

Relational Database Management for Epidemiologists: Starting the Process

Wayne Enanoria, PhD, MPH
Center for Infectious Disease Preparedness
University of California at Berkeley

Outline

- Seven Steps to Designing a Database
- Mission Statement
- Mission Objectives
- Field Types

Phases of Database Design

- Define Mission Statement and Objectives
- Analyze the current database
- Create data structures
- Establish table relationships
- Define business rules
- Determine and establish views
- Review data integrity

Elements of Design

- Epidemiologic Study
 - Research Question
 - Methods
 - Target Population
 - Recruitment
 - Analysis
- Database Design
 - Mission Statement and Objectives
 - Methods
 - Fields
 - Table Structure
 - Analysis

Developing a Mission Statement

- A mission statement declares the specific purpose of the database; defined at the beginning of the design process.
 - ◆ Focuses design efforts
- What's in a good mission statement?
 - ◆ It's very general, succinct, and to the point
 - ◆ It should *not* describe specific tasks
 - ◆ It should define the purpose for the database

Good Example

- The purpose of the West Nile virus database is to maintain the data we receive on potential cases and to provide information that supports the disease control activities we provide to the public health community.

Bad Example

- The purpose of the acute hepatitis A database is to keep track of cases for public health use, maintain data on cases, keep a record of all laboratory tests, keep a record of all results, keep a record of all exposures in the 4 weeks before symptom onset, and maintain data about travel in the 4 weeks before symptom onset.

Example Revised

- The purpose of the acute hepatitis A database is to maintain the data on potential cases in order to determine who meets the case definition and ensure proper follow-up of all identified cases so as to prevent further cases of disease.

Developing Mission Objectives

- Mission Objectives are statements that represent the general tasks supported by the data maintained in the database.
- Your objectives should:
 - ◆ Each clearly define a single general task
 - ◆ Be free from excess detail
 - ◆ Be easy to understand

Good Examples

- We need to maintain complete patient address information.
- We need to keep track of all identified contacts.
- We need to make sure an investigation is completed on all identified contacts.

Bad Example

- We need to keep track of the close contacts of cases and their occupations, as well as the contacts of these contacts.

Example Revised

- We need to maintain complete information on contacts of the cases.
- We need to keep track of all contacts of these contacts.

Conducting Interviews

- Purpose: to find out about the department, its needs, and why a database should be created.
 - ◆ Involve management and users
 - ◆ Involves everyone in the process of database design; can save time and effort in the long run.
 - ◆ Identify user capabilities and features of the database that are needed

Mission Statement Questions*

- How would you describe the purpose of the organization to a new client?
- What would you say is the purpose of your organization?
- What is the major function of your organization?
- How would you describe what your organization does?
- What is the main focus of your organization?

* from Hernandez MJ. Database Design for Mere Mortals. Second Edition, Addison-Wesley 2003.

Mission Objective Questions*

- What kind of work do you perform on a daily basis?
- How would you define your job description?
- What kind of data do you work with?
- What types of reports do you generate?
- What types of things do you keep track of?
- What types of services does your organization provide?
- How would you describe the type of work you do?

* from Hernandez MJ. Database Design for Mere Mortals. Second Edition, Addison-Wesley 2003.

The “Legacy” Database

- Goals:
 - ◆ To determine whether the database supports the organization’s current information requirements
 - ◆ To uncover existing structural deficiencies
 - ◆ To determine how the database needs to evolve so that it will support the organization’s future information requirements.

Restructuring the “Legacy”

- One can use the existing database as a resource for developing a new database.
- Judge carefully which aspects remain useful and which should be discarded.
 - ◆ What types of data does the organization use?
 - ◆ How does the organization use that data?
 - ◆ How does the organization manage and maintain that data?
- This judgement can help define the preliminary field and table structures.

Remember

- Do not adopt the current database structure as the basis for the new database structure!
 - ◆ Hidden problems with current structure will be transferred over to the new database.
- If the job of overhauling the database is too big, determine the resources needed and possible solutions.

Conducting the Analysis

- Three steps:
 - ◆ Review the way data is collected
 - ◆ Review the manner in which information is presented
 - ◆ Conduct interviews with users and management

Data Collection

- Review paper-based items
 - ◆ Gather all sample forms (hard copy, paper-based) and make a copy of each form.
- Review all of the computer software programs used by the organization/department
 - ◆ Gather a set of sample screen shots

Presentation of Information

- Reports
- Presentations with summaries, statistics, etc.
- Journal Articles or Publications
- Web pages

Conducting Interviews

- Determine how the organization uses its data; get into specifics!
- Interviews:
 - ◆ Provide details about the samples you assembled during previous reviews
 - ◆ Provide information on the way the organization uses its data
 - ◆ Define preliminary field and table structures
 - ◆ Define future information requirements

Identifying Subjects

- “As an account representative, I’m responsible for 10 clients. Each of my clients makes an appointment to come into the showroom to view the merchandise we have to offer for the current season. Part of my job is to answer any questions they have about our merchandise and make recommendations regarding the most popular items. Once they make a decision on the merchandise they’d like to purchase, I write up a sales order for the client. Then I give the sales order to my assistant, who promptly fills the order and sends it to the client.”

List of Subjects

- Account
- Representative
- Appointment
- Assistant
- Clients
- Items
- Job
- Merchandise
- Sales Order
- Season
- Showroom

Identifying Subjects

- “As a disease control investigator, I’m responsible for collecting reports of positive laboratory tests for viral hepatitis from laboratories and clinicians. With each report from clinicians, we fax the clinician in order to find out the reason the person was tested.

List of Subjects

- Disease Control Investigator
- Reports
- Person
- Laboratory Tests
- Laboratories
- Clinicians
- Fax

Subjects

- One can use the list of subjects to define tables later in the design process.
- Table: chief structures in the database and each table always represents a single, specific subject (object or event).

Identifying Characteristics

- “Well, I enter all the client information first, such as the client’s name, address, and phone number. Then I enter the items the client wants to purchase. After I’ve entered all the items, I tally up totals and I’m done. Oh, I forgot to mention that I enter the client’s fax number and shipping address – if they have one.”

List of Characteristics

- Address
- Fax Number
- Name
- Phone Number
- Shipping Address
- Totals

Characteristics

- One can use the list of characteristics to define fields of tables later in the design process.
- Field (attribute): represents a characteristic of the subject of the table to which it belongs
 - ◆ Multipart fields
 - ◆ Multivalued fields
 - ◆ Calculated fields.

Review the Samples

- Identify how the objects represented by the samples are used
- Clarify aspects of the samples you don't understand
- Assign a description to each sample

Review Information Requirements

- Determine whether individual users/management receive information based on data they don't directly control or maintain
- Determine what types of additional information they need
- Determine what types of information they can foresee themselves needing in the future.

Compiling a Complete List of Fields

- Based on:
 - ◆ Analysis of the current database
 - ◆ Interviews with users and management
- Consists of organization's fundamental data requirements and constitutes the core set of fields for the database

Review and Refine

- Identify and remove duplicate fields (characteristics).
 - ◆ Remove same characteristic of the same subject with two or more different names (eg, ProductNo and ProductNumber)
 - ◆ Rename same characteristic of different subjects; do not delete (eg, NamePt, NameMD, NameTest)
- Make sure that each item on your list represents a characteristic (and not subjects).

Characteristics versus Subjects

- Can this word be used to describe something?
- Does this word represent a component or something else?
- Does this word represent a collection of things?
- Can this thing be broken down into smaller pieces?

Determine New Characteristics

- Examine the samples you gathered throughout the analysis process.
 - ◆ Examine each characteristic and determine if it's on the preliminary field list
 - ◆ Determine whether any of the remaining characteristics has the same meaning as an existing field



Mike's Elke Shop

File Edit View Insert Format Records Tools Window Help

Supplier Information

Company:	Acme Power Tools	Status:	Active
Address:	635 Montana Ave	Office Phone:	598-4455
City:	El Paso	FAX Number:	598-5715
State:	TX	Zip:	79925

Contacts

Name:	George Barlett	Phone No.:	532-9228
Name:		Phone No.:	

< Back Next > Edit Save

Figure 6.15. A sample with highlighted characteristics

SOURCE: Michael J. Hernandez. Database Design for Mere Mortals. Second Edition, Addison-Wesley 2003.

Value Lists

- Record the name of each characteristic that incorporates a value list (enumerated list).
 - ◆ A value list specifies the acceptable range of values for a particular characteristic and enforces constraints on the field.
 - ◆ Examples: gender, race/ethnicity, disease control investigator, etc.
- This list of names with value lists will help you define field specifications later.

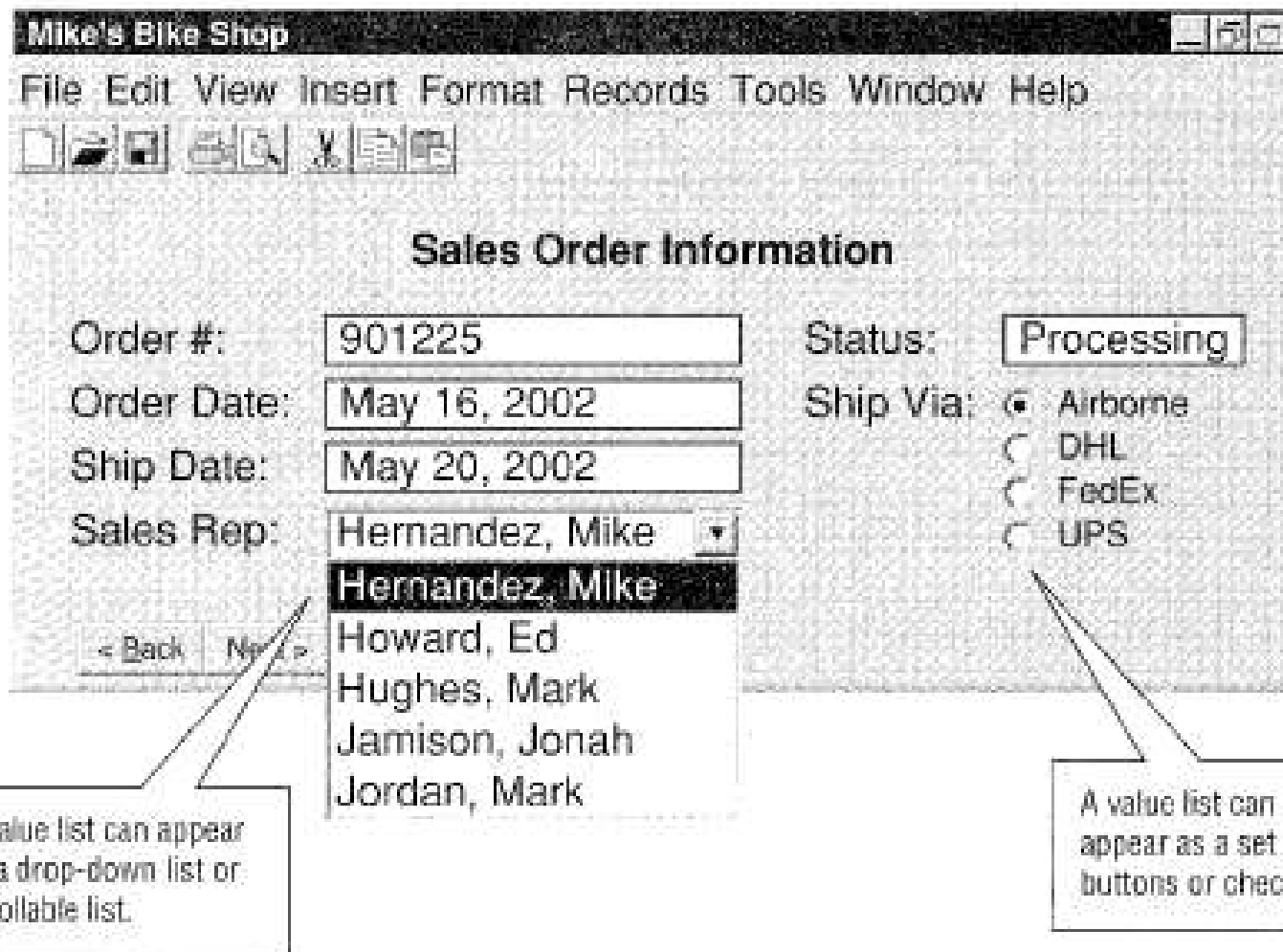


Figure 6.16. A database screen with two value lists.

SOURCE: Michael J. Hernandez. Database Design for Mere Mortals. Second Edition, Addison-Wesley 2003.

Calculated Field List

- A calculated field is one that stores the result of some operation or function of other fields; you do not enter its value directly.
- Remove every calculated field and place it on a separate list; you'll use these later in the design process.

At the End of the Day...

- Engaged users and management in the design process
- Reviewed existing databases and organizational structure
- Lists
 - ◆ Subjects
 - ◆ Preliminary Fields (Characteristics)
 - Fields with Value Lists
 - Calculated Fields

Next Time

- Entity-Relationship Diagrams
- Planning Table Structure