8

Kendall & Kendall
Systems Analysis and Design, 9e

Analyzing Systems Using Data Dictionaries

Learning Objectives

- Understand how analysts use of data dictionaries for analyzing data-oriented systems.
- Understand the concept of a repository for analysts' project information and the role of CASE tools in creating them.
- Create data dictionary entries for data processes, stores, flows, structures, and logical and physical elements of the systems being studied, based on DFDs.
- Recognize the functions of data dictionaries in helping users update and maintain information systems.

Cataloging

- Data flow diagrams can be used to catalog:
  - Data processes
  - Flows
  - Stores
  - Structures
  - Elements
- Cataloging takes place with the data dictionary

Major Topics

- The data dictionary
- The data repository
- Defining data flow
- Defining data structures
- Defining data elements
- Defining data stores
- Using the data dictionary

The Data Dictionary

- A reference work of data about data (metadata)
- Collects and coordinates data terms, and confirms what each term means to different people in the organization

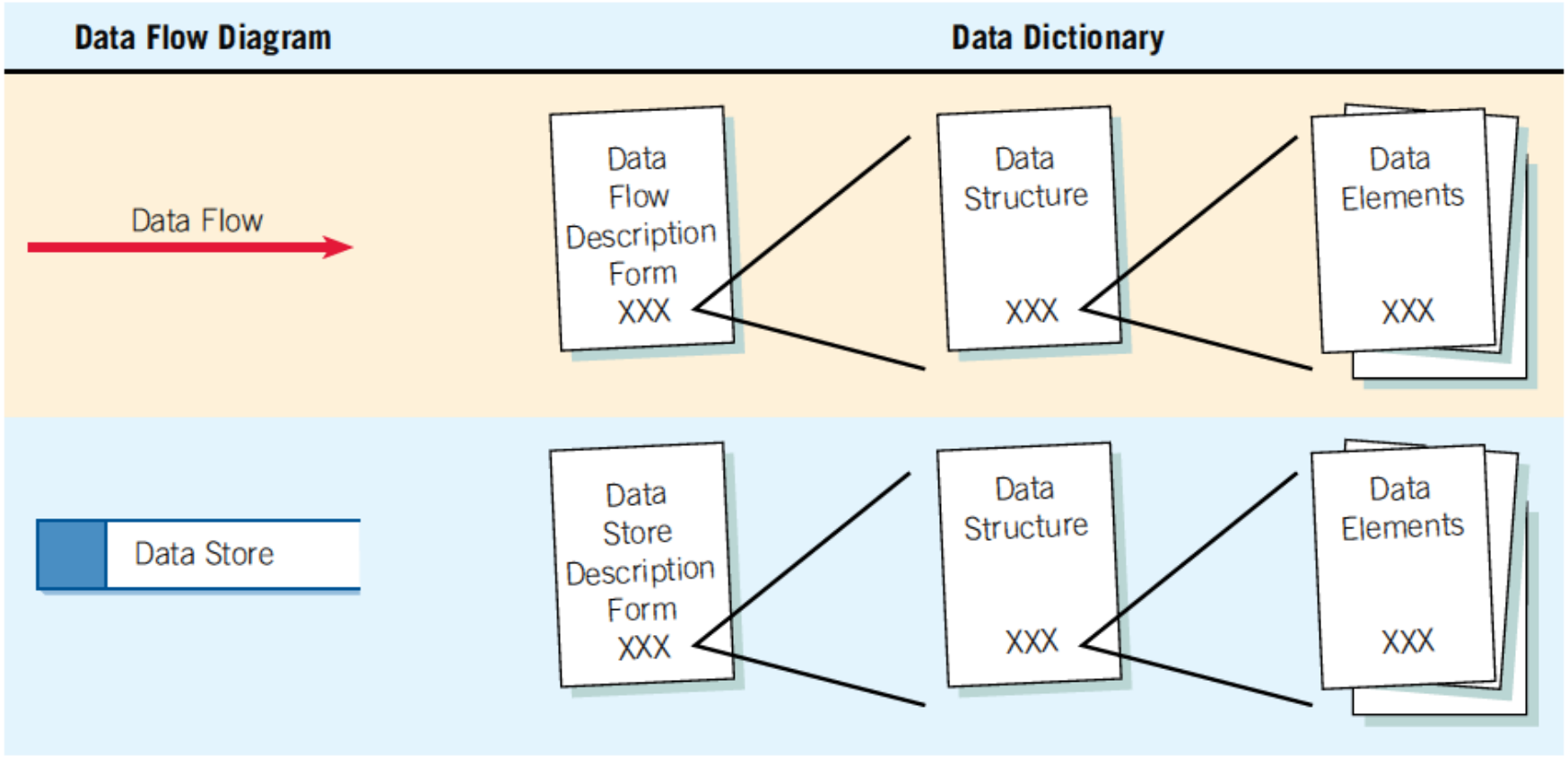
Need for Understanding the Data Dictionary

- Provide documentation
- Eliminate redundancy
- Validate the data flow diagram
- Provide a starting point for developing screens and reports
- Determine the contents of data stored in files
- To develop the logic for DFD processes

The Data Repository

- A data repository is a large collection of project information
- It includes:
  - Information about the data maintained by the system
  - Procedural logic and use cases
  - Screen and report design
  - Data relationships
  - Project requirements and final system deliverables
  - Project management information

How Data Dictionaries Relate to Data Flow Diagrams (Figure 8.1)



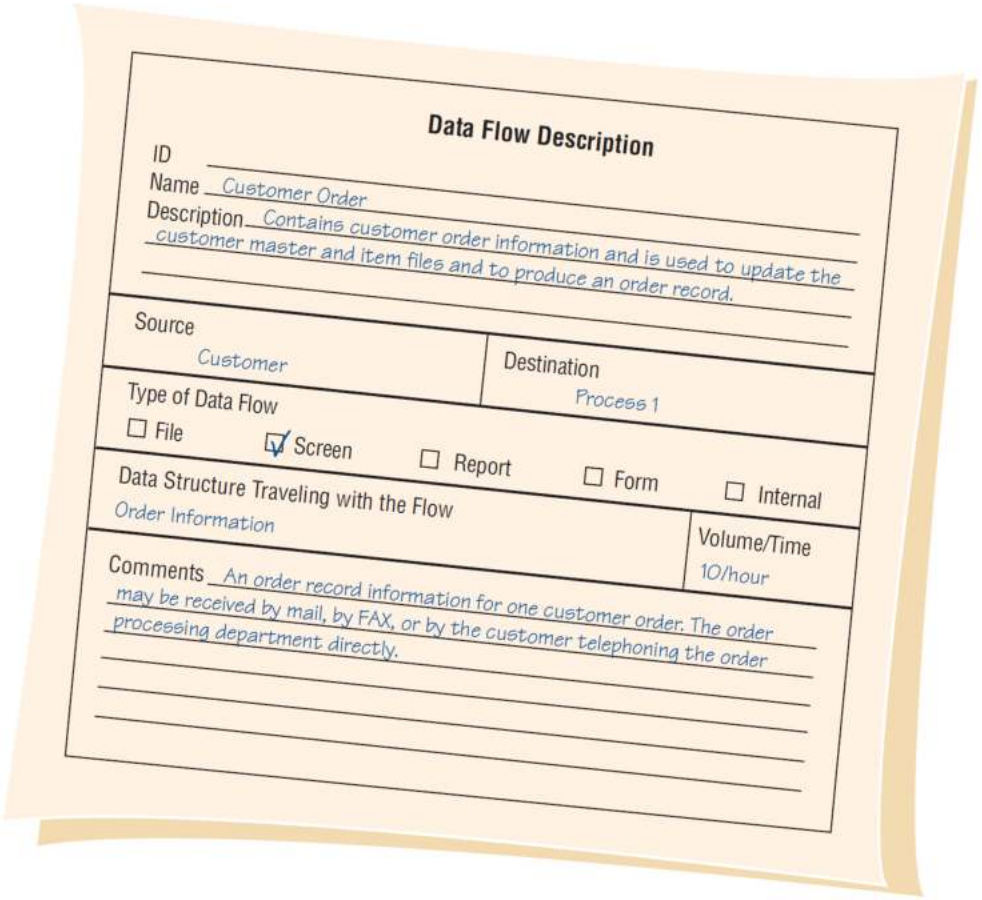
Data Dictionary Categories

- Data flows
- Data structures
- Elements
- Data stores

Defining the Data Flow

- ID—identification number
- Unique descriptive name
- A general description of the data flow
- The source of the data flow
- The destination of the data flow
- Type of data flow
- The name of the data structure describing the elements
- The volume per unit time
- An area for further comments and notations

An Example of a Data Flow Description from World's Trend Catalog Division (Figure 8.3)



Data Flow Description

ID

Name Customer Order

Description Contains customer order information and is used to update the customer master and item files and to produce an order record.

Source: Customer

Destination: Process 1

Type of Data Flow: ☐ File ☑ Screen ☐ Report ☐ Form ☐ Internal

Data Structure Traveling with the Flow: Order Information

Volume/Time: 10/hour

Comments An order record information for one customer order. The order may be received by mail, by FAX, or by the customer telephoning the order processing department directly.

Describing Data Structures

- Data structures are made up of smaller structures and elements
- An algebraic notation is used to describe data structures

Algebraic Notation

- Equal sign means "is composed of"
- Plus sign means "and"
- Braces {} mean repetitive elements
- Brackets [] for an either/or situation
- Parentheses () for an optional element

Data Structure Example for Adding a Customer Order at World's Trend Catalog Division (Figure 8.4)

Customer Order = Customer Number +
Customer Name +
Address +
Telephone +
Catalog Number +
Order Date +
{Available Order Items} +
Merchandise Total +
(Tax) +
Shipping and Handling +
Order Total +
Method of Payment +
(Credit Card Type) +
(Credit Card Number) +
(Expiration Date)

Customer Name = First Name +
(Middle Initial) +
Last Name

Address = Street +
(Apartment) +
City +
State +
Zip +
(Zip Expansion) +
(Country)

Telephone = Area Code +
Local Number

Available Order Items = Quantity Ordered +
Item Number +
Item Description +
Size +
Color +
Price +
Item Total

Method of Payment = [Check ¦ Charge ¦ Money Order]

Credit Card Type = [World's Trend ¦ American Express ¦ MasterCard ¦ Visa]

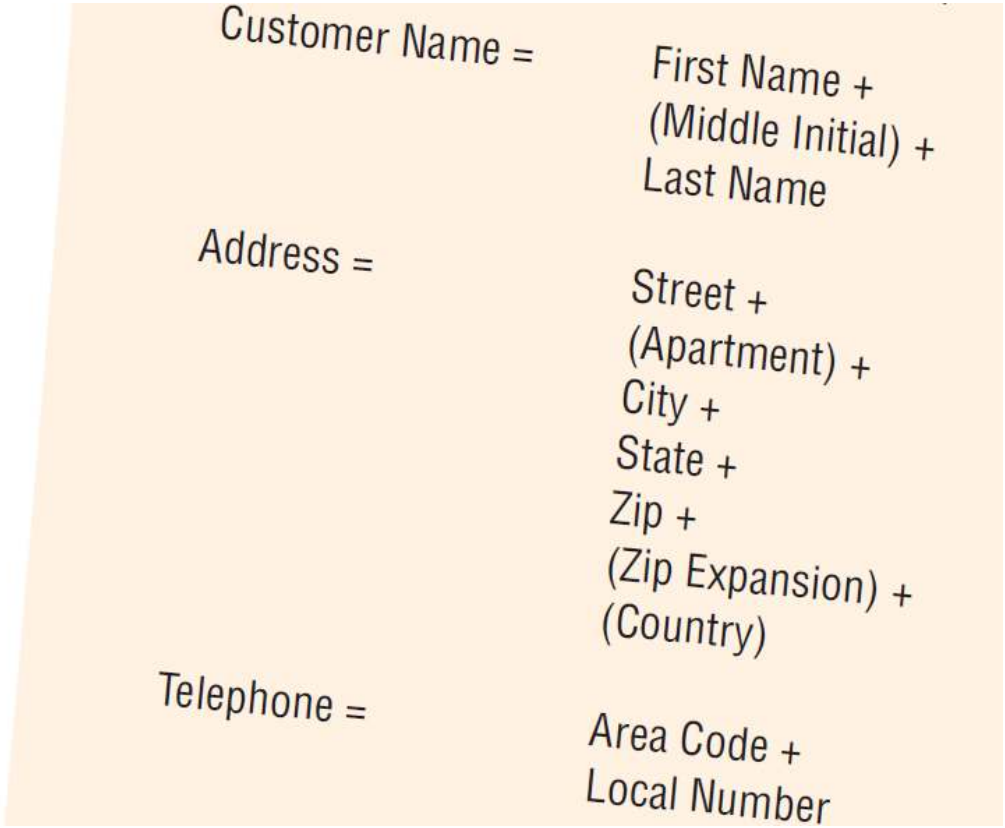
Structural Records

- A structure may consist of elements or structural records
- These are a group of elements, such as:
  - Customer name
  - Address
  - Telephone
- Each of these must be further defined until they are broken down into their component elements

Structural Records Used in Different Systems

- Structural records and elements that are used within many different systems are given a non-system-specific name, such as street, city, and zip
- The names do not reflect a functional area
- This allows the analyst to define them once and use in many different applications

Structural Record Example



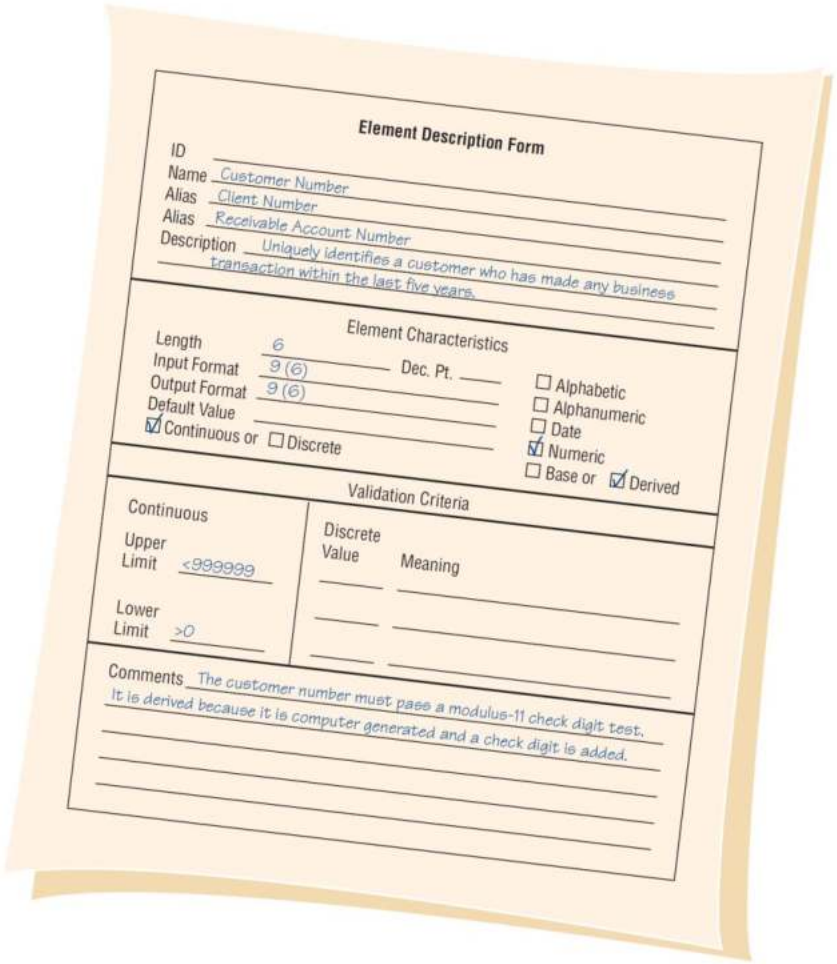
Logical and Physical Data Structures

- Logical:
  - Show what data the business needs for its day-to-day operations
- Physical:
  - Include additional elements necessary for implementing the system

Physical Data Structures

- Key fields used to locate records
- Codes to identify record status
- Transaction codes to identify different record types
- Repeating group entries
- Limits on items in a repeating group
- Password

An Element Description Form Example from World's Trend Catalog Division (Figure 8.6)



Element Description Form

ID

Name Customer Number

Alias Client Number

Alias Receivable Account Number

Description Uniquely identifies a customer who has made any business transaction within the last five years.

Element Characteristics

Length 6 Dec. Pt. ____

Input Format 9(6)

Output Format 9(6)

Default Value

☑ Continuous or ☐ Discrete

☐ Alphabetic
☐ Alphanumeric
☐ Date
☑ Numeric
☐ Base or ☑ Derived

Validation Criteria

Continuous

Upper Limit <999999

Lower Limit >0

Discrete

Value | Meaning

Comments The customer number must pass a modulus-11 check digit test. It is derived because it is computer generated and a check digit is added.

Data Element Characteristics

- Element ID
- The name of the element
- Aliases
- A short description of the element
- Element is base or derived
- Element length
- Type of data
- Input and output formats
- Validation criteria
- Default value
- An additional comment or remark area

Element ID

- Optional entry
- Allows the analyst to build automated data dictionary entries

The Name of the Element

- Should be:
  - Descriptive
  - Unique
- Based on what the element is commonly called in most programs or by the major user of the element

Aliases

- Synonyms or other names for the element
- Names used by different users in different systems
- A CUSTOMER NUMBER may also be called a RECEIVABLE ACCOUNT NUMBER or a CLIENT NUMBER

Short Description of the Element

- An example might be:
  - Uniquely identifies a customer who has made any business transactions within the last five years

Element Is Base or Derived

- A base element is one that has been initially keyed into the system
- A derived element is one that is created by a process, usually as the result of a calculation or a series of decision-making statements

Element Length

What should the element length be?

- Some elements have standard lengths, state abbreviations, zip codes, or telephone numbers.
- For other elements, the length may vary and the analyst and user community must decide the final length.

Element Length Considerations

- Numeric amount lengths
- Name and address fields
- Other fields

Name and Address Length

Element	Length	Percent of data that will fit (United States)
Last Name	11	98
First Name	18	95
Company Name	20	95
Street	18	90
City	17	99

Data Truncation

- If the element is too small, the data will be truncated
- The analyst must decide how this will affect the system outputs
- If a last name is truncated, mail would usually still be delivered
- A truncated email address or web address is not usable

Type of Data

- Alphanumeric or text data
- Formats
  - Mainframe: packed, binary, display
  - Microcomputer (PC) formats
  - PC formats, such as Currency, Number, or Scientific, depend on how the data will be used

Some Examples of Data Formats Used in PC Systems (Figure 8.7)

Data Type	Meaning
Bit	A value of 1 or 0, a true/false value
Char, varchar, text	Any alphanumeric character
Datetime, smalldatetime	Alphanumeric data, several formats
Decimal, numeric	Numeric data that are accurate to the least significant digit; can contain a whole and decimal portion
Float, real	Floating-point values that contain an approximate decimal value
Int, smallint, tinyint	Only integer (whole digit) data
Currency, money, smallmoney	Monetary numbers accurate to four decimal places
Binary, varbinary, image	Binary strings (sound, pictures, video)
Cursor, timestamp, uniqueidentifier	A value that is always unique within a database
Autonumber	A number that is always incremented by one when a record is added to a database table

Format Character Codes (Figure 8.8)

Formatting Character	Meaning
X	May enter or display/print any character
9	Enter or display only numbers
Z	Display leading zeros as spaces
,	Insert commas into a numeric display
.	Insert a period into a numeric display
/	Insert slashes into a numeric display
-	Insert a hyphen into a numeric display
V	Indicate a decimal position (when the decimal point is not included)

Validation Criteria

- Ensure that accurate data are captured by the system
- Elements are either:
  - Discrete, meaning they have fixed values
  - Continuous, with a smooth range of values

Default Value

- Include any default value the element may have
- The default value is displayed on entry screens
- Reduces the amount of keying
  - Default values on GUI screens
    - Initially display in drop-down lists
    - Are selected when a group of radio buttons are used

Comment or Remarks Area

- This might be used to indicate the format of the date, special validation that is required, the check-digit method used, and so on

Data Stores

- Data stores are created for each different data entity being stored
- When data flow base elements are grouped together to form a structural record, a data store is created for each unique structural record
- Because a given data flow may only show part of the collective data that a structural record contains, many different data flow structures may need to be examined to arrive at a complete data store description

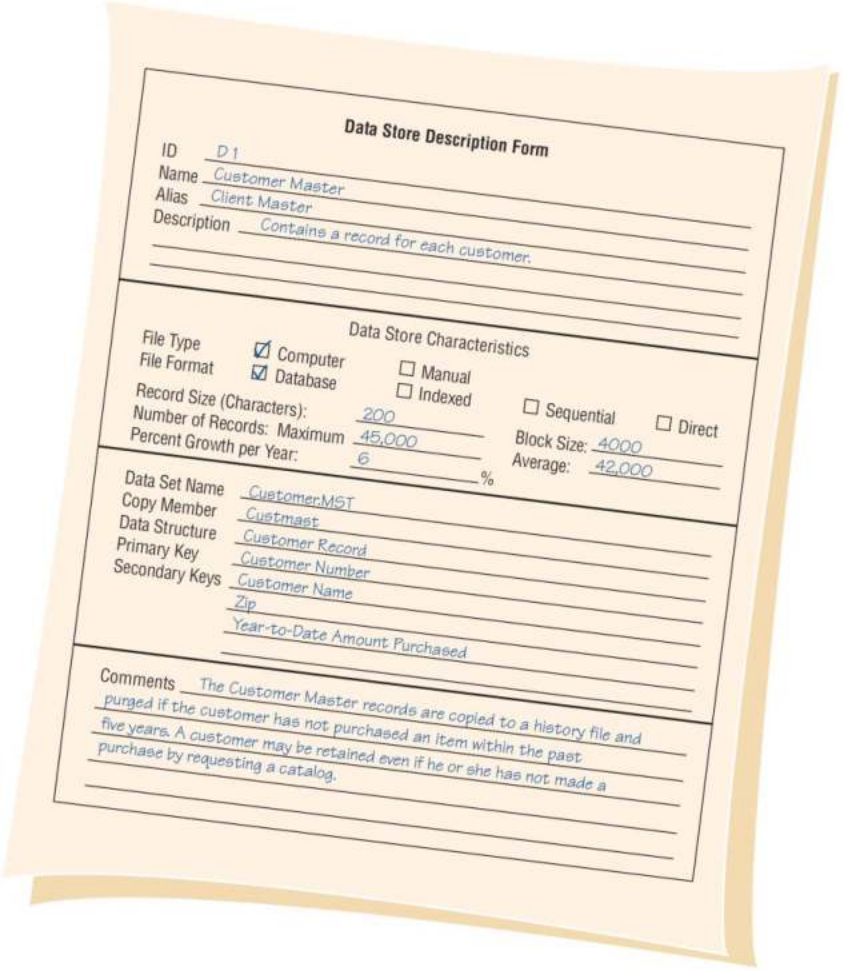
Describing the Data Store

- The data store ID
- The data store name
- An alias for the table
- A short description of the data store
- The file type
- File format

Describing the Data Store (continued)

- The maximum and average number of records on the file as well as the growth per year
- The file or data set name specifies the file name, if known
- The data structure should use a name found in the data dictionary
- Primary and secondary keys
- Comments

Example of a Data Store Form for World's Trend Catalog Division (Figure 8.9)



Data Store Description Form

ID D1
Name Customer Master
Alias Client Master
Description Contains a record for each customer.

Data Store Characteristics

File Type ☑ Computer ☐ Manual
File Format ☑ Database ☐ Indexed ☐ Sequential ☐ Direct
Record Size (Characters): 200
Number of Records: Maximum 45,000
Block Size: 4000
Percent Growth per Year: 6 %
Average: 42,000

Data Set Name CustomerMST
Copy Member Custmast
Data Structure Customer Record
Primary Key Customer Number
Secondary Keys Customer Name
Zip
Year-to-Date Amount Purchased

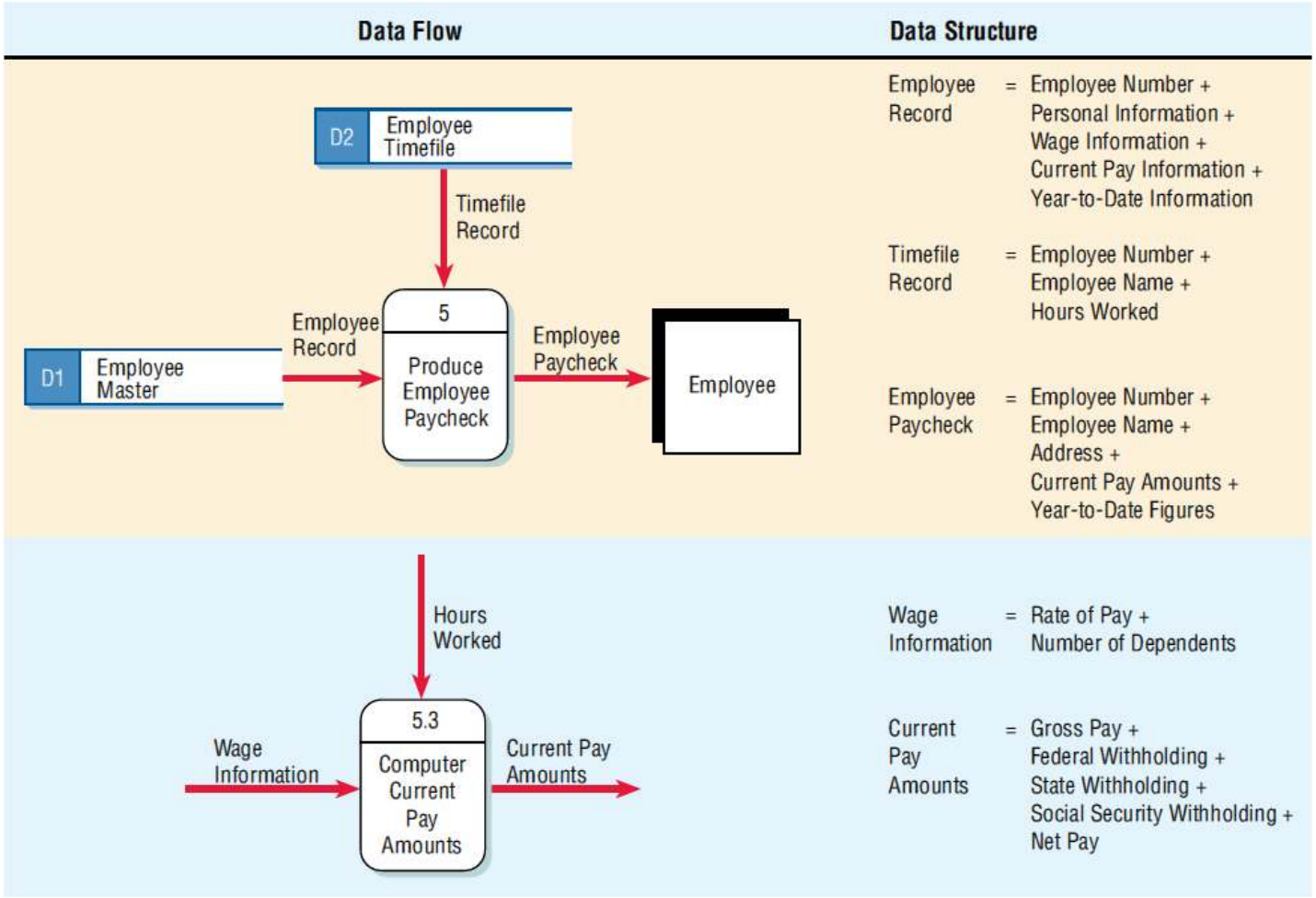
Comments The Customer Master records are copied to a history file and purged if the customer has not purchased an item within the past five years. A customer may be retained even if he or she has not made a purchase by requesting a catalog.

Creating the Data Dictionary

- Data dictionary entries
  - Created after the data flow diagram is completed

  or
  - Created as the data flow diagram is being developed
- Created using a top-down approach

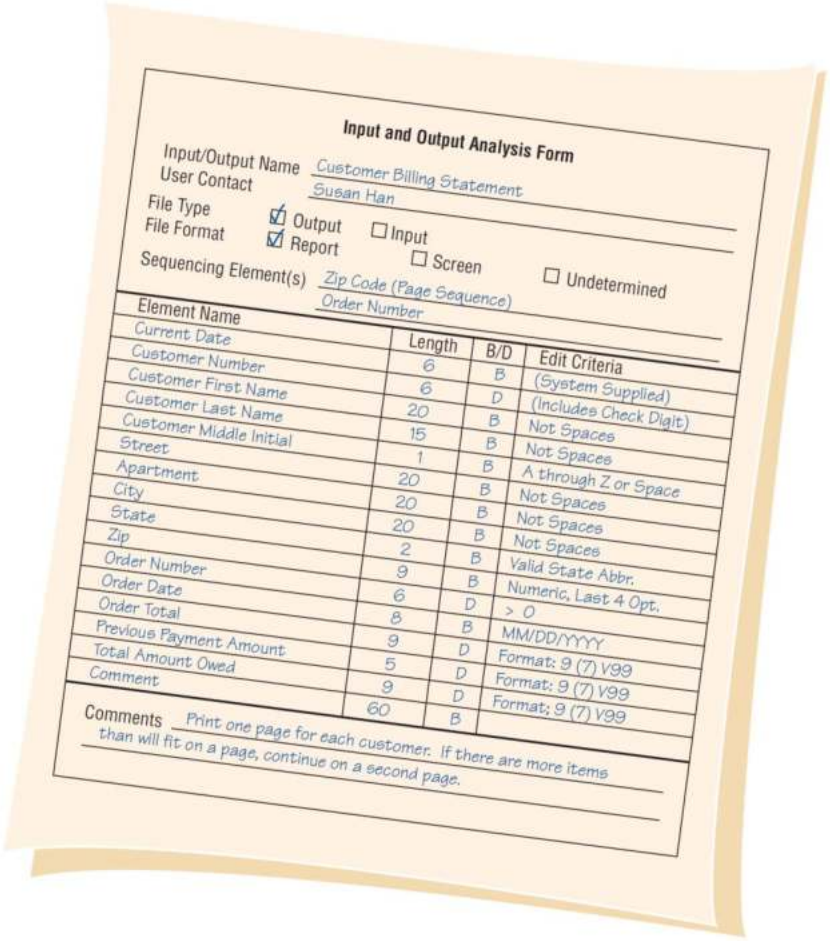
Two Data Flow Diagrams and Corresponding Data Dictionary Entries for Producing an Employee Paycheck (Figure 8.11)



Analyzing Input and Output

- A descriptive name for the input or output
- The user contact responsible
- Whether the data is input or output
- The format of the data flow
- Elements indicating the sequence of the data on a report or screen
- A list of elements

An Example of an Input/Output Analysis Form for World's Trend Catalog Division (Figure 8.12)



Input and Output Analysis Form

Input/Output Name: Customer Billing Statement
User Contact: Susan Han
File Type: ☑ Output ☐ Input
File Format: ☑ Report ☐ Screen ☐ Undetermined
Sequencing Element(s): Zip Code (Page Sequence), Order Number

Element Name	Length	B/D	Edit Criteria
Current Date	6	B	(System Supplied)
Customer Number	6	D	(Includes Check Digit)
Customer First Name	20	B	Not Spaces
Customer Last Name	15	B	Not Spaces
Customer Middle Initial	1	B	A through Z or Space
Street	20	B	Not Spaces
Apartment	20	B	Not Spaces
City	20	B	Not Spaces
State	2	B	Valid State Abbr.
Zip	9	B	Numeric, Last 4 Opt.
Order Number	6	D	> 0
Order Date	8	B	MM/DD/YYYY
Order Total	9	D	Format: 9 (7) V99
Previous Payment Amount	5	D	Format: 9 (7) V99
Total Amount Owed	9	D	Format: 9 (7) V99
Comment	60	B	

Comments: Print one page for each customer. If there are more items than will fit on a page, continue on a second page.

Developing Data Stores

- Represent data at rest
- Contain information of a permanent or semipermanent (temporary) nature
- When data stores are created for only one report or screen, we refer to them as "user views"

Using the Data Dictionary

- To have maximum power, the data dictionary should be tied into a number of systems programs
- May be used to
  - Create screens, reports, and forms
  - Generate computer language source code
  - Analyze the system design, detecting flaws and areas that need clarification

Create Screens, Reports, and Forms

- Use the element definition and their lengths
- Arrange the elements in a pleasing and functional way using design guidelines and common sense
- Repeating groups become columns
- Structural records are grouped together on the screen, report, or form

Analyze the System Design, Detecting Flaws and Areas that Need Clarification

- All base elements on an output data flow must be present on an input data flow to the process producing the output
- A derived element should be created by a process and should be output from at least one process into which it is not input
- The elements that are present in a data flow coming into or going out of a data store must be contained in the data store

Summary

- The data dictionary
  - A reference work containing data about data
  - Includes all data items from data flow diagrams
- Repository
  - A larger collection of project information
- Defining data structures
- Defining elements