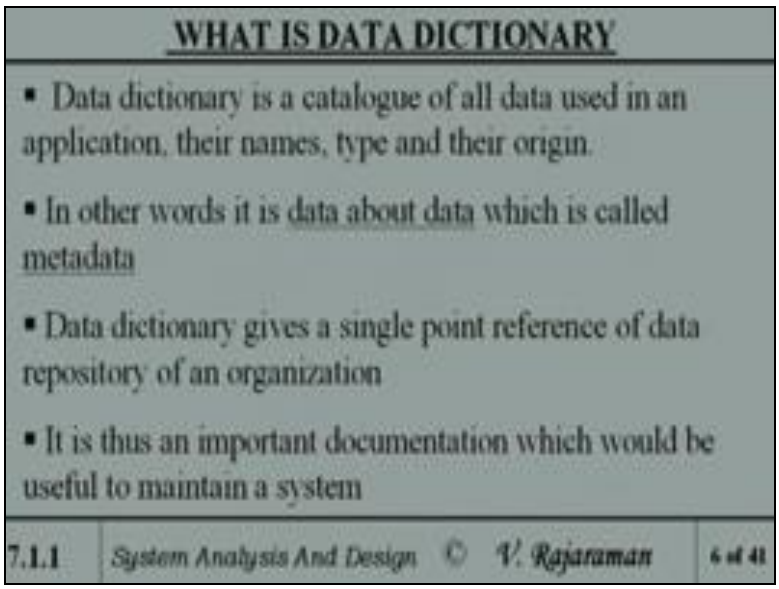
National Programme
Prof. V. Rajaraman
Department of Super Computer Education and Research
Indian Institution of Science, Bangalore

Lecture – 20
System Analysis and Design

We talked about the need for a Data Dictionary. I will recall, what we said? Data Dictionary is a catalogue of all data, which will be used in an application, their names, type and their origin.

(Refer Slide Time: 01:28)



In other words, it is data about data which is also called metadata. Data dictionary gives a single point reference of the data repository of an organization. So, it is a very important documentation which would be useful to maintain a system. As I said, very often most important part of a system which is designed is it is longevity. And, the statistical data tell us that in fact, 90 percent of the time in the life of a program of our system is spent in maintenance. Other words 10 percent is taken for actual development and start using it.

After you start using it, if it is long enough living in the sense that it is used for say 5 or 6 or 7 or 8 years. Then during that period, the amount of time spent on maintenance is quite considerable. The maintenance can be because of two reasons; one maintenance

issue is just remove bugs. And, normally if a system is well designed, the bug removal will not be the major component.

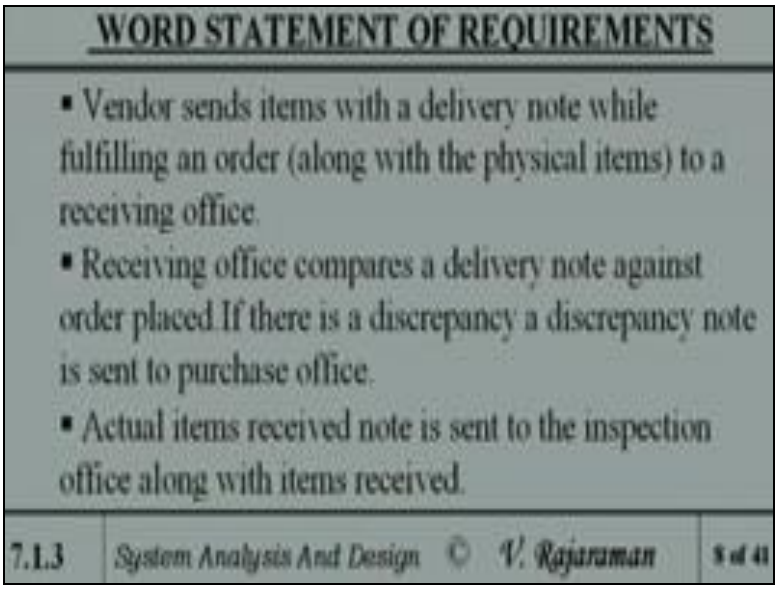
But, the major component is in terms of the fact that, there is some misunderstanding of the requirements or there are some changes in requirements which the user himself makes it known. So, when these are made known, then you have to add these requirements. And, you cannot start having an issue from the beginning of a system. We have use whatever we have currently. And, the current system has a good data dictionary. Then about, what all data is there and what their meanings are? Where it originates, where it ends, how it is used and so on.

And this great aid to the person, who is maintaining; because the developer is not normally the maintainer. Because, as I said lifetime of a system may be 5 6 7 years and the person, who actually develops the system, in fact in many cases may not be in the organization anymore. He may have left the organization. And, it will leave up to the other people to carry the ball and continue with this system.

And in fact, one of the major revenue owners of all software companies is not just the development, it is the maintenance also. Because, for maintenance they give you charge. After every fix, they make not a bug fix. But, a fix which is due to the users requirements change that is charged for. So that is the reason why the continuous revenue model is there. As far as the software company is concerned.

So, it is a one-time thing. Then, once development is over, then that revenue stops that normally fortunately. For the software companies does not happen, because it continues to evolve over a period of time.

(Refer Slide Time: 04:50)

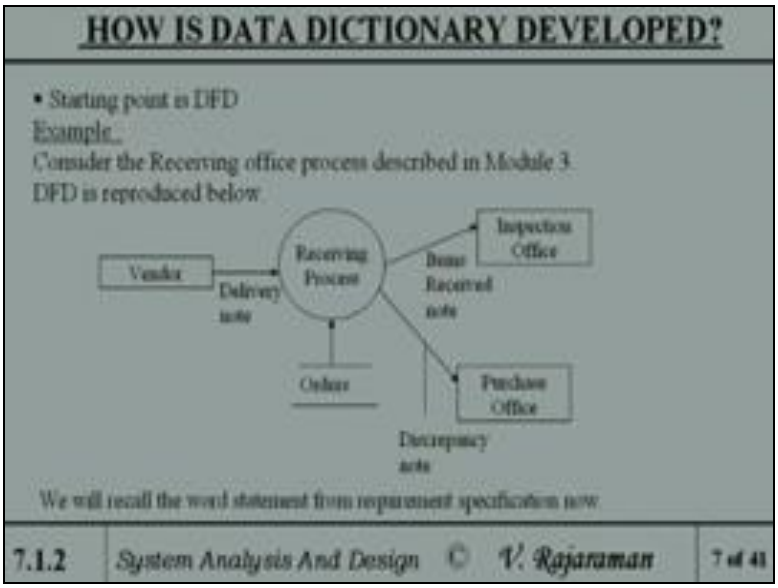


Now, let us recall word statement of a problem. And, I will go back to the data flow diagram corresponding to that. The word statement requirements, this is of course a toy problem which I can discuss in a class. In a real life problem of course, this huge requirement specification will be there. And of course, there will be a plenty of different data flows. And then, as we find it out in that part of the module parsing and so on whichever take place.

But let me; just take one such small system to for the sake of discussion. And, on how to create a Data dictionary. So, vendor sends items with a delivery note while fulfilling an order along with physical item to a receiving office. And, the receiving office compares the delivery note against an order placed. If there is a discrepancy, a discrepancy note is sent to the purchase office.

Actual items received note is sent to for the inspection along with the physical items which are received. And, from this word statement, we derive. In fact, we can show that go back and look at the dataflow diagram.

(Refer Slide Time: 06:11)

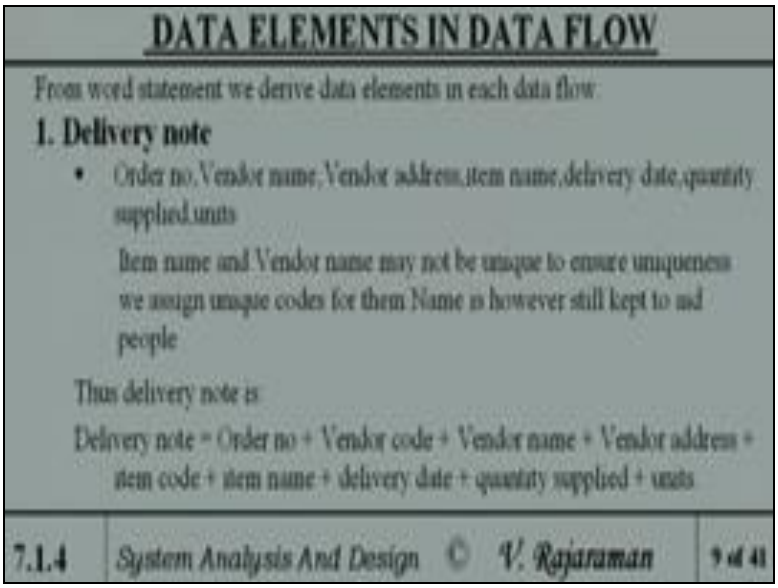


The diagram says that the vendor sends a delivery note along with the items. That is the receiving process which looks at the orders. Against and compares it against the delivery note, orders are in a file. And if there is a discrepancy, that is something which is ordered which is either short supply or extra supply or something which is ordered which never came or something no, which never came of course, I cannot find out. If something which I did not order came, then I will find out. Then, that will be in the discrepancy note.

And, so that will be sent to the purchase office, to tell the purchase office that, what was ordered is not delivered. If there is a short fall, then purchase office may decide whether to buy extra or just leave that rate at that. And then, when the items received note goes also to an inspection, which is a physical inspection to see if the items are of reasonable quality and also physical counting of the items. So, all these things are done by the inspection office.

So, this essentially what the origin of the data is given here. The destination of the data is given here, what is there, what files are there is given here and there is a process. As far as the dictionary is concerned, it is not really concerned about the actual process. Process specification is something different, what it is concerned about is, what is contained in the delivery note? What is contained in the items received note? And, what is contained in the discrepancy note? And, what is in the order file; because these are the things which are of concern to the data dictionary.

(Refer Slide Time: 08:04)



So, if I take one by one if I take delivery note. The delivery note should have the order number. Because, order number is the key item that is, because the order number is the key against which the delivery will be compared by the receiving office. Because, the file contain the order number as a key and the rest as information, because the vendor name normally you also put a vender code along with the vendor name.

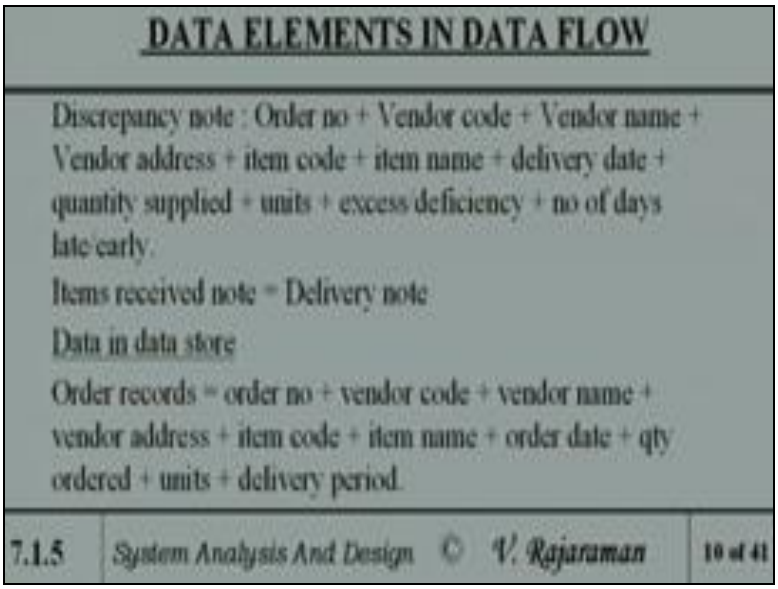
Because, you also require, because there could be two vendors with the same name, but of course same name and same address is rare. So, one may be satisfied with vendor name only. But in abundant caution one would also have that vendor code. So that, they quantity of vendor is there. And, I do not have to compare the names and code comparison is much faster. So, order number, vender code, vender name, vendor address item name delivery note quantity supplied and units.

In fact, I deliberately did not put the vendor code here. Just to illustrate the point that at the time you kind of look at the data which is going to be there, you suddenly realize that the code may be useful thing to have, particularly if you want to compare. And so, it is inserted at the time of creating this delivery note. In other words, I might at this time doing a design and not really implementing. At the design stage, I will create the data dictionary and that will be used in the implementation stage. Because the number of other places, where it will be used later on in the design process.

And, the item name and vendor may not be unique to ensure uniqueness we assign unique codes for them. So, name is however kept, still kept as a need to for the people. So, I have given here down below. Delivery note the order number plus vender code plus, vender name vender address item, code item name, delivery date quantity supplied and units. So, item also the item name is not sufficient.

You require a code, because again quantification of items is required for unique identification of the item and when delivery note is to be compared against orders. The item name comparison will take longer and of course, it may not be unique. So, item code is what is important in terms of anything which is compared.

(Refer Slide Time: 10:54)



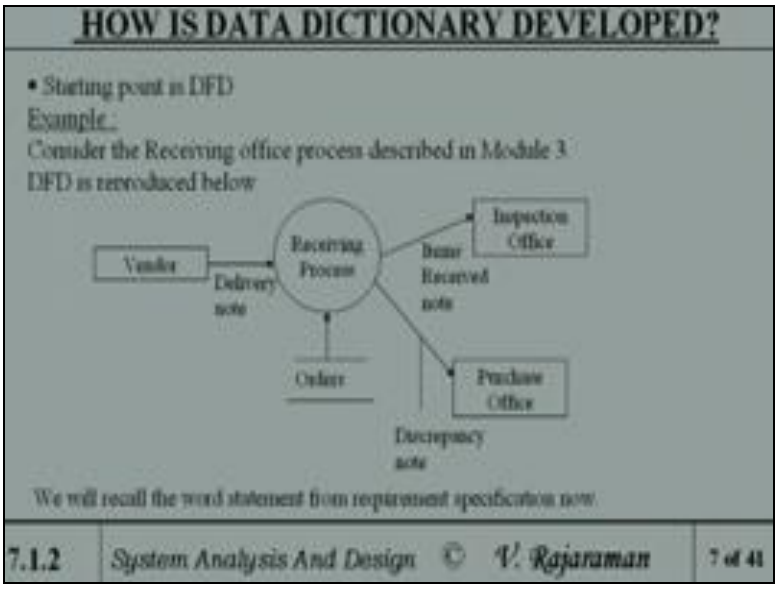
So, delivery not is one item. And, discrepancy note is a difference between the, what has come in and what is in the order. So again, you give the order number, vender code, vender name, vender address item code item name delivery date actually date of delivery, which is important in this case. Because you have to find out if there is a late delivery or early delivery. Because, in the order you have a requirement of the delivery date, how long, when is to be delivered.

Quantity supplied units excess and deficiency number of days late or early. So; that means, is it come late or has it come too early. At an item received note is actually the same as the delivery note. Delivery note and the item received note should, there are in fact. The effectively or two different names you might say for the same logical data item.

And, data stored where you have an order data stored. The order records will have order number, vender code, vender name vender address item code item name, order date quantity ordered, units and delivery period, how long how much of time is given for delivery. So, these are all part of the data in the data store.

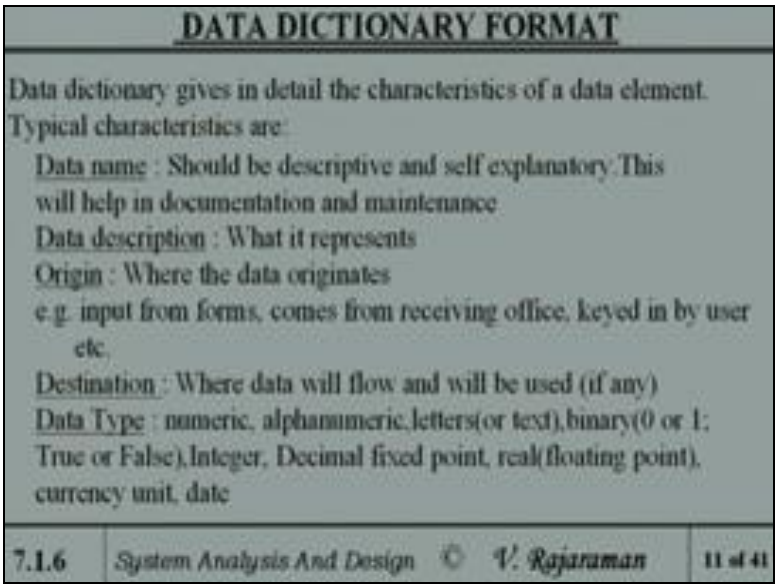
So data dictionary, now having come up with these things, I think delivery note discrepancy note and I got data in data store, what does now the, if I look at the data flow diagram, I have an item received note.

(Refer Slide Time: 12:53)



Item received note is nothing but the delivery note, because I am just going to take a copy of the delivery note. I am going to attach to the physical item box or whatever it is. And, send it to the inspection office. So, they are not really different ones. That is the reason I say they are the same.

(Refer Slide Time: 13:17)



So once I get this, then the dictionary has a certain format. As I said dictionary is data about data. And, dictionary gives detail characteristics of each data element and typical characteristics are data name. It should be descriptive, I mean just like a program, when you write a program. We always ask you to name variables in the program to have some meaning. So, that it is easy to read later on and it is part of the documentation and what not.

So, same reason here also, the name should have should be descriptive and to some extent self explanatory. It should essentially say what it is. You can again say for an instance. It could be a Pentium 5 mother board that could be an item name or it could be dim memory module number so and so, so and so. So this could be different item names. And, so the description should really say exactly what it is, because that is what is going to be used by inspection to actually do the inspection.

So, data name to be descriptive self explanatory. It will help in documentation and maintenance. And, data description what it represents, what this, what does this data represent? And origin where does the data originate? For in exact it does it come from input forms or it come from the receiving office or is it keyed in key in by user so on, what is the origin of the data and where is, what is the destination? Where will the data flow and what where will that be used? Which process uses for instance, if we are get a

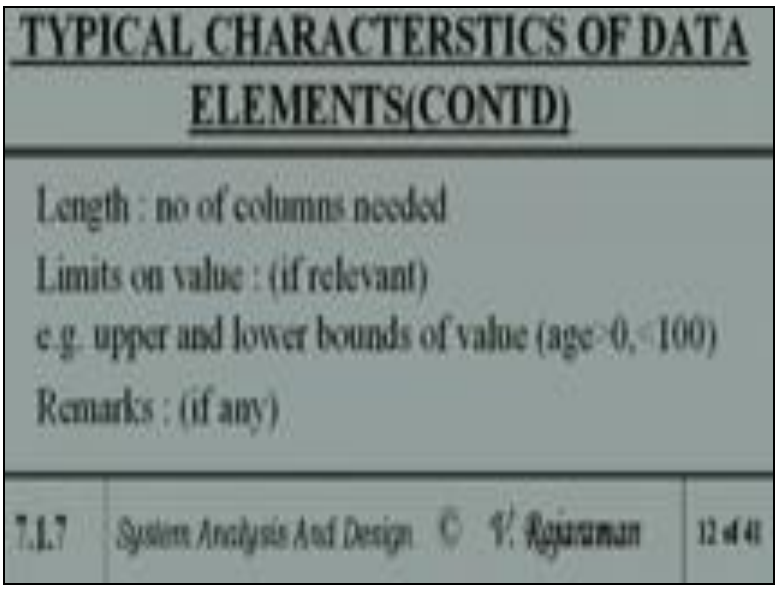
have a data flow diagram? The data will flow into some process and that process is going to use that data.

So the destination in the state would be a process where data will flow and will be used and data type? Data type of course, is important from the point of view of the program. And is it numeric, is it alphanumeric or is it letters or text only. Is it binary or 0 1 or decimal fixed point or it is a real. Is it a real number is it a currency in which case I have to get a currency note, like a dollar, rupees or euro or yen or whatever it is.

And then, the date, date is the data type also I put as a date because, date will be an important data type, in many applications in data processing. So type of data, the reason why this data type is useful is when you do a program later on to check the input data. If some data which is supposed to be alpha pure alphabetic field, you find out there in alphabetic field, numbers are coming.

And in numeric field, you got alphabets in it or letters in it. Then there is an error. So, that can be thrown out at the very beginning at the early stage and so it is very useful both from the point of view of programming and from the point of view of checking the data.

(Refer Slide Time: 16:56)



So, typical characteristics of data elements are length, number of columns needed. Particularly if, once I say is an integer, how many columns have to be reserved for

integer, limits on the value. Because, the limits on the value is sometimes very relevant like if you are entering marks and you have a limit.

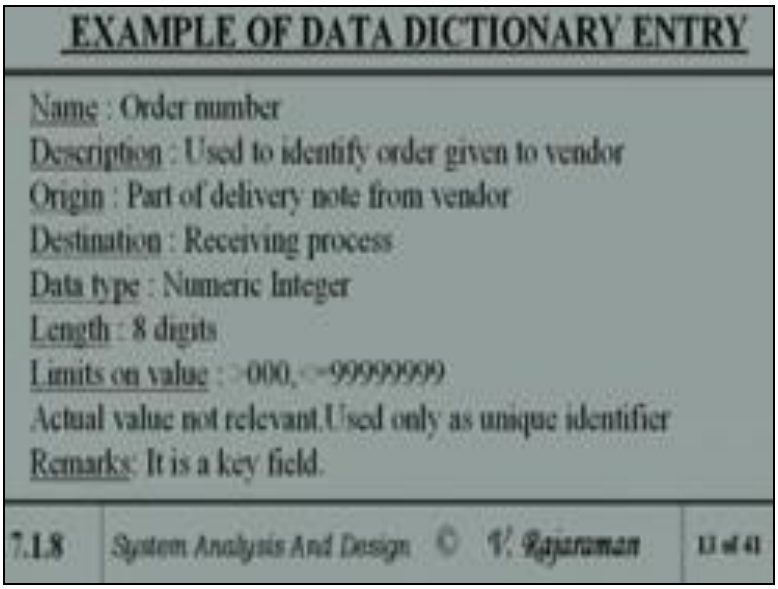
It cannot normally be below 0. It cannot be over 100. Example, yesterday I gave of a girl getting 126 and marks out of 100 would not have occurred. If they had put that limit and checked that limit, while inputting the data. It would have thrown it out. Thrown out the data right at the before the processing stage. So, you have to get give limits on data items and remarks if any.

And sometimes, it also useful to have the relationship between different data items in a particular data. Apart from limits, sometimes you also put the remarks if any. Normally you put that this data item is related to some other item. Like for instance you might say marks field of three different subjects are related. And, we have to cross check the relationship saying that. Suppose in two subjects the marks are very very high and one subject mark is very very low. Then this relationship will warn you, that possibly there is a data entry error.

So in other words the relationship saying that, if many marks are high and one is low. Bring it to the notice of the person, who is actually inspecting the input data. In other words, it is an exception report that is generated. The exception report will go to a supervisor, to see if there is any data entry error. Because, what may happen is that if you see, if you enter 36 instead of a 63. As far as the machine is concerned both are numeric. Both are two digits long It does not know the difference between 36 and 63.

But as a human being, they can make a difference of life or death for a student. So, that is important in terms of no, this kind of a fact that, there is a transfersation error which is occurred. If the supervisor sees that, immediately he will go back to the original. And, check whether there is any input error which is happened. So, these are important checks which can make as. And, these checks the origin is the dictionary.

(Refer Slide Time: 19:54)

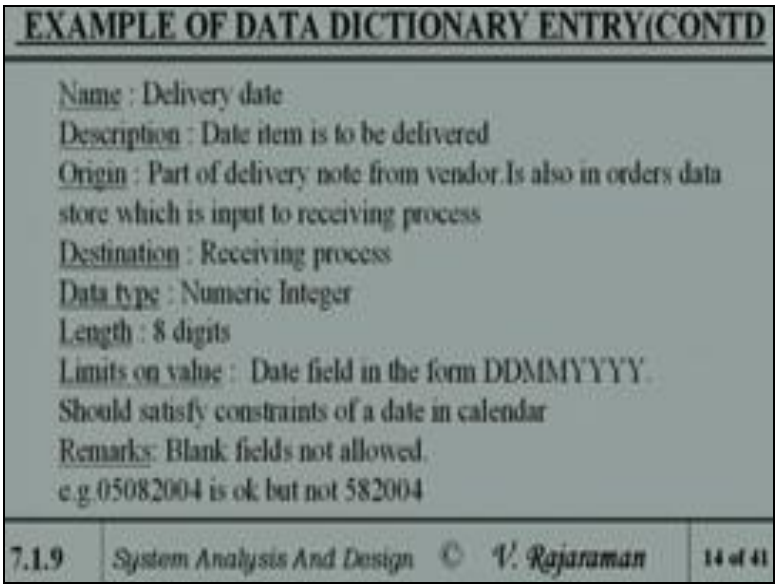


Like the order number is another data dictionary item. Description used to identify order given to a vendor that is the description of it. Origin, part of a delivery note from a vendor and it is also another origin would be that at the time of order placement. The order could be at order placement time also. But, in my restricted data flow diagram, it is the origin is the part of the delivery note from a vendor.

Destination is the receiving process. And, data type is a numeric integer. I am assuming that order number is a numeric integer. It need not be so. But in this case, there is a probably that is the constraint which the user has put that is a numeric item. And, length is 8 digits. It cannot be more than 8 digits and limits on value. Some value limits can be given. It cannot be less than 0; obviously and item code. It cannot be greater than 8 9s. Sometimes, they may even be even though you allow it.

In a given situation more than 6 digits may not occur. And, you may put a limit or say less than or equal to 9 9 9 9 9 9 that is 6 9s. Actual value is not relevant. Used only as a unique identifier. The order number is not an important. It is not used in arithmetic. It is only used in as identification. And remarks, you say is a key field. In other words, it is a key which is going to use for retrieval and so on. So, you got to have certain constraints in terms of key. Key as got to be unique and there are many other characteristics of key items which I will talk about later on.

(Refer Slide Time: 21:47)



Another example, of delivery note is another one. Data item is description, data item which is to be delivered. Origin, actually name is delivery date, is not it. Not delivery note. The name of the data item is delivery date. Delivery date is the sub part of the delivery note. Date item is to be delivered is description. And, origin part of delivery note from vendor. It also is in orders data store which is input to the receiving process.

In other words the delivery date is occurs in two places. One in delivery note as well as in the order, two plus itself it occurs, destination receiving process. And, data types numeric integer. It is 8 digits, I am receiving 8 digits. And, it has got to be in the format, D D M M Y Y Y Y; that means date month and year and should satisfy constraints of day in a calendar. In other words, whenever a date occurs in any data item and occurs in a data dictionary also. Normally, you have related function to check the validity of the date.

For instance, you cannot have say 31st of April, which are wrong date obviously. So, those kinds of things had to be automatically checked. And that of course, is a very simple program. It can be canned program of a function say checking the delivery date of a thing. And that is a given as a very very common exercise to computer science students in a programming course. Say, write a program to find out the date, years valid date or invalid date.

And use all the rules, including the leap year rule. Saying that if it is divided by 4, if it is a non century then it is divisible where it is a leap year. If it is century like 2000, then you

have to divide by 400. Fortunately, the year 2000 problem, part of it was solved primarily, because 2000 happen to be a leap year whereas the year 1900 was not a leap year. And, the year 2100 will not be a leap year and after the year 2000 problem, people have woken up.

And then, they have nowadays always put 4 digit year and not just 2 digit year. Because, it may require, of course I do not expect to live up to that. But it may require in 2099 that year, it may again occur. So, anyhow have the reason why 2 digits are used for year in the earlier systems was that storage was at a premium. The storage available on computers like IBM 14 Rom which was used for data processing extensible was a total main memory of only 16 k digits, that too. 16th k digits; obviously, you try to kind of say whatever digits you could do.

And so, you say those 219 as redundant. And it supporting 19, you just put 66 67 after all computer started only in the year in a serious way. Only in the year may be 50s 57 58 and so. One did not worry about the year 2000 in 57 58 and it carried on like that. And, suddenly people woke up in the1994, only 6 years to D-day and so all kinds of problems arose.

But of course, now it is so forgotten, because we are now in 2006, 6 years ago all this happened and that is before your, many of you started college. So, you do not have even heard about the so called Y to K problem, which the entire world worried about for at least one year. And the midnight of 1999, everybody was worried, whether the plains next day will take off or thinks like that because, by that time computers had been used extensively in many many things.

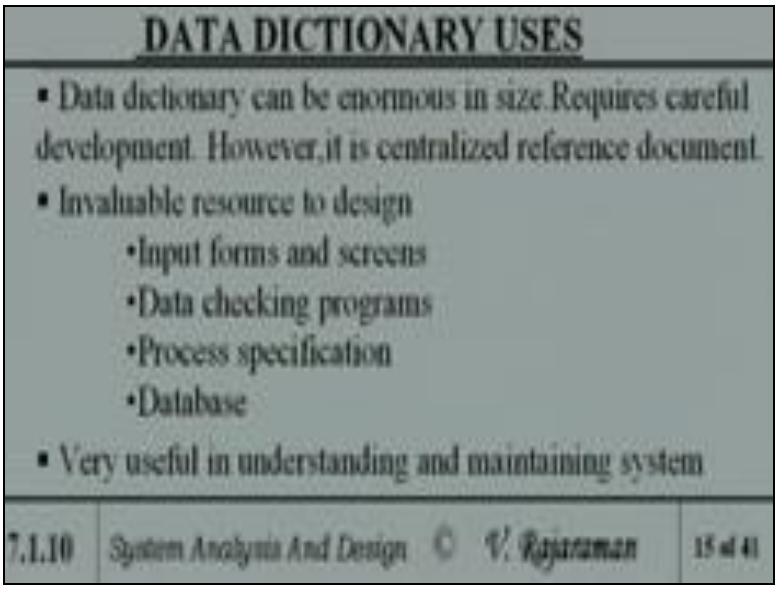
In any case, now we have learnt. And also, it is important to specify the order, the day, month and year. Because, that is the format we normally use in India, in England and other places. But, Americans have their own way. They put month before the day. Month, M M D D Y Y Y Y, there is no either merit or demerit for either one of them. Because, of some habit as far as we are concerned. We are habituated to, day, month and year. So, let us continue with day month and year.

Unless you are writing the program for an American customer, in which case he will expect you to put it as M M D D and Y Y Y Y. So, it is extremely important. And, once you put this, it also the remarks you will say that blank field are not allowed. Because, I

am going to check it, the date and so on by computer program and program is done, the program is done, it will always assume. The first two digits is a date, if the format is that.

So, no blank fields and that means, if it is a 5th of August, you have to say 05 08 not just 5 8. Then, if you say 5 8 2 0 0 4, then the machine will throw up it is hands. And say, it is a wrong dates, because the program will come with a wrong date. So, you have to write it. It is not right to put this.

(Refer Slide Time: 28:19)



So, these are important points which are specified in the data dictionary. Data dictionary can be enormous in size, because if we are going to catalogue, every data item which is going to occur in a large system. It can grow every a very small data flow diagram with one process. One input and two outputs and one file we have so much of detail.

But as and when you have developed the data flow diagram, if you bend up the data dictionary then and there, then you will not have to wait till the last moment to come up with an enormous task of creating the data dictionary, because it will have than huge number. So, this way even before when you pass, you can very clearly see, what dictionary is relevant to, what parsed part of a data flow diagram. So, it is a centralized reference document. Normally, it is alphabetized and so on.

So, whenever you have a program and one variable is used. And, you do not understand what the variable is and what it is used then you go back to the dictionary. Of course, as

dictionary will be kept in a nicely indexed file in a computer. So, you just type in the name and immediately it will go and search. And, pick out that and give you all the details about that data item.

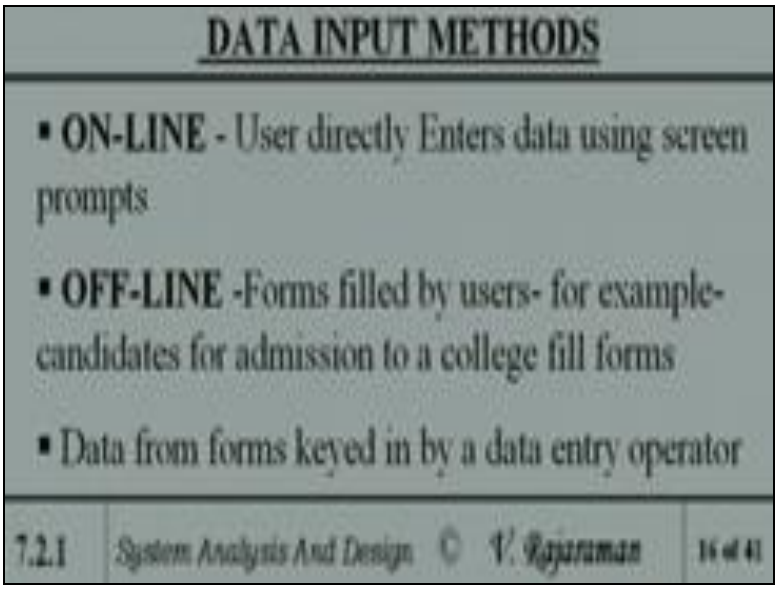
So that in as much as the documentations are created and it had been made into a database for reference. It is extremely useful for a person to maintain the system and do not ignore it. It is a clerical chore; it is a clerical chore to create that particular data dictionary and so on. But, it is a very important part and normally managers should really insist unless data dictionary prepared for a given application.

With the data flow diagram or whatever is equivalent, he is not going to accept it when you. As the before programmer signs off or not in this case, system designer or system analyst before it is goes to a program again that stage itself. And, invaluable resource to design input forms and screens data is checking programs as we find it out repeatedly. Process specification and in a database, creation of a database also, it is very useful resource, very useful in understanding and maintaining a system.

Now, having looked at the data dictionary, of course I will it as an exercise to you to look at the full problem which has we discussed as case studies and of course in the book which I have been using text book systems analysis and design. There are in the earlier chapter there is a case study of hostel information system. Also a greater details this particular problem where stores and other things are involved. So, much larger data flow diagram.

In that, we will have an enormous size and I would suggest that, you just create it as an exercise. And, try to create a proper dictionary format for storing in a computer. And, try to kind of make a file for the machine and try to have a little retrieval program. To retrieve given a name, it will be a nice exercise for you.

(Refer Slide Time: 32:27)



Now, data input methods are the next topic in this module. And, there are two ways of inputting data. One is on online, that is user directly enters data using a screen prompts. And this is valid, in fact most often user does online data entry in an electronic commerce.

Nowadays, you can book tickets for trains or planes or even buses, nowadays on internet; that means, it user can inputs the data on the requirements in a given format. And, the format of course, user does not know. So, we have to get a prompt from the program and how to enter it. Because then, it will be processed by the reservation program. And then, even a ticket will be generated on your computer. So in such cases, it is the user directly enters.

Or it could be as I said if it is a user in a very wide sense. It is not necessarily just a customer trying to buy a ticket. It also can be a clerk in a store who is checking out your data, your items when you purchase a set of items from a supermarket. And, you go with the set of items to the girl at the counter and the girl at the counter is essentially going to enter the item name and stuff like that online. It is- online data entry and offline data entry is one where forms are filled in by users.

For example candidates for admission to a college, if you are going to appearance ever in C E T or appear for some other A I A I W all India entrance exam or whatever there is a form which they give and nowadays those forms. The machine itself processes. The

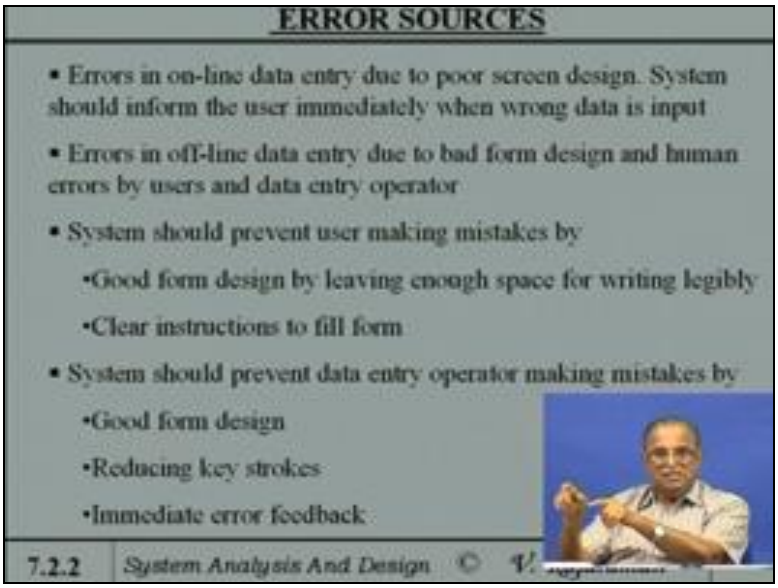
forms will be in such a way that you kind of fill in fairly in a way. So, that the processing of the forms can be done automatically; that means the forms which are filled in by the user have got to be keyed in by the data entry operator. The data entry operator is a professional typist kind of person, who enters in fairly fast rate.

And normally, they are paid by piece rate. If you enter 100 forms in a certain period of time you get paid. So, much if you enter 110 I will pay you put more extra. If you go below that standard of 100 then you payment is gets less. So that is always hurried on their part. So; that means, if somebody hurries an inputs. The probability of error is high. So, you must have some method of automatically checking such errors, because data entry operators.

And, there are that is one of the most important things which system is supposed to do. But, in offline the volume is extremely large. Volume of data is going to be taken and processed is very large. And so, there are keyed by more than one operator, if we have 4, 00,000 entries to be made for say entrance examination for some entrance examination or 4, 00, 000 results to be published.

Then, you have that large number to be keyed in a fixed time. So, you have to give it to many people. And there can be uneven quality, so all that has got to be checked properly. So, it is called offline data entry. Online normally you only do one at a time and one advantage of online is that have and when data is entered it can be all checked automatically at the entry stage itself whereas offline normally it is difficult to check while entering. It is good idea there are also it go screens and while you enters something which is absolute incorrect. It should actually ask him to reenter.

(Refer Slide Time: 37:03)



Errors in online data entry due to poor screen design that is when you design screen online is all on the screen. That is you use a keyboard and you see the screen when you enter it. Error in offline data entry is due to bad form design and human errors by users and data entry operators. The users also are very often to blame.

They have entered wrongly in the form even if you, because the form sometimes is not properly designed and so we cannot really blame the user. But sometimes, the form is well-designed even then the users make an error. And this has got to be checked. System should prevent user making mistakes by good form design by leaving enough space for writing legibly.

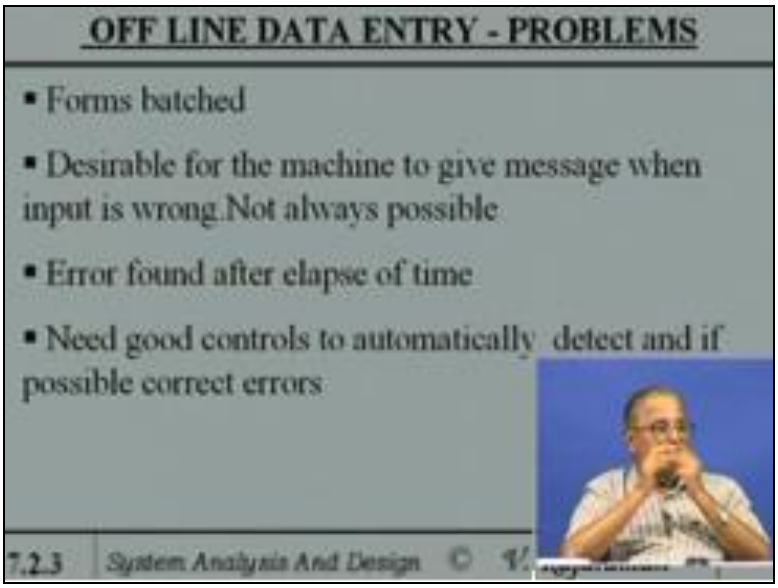
Because otherwise, they just scratch and denote. It is very difficult for data entry operator later on to key in something which is scratched and written in very bad handwriting and so on. Particularly, if it is scant very little space is given is forced to scrap it. So, you should give enough space for writing legibly clear instructions to fill forms. Very often instructions are not clear enough.

For example, if you had entered the date as D D M M Y Y Y Y. You have to say enter date first month later year last. And, you have to say Date 5th has got to be entered as 05 and not just 5 in case of these instructions to be very clearly marked or written there. And of course, you hope that the reader. The user will read that if there are for entering. But, there is sometimes an assumption. People do not read anything. They just go ahead and

there are overconfident and fill up. Because that is poor audio, because it will get rejected.

And so, it is very important for the user also to follow the instructions. Once instructions are clear, system should prevent data entry operator making mistakes by good form design reducing key strokes. Key strokes have to be reduced, something which got because each key stroke is a potential source of error and immediate error feedback. We can feedback immediately as soon as the- entry is made. Then, correction is very easy.

(Refer Slide Time: 39:36)

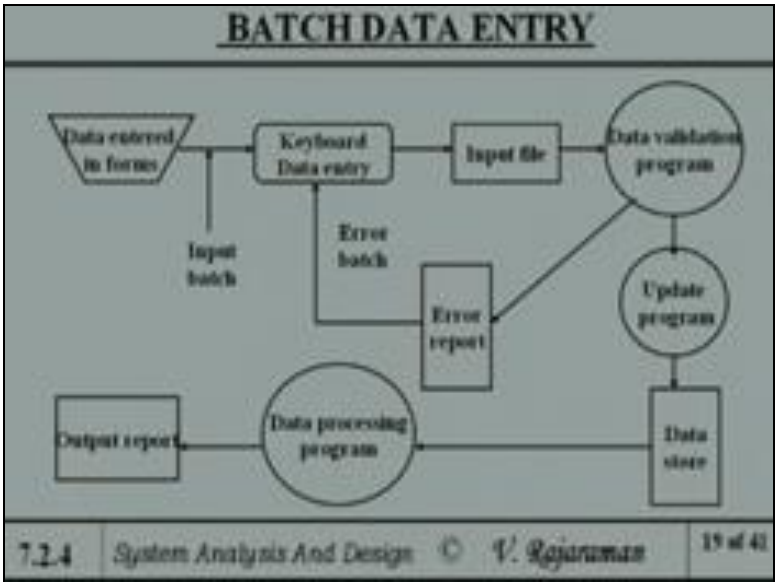


Offline forms are batched desirable for the machine to give message. When the input is wrong not always possible. Because, we are doing huge number, it is not always possible. Error found after elapse of time that is you take a batch. Enter a batch and after entering the batch you check this batch; that means you write a program to find out whether the fields are correctly placed whether there are any incorrect entries.

So, there is a little delay and then you feed back to the entry operator with the incorrect data to be in keying. It is always a little time delay need good controls to automatically it detect and if possible correct errors. Because, there is a delay to avoid this delay you would like to be able to detect at least errors before they get into the edit state. At the time the data entry operator enters itself.

And, if you can correct well and good I mean it is a it is a bonus. But as I said correction is not very often done, because it is very expensive in terms of number of digits to be used and so on.

(Refer Slide Time: 40:54)



Now, the batch data entry proceeds this way. Data- entered in data entered in form is taken. So, it is an input batch. There is very important reason, why the forms are batched together in small batches.

Because, if suppose you give 1000 forms together and you find some error in 2 or 3 of them. You have to go through 1000 forms to find out where they are occurring- error occurred whereas you batch it in 50s. As soon as the 50 is completed you can check it out and so you can go back and check that 50 where it is. And also, there is something called control totals which are used which I will talk about later on.

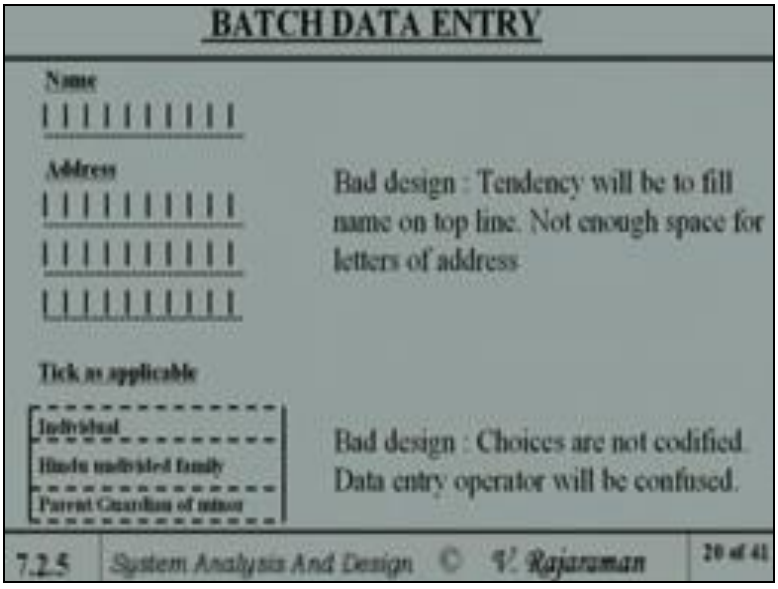
Control totals are essentially manually calculated for this set of small number. The control totals are used for error detection that is, what is done by a data validation program. So, you keyed by data entry operator keys input file created. And, the data validation program which as the uses the data dictionary entry to find out the, what is the format of the data and what fields it should be and so on. And, does an edit check that is called an edit check are cleaning up the data.

Data validation program and once it finds out some items are not valid. It generates an error report. The error report goes back as an error batch back to the keyboard entry operator and those incorrect entries got reentered. So that is little delay in this feedback So, then once it is corrected, we have define and put in the data store. And then, that data stores one which is used by the data processing program.

And then, you get the output report. So, that is data entry as this spirit. In fact, in the one of the things which most people when they join a company as a software developer? Most often the program which is given to them as a first program to develop is a delete program. Because, delete program development is quite challenging. It also tests your perseverance in terms of, how you are able to find out all types of errors in data entry.

At least, I know then the early days of T C S, they every person who used to join. The first six months they were given the assignment of writing an edit program, because in those days batch data entries were very very common. Interactive terminals are not there in the seventies and so on. Interactive terminals are there only later on. Only after 80s interactive terminals are coming in. Once the P C's came into the market, before that it is all batch data entry. In fact, every page taking on punched cards and so on.

(Refer Slide Time: 44:21)



And anyhow that is all history. Now, this is a bad design. Name, first of all we have not given enough space for name and also will be tendency on the part of a person sometime even though you have given space this below name. We will write name another name.

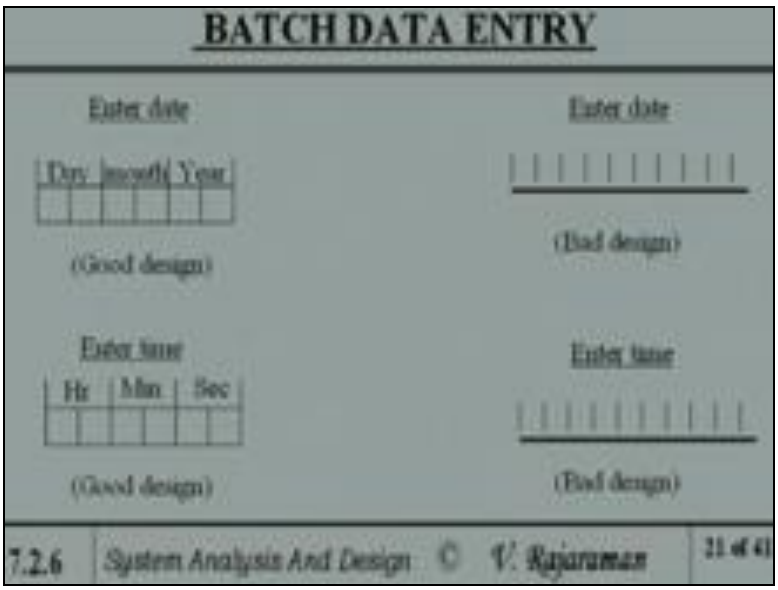
As soon as the name, it does not see below. He just writes name on top. Because, everybody is very happy to write his name or her name, they will write their name on top.

And then, addresses, they will start next one and then they say address they will repeat their name and then put the address. So, redundantly name occurs twice and so it is a very bad design. Tendency will be filling on top of line. Not enough space for letters and not enough space for address.

Now, you will like to have the address in a very specific form. Like nowadays every form should specify, what pin code it is. Because, pin code is very important for postal addresses and very often postal addresses are one. For instance, if it is a CT exam for which you have applied. The admission letter has to do to your address. The address unless it has got a pin code, it will reach you after the exams are over. So, it is in your interest to put the pin code and it is in interest of the person to also form designer should design it in a proper way.

And there are three categories individual Hindu undivided family, parent guardian, I know check pick as applicable. Choices are not codified. And, data entry operator will be confused. Suppose somebody ticks individual, I agree operator will ask the question do you type individual or and so on. If you codified say 1 2 3 and if you tick 1, the entry operator will just put 1, he will enter 1. So, it is a it is a bad design from that point of view.

(Refer Slide Time: 46:35)



A good design for instance. Let us enter date day month and year. In this case in some sense, I would not call it a very good design. Because, four places are not given for the year, I gave only 2. And of course, you might say that the problem will not arise till the year 2099 till, then this is not will not occur. Then when and caution why get cursed by the boxes ((Refer Time: 47:16)). So, you might well enter 4 digits, because it is not anymore no. Storage is no more it is premium.

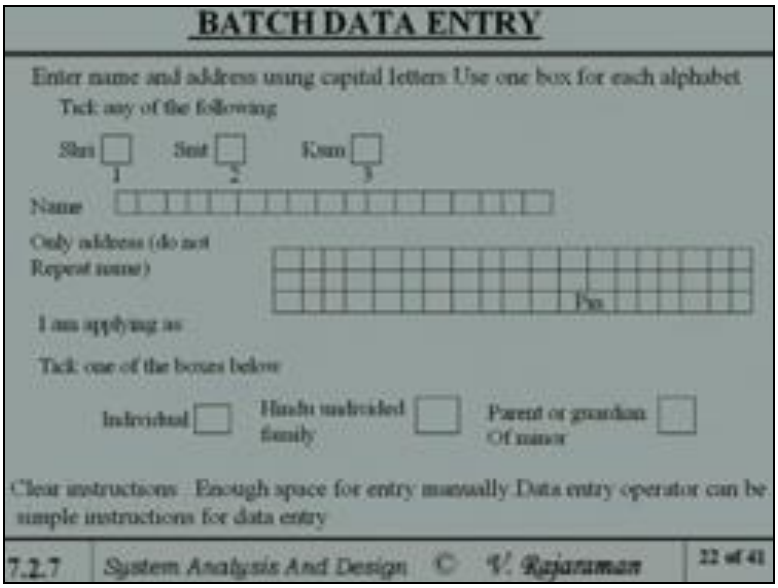
Of course, it depends on the application. If you, if the application is very short-lived. Like for instance if it is a reservation of a train. It is three months validity only. So, 2 digits not more 2 2 2 2 digit year is more than sufficient. 4 digit years is not required in that such a case. Because, the life is very small, so you have to use common sense. It is not hard and fast rule that you should always use 4 digits for a year.

But, it is a good idea to use 2 for day for month. Otherwise, it is a confusion possibility there and enters time. Now again, this is looks like a good design. But the right hand side is a bad design to begin with, because it does not say in what way I enter hour, minutes or second or each person will enter in his own way. Similar enter date, you do not specify some people may enter month, date and year. So, it is very bad design.

At least in this design the left hand side design, I have effectively given where to enter. But, if you look at the clock there is still confusion, if you say hours minutes and seconds. Suppose, it is 1 P M do you have to write it 130000 or do I put 010000. So, there could be a mistake, because there is no place for P M in this case. So, it is a common sense dictates that you should use 24 hour clock. But, common sense is a very precious commodity.

Lot of people does not have it alright. So, they will essentially write 0 1 and whatever comes to their mind. We really require 24 hour clock. And, you have to say at the bottom enter time or when you say enter time and within brackets you say use 24 clock. Then, immediately your person knows and you use 24 clock depending on the P M or A M.

(Refer Slide Time: 49:41)



So, in a batch data entry it is important to have similarly in this case there is a name. See there is- only you know shri shrimathi and kumari is there in this case. Nowadays the- use just- shri and some place they use mister and M S. M S is for purpose. And, nowadays people do not like to be called kumara everybody will like to be called shrimathi whatever it is. So, I might give two choices only. So, this is depends alright.

So in this case, I have of course, re- recharged those given and then the name as to be written again. The name wills often there is confusion. See name how do we enter. Do we do we enter the first name first and last name last or last name first- first name last where is the middle name and so on. At least occur in America and England and all that the concept of first name middle name and last name is very commonly prevent.

In India there is no such things as first name nobody understand what is the first name and last name. And you when I say V Rajaraman, Rajaraman is my first name from the point of view of the Americans. I do not have a family name which is in my name. If I put Rajaraman something something else, then it would be a even then it may not be a family name. It could be a caste name like rao or iyengar or whatever it is alright.

So, in- whatever it is the point is you have to specify saying that. If you do not want you do not care with first name last name. You just leave it as it is name. And, in India it may not be a bad idea. Because, very often people got g- get confused. When you say last name first name last because you do not have a last name.

So, what do I do and I have a I have a friend of mine who went to America. His name is Aravind and his father purposefully did not give him a last name did not give an initial. So, his name is Aravind. So, it is true step, it is not. I am not making it up. And so, he will pick up a phone and say that I want to book a ticket. There are whose names the girl will ask Aravind first name, last name middle name says all and they are confused.

Similarly, I think when he fills up a form you will occur problem where to put. But, he has somehow managed everything. He is not changes his name. He is still Aravind. Single name and in one case somebody asked, what is your first name middle name and last name. He said Aravind he said just Aravind basics plain Aravind. If I put plain as his last name, Aravind his first name, Aravind plain. So that also happens.

Because in human communication, there could be misinterpretation, when on the form somebody ask you. And because, this case is a very peculiar case, but it did not happen on practice. And he got some credit card or something like that as saying Aravind plain and of course, completely stumped, while he got a card- like that.

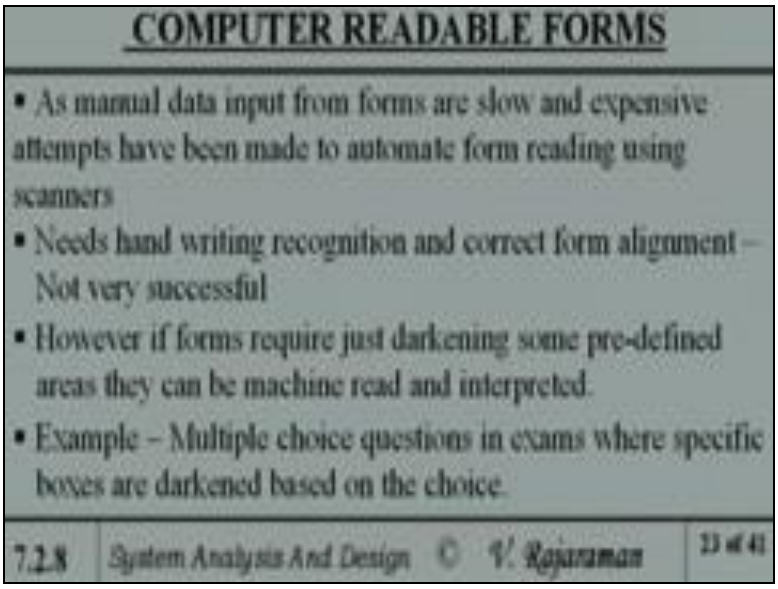
Then, we say the address, only the address and do not repeat the name. Otherwise, people have a tendency to put their name again and so you have given enough space. In fact, it- it may be well, the address is very diffuse quantity in India particularly. Because, if you look an address in Hyderabad 1 slash 100 slash ((Refer Time: 53:49)) 5 slash 4 8 4 and so on. It will be stupid looking address.

If you look at addresses in Bangalore it will always be 6h main 4th cross something like that. That it is many names and so it is very difficult to talk about address in a uniform way for the whole country. It is just leave it at that see. You normally request a person to write the address as you would normally write. It is postal address and so the address is given.

Of course in this case pin is specified. Pin code as you know in India is 6 digits. It is standardized in 6 digits. So, enter the pin code and I am applying as an individual. There are 3 places are there. Tick the one box below, tick one of the boxes and clear instructions in terms of what box is to be ticked. The box is ticked below that box, you normally also put a number. If data entry operator, if box one is ticked you enter 1. If box 2 is ticked you will enter 2 and so on.

So, it is also a good idea to the point of view of the data entry operator. To put the numbers below it and so just like shri shrimathi kumara. I have put the numbers 1 2 3 below it. When you tick one of them the entry operator will just put 1 and the machine will actually translate 1 to 3 and same way in this case.

(Refer Slide Time: 55:35)



As manual data input from forms are slow and expensive. Attempts have been made to automate form reading using scanners. In other words, you have put here form. If it is a well designed form, you expect if the person also wrote legibly in that form. In the plenty of space and all that you would like to have situation where you have a scanner. Into that as you know scanners are commonly very commonly available. You put the form in the scanner.

The scanner reads entire thing. Converts that into appropriate data, instead of manual entry and then creates a file. It is a very very interesting topic of research people have been doing. There is no solution as of today fortunately unfortunately. Because, even if you s- the form a little bit that is if you do not put it correctly. And, even if suppose the writing is smudged or some ink smudge is there and so on, so forth.

Recognition of hand-written characters automatically to clear a file ends up with a file with lot of errors. I know at least one all India of a data processing company. To spend at least two years trying to kind of automate this process, but they just did not succeed.

There are at least few cases where of course, there has been success stories like in one of the success stories in terms of this recognition as been an automatic sorting of letters.

If want to s- sorting of letters as a manual process, when the letter is given to the post office. The post man has got to sort it manually and that can take time. So, the automatic sorting machine at least, what it tries to do, is tries to dictate in that address the pin code. Are equivalent in America as a zip code.

Once in deducts the zip code t- ignores the rest field. It just picks up the zip code which is numeric and of a certain number of digits long. And then, it kind of sorts it using that zip code or in this in our case similar trials are being made charged by pin code, charged by pin code then, at least you have solved part of the problem after that it is kind of amount of work is reduced a little bit. So, it has been lot of work as been done.

But current alignment, because of wrong alignment equal the line correctly and hand recognitions difficult it is not success however forms require just darkening some predefined areas. They can be machine read like. Many of the multiple choice question examinations where you just darken with a graphite pencil that can be read by UPS examination, mini examinations use it. Multiple choice questions are over.

So this is, where I think I have to stop and we will continue with Interactive data input next time. So, we conclude by saying that the reliable forms come automatic reading as not yet been full success except some special cases.