Access 2007 - 2010) - Microsoft Access
File
Home
Create
External Data
Database Tools
Fields
Table
View
Text
Number
Currency
Date & Time
Yes/No
More Fields
Delete
Name & Caption
Default Value
Field Size
Modify Lookups
Modify Expression
Memo Settings
Data Type:
Format:
Formatting
Required
Unique
Indexed
Validation
Views
Add & Delete
Properties
Formatting
Field Validation
All Access O
Search...
Tables
Table1
Table1
ID
Click to Add
(New)
1
Select “Design View” instead of “Dataset View”
Save As
Table Name:
Personnel
OK
Cancel
2
Name the Table “Personnel”

Define the Personnel Table

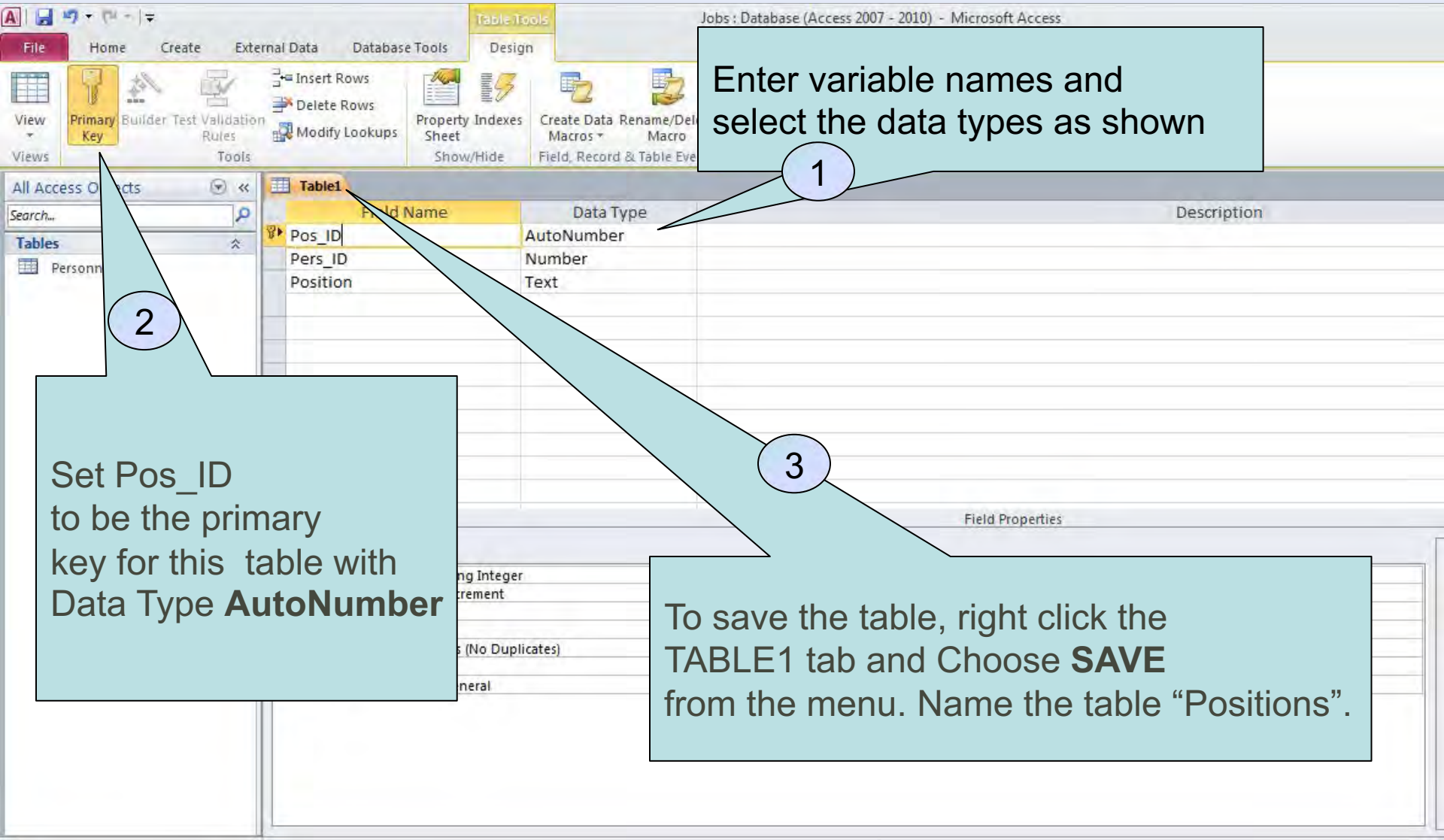
- Enter variable names and select the data types as shown
- Note that field size and other properties can be changed in the "Field Properties" area
- If there is NO key, next to **Pers_ID**, Right click on the box to the left of Pers_ID and select *Primary Key*
- Click the disk icon to save the changes to the Personnel table
- Close the table by clicking on the X on the right-hand side



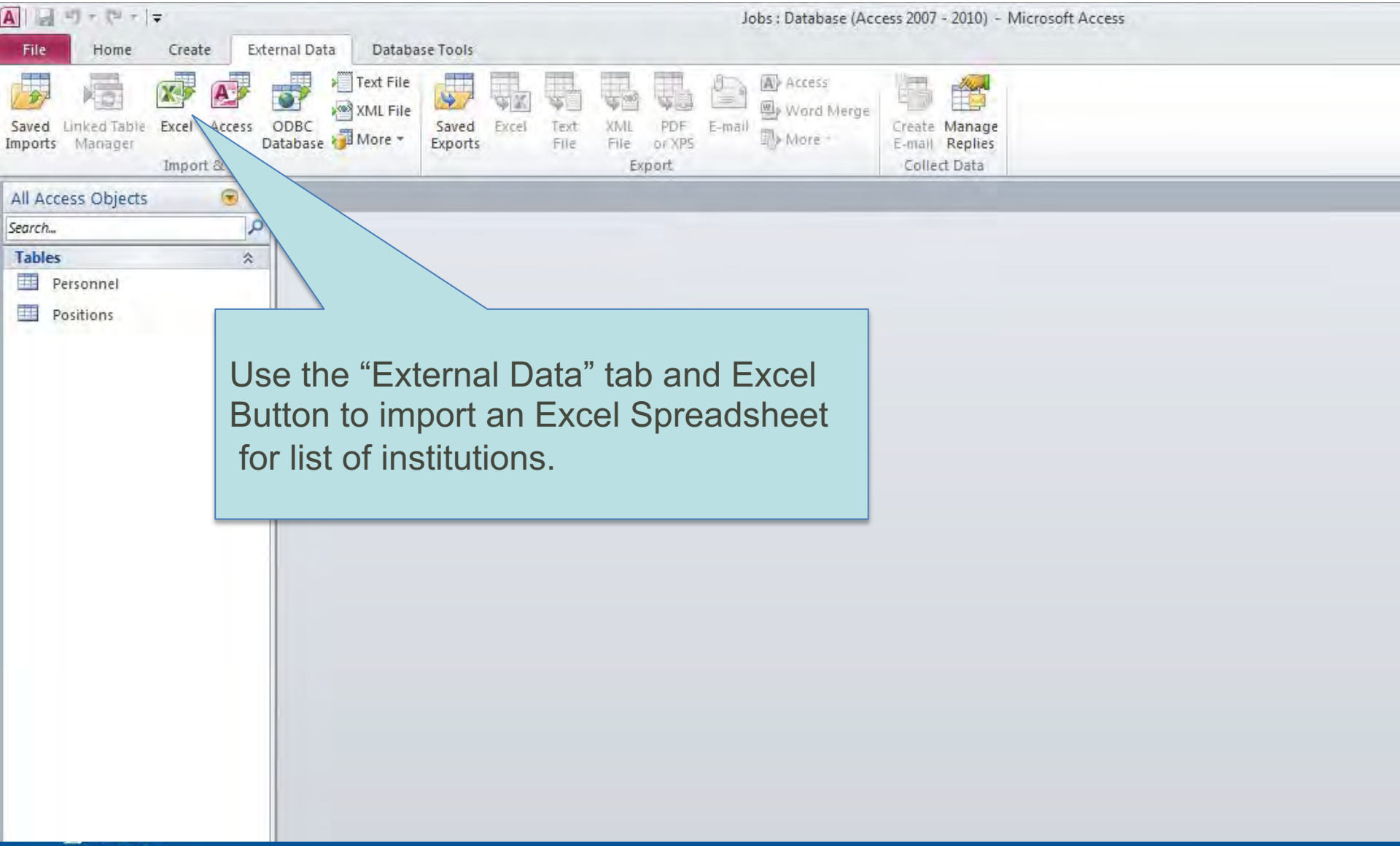
Define Positions Table



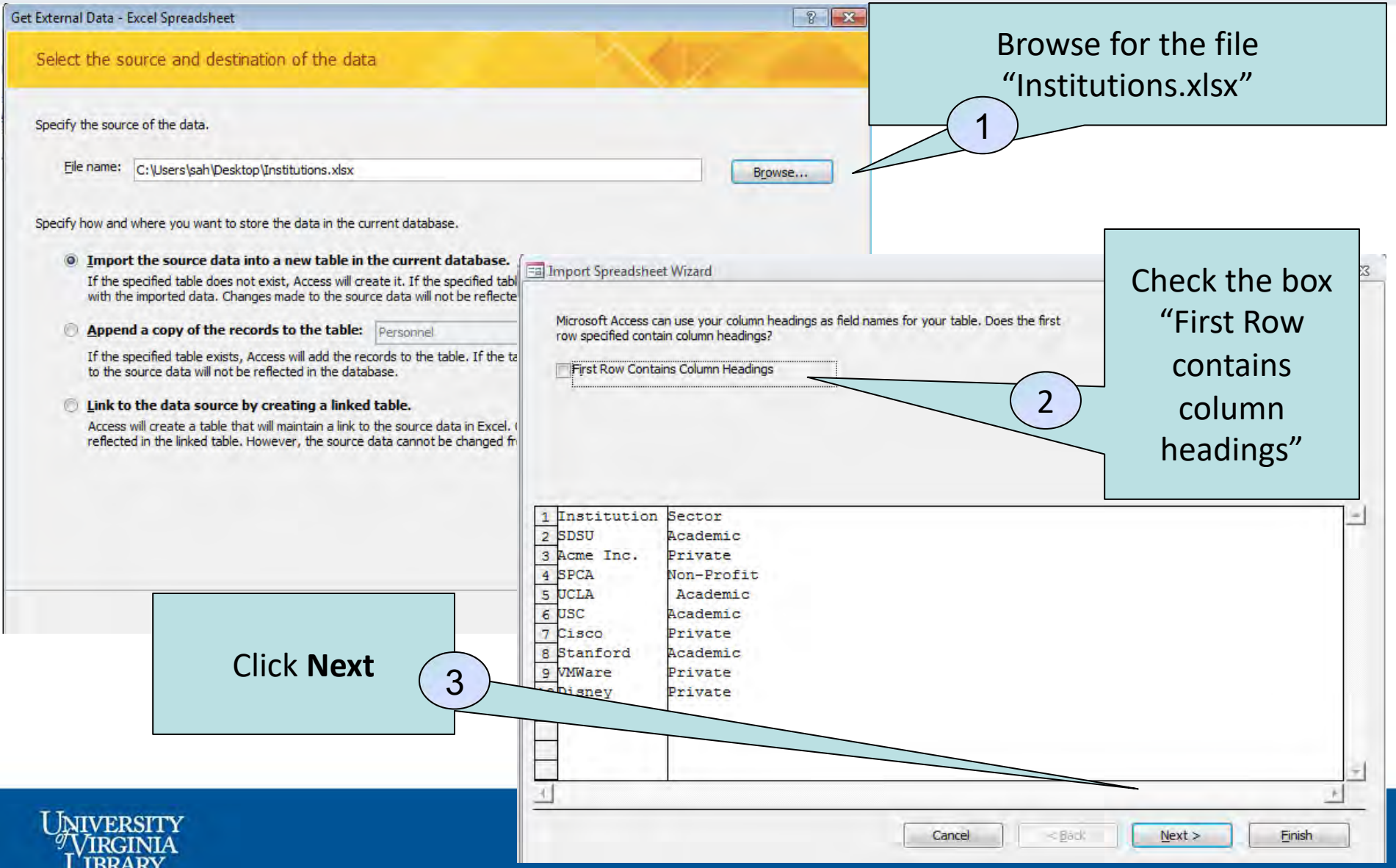
Define Positions Table

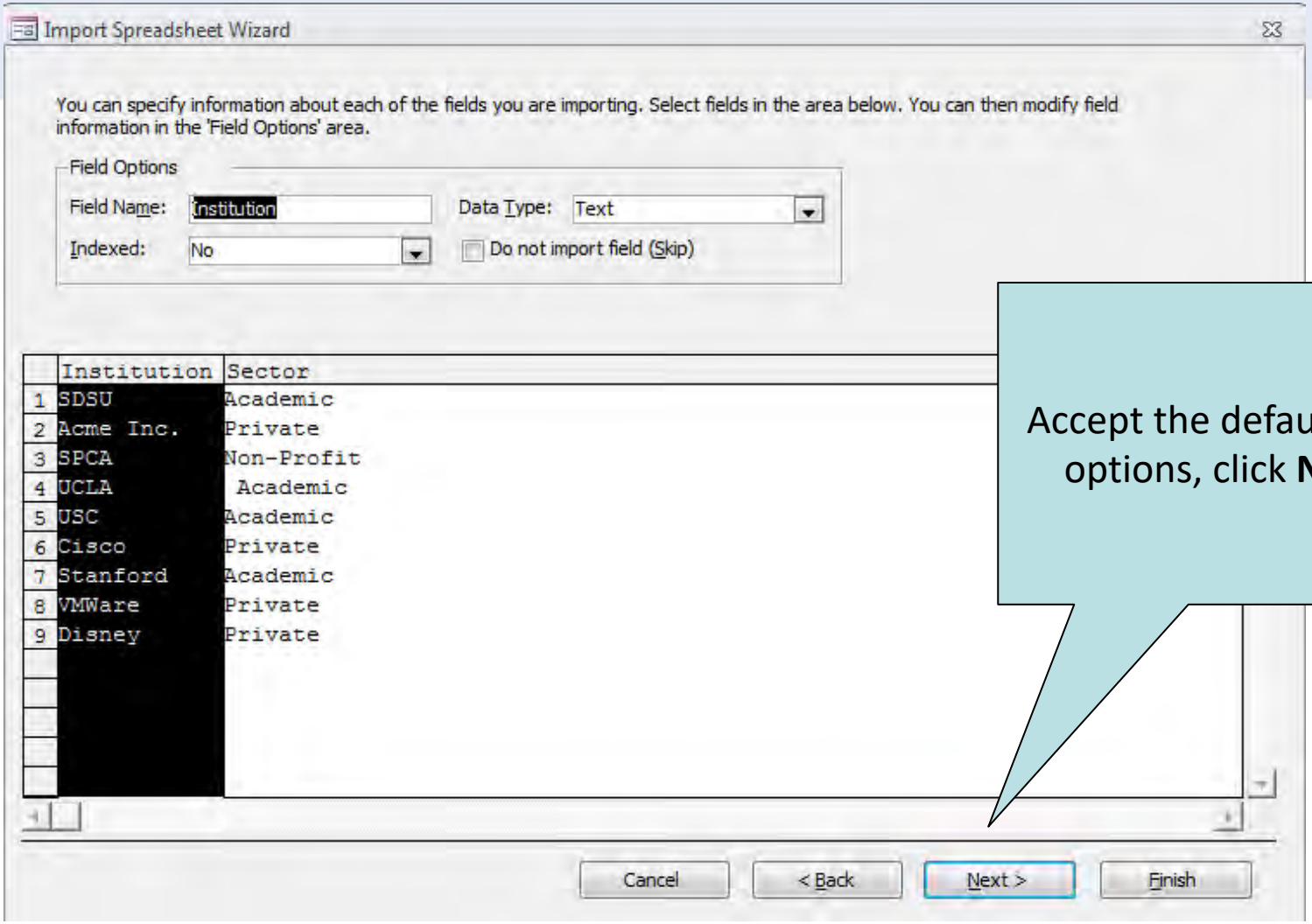


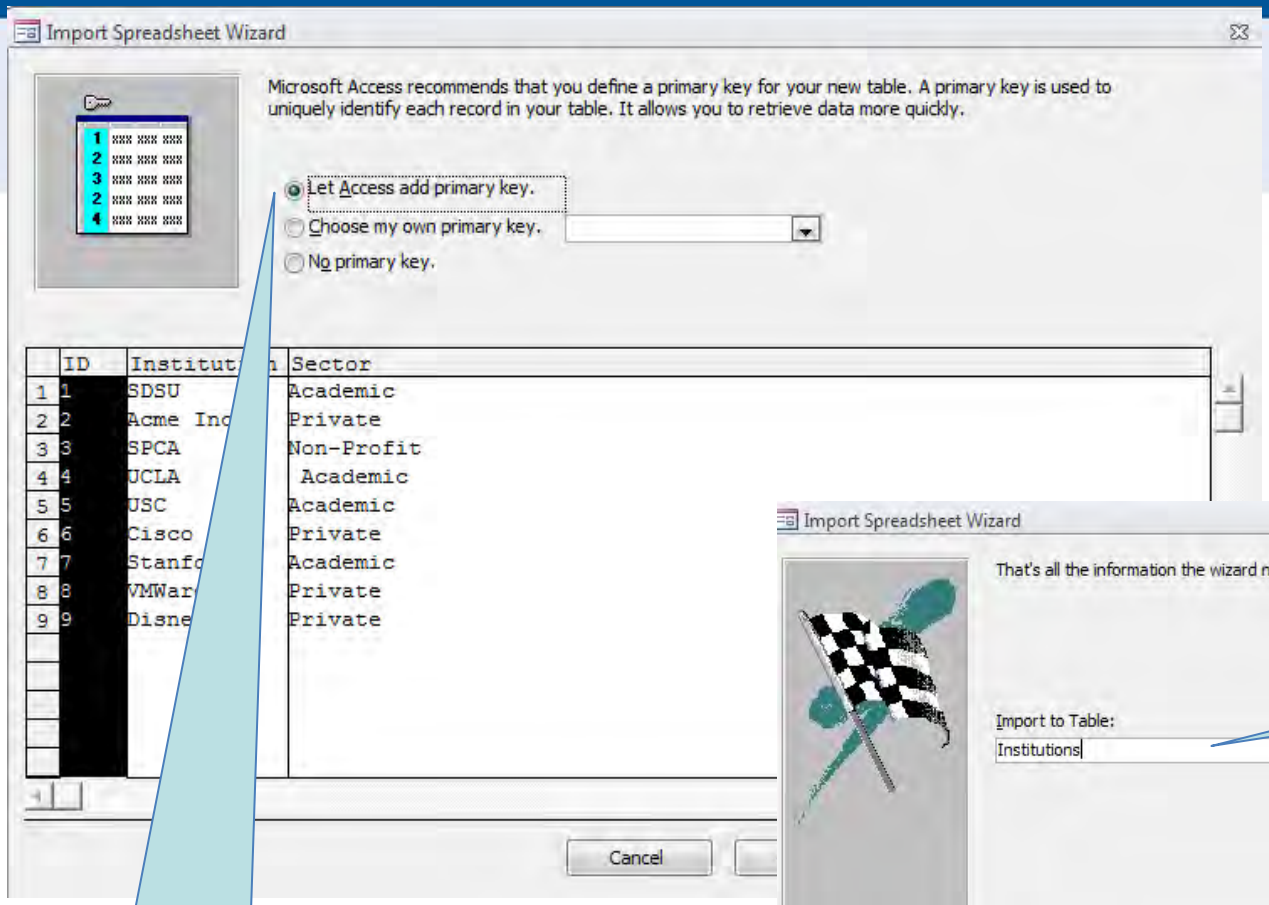
Import Institutions Table



Import Institutions Table

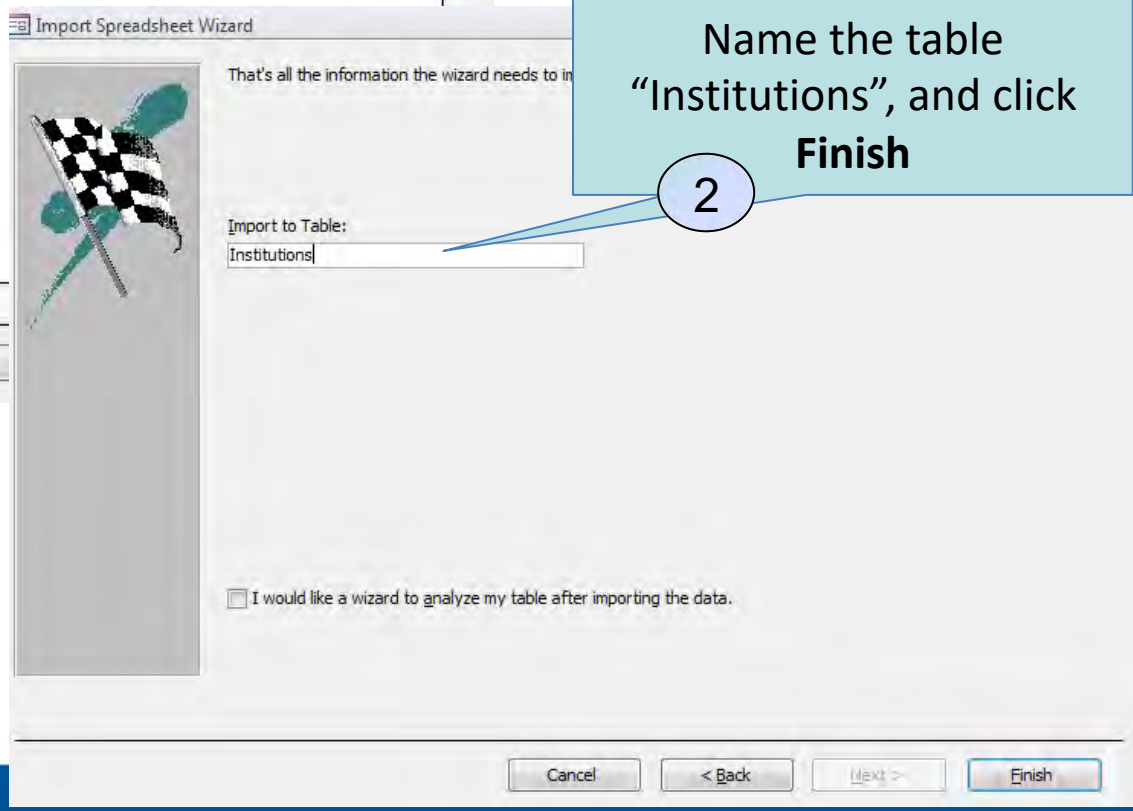



Import Spreadsheet Wizard
You can specify information about each of the fields you are importing. Select fields in the area below. You can then modify field information in the 'Field Options' area.
Field Options
Field Name: Institution
Data Type: Text
Indexed: No
Do not import field (Skip)
Institution Sector
1 SDSU Academic
2 Acme Inc. Private
3 SPCA Non-Profit
4 UCLA Academic
5 USC Academic
6 Cisco Private
7 Stanford Academic
8 VMWare Private
9 Disney Private
Cancel
< Back
Next >
Finish
Accept the default field options, click **Next**.



1

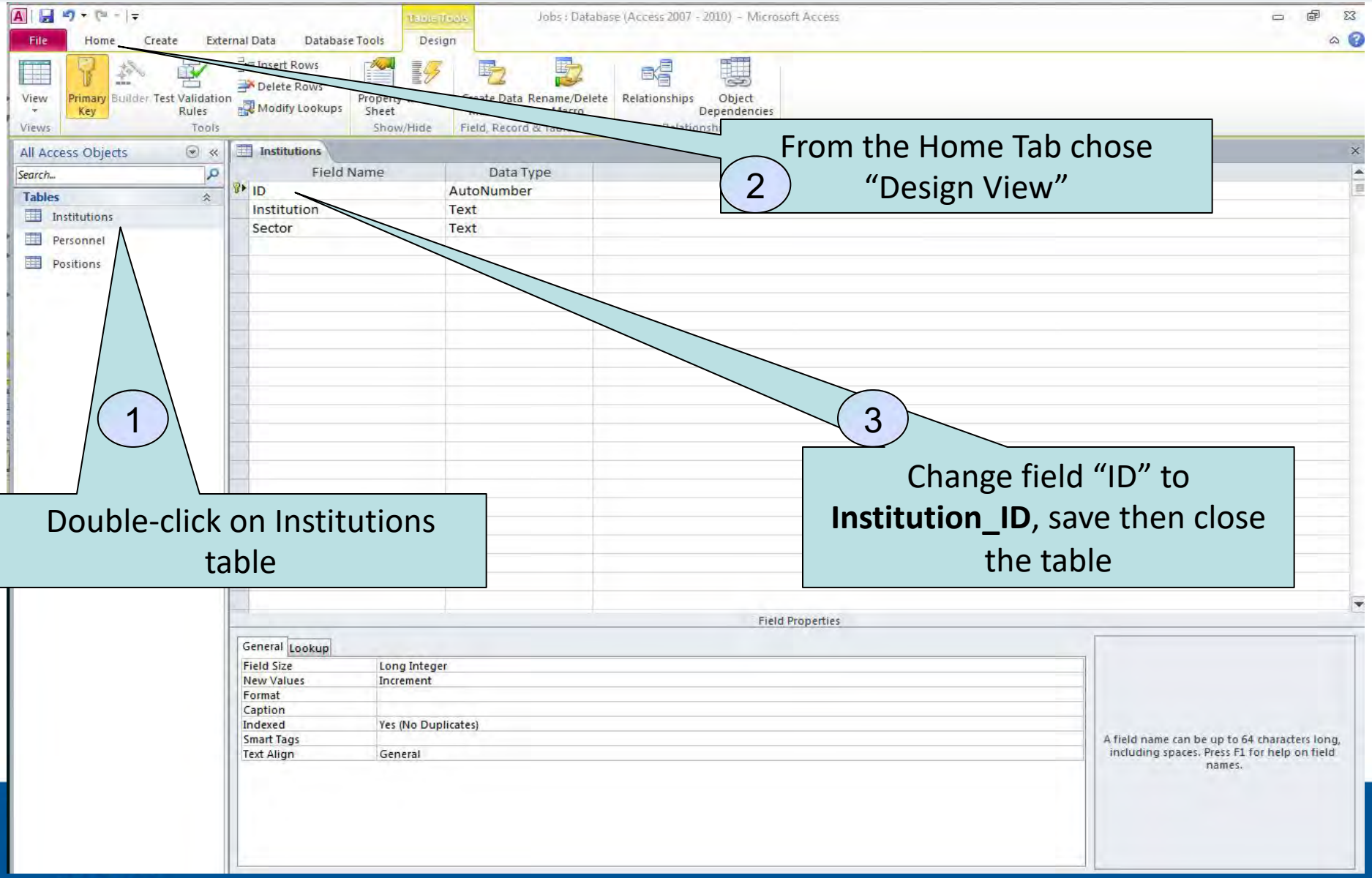
Let access add primary key (accept as is), click **Next**



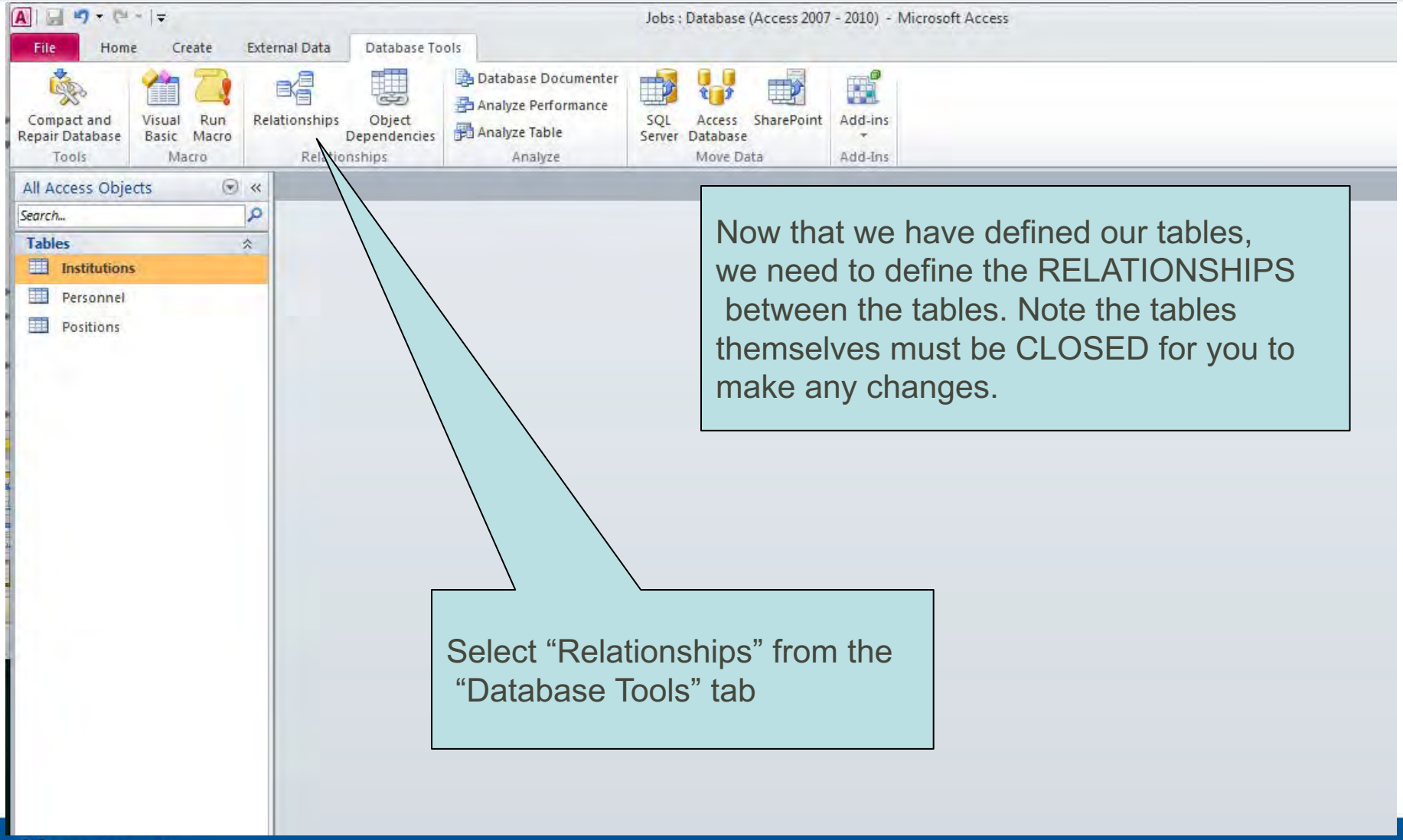
2

Name the table "Institutions", and click **Finish**

Finish Institutions Table

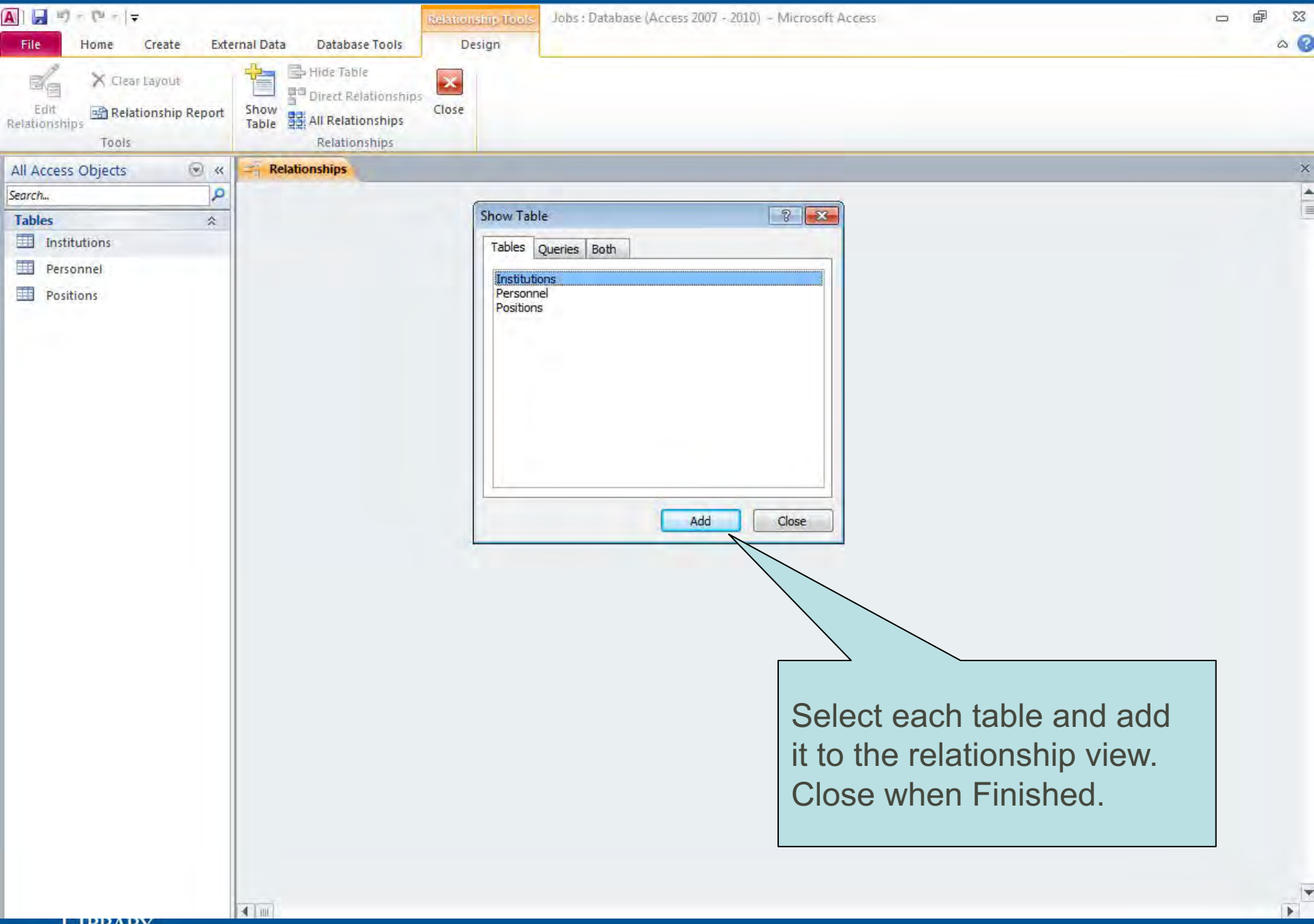


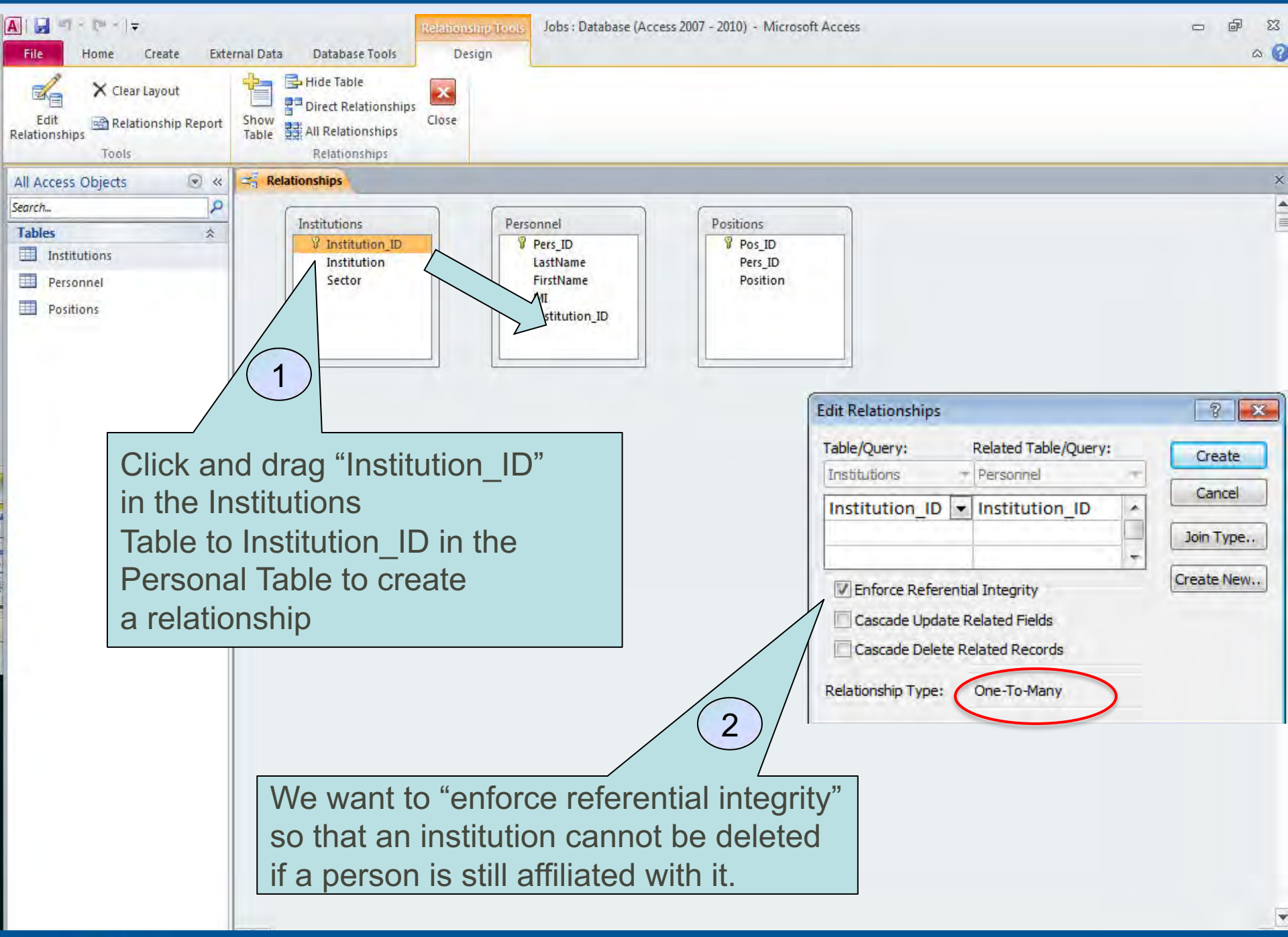
Relationships

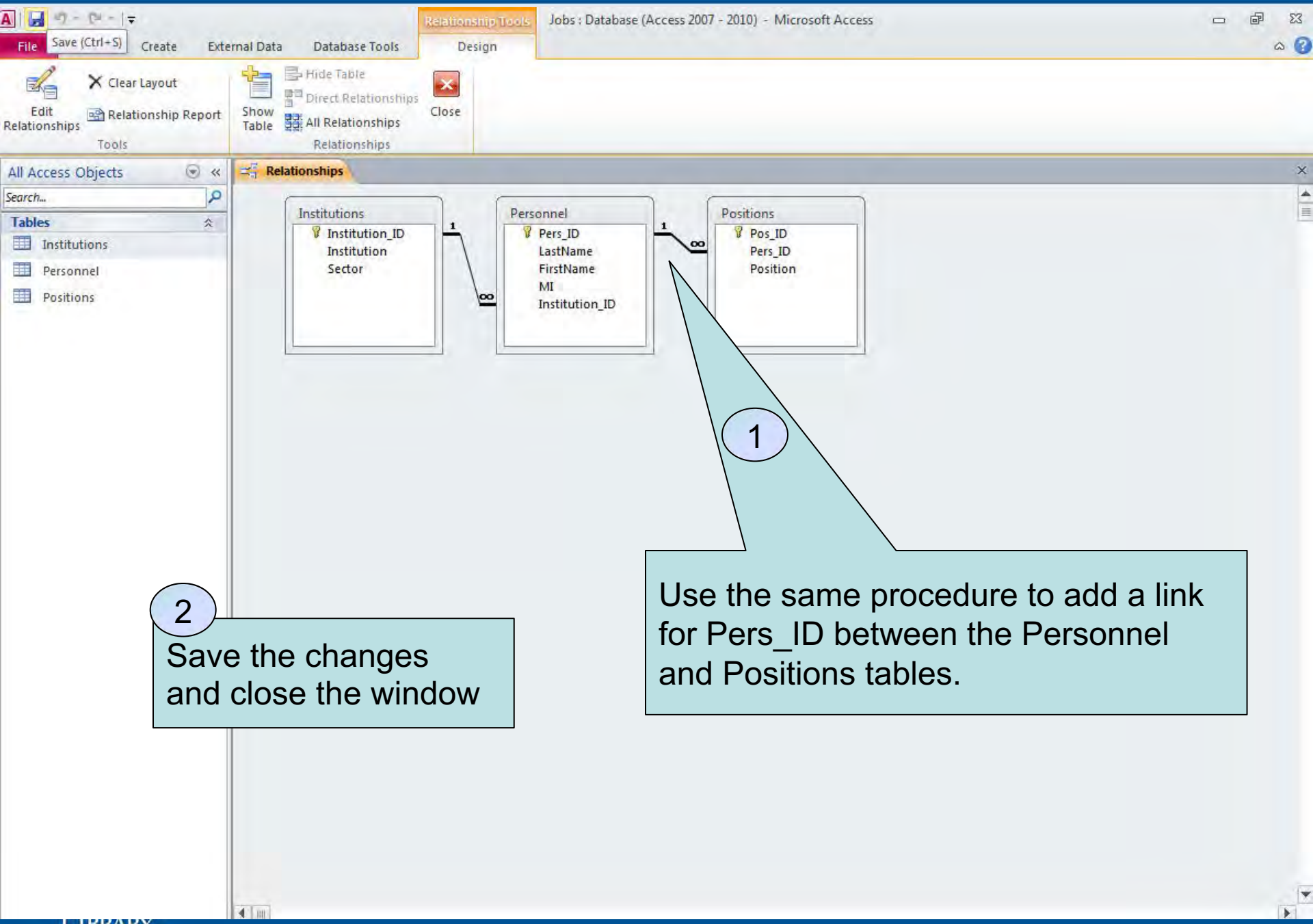


Now that we have defined our tables, we need to define the RELATIONSHIPS between the tables. Note the tables themselves must be CLOSED for you to make any changes.

Select “Relationships” from the “Database Tools” tab


Relationship Tools
Jobs : Database (Access 2007 - 2010) – Microsoft Access
File
Home
Create
External Data
Database Tools
Design
Edit Relationships
Clear Layout
Relationship Report
Tools
Show Table
Hide Table
Direct Relationships
All Relationships
Close
Relationships
All Access Objects
Search...
Tables
Institutions
Personnel
Positions
Relationships
Show Table
Tables
Queries
Both
Institutions
Personnel
Positions
Add
Close
Select each table and add it to the relationship view. Close when Finished.


Relationship Tools
Jobs : Database (Access 2007 - 2010) - Microsoft Access
File
Home
Create
External Data
Database Tools
Design
Edit Relationships
Clear Layout
Relationship Report
Tools
Show Table
Hide Table
Direct Relationships
All Relationships
Relationships
Close
All Access Objects
Search...
Tables
Institutions
Personnel
Positions
Relationships
Institutions
Institution_ID
Institution
Sector
Personnel
Pers_ID
LastName
FirstName
Position
Positions
Pos_ID
Pers_ID
Position
1
Click and drag “Institution_ID” in the Institutions Table to Institution_ID in the Personal Table to create a relationship
Edit Relationships
Table/Query:
Related Table/Query:
Institutions
Personnel
Institution_ID
Institution_ID
Create
Cancel
Join Type..
Create New..
Enforce Referential Integrity
Cascade Update Related Fields
Cascade Delete Related Records
Relationship Type:
One-To-Many
2
We want to “enforce referential integrity” so that an institution cannot be deleted if a person is still affiliated with it.


Relationship Tools
Jobs : Database (Access 2007 - 2010) - Microsoft Access
File
Save (Ctrl+S)
Create
External Data
Database Tools
Design
Edit Relationships
Clear Layout
Relationship Report
Tools
Show Table
Hide Table
Direct Relationships
All Relationships
Close
Relationships
All Access Objects
Search...
Tables
Institutions
Personnel
Positions
Relationships
Institutions
Institution_ID
Institution
Sector
Personnel
Pers_ID
LastName
FirstName
MI
Institution_ID
Positions
Pos_ID
Pers_ID
Position
1
∞
1
∞
1
Use the same procedure to add a link for Pers_ID between the Personnel and Positions tables.
2
Save the changes and close the window

Demos

- Use of Lookup Table
- Forms for data entry

More Information on Databases

- Database Normalization Basics
  - http://databases.about.com/od/specificproducts/a/normalization.htm
- Step-by-step Guides to Using Databases
  - http://www.geekgirls.com/menu_databases.htm
- Comparison between Access & Excel
  - http://office.microsoft.com/en-us/access/HA102101951033.aspx

ACCESS Help

- ACCESS has an extensive “help” system.
- Microsoft Online Tutorials
  - Access 2010 Training Courses

http://office.microsoft.com/en-us/access-help/training-courses-for-access-2010-HA104039037.aspx