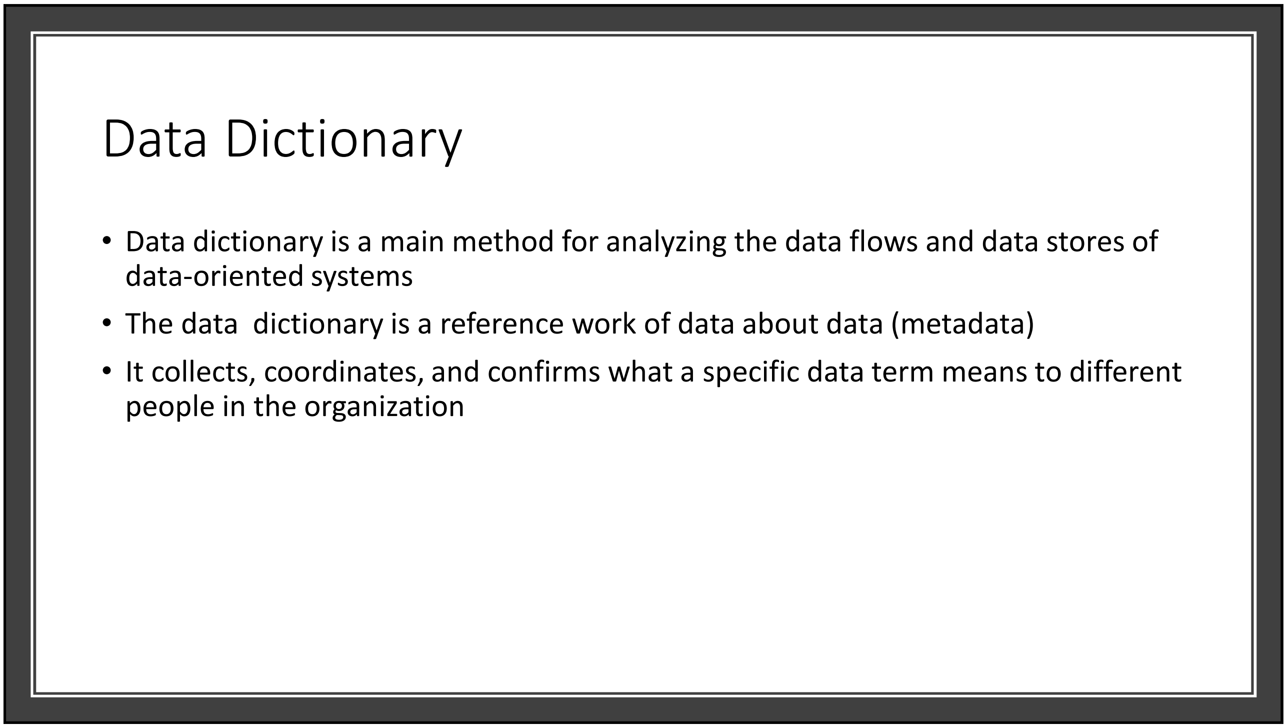
Data Dictionary

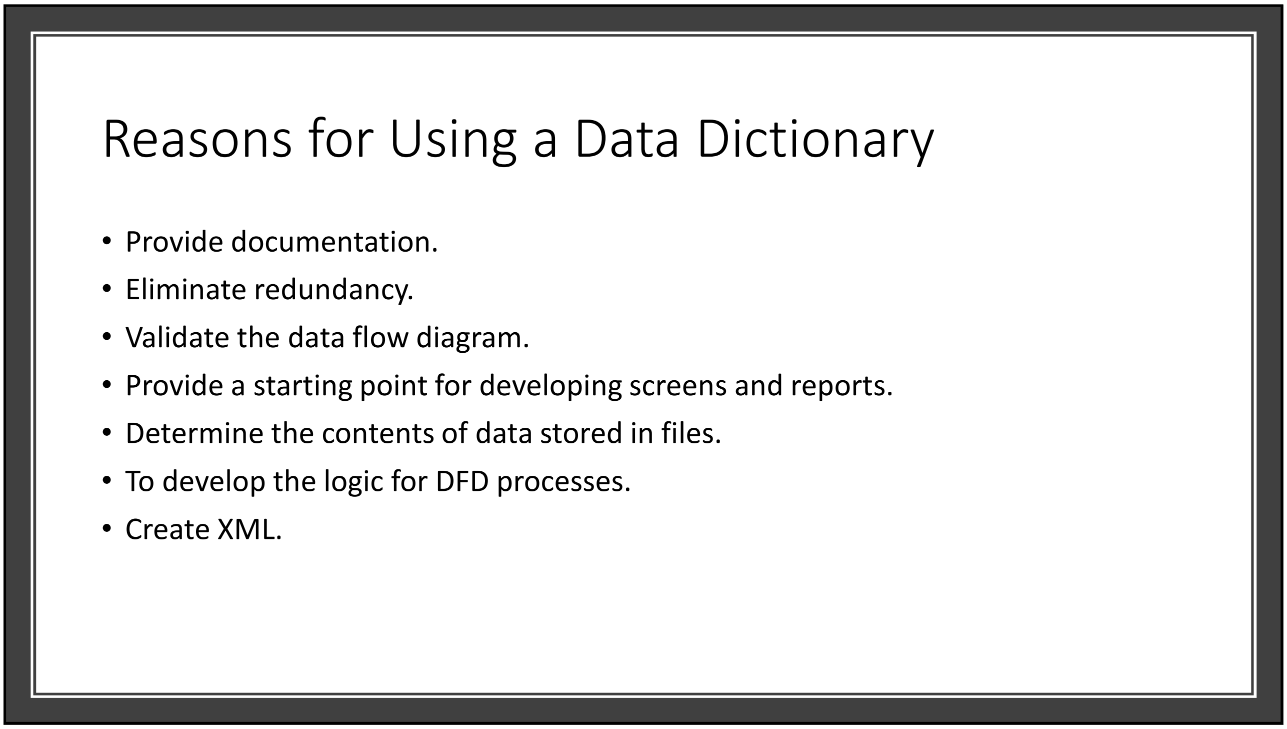
169



Data Dictionary

- Data dictionary is a main method for analyzing the data flows and data stores of data-oriented systems
- The data dictionary is a reference work of data about data (metadata)
- It collects, coordinates, and confirms what a specific data term means to different people in the organization

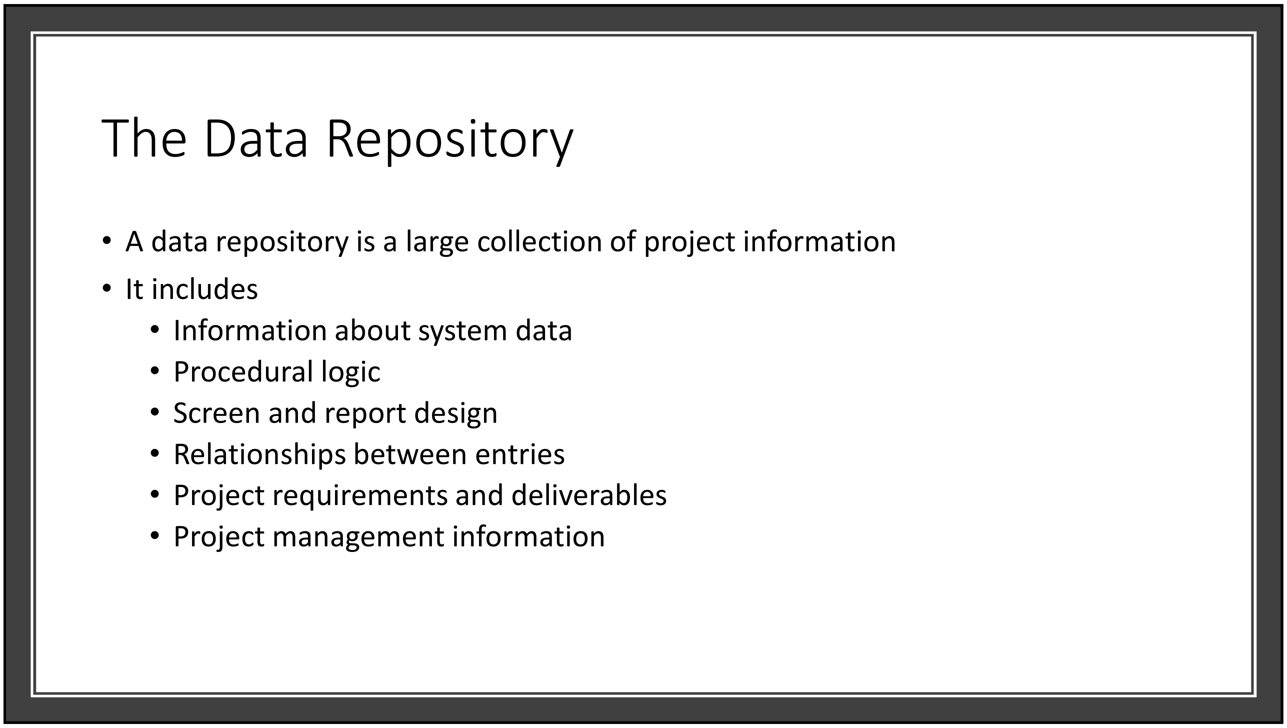
170



Reasons for Using a 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.
- Create XML.

171

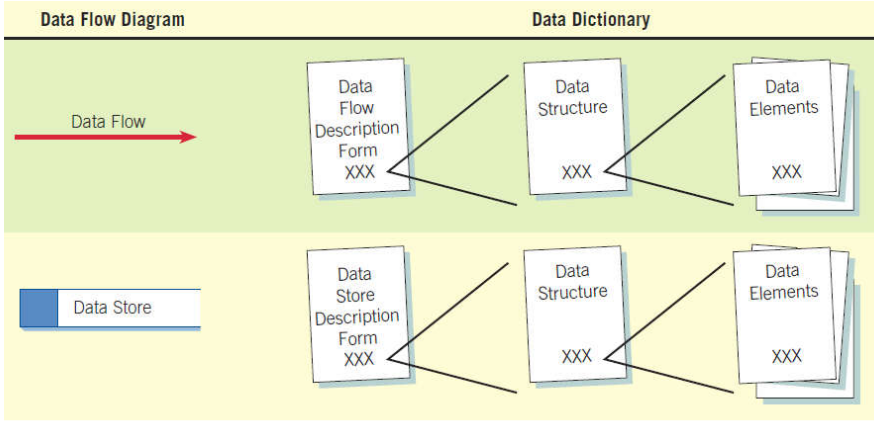


The Data Repository

- A data repository is a large collection of project information
- It includes
  - Information about system data
  - Procedural logic
  - Screen and report design
  - Relationships between entries
  - Project requirements and deliverables
  - Project management information

172

How Data Dictionaries Relate to Data Flow Diagrams

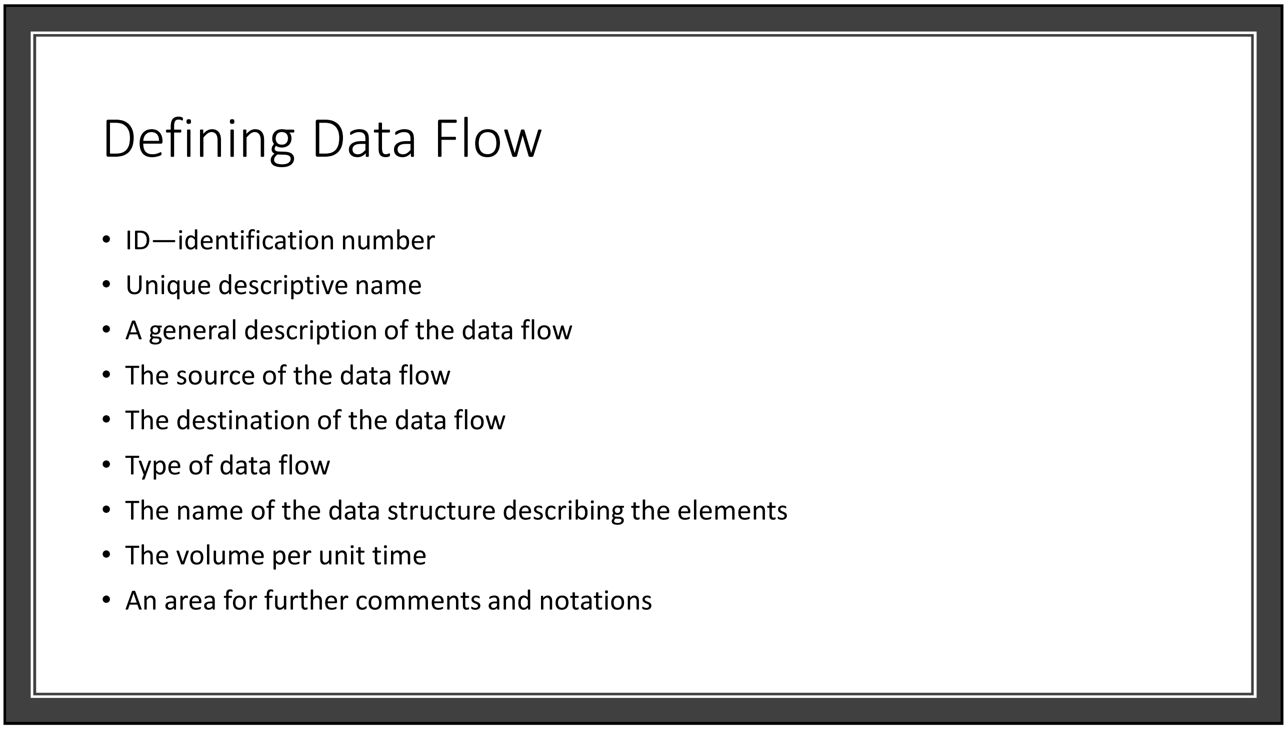


173

Data Dictionary Contents

- Data flows
- Data structures
- Elements
- Data stores

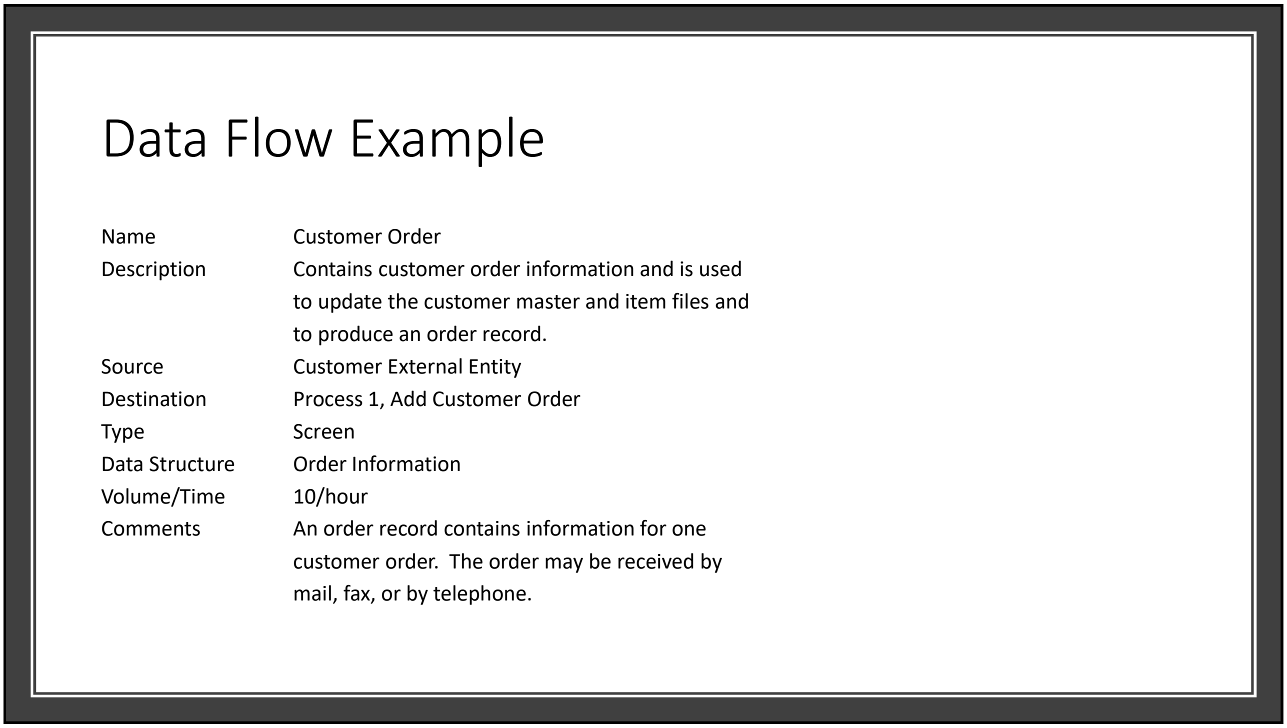
174



Defining 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

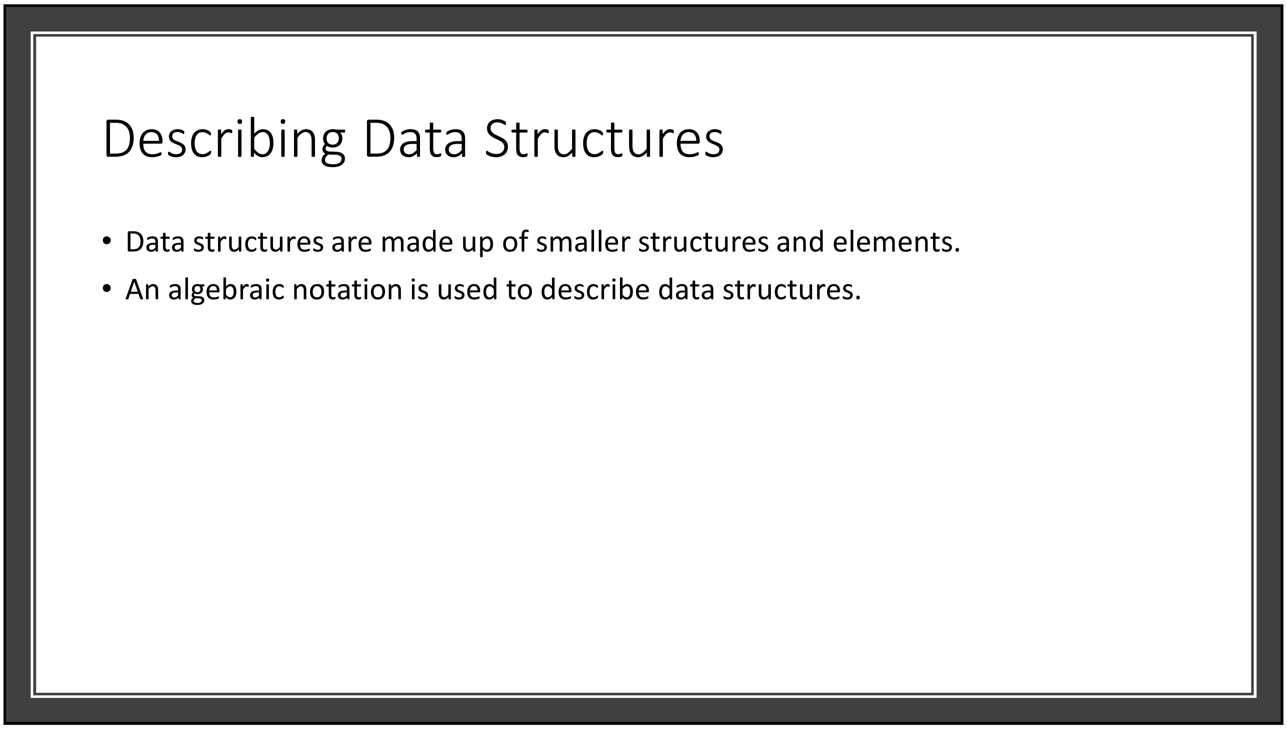
175



Data Flow Example

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 External Entity
Destination	Process 1, Add Customer Order
Type	Screen
Data Structure	Order Information
Volume/Time	10/hour
Comments	An order record contains information for one customer order. The order may be received by mail, fax, or by telephone.

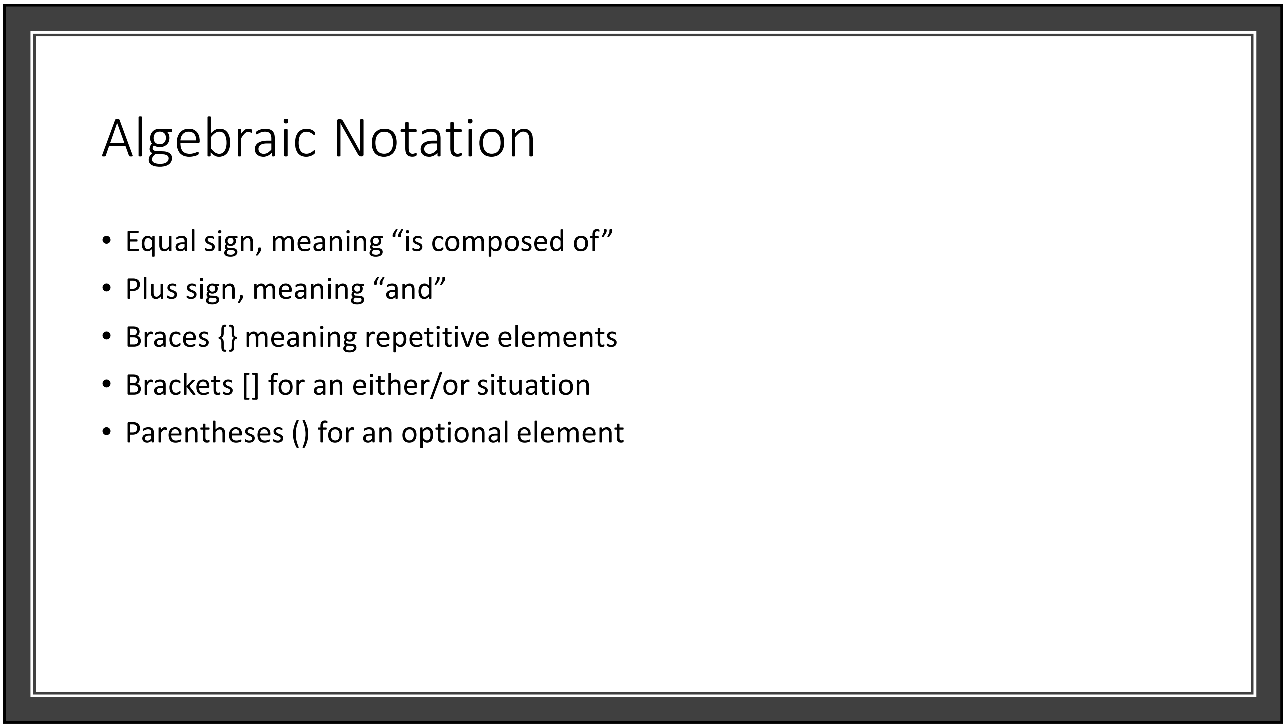
176



Describing Data Structures

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

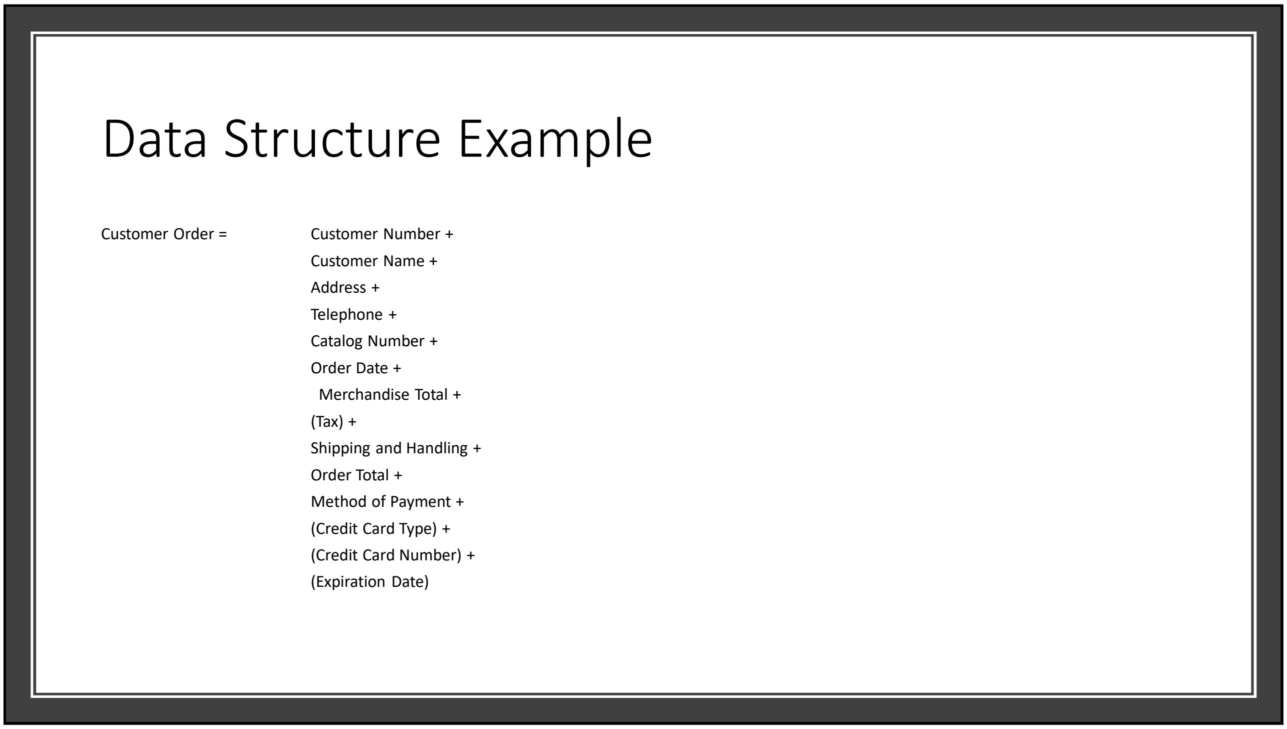
177



Algebraic Notation

- Equal sign, meaning “is composed of”
- Plus sign, meaning “and”
- Braces {} meaning repetitive elements
- Brackets [] for an either/or situation
- Parentheses () for an optional element

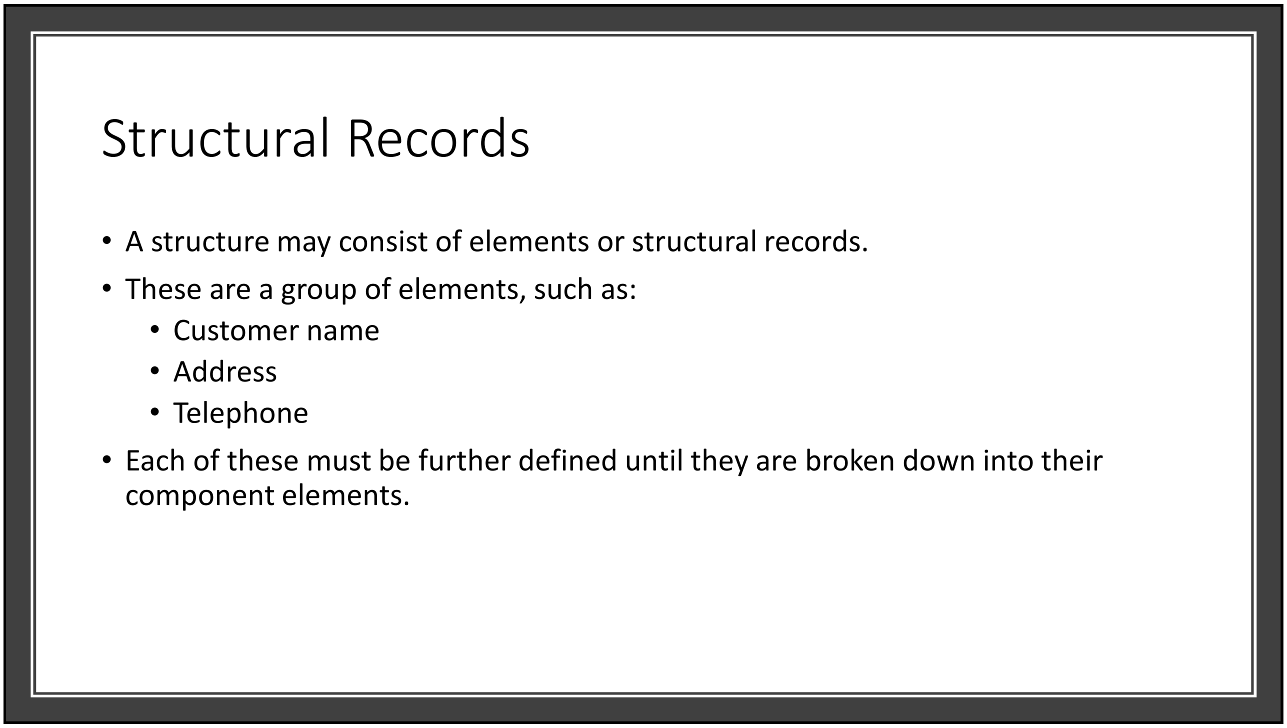
178



Data Structure Example

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

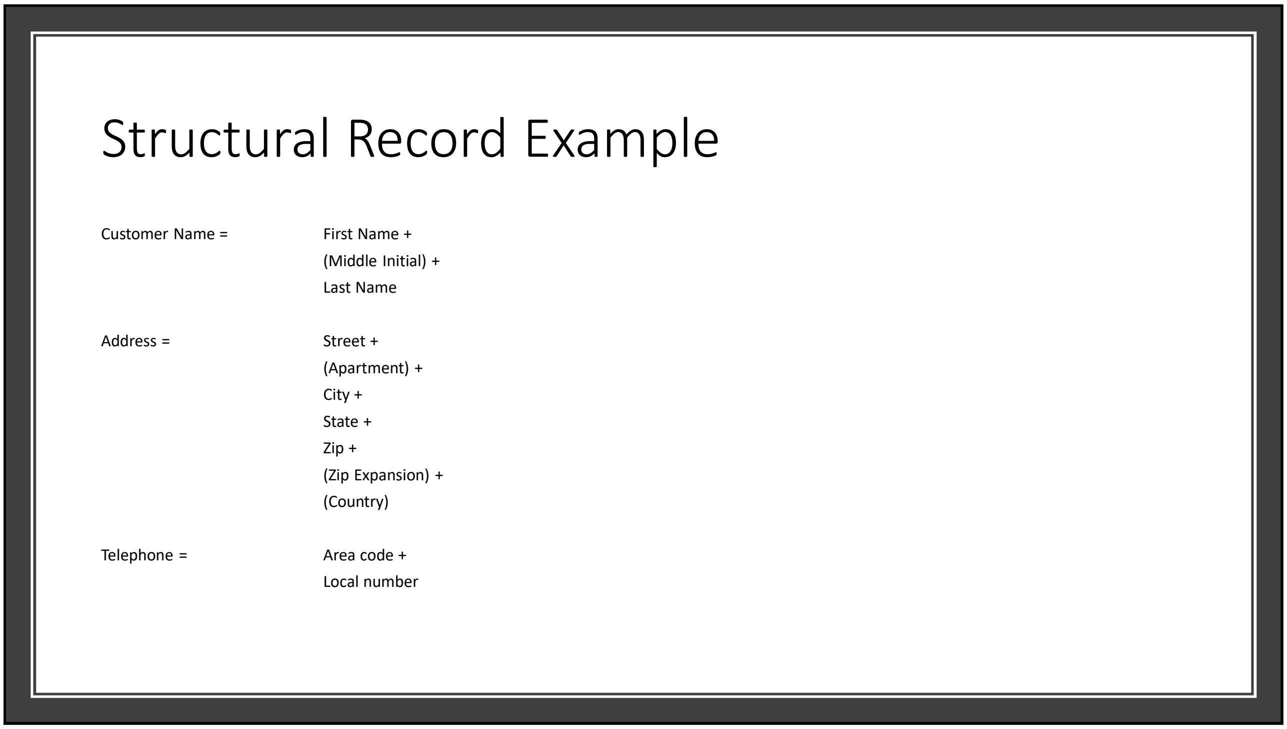
179



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.

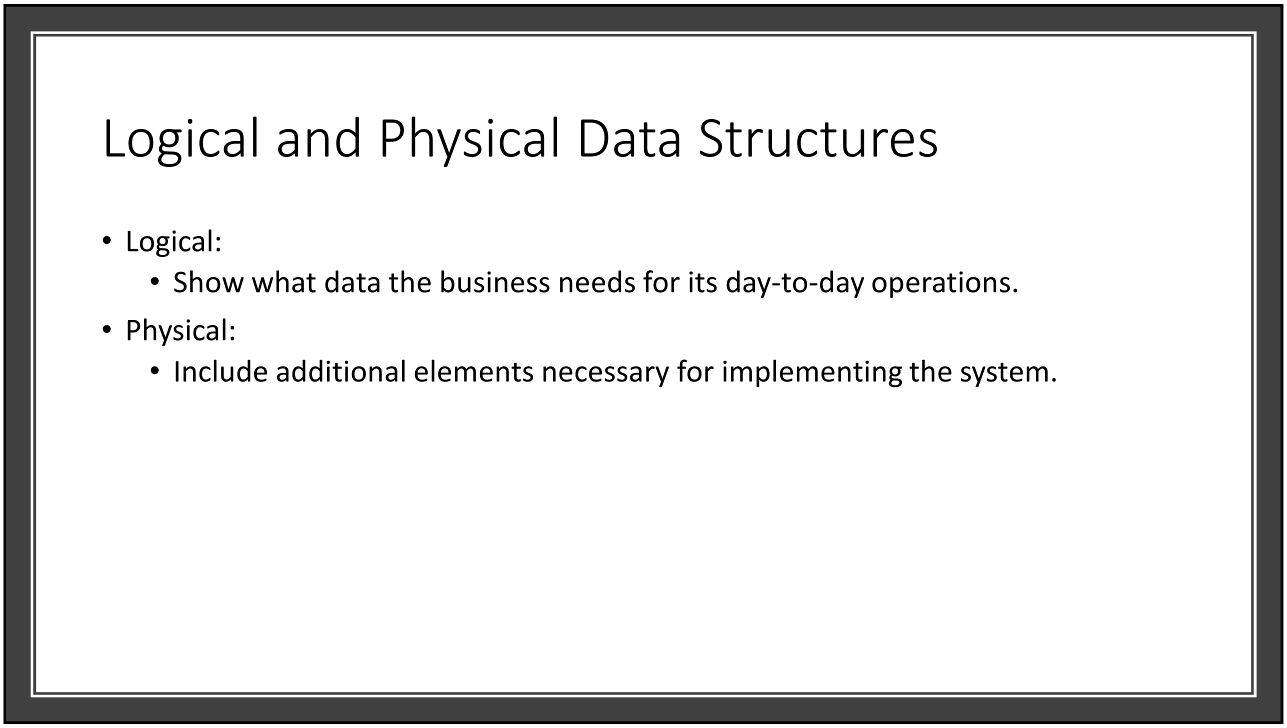
180



Structural Record Example

Customer Name =	First Name + (Middle Initial) + Last Name
Address =	Street + (Apartment) + City + State + Zip + (Zip Expansion) + (Country)
Telephone =	Area code + Local number

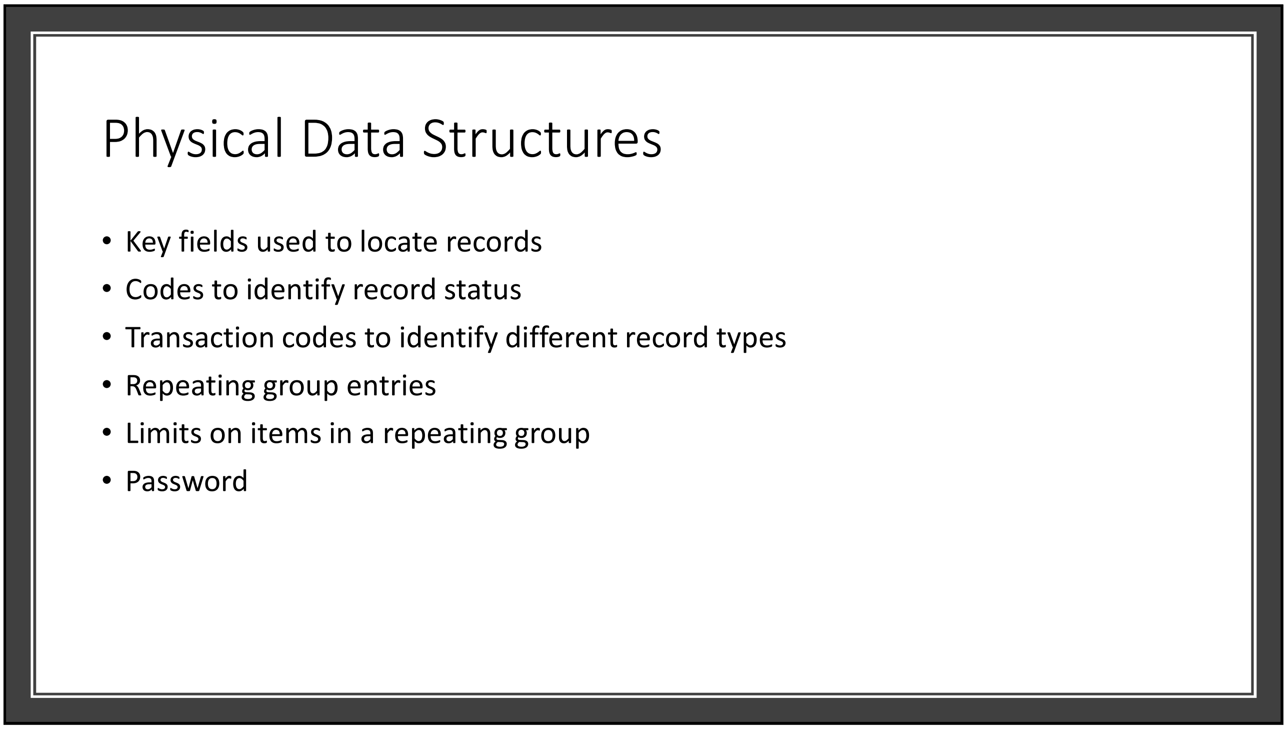
181



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.

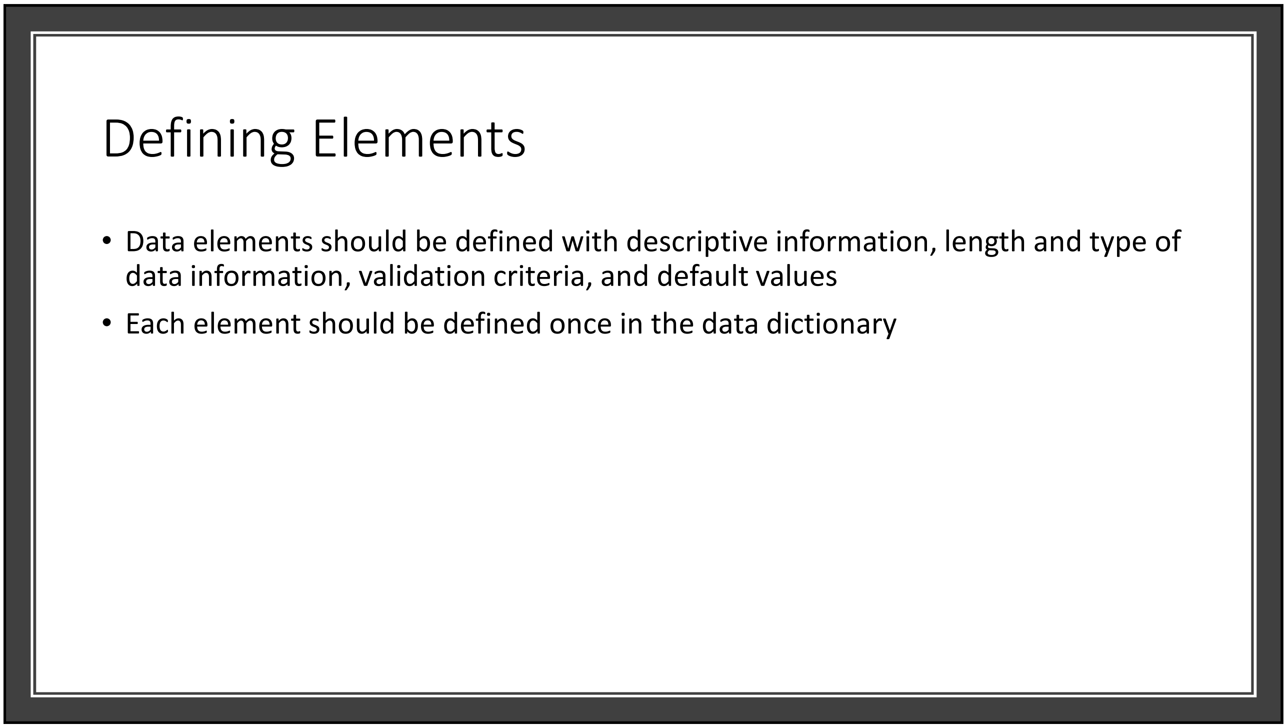
182



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

183



Defining Elements

- Data elements should be defined with descriptive information, length and type of data information, validation criteria, and default values
- Each element should be defined once in the data dictionary

184

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

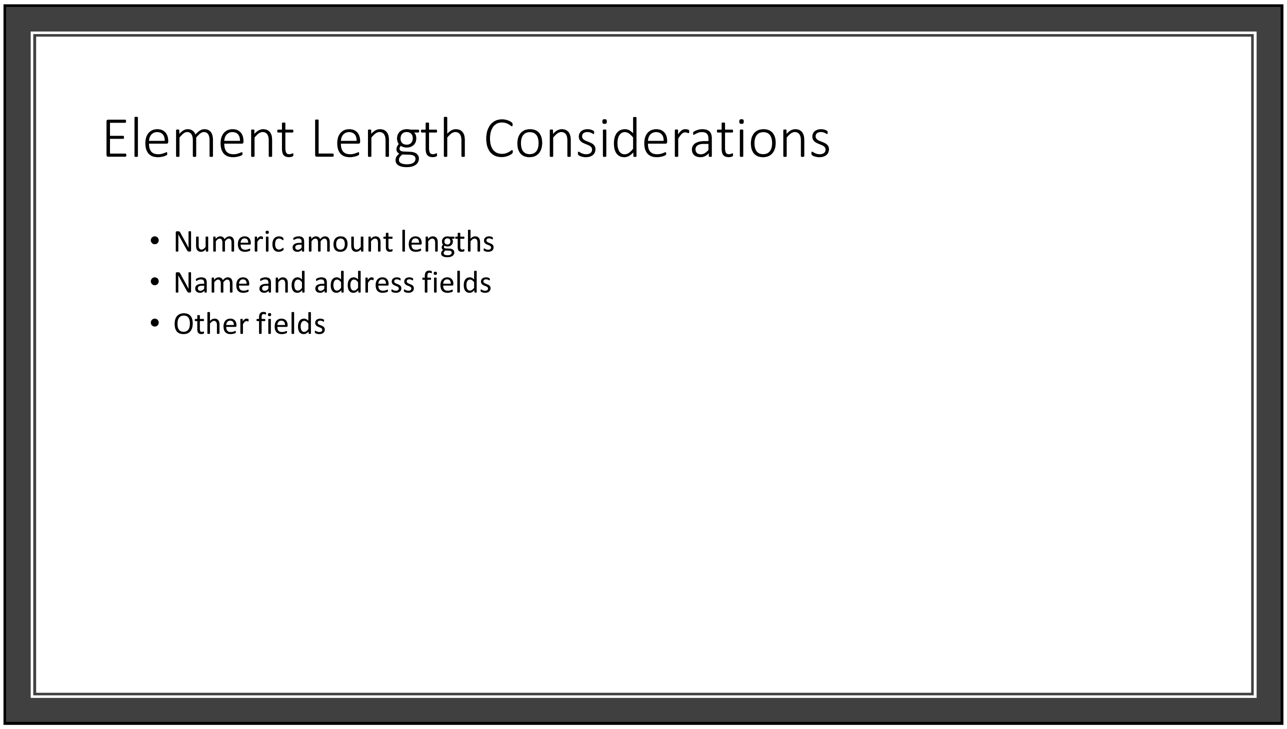
185

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.

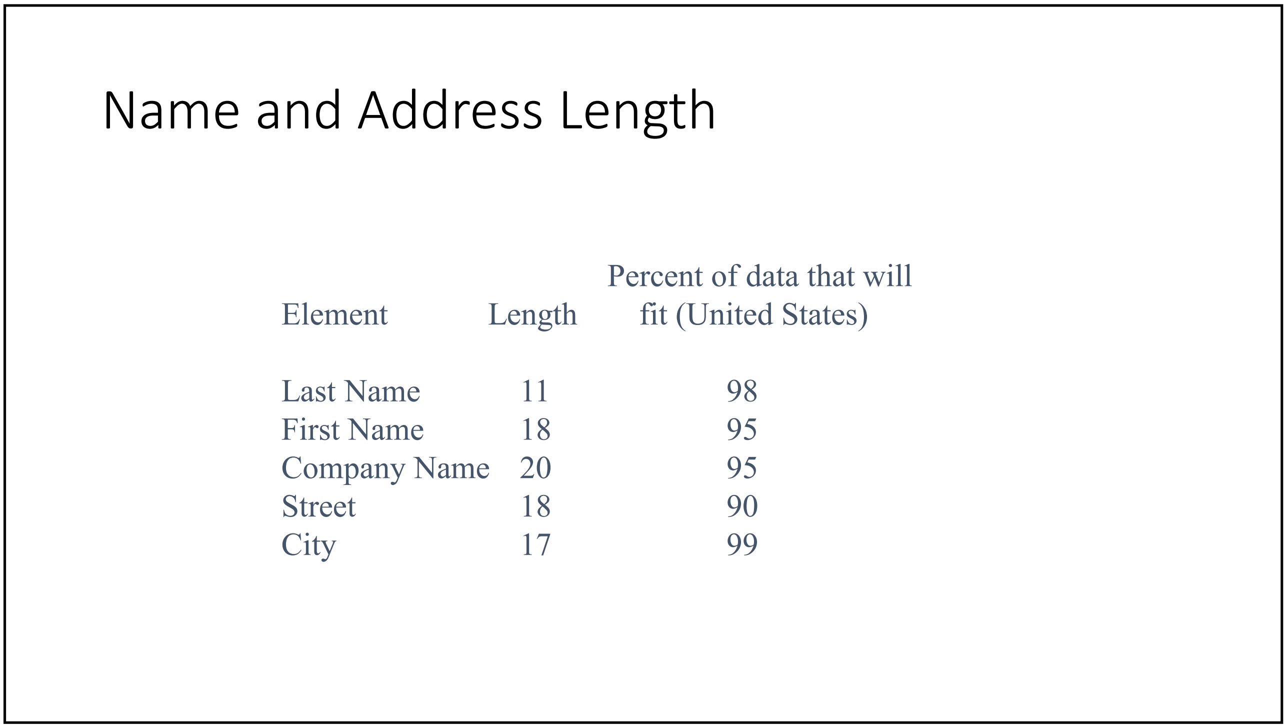
186



Element Length Considerations

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

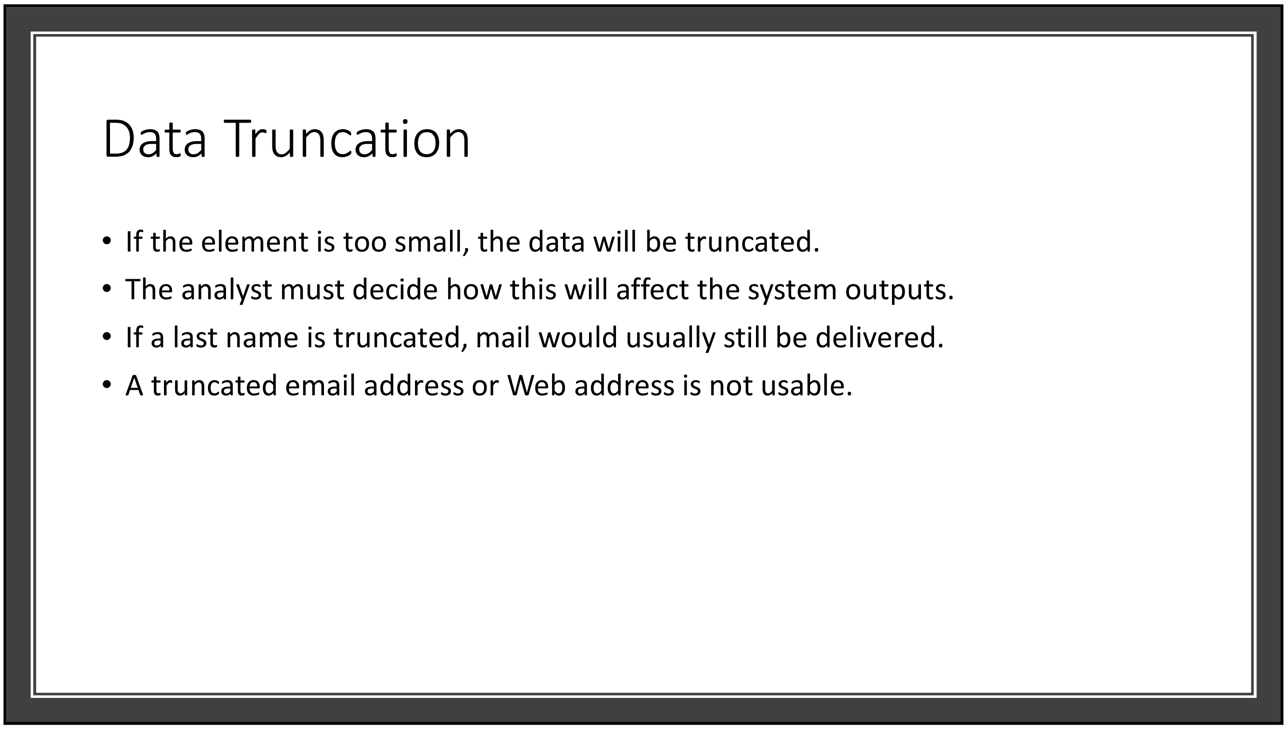
187



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

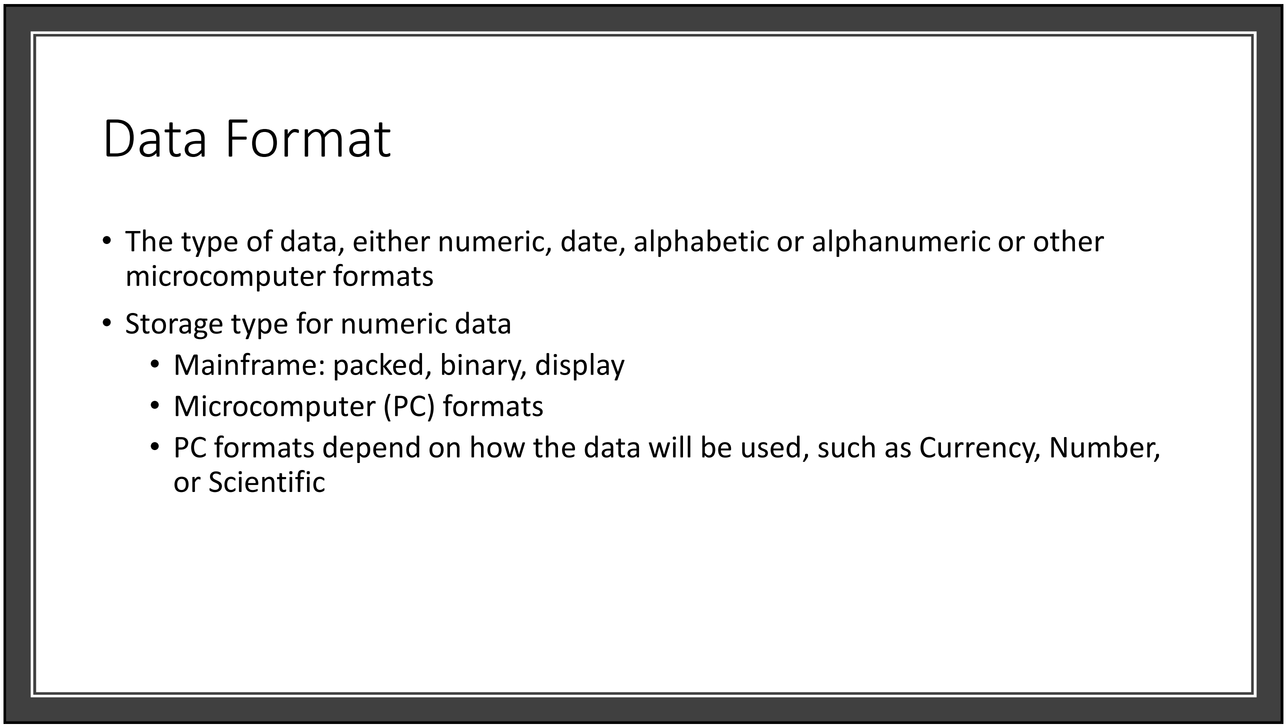
188



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.

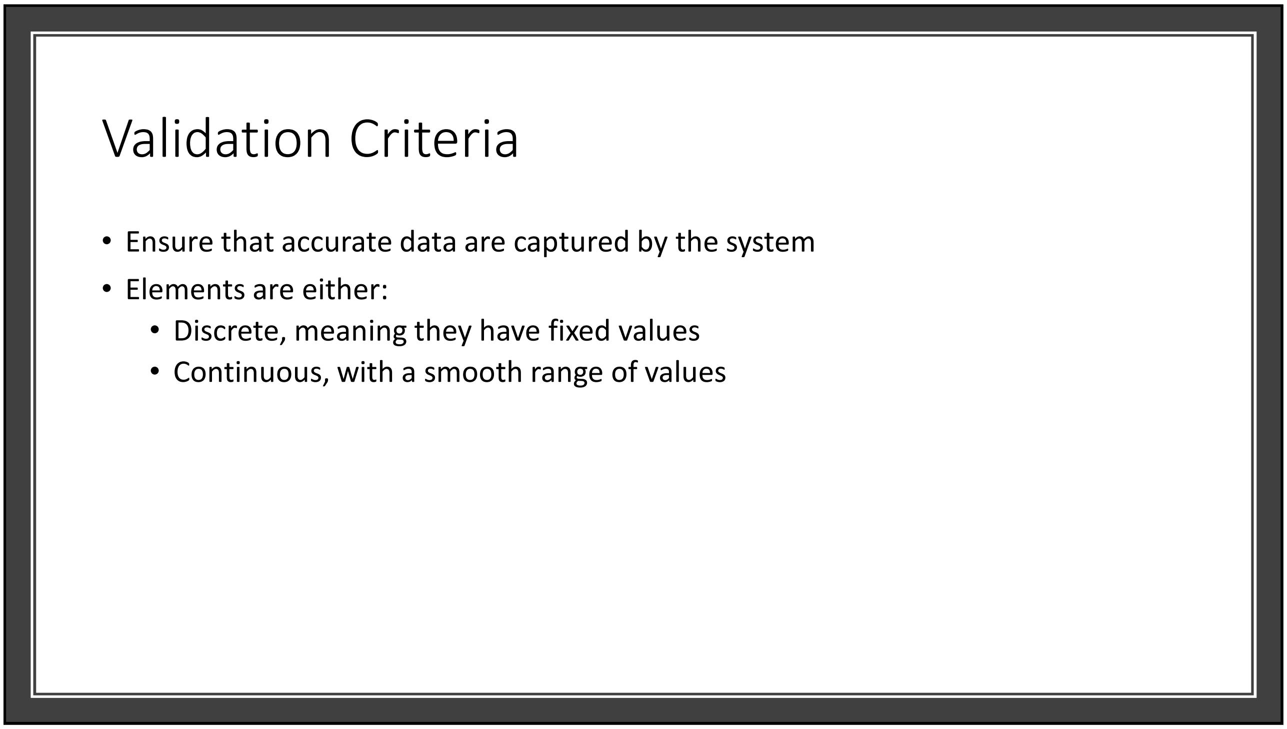
189



Data Format

- The type of data, either numeric, date, alphabetic or alphanumeric or other microcomputer formats
- Storage type for numeric data
  - Mainframe: packed, binary, display
  - Microcomputer (PC) formats
  - PC formats depend on how the data will be used, such as Currency, Number, or Scientific

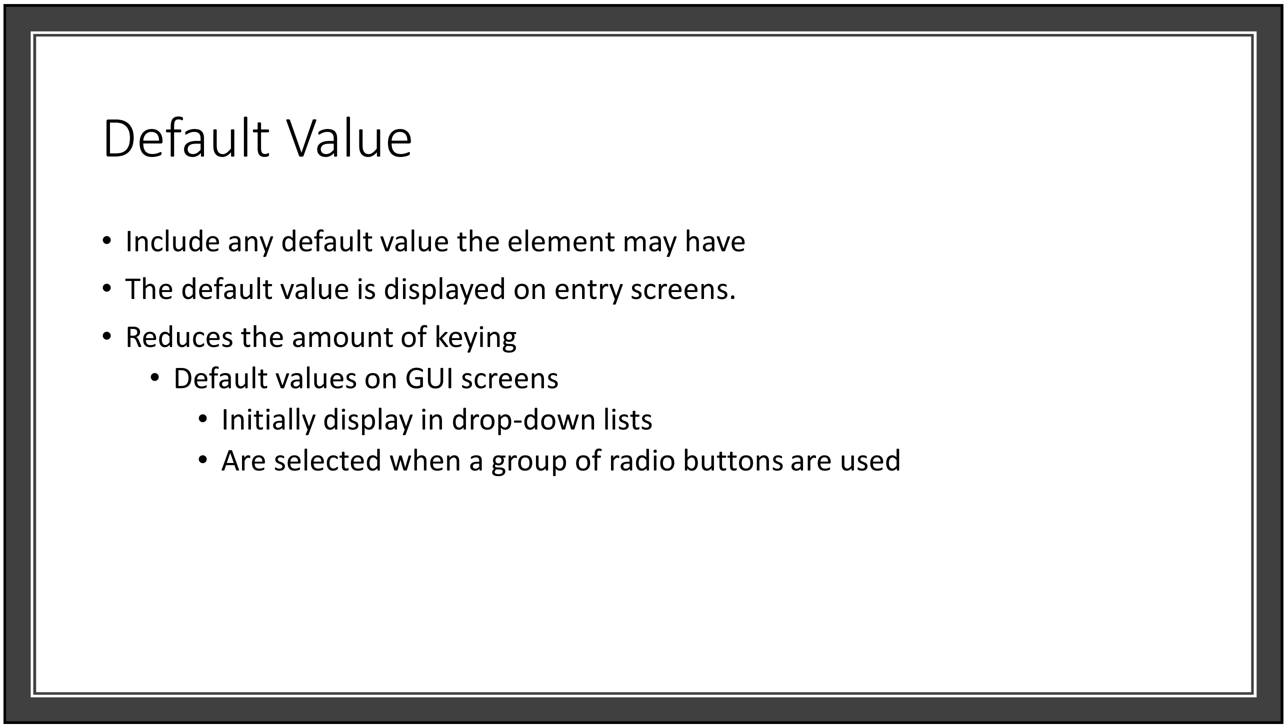
190



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

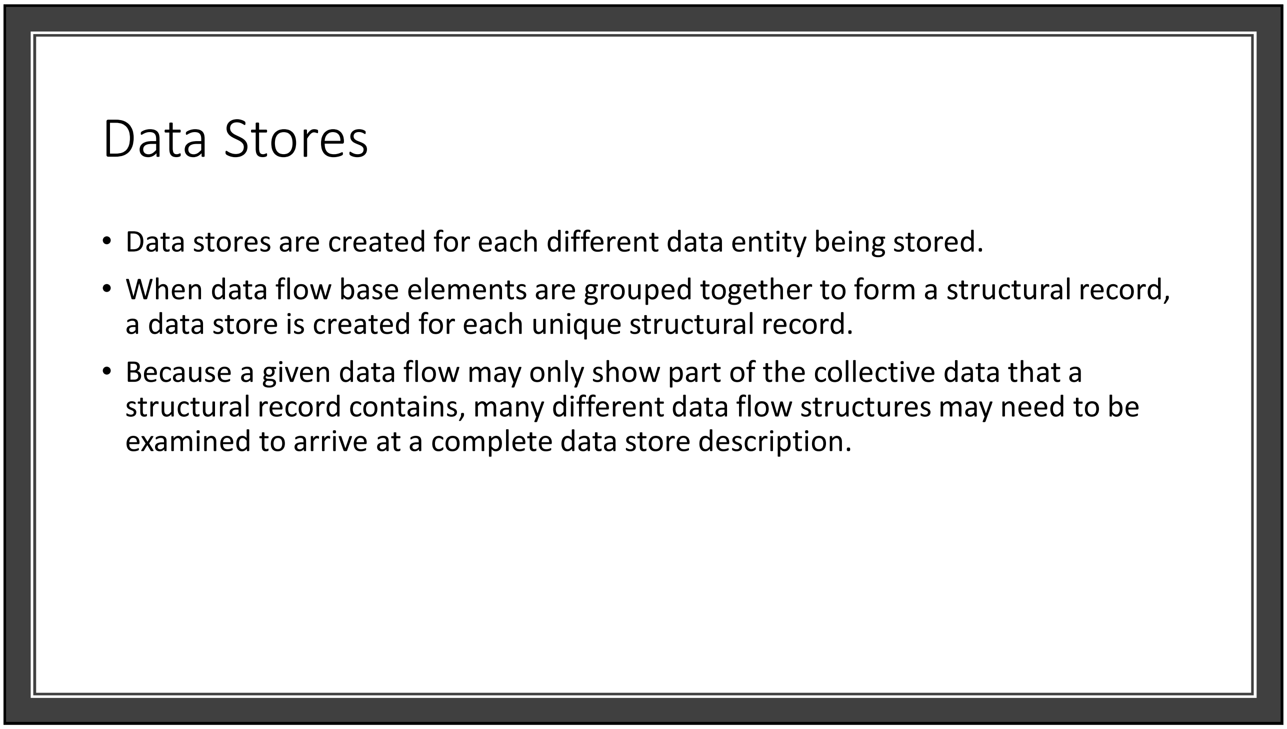
191



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

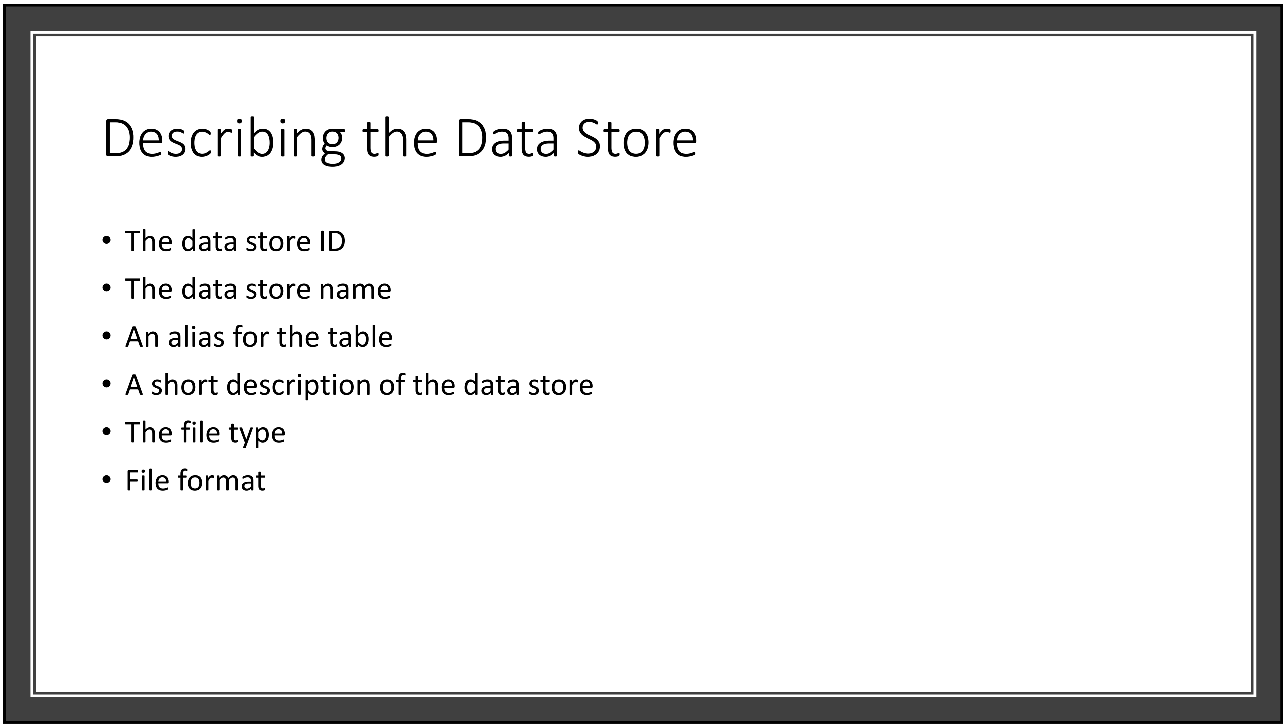
192



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.

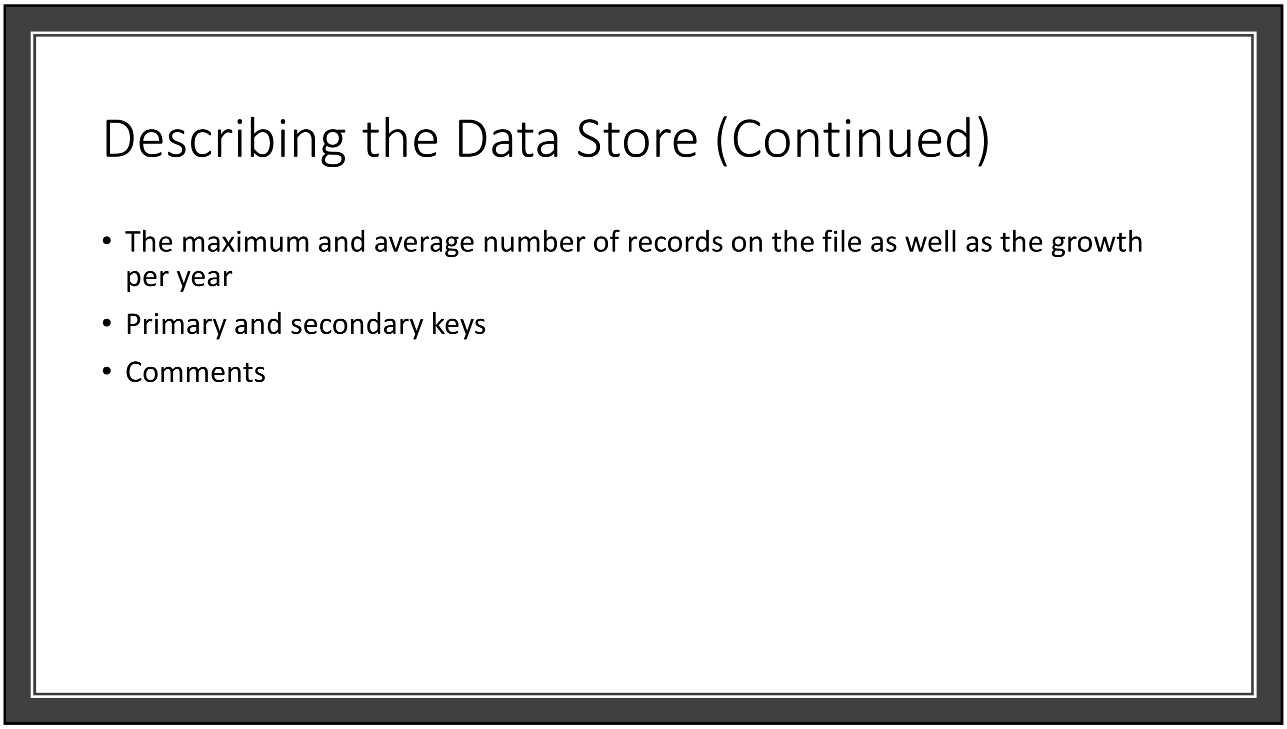
193



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

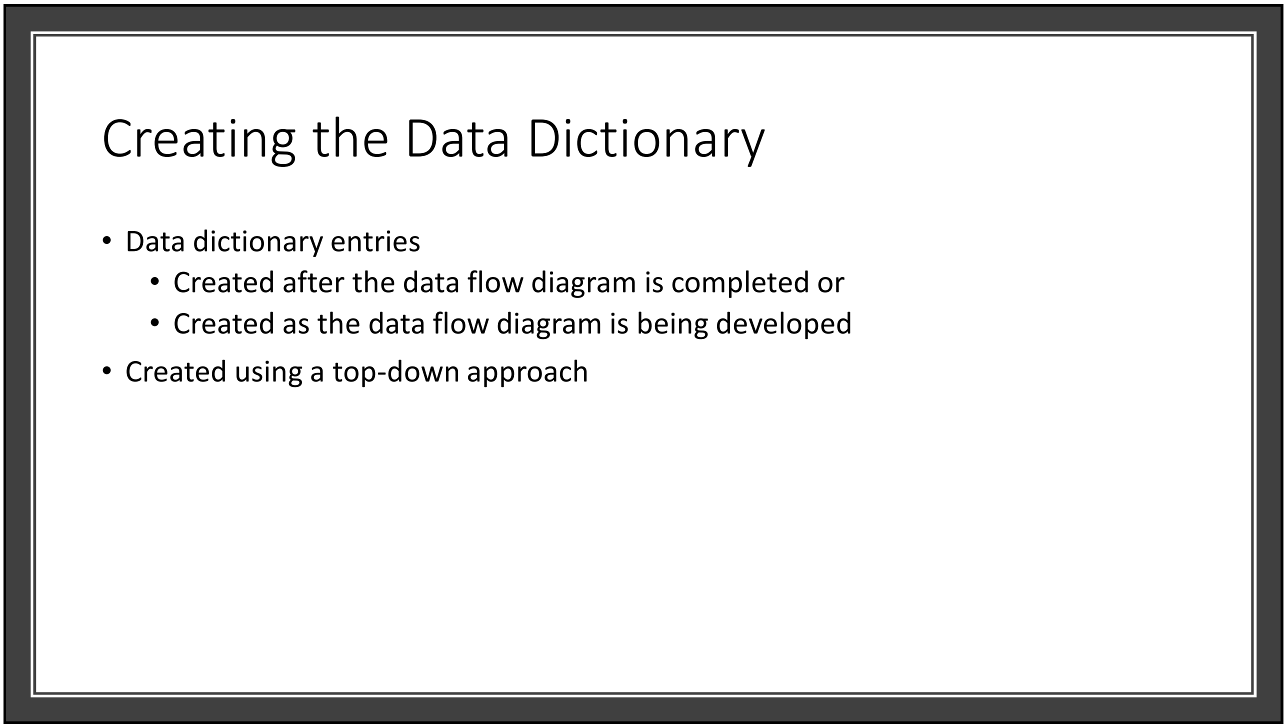
194



Describing the Data Store (Continued)

- The maximum and average number of records on the file as well as the growth per year
- Primary and secondary keys
- Comments

195

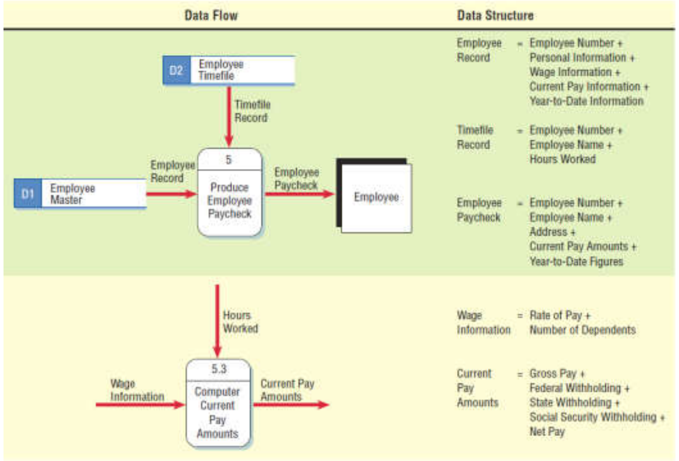


Creating the Data Dictionary

- 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

196

Two Data Flow Diagrams and Corresponding Data Dictionary Entries for Producing an Employee Paycheck



Data Structure

Employee Record = Employee Number +
Personal Information +
Wage Information +
Current Pay Information +
Year-to-Date Information

Timefile Record = Employee Number +
Employee Name +
Hours Worked

Employee Paycheck = Employee Number +
Employee Name +
Address +
Current Pay Amounts +
Year-to-Date Figures

Wage Information = Rate of Pay +
Number of Dependents

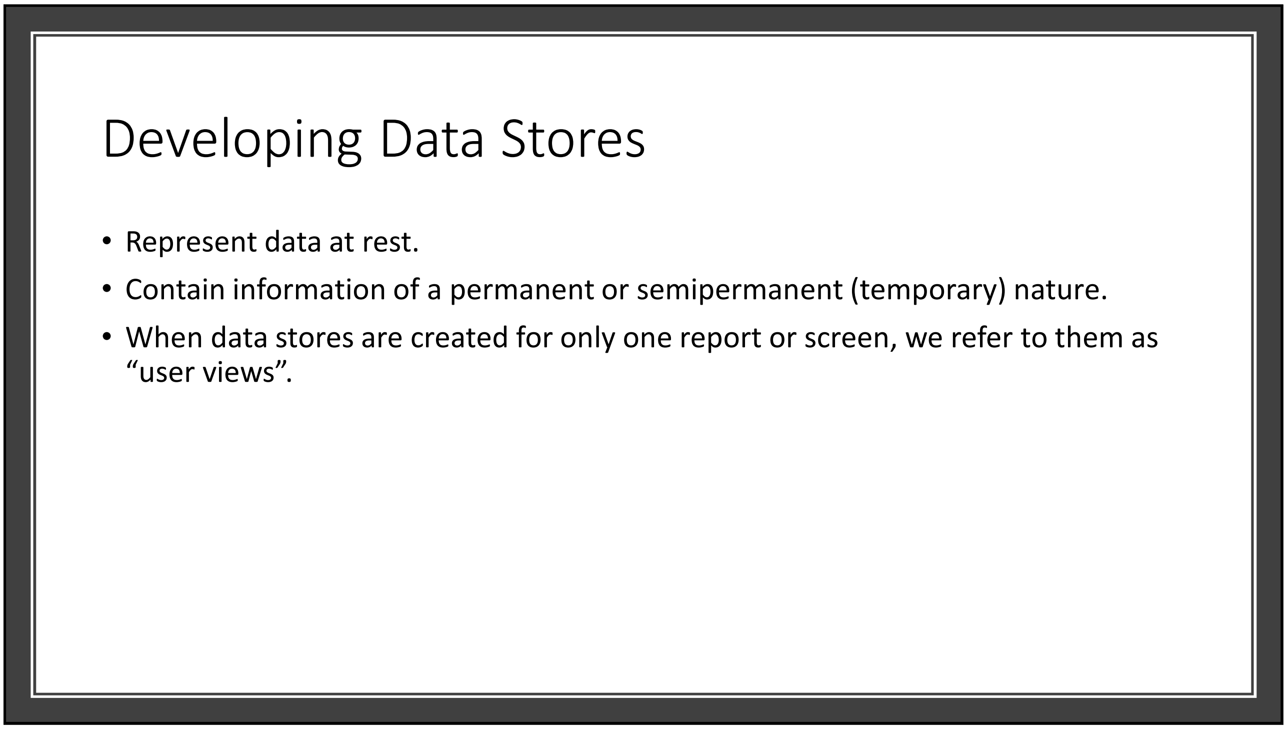
Current Pay Amounts = Gross Pay +
Federal Withholding +
State Withholding +
Social Security Withholding +
Net Pay

197

Creating Data Dictionaries

1. Information from interviews sessions is summarized on Input and Output Analysis Forms
   - This provides a means of summarizing system data and how it is used
2. Each structure or group of elements is analyzed
3. Each element should be analyzed by asking the following questions:
   - A. Are there many of the field?
     - If the answer is yes, indicate that the field is a repeating field using the { } symbols
   - B. Is the element mutually exclusive of another element?
     - If the answer is yes, surround the two fields with the [|] symbols
   - C. Is the field an optional entry or optionally printed or displayed?
     - If so, surround the field with parenthesis ( )
4. All data entered the system must be stored
   - Create one file or database file for each different type of data that must be stored
   - Add a key field that is unique to each file

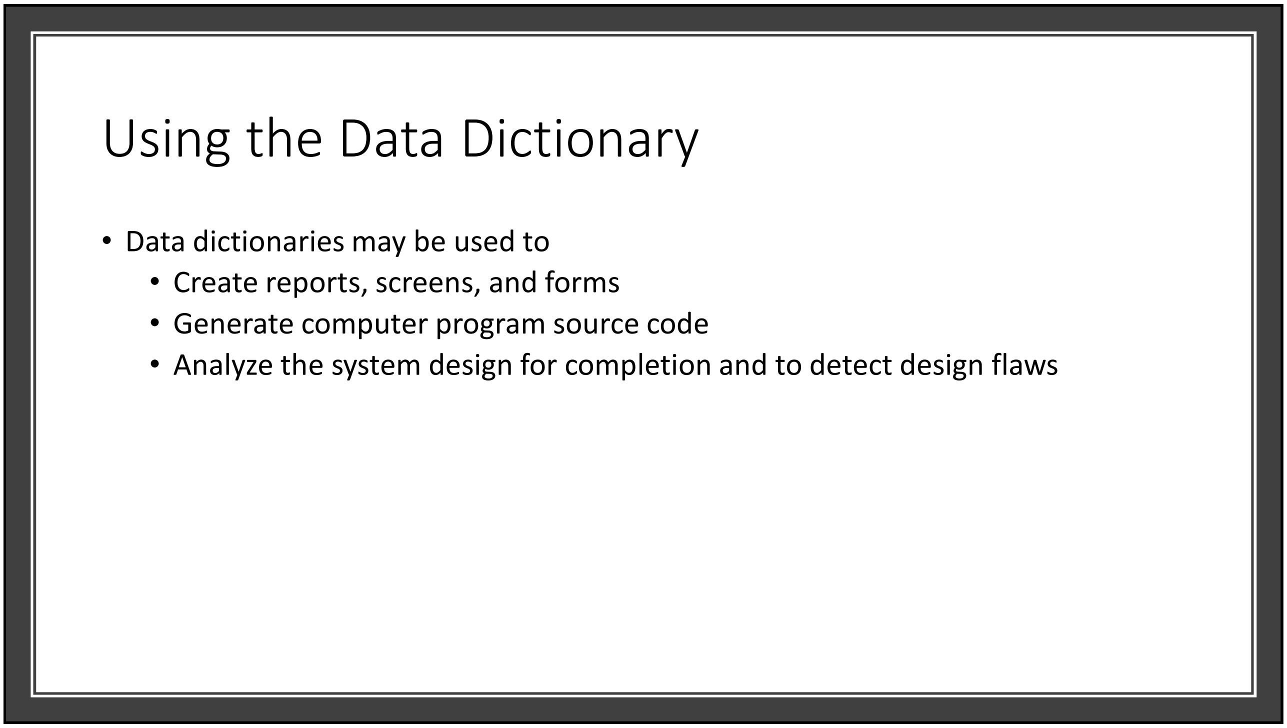
198



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".

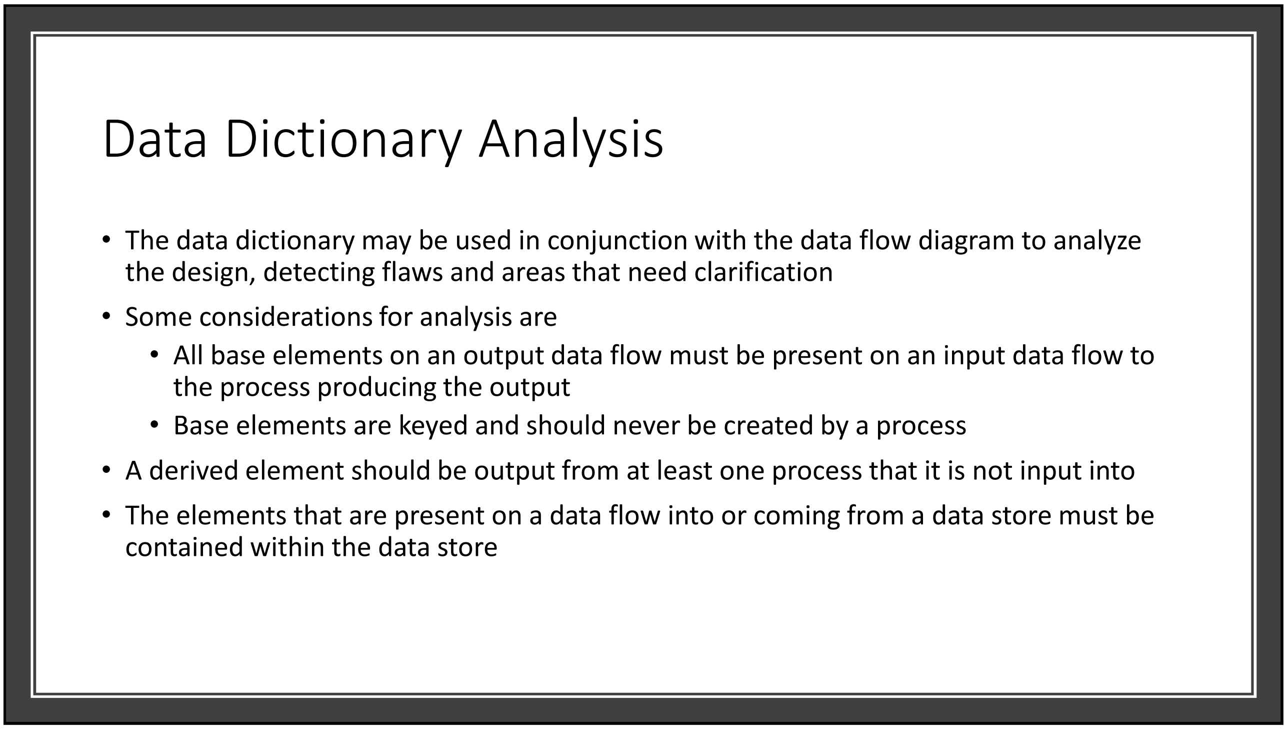
199



Using the Data Dictionary

- Data dictionaries may be used to
  - Create reports, screens, and forms
  - Generate computer program source code
  - Analyze the system design for completion and to detect design flaws

200



Data Dictionary Analysis

- The data dictionary may be used in conjunction with the data flow diagram to analyze the design, detecting flaws and areas that need clarification
- Some considerations for analysis are
  - All base elements on an output data flow must be present on an input data flow to the process producing the output
  - Base elements are keyed and should never be created by a process
- A derived element should be output from at least one process that it is not input into
- The elements that are present on a data flow into or coming from a data store must be contained within the data store