MODULE 7

DATA INPUT METHODS

Contents

DATA INPUT METHODS

MOTIVATION

During systems analysis it is essential for an analyst to decide the necessary and sufficient data for designing an application. DFD gives the dataflows and stores of a system. Individual data elements of dataflows and stores can be catalogued. Such a catalogue with description of each element and their types will be an invaluable aid while designing a system. A catalogue will also bring out if any data is duplicated/missed. A catalogue will also be an invaluable documentation of a system. Such a catalogue is called Data dictionary-It is actually metadata, i.e., data about data. After data dictionary is designed one needs to determine how the data is to be input. Data input methods depend on whether the data is filled in by customers in forms manually and later input by data entry operators or data is directly input by users on PC's. We thus need to understand both these methods.

Unless data input is correct, results will be unreliable. Information systems normally have a large volume of data. Because of large volume special controls are needed to ensure correctness of data input - otherwise it is difficult to find which data is incorrect. Thus it is important to design appropriate data input methods to prevent errors while entering data. Key data elements are important to identify records. They need to be unique, concise and understandable by users. Thus we need to study methods of coding key data element

LEARNING GOALS

At the end of this module you will know
1.The need for a data dictionary for an application
2.How to develop a data dictionary for an application
3.Design of forms and screens for data input.
4.Need and methods of coding data elements.
5.Coding schemes for automatic error detection while inputting data
6.Need for and design of input data validation methods.

# Data Dictionary : its development and use

WHAT IS DATA DICTIONARY

Data dictionary is a catalogue of all data used in an application, their names, type and their origin. In other words it is <u>data about data</u> which is called <u>metadata.</u> Data dictionary gives a single point reference of data repository of an organization. It is thus an important documentation which would be useful to maintain a system

HOW IS DATA DICTIONARY DEVELOPED?

The Starting point of developing a data dictionary is a DFD.

<u>Example:</u>
Consider the Receiving office DFD.
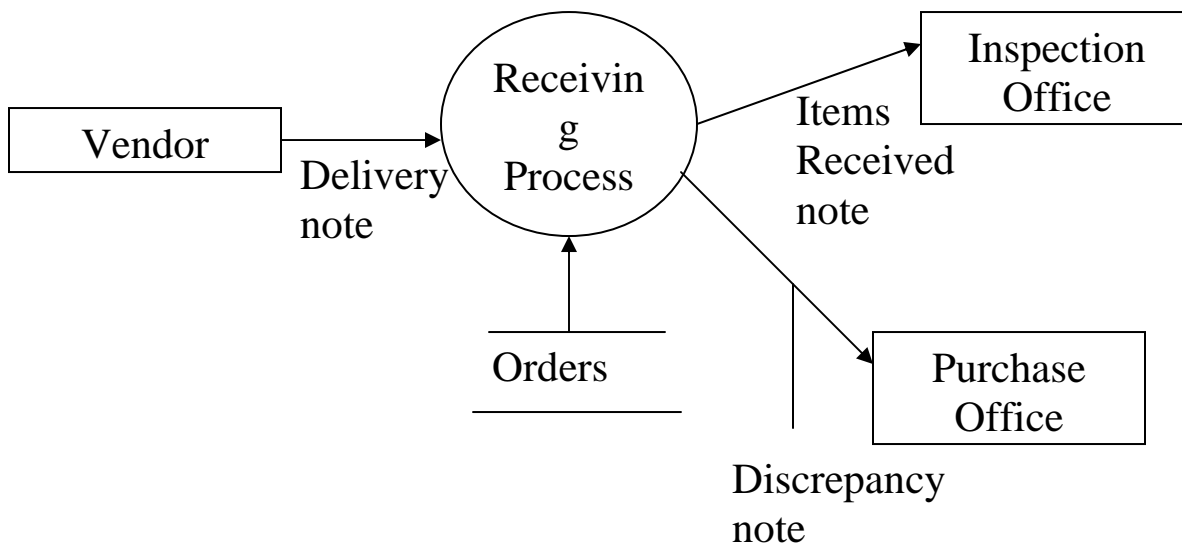
WORD STATEMENT OF REQUIREMENTS FOR THE ABOVE DFD

▪ Vendor sends items with a delivery note while fulfilling an order (along with the physical items) to a receiving office.
▪ Receiving office compares a delivery note against order placed. If there is a discrepancy a discrepancy note is sent to purchase office.
▪ Actual items received note is sent to the inspection office along with items received.

DATA ELEMENTS IN DATA FLOW

From word statement we derive data elements in each data flow.

▪Order no,Vendor name,Vendor address,item name,delivery date,quantity supplied,units
> Item name and Vendor name may not be unique. To ensure uniqueness
> we assign unique codes for them. Name of item is however still kept as
> it is to aid
> people.

Thus delivery note is:
Delivery note = Order no + Vendor code + Vendor name + Vendor address + item code + item name + delivery date + quantity supplied + units.

Discrepancy note : Order no + Vendor code + Vendor name + Vendor address + item code + item name + delivery date + quantity supplied + units + excess/deficiency + no of days late/early.

Items received note = Delivery note

Data in data store

Order records = order no + vendor code + vendor name + vendor address + item code + item name + order date + qty ordered + units + delivery period.

TYPICAL CHARACTERSTICS OF DATA ELEMENTS(CONTD)

Data dictionary gives in detail the characteristics of a data element.

Typical characteristics are:

Data name : Should be descriptive and self explanatory.This will help in documentation and maintenance
Data description : What it represents
Origin : Where the data originates
   e.g. input from forms, comes from receiving office, keyed in by user
     etc.
Destination : Where data will flow and will be used (if any)
Data Type : numeric, alphanumeric,letters(or text),binary(0 or 1; True or False), Integer, Decimal fixed point, real(floating point), currency unit, date
Length : no of columns needed
Limits on value : (if relevant)
e.g. upper and lower bounds of value (age>0,<100)
Remarks : (if any)

EXAMPLE OF DATA DICTIONARY ENTRY

1)
Name : Order number
Description : Used to identify order given to vendor
Origin : Part of delivery note from vendor
Destination : Receiving process
Data type : Numeric Integer
Length : 8 digits
Limits on value : >000,<=99999999
Actual value not relevant.Used only as unique identifier
Remarks: It is a key field.

2)

Name : Delivery date
Description : Date item is to be delivered
Origin : Part of delivery note from vendor.Is also in orders data store which is input to receiving process
Destination : Receiving process
Data type : Numeric Integer
Length : 8 digits
Limits on value : Date field in the form DDMMYYYY.
Should satisfy constraints of a date in calendar
Remarks: Blank fields not allowed.
e.g.05082004 is ok but not 582004

DATA DICTIONARY USES

Data dictionary can be enormous in size. Requires careful development. However, it is centralized reference document. It is an invaluable resource to design input forms, screens, data checking programs, process specification and database. It is very useful in understanding and maintaining system

Data input methods : Batch and Interactive

ON-LINE - User directly Enters data using screen prompts

OFF-LINE -Forms filled by users- for example-candidates for admission to a college fill forms

ERROR SOURCES

- Errors in on-line data entry due to poor screen design. System should inform the user immediately when wrong data is input
- Errors in off-line data entry due to bad form design and human errors by users and data entry operator.
- Using a form which leaves enough space for writing legibly and has clear instructions prevents user from making mistakes.

OFF LINE DATA ENTRY – PROBLEMS

It is not always possible for the machine to give message when input is wrong, error may be found after elapse of time period. Therefore good controls to automatically detect and if possible correct errors is required.

BATCH DATA ENTRY

```
Data entered        Keyboard           Input file        Data
in forms      →→   Data entry    →→                →→   validation
                                                         program

         ↑                                                  │
      Input              Error                              ↓
      batch              batch
                                    Error              Update
                                    report             program

                                                          │
                                                          ↓
   Output report  ←←   Data       ←←   Data      ←←    Data
                       processing       store            store
                       program
```
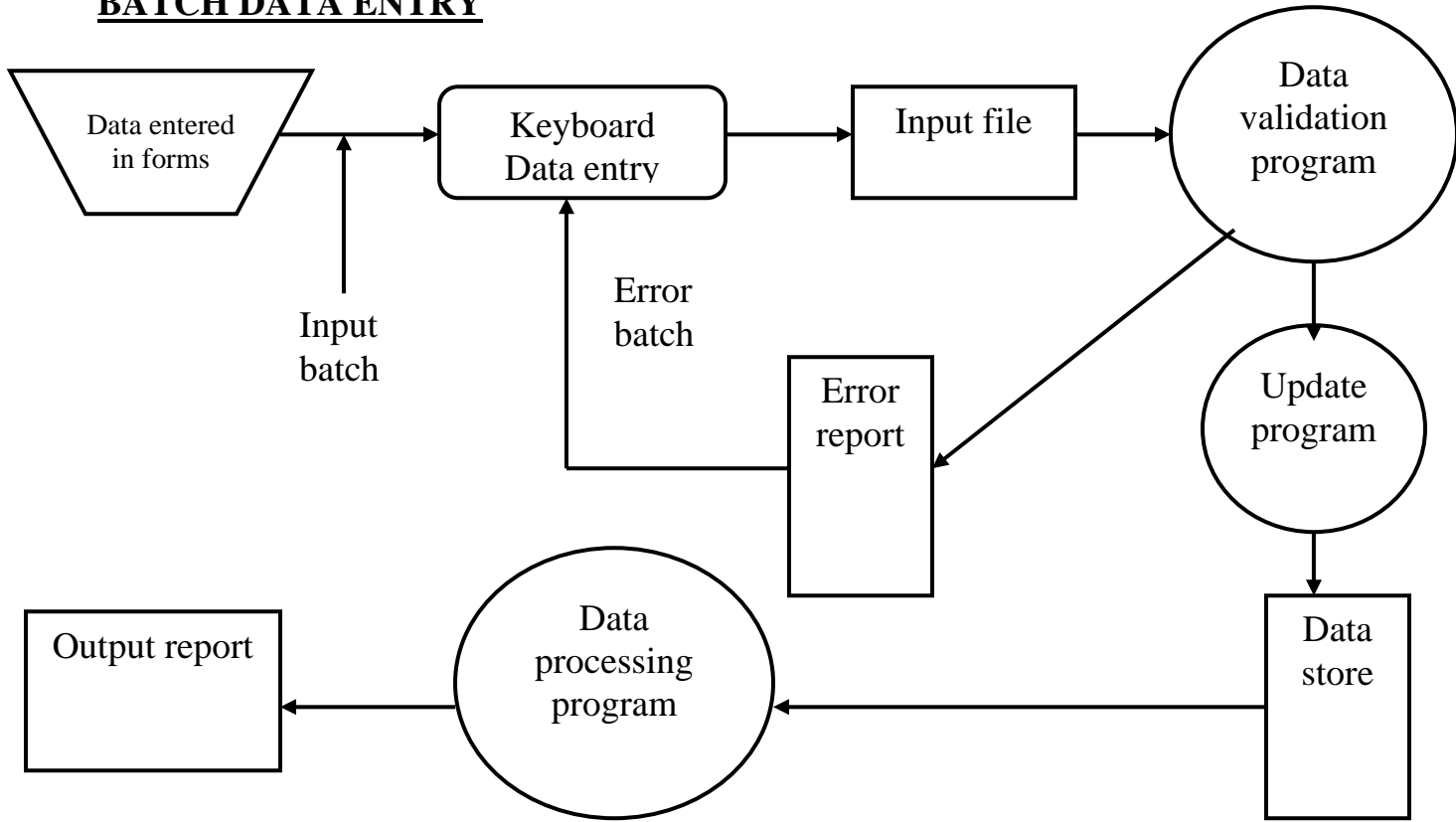
Name

| | | | | | | | | | |
....

Address

| | | | | | | | | | |
....

| | | | | | | | | | |
....

| | | | | | | | | | |
....

Bad design : Tendency will be to fill name on top line.
Not enough space for letters of address

Bad design : Choices are not codified.

Tick as applicable

Data entry operator will be confused.

```
┌ ─ ─ ─ ─ ─ ─ ─ ─ ┐
Individual
├ ─ ─ ─ ─ ─ ─ ─ ─ ┤
Hindu undivided family
├ ─ ─ ─ ─ ─ ─ ─ ─ ┤
Parent/Guardian of minor
```

Enter date

Day	month	Year

(Good design)

Enter date

| | | | | | | | | | | |

(Bad design)

Enter time

Hr	Min	Sec

Tick any of the following

Shri ☐   Smt ☐   Kum ☐
  1   2   3

Enter time

Enter name and address using capital letters Use one box for each alphabet

| | | | | | | | | | | |

(Bad design)

(Good design)

Name | | | | | | | | | | | | | | | | | | | | | | | |

Only address (do not Repeat name)

											Pin					

I am applying as:

Tick one of the boxes below. Clear instructions. Enough space for entry manually.

Individual ☐   Hindu undivided family ☐   Parent or guardian Of minor ☐

COMPUTER READABLE FORMS

As manual data input from forms are slow and expensive, attempts have been made to automate form reading using scanners, but this needs hand writing recognition and correct form alignment, which is not always successful. However, if forms require just darkening some pre-defined areas they can be machine read and interpreted.

Example – Multiple choice questions in exams where specific boxes are darkened based on the choice.

INTERACTIVE DATA INPUT

Advent of PC's and client/server model in computer networks, interactive data input is now widely used

Advantages are instant response when data is input so that errors are immediately corrected, flexibility in screen design which minimizes manual effort. And use of mouse and icons simplifies pre-determined choices of data

Three main models of interactive data input :

- Menus
- Templates
- Commands

MODELS OF DATA INPUT

MENUS

User presented several alternatives and asked to type his/her choice

EXAMPLE

SELECT ALTERNATIVE

Type 1 For entering new student record
Type 2 For deleting student record
Type 3 For changing student record

Your choice

TEMPLATE

Template is analogous to form. It has features to reject incorrect data input using built-in program and is user friendly

Example

Roll no [   ]

Name [   ]

       FIRST NAME/INTIALS   LAST NAME

Dept code [   ] | CE | CS | ME | EE | IT |

Year [   ]

Hostel code [   ] | A | B | C | D |

Pre-programmed to reject incorrect Roll no,Dept code,Year, Hostel code

Interactive commands guides user through alternatives

Example
Computer : Did you request deletion of record ?
        Type Y or N
User   : Y
Computer : Give student roll no
User   : 56743
Computer : Is name of the student A.K.Jain?
        Type Y or N
User   : Y
Computer : Is he 1st year student
        Type Y or N
User   : Y
Computer : Shall I delete name?
User   : Y

Normally all three models will occur together in application .In other words Menu, Forms and Commands are not mutually exclusive. In Graphical user interface design use of languages such as Visual Basic simplifies design of user interface.

LEARNING UNIT 3

Coding technique for unique data representation.

WHY DO WE NEED CODES?

- UNIQUE IDENTIFIER
  - -Example Roll no instead of name

- CROSS REFERENCING BETWEEN APPLICATIONS
  - -unique Roll no may be used in examination records, accounts, and health centre
- EFFICIENT STORAGE AND RETRIEVAL
  - - Codes concise- a long name will have a shorter roll no

WHAT ARE THE REQUIREMENTS OF A GOOD CODE?

- CONCISE - Smallest length to reduce storage and data input effort
- EXPANDABLE - Add new members easily
- MEANINGFUL- Code must convey some information about item being coded
- COMPREHENSIVE - Include all relevant characteristics of item being coded
- PRECISE - Unique, unambiguous code

WHAT METHODS DO WE USE TO CODE

1) SERIAL NO: Assign serial number to each item
2) BLOCK CODES: Blocks of serial numbers assigned to different categories.
3) GROUP CLASSIFICATION CODE- Groups of digits/characters assigned for different characteristics

Roll no	87	1	05	2	465
	↓	↓	↓	↓	↓
	Year admitte	Term admitte	Dept	Status UG/PG	Serial no In dept
(use meaningful characters)	87	1	CS	UG	465

4) SIGNIFICANT CODES - Some or all parts given values

Roll no	BA	1	95	C	B	R
	↓	↓	↓	↓	↓	↓
	Banian	Male	Chest size cms	Cotton	Color (blue)	Style (Round neck)

CHARACTERSTICS OF CODES

Characteristics ──────────────→

Codes	Concise	Expandable	Meaningful	Comprehensive	Precise
SERIAL NO	Yes	Yes	No	No	Yes
BLOCK CODES	Moderate	Yes	No	No	Yes
GROUP CLASSIFICA-TION CODE	No	Yes	Yes	Yes	Yes
SIGNIFICANT CODE **ERROR DETECTION CODE**	No	Yes	Yes	Yes	Yes

▪Incorrect data entry can lead to chaos.Mistakes occur as volume of data processed is large. Therfore its necessary to detect and if possible correct errors in data entry. Error can be detetected by introducing controlled redundancy in codes.

MODULUS 11 CHECK DIGIT SYSTEM

Error detection digit added at the end of a numeric code
Code designed in such a way as to detect all single transcription and single transposition errors which is 95% of all errors
Single transcription error-→49687→4<u>8</u>687

Single transposition error→ 45$\underline{68}$7→48$\underline{65}$7
Given code 49687 modulus check digit obtained as follows:
Multiply each digit by
Weights of 2,3,4 etc starting with least significant digit
7*2+8*3+6*4+9*5+4*6=131
             131/11=11,remainder 10; or 131 mod (11) =10;
(11-10)=1 append it to the code
The code with check digit=496871
If remainder is 1 then append(11-1)=10 code as X

WHY DOES MODULUS 11 CHECK DIGIT WORK

Given $d_n, d_{n-1}, \ldots \ldots d_1$ where d_1 is the check digit

$$\left(\sum_{i=1}^{n} W_i d_i\right) \mod N = 0 \text{ by design}$$

What should be the values of N & W_is
Single transcription error: d_k become t

$$\left(\sum_{i=1}^{n} W_i d_i\right) = \left(\sum_{i=1}^{n} W_i d_i\right) + t\, W_k - W_k\, d_k$$

As $\left(\sum W_i d_i\right) \mod N = 0\ (\mathbf{t-d_k})\, \mathbf{W_k} \mod N \neq 0$

OTHER CHECKING SYSTEMS
$\mathbf{(t-d_k)\, W_k \neq p.N}$ **where p is any integer**

Use modulo n check with n prime > largest code character value
<u>Conditions</u>

For hexadecimal codes symbols = 16, n =17
* For alphanumeric codes   26 letters $0 < W_k < N$
* As [t-dk] < 10 and W_k < N, N>10
* Product of integer not a prime 10 digits prime
* Smallest prime > 10 =11 => N=11

                            36 symbols

Therefore n=37.

VALIDATING INPUT DATA

When large volume of data is input special precautions are needed to validate data
- validation checks methods:
- sequence numbering - detects missing record
- batch control - use batch totals
- data entry and verification-dual input
- record totals-add individual values for checking
- modulus 11 check digit

CHECKS ON INDIVIDUAL FIELDS

- Radix errors - For example seconds field cannot exceed 60,month field cannot exceed 12
- Range check - Fields should be within specified range
- Reasonableness check - Telephone bill cannot be more than 10 times average bill of last few months
- Inconsistent data - For example : 31-04-99
- Incorrect data- Batch total checks this
- Missing data - Batch control data checks this
- Inter field relationship check -
  For example - Student of 8^{lh} class cannot have age > 25

REFERENCES

1. Most of the material in this module has been adapted from the book "Analysis and Design of Information Systems", 2nd Edition, by V.Rajaraman, Prentice Hall of India, 2003. Chapter 5 (pp. 49-52) and Chapter 11 (pp.154-170).

2. Good material on Data Dictionary is found in K.E.Kendall and J.E.Kendall , "Systems Analysis and Design", 5th Edition, Pearson Education Asia, 2003.Chapter 10 on Data Dictionaries. Chapter 16 Designing Effective Input.