

PhUSE 2011

Managing Data Issues



 Shafi Consultancy Ltd.

Overview

- How do we manage data issues now?
- What is the problem?
- The new approach
- Benefits of the new approach
- Conclusion

How do we manage data issues now?

- Send messages to the LOG window

```
*** Data Issue: Patient older than 75 - PATIENT=130164 AGE=78 SEX=Male RACE=White  
*** Data Issue: Patient older than 75 - PATIENT=130186 AGE=80 SEX=Male RACE=White  
*** Data Issue: Patient older than 75 - PATIENT=130204 AGE=77 SEX=Male RACE=White  
*** Data Issue: Patient older than 75 - PATIENT=130302 AGE=82 SEX=Male RACE=White  
*** Data Issue: Patient older than 75 - PATIENT=130401 AGE=77 SEX=Male RACE=White
```

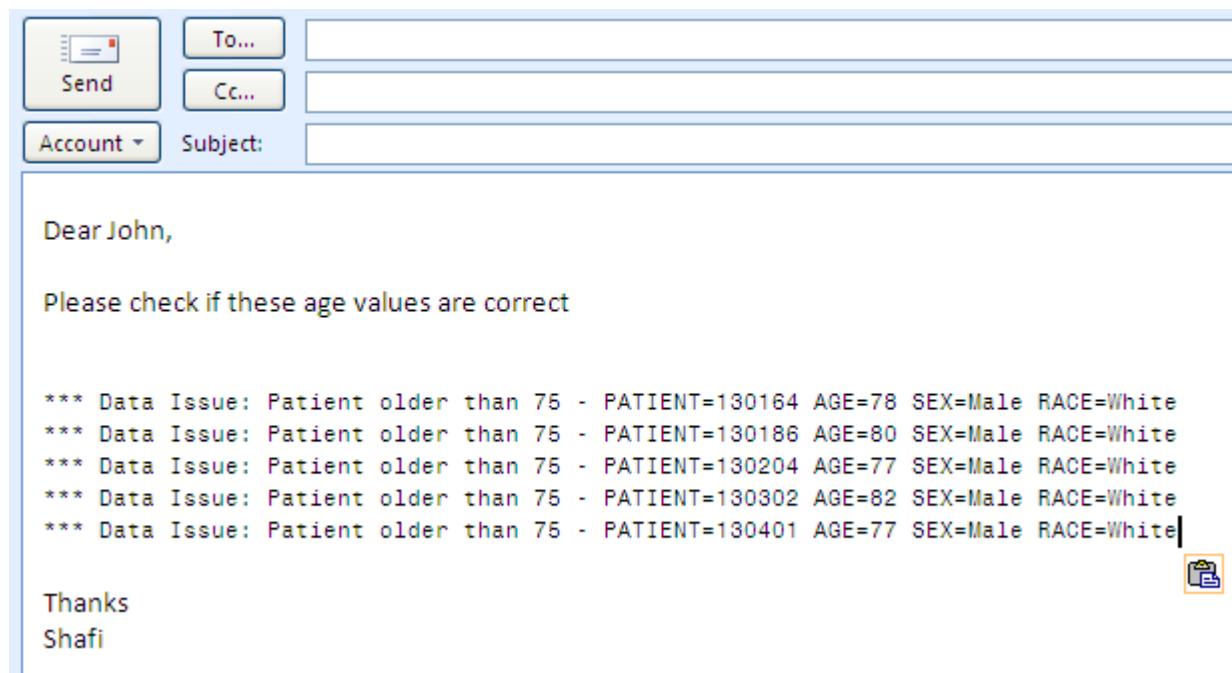
- Print dirty data in the OUTPUT window

```
Data Issue: Patient older than 75
```

Obs	PATIENT	AGE	SEX	RACE
1	130164	78	Male	White
2	130186	80	Male	White
3	130204	77	Male	White
4	130302	82	Male	White
5	130401	77	Male	White

How do we manage data issues now?

- Copy and paste the issues into e-mails and send to the Data Manager



The screenshot shows an email client interface with a header bar containing a 'Send' button, 'To...' and 'Cc...' fields, an 'Account' dropdown, and a 'Subject:' field. The email body contains the following text:

Dear John,

Please check if these age values are correct

*** Data Issue: Patient older than 75 - PATIENT=130164 AGE=78 SEX=Male RACE=White
*** Data Issue: Patient older than 75 - PATIENT=130186 AGE=80 SEX=Male RACE=White
*** Data Issue: Patient older than 75 - PATIENT=130204 AGE=77 SEX=Male RACE=White
*** Data Issue: Patient older than 75 - PATIENT=130302 AGE=82 SEX=Male RACE=White
*** Data Issue: Patient older than 75 - PATIENT=130401 AGE=77 SEX=Male RACE=White

Thanks
Shafi

A small icon of a clipboard with a document is visible to the right of the email body text.

How do we manage data issues now?

- When we re-run the program a week later and the issues are in the LOG again:
 - New issues or mostly the same ones?
 - Send all the issues to the DM again?
 - Don't send, wait for a future run?
 - Discuss with the DM if something will be done about these issues?

What is the problem?

- Same issues sent to the DM in different e-mails
- Not clear which are new and which have been sent previously
- The DM frustrated by having to go through the list and first discount all those they have already looked into
- The DM starts to ignore e-mails from the programmer!



New approach

- Don't send issues to the DM
- SHARE issues with the DM
- Store them in ONE CENTRAL location
- Use dates to identify new and old issues
- Group the resolved and unresolved issues

New approach

- Use ONE Excel file to SHARE issues with the DM

The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D	E	F	G	H	I
	COL1	COL2	COL3	COL4	COL5	COL6	COL7	COL8	COL9
1	COL1	COL2	COL3	COL4	COL5	COL6	COL7	COL8	COL9
2									
3									
4	TITLE: Patient older than 75								
5									
6	ISSUE_CRDATE	ISSUE_LAST_UPDATED	ISSUE_RESOLVED	PATIENT	AGE	SEX	RACE		
7	20SEP11:18:52:01	01OCT11:18:58:04	Yes	130164	78	Male	White		
8	20SEP11:18:52:01	01OCT11:18:58:04	Yes	130186	80	Male	White		
9	20SEP11:18:52:01	01OCT11:18:58:04	No	130204	77	Male	White		
10	01OCT11:18:58:04	01OCT11:18:58:04	No	130302	82	Male	White		
11	01OCT11:18:58:04	01OCT11:18:58:04	No	130401	77	Male	White		
12									
13									
14	TITLE: HbA1c Values above 13								
15									
16	ISSUE_CRDATE	ISSUE_LAST_UPDATED	ISSUE_RESOLVED	PATIENT	VISIT	SAMP_DT	SAMP_TM	RESULT	UNIT
17									
18									
19	20SEP11:18:52:01	01OCT11:18:58:04	Yes	12191	10.00	15MAR2011	15:50	17.00 %	
20	20SEP11:18:52:01	01OCT11:18:58:04	Yes	11929	10.00	29JUN2011	10:58	14.40 %	

- Issues from each program in a new sheet
- A summary sheet to show all programs and the number of issues resolved and unresolved

New approach

- Summary sheet

	A	B	C	D	E
1	COL1	COL2	COL3		
2					
3	Sheet	Number of issues	Number of issues unresolved		
4	PROG1	18	6		
5	PROG2	14	3		
6					
7					
8					

New approach

- To send the issues to Excel:
 - Create a dataset with the issue
 - Set the LABEL as the title
 - Specify a KEEP statement inside the label to print the variables in a specific order

```
DATA age_gt75 (LABEL='Data Issue: Patient older than 75 KEEP=patient age sex race');  
  SET demo;  
  WHERE age GT 75;  
RUN;
```

```
%data_issue(pgmname=&progname., indata=age_gt75, send2xls=y, firstorlast=first);
```

Benefits of the new approach

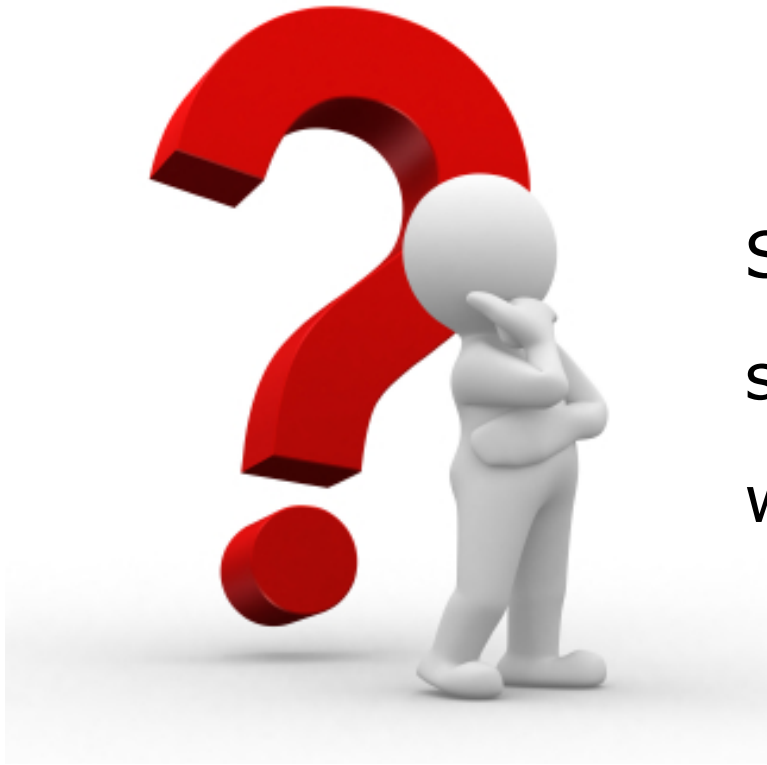
- No need to cut, paste and send e-mails
- One central location with issues from ALL programs
- Easy to see the number of unresolved issues
- Can see how long some issues take to resolve
- The DM is happy and resolves the issues



Conclusion

- It is OK to help the DM and inform them when we find an issue...and we do find issues
- Everyone is familiar with Excel, so it is easy to use this as a sharing tool
- It is not any more work for programmers or Data Managers
- It will help to manage the data issues in a more structured manner

Questions or Comments



Shafi Chowdhury

shafi@shaficonsultancy.com

www.shaficonsultancy.com



Shafi Consultancy Ltd.