

Making the leap from Excel to a Database used for Reporting

Shafi Chowdhury, Shafi Consultancy Ltd.

Md. Zobair, Shafi Consultancy Bangladesh

Abstract

In the world of science we do not use Excel as it is not trusted. Yet day by day, Excel seems to be creeping into every part of our process except the analysis. As it is in our everyday working life, using it as a springboard to populate a database would seem to be the natural progression. This paper will look at how a simple Excel file can help us save time and effort to enter the world of Meta data, leading to faster production, validation and overall management of the reporting process, programs and output in a controlled environment. This will also document who produced what, when, who validated it, and even what is left to do.

Introduction

Most Pharmaceutical companies and CROs use Excel for study management and evaluation, event site feasibility. Moving further into the clinical study life cycle, Excel is still widely used for the tracking of study startup (SSU) processes. Ironically, the top Sponsor/CRO pain points are lack of operational oversight with no availability of real-time reporting on clinical trial status or CRO performance, lack of project management standards, lack of integration of systems for feasibility, activation and document management.

This paper focuses on the reporting part of the clinical trial process, investigate how to set up procedures that are easy to follow and save time. Hundreds of outputs are produced, but the management and status are often still controlled manually by using Excel. Who has written the associated program, who validated it and when, is it finished? These are often manually entered into a spreadsheet that is used for tracking purpose. Can these data be stored into a database and tracked automatically? If the manual aspects of documentation can be removed, then programmers and managers simply perform the tasks they need to do, and the documentation

can take care of itself in the background. This is the ideal scenario that will make the process transparent, current, and easy to use.

Excel Used In Clinical Trial Process / Project Management

Excel is widely and regularly used in most of organizations to manage the reporting part of the process. The basic details required for management and study tracking include titles, program names and locations, and the validation status of individual programs or outputs. There are times when additional details such as table shell names, dates and times of programs or outputs, as well as the names of programmers and validators are also stored in the spreadsheets. The users need to update the content of the Excel file manually after each change, but it allows everyone to easily see the current status. This simplicity is its attraction.

tifnum	Title	pgmloc	pgmname	outno	outloc	outname
1.1.1	Patient analysis sets and reasons for exclusion - RS	pgm/ctr/tpopu.sas	tpopu.sas	T1	lst/tpopu_t1.lst	tpopu_t1.lst
1.2.1.1	Demographic data - TS	pgm/ctr/ct/demo.sas	demo.sas	T1	lst/demo_t1.lst	demo_t1.lst
1.2.1.2	Demographic data - FAS	pgm/ctr/ct/demo.sas	demo.sas	T2	lst/demo_t2.lst	demo_t2.lst
1.2.1.3	Concomitant diagnoses by MedDRA SOC and preferred term - TS	pgm/ctr/condiag.sas	condiag.sas	T1	lst/condiag_t1.lst	condiag_t1.lst
1.2.1.4	Concomitant therapies during screening by WHO INN - TS	pgm/ctr/conmed.sas	conmed.sas	T1	lst/conmed_t1.lst	conmed_t1.lst
1.2.1.5	New concomitant therapies during study part A by WHO INN - TS	pgm/ctr/conmed.sas	conmed.sas	T2	lst/conmed_t2.lst	conmed_t2.lst
1.2.1.6	New concomitant therapies during study part B by WHO INN - TS2	pgm/ctr/conmed.sas	conmed.sas	T3	lst/conmed_t3.lst	conmed_t3.lst
1.2.1.7	Use of ASA, antihypertensives, lipid lowering drugs or P-gp and CYP 3A4 inhibitors during screening - TS	pgm/ctr2/cttass.sas	cttass.sas	T1	lst/cttass_t1.lst	cttass_t1.lst
1.2.1.8	New use of ASA, antihypertensives, lipid lowering drugs or P-gp and CYP 3A4 inhibitors	pgm/ctr2/cttass.sas	cttass.sas	T2	lst/cttass_t2.lst	cttass_t2.lst
1.2.1.9	New use of ASA, antihypertensives, lipid lowering drugs or P-gp and CYP 3A4 inhibitors	pgm/ctr2/cttass.sas	cttass.sas	T3	lst/cttass_t3.lst	cttass_t3.lst

Databases store information in electronic records that may be searched, retrieved and organized in countless ways. Having information in a database, instead of in spreadsheets, not only saves time and preserves vital information; it also allows users to get clear overviews and use the database to perform the actions they need to carry out in their daily tasks.

Study Name : study_002

Total Number of Outputs: 15

■ No Program 15
 ■ In Development 0
 ■ To Be validated 0
 ■ Validated 0
● No Output 15
 ● Old Output 0
 ● Current Output 0

If the database data is relayed through a web application or portal, then the project team can check the progress easily using a web browser. They would also be able to see each output or program, who is updating the program and when. Depending on what is retained in the database, they can also track the program history to see who made what changes and when, and perhaps even view old versions of the programs and outputs. The important thing to keep in mind about any web application is that it should allow the programmers to perform their duty, and everything else should be in the background. Any

additional tasks will make the tool less desirable and may lead to it being underused and not maximized to its potential.

Benefits of Using Database Instead of Using Spreadsheet

Study Name: study_002

Final Status Of Outputs By Table Number

10 records per page Search:

TLF Number	Title	Program Name	programmer Name	Date of Last Program Update	Date of Last Output Update	Validator Name	Current Status	Date of Last Status Update
Table 1.1.1	Patient analysis sets and reasons for exclusion - RS	tpopu.sas	dummy_user2	02 Aug 2016 18:56:45		dummy_user1	In Development	02 Aug 2016 18:56:45
Table 1.2.1.1	Demographic data - TS	demo.sas	dummy_user2	21 May 2016 19:24:20		dummy_user2	In Development	21 May 2016 19:24:20
Table 1.2.1.2	Demographic data - FAS	demo.sas	dummy_user2	21 May 2016 19:24:20		dummy_user2	In Development	21 May 2016 19:24:20
Table 1.2.1.3	Concomitant diagnoses by MedDRA SOC and preferred term - TS	condiag.sas						
Table 1.2.1.4	Concomitant therapies during screening by WHO INN - TS	conmed.sas						

Excel is not a relational database; it is a spreadsheet. A relational database contains tables where each column holds the value for an attribute. Every value in that column holds the same attribute, but for a different entity. Each row holds the attributes for a single entity. The values in each row are tied together as a single unit. Each column must have a unique name and can only contain a specific type of data ('Integer', 'Text', 'Date', etc).

The following are the problems we will find when using Excel as a relational database:

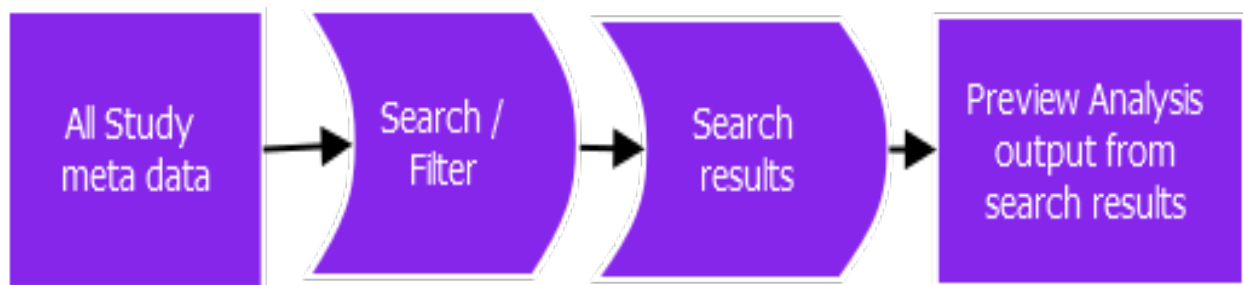
- Excel sets data types automatically based on the data entered.
- Data types are set at the cell level rather than at the column level.
- A worksheet holds no more than 256 columns and 65,536 rows.
- A cell can contain no more than 32,767 characters. (Okay, this seems to be a good thing.)
- Excel dates go back only to 1 January 1900. (Well, Excel does claim that there was a 0 January 1900!) and Excel believes that 1900 was a leap year.)
- Excel has one data type (date-time) to store both dates and times of day. (This is not really an Excel shortcoming; it is just not the SAS standard.)

- Excel puts a dollar sign at the end of its worksheet name. (Again, this is simply an annoyance from the SAS perspective.)

The following are what we can do using a database:

- Traceability
- Summarization of reporting facility
- Substantial time savings
- More accessible information
- Ability to link data
- Ability to access email
- Reduce the amount of time spent on managing data
- Give the ability to analyze data in a variety of ways
- Promote a disciplined approach to data management
- Turn disparate information into a valuable resource
- Improve the quality and consistency of information

Databases save time and provide many different ways to look at data. Retrieving a single record, which may take several minutes of digging in multiple Excel files, can be done with just a few clicks in a computerized database using advance searching options. This means you do not have to spend time wading through other records to check spellings, addresses, and historical data such as previous records. Using a database through the system, a single user like a Team leader can coordinate with every other team members at the same time.

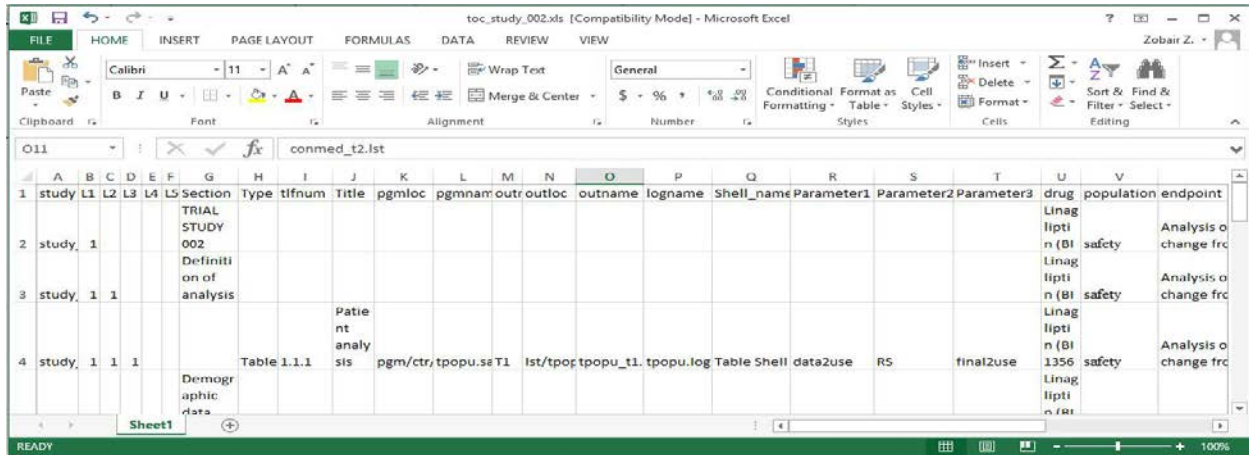


Database review tool

Most of the research organizations use Excel files to create some items like table, listing, and figure to draw the inference of the research output. Sometimes they need midterm reports of a research project and sometimes they need some slight changes in those reports. In all of these cases the research group needs to create the whole report documents repeatedly, which can be very time consuming..

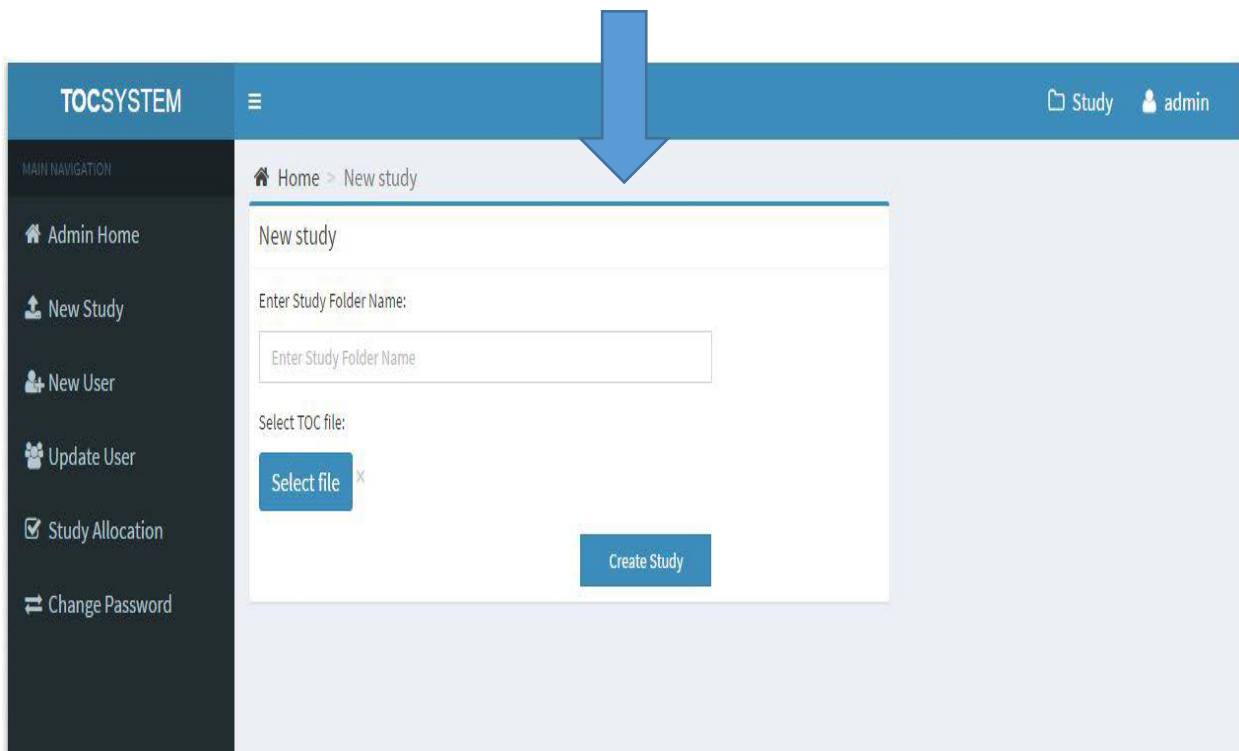
PhUSE 2016

To make performance quicker and generic, we are trying to provide a web-based service where a viewer can see all the updated items which come from structural database through the system. The database keeps all the records and history that was performed by a user. The system converts Excel data into a database through the study management process.



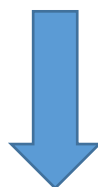
The screenshot shows an Excel spreadsheet titled 'conmed_t2.lst'. The columns are labeled A through V. The data includes study identifiers, section names, patient analysis details, and various parameters like drug, population, and endpoints.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V		
1	study	L1	L2	L3	L4	L5	Section	Type	tfnum	Title	pgmloc	pgmnam	out	outloc	outname	logname	Shell_name	Parameter1	Parameter2	Parameter3	drug	population	endpoint
2	study	1				TRIAL STUDY 002															Linag		
3	study	1	1			Definition of analysis				Patient analysis											Linag	1356	Analysis o
4	study	1	1	1		Demographic data	Table 1.1.1			pgm/ctr,tpopu.s	T1	lst/tpo;	tpopu_t1.	tpopu.log	Table Shell	data2use	R5		final2use		Linag		Analysis o



The screenshot shows the TOCSYSTEM web application interface. The top navigation bar includes 'TOCSYSTEM', a menu icon, and 'Study admin'. The main navigation sidebar lists: Admin Home, New Study, New User, Update User, Study Allocation, and Change Password. The main content area shows the 'New study' form with the following fields and buttons:

- Home > New study
- New study
- Enter Study Folder Name:
- Select TOC file:
-



1 TRIAL STUDY 002

- 1.1 Definition of analysis sets
 - Table 1.1.1 - Patient analysis sets and reasons for exclusion - RS
- 1.2 Demographic data and baseline characteristics
 - 1.2.1 Demographic data, concomitant diagnoses and therapies
 - Table 1.2.1.1 - Demographic data - TS
 - Table 1.2.1.2 - Demographic data - FAS
 - Table 1.2.1.3 - Concomitant diagnoses by MedDRA SOC and preferred term - TS
 - Table 1.2.1.4 - Concomitant therapies during screening by WHO INN - TS
 - Table 1.2.1.5 - New concomitant therapies during study part A by WHO INN - TS
 - Table 1.2.1.6 - New concomitant therapies during study part B by WHO INN - TS2
 - Table 1.2.1.7 - Use of ASA, antihypertensives, lipid lowering drugs or P-gp and CYP 3A4 inhibitors during screening - TS
 - Table 1.2.1.8 - New use of ASA, antihypertensives, lipid lowering drugs or P-gp and CYP 3A4 inhibitors
 - Table 1.2.1.9 - New use of ASA, antihypertensives, lipid lowering drugs or P-gp and CYP 3A4 inhibitors
 - 1.2.2 Diabetic baseline characteristics, antidiabetic therapies and diabetic medical history
- 1.3 Compliance data
 - Table 1.3.1 - Compliance data during study part A - FAS
 - Table 1.3.1 - Compliance data during study part B - TS2

The system can be used in different studies to present the analysis as well as the outcomes in a structural way which is secure and access controlled by the administrators. The system should allow access to the file server so that the programs can be updated, outputs can be reviewed and perhaps even submit the programs to produce the outputs.

If the meta data that comes from users choosing to update or run programs are stored, then this can be used to generate reports that can be used to show status and program history. They can perhaps also be used to store old programs, so that they can also be retrieved as and when required.

Program History: disp.sas Study Name : study_001

Date/time	User Name	Comment	Final Status	Action
09 Jun 2016 21:55:24	dummy_user1	failed	In Development	download
08 Jun 2016 12:03:37	dummyuser3		Validated	download
01 Jun 2016 11:52:05	dummy_user1	Edited	In Development	download
01 Jun 2016 11:51:31	dummy_user1		Validated	download
01 Jun 2016 11:43:12	dummy_user1	Edited	In Development	download
01 Jun 2016 11:42:09	dummy_user1	validated again	Validated	download
01 Jun 2016 07:51:00	dummy_user1	Uploaded	In Development	download

Conclusion

Excel is a very easy tool to use and everyone knows how to use it in its basic form. Everyone has access to it and the spreadsheets used can be sent by e-mail or shared with each other in a defined location. It is therefore not difficult to see what attracts us to almost instinctively reach for Excel whenever we need to produce a list of items of anything. However, what Excel lacks is control. There is no control over the sheet, it is in the main manually driven and the data is only as current or as accurate as the team's willingness to keep it up to date.

Database can be created to contain all the same data that is stored in the spreadsheet. In being a dataset there are some inherent properties. It can document all the actions performed on it, who performs what and when can therefore be easily monitored and controlled. A relational database can mean that details only need to be entered once and it can be used in many places, without worrying about integrity. It allows a front end to be placed up on it for viewing and editing the data within the database. Additional functionalities can be placed on the front end to also perform tasks such as running programs and reviewing the log and output, perhaps even allowing the user to view old versions of programs.

Using a database to store these details therefore is more efficient as all management and documentation tasks are automated, it is transparent due to audit trails, and a front end allows users to retrieve data as and when required. Validation reports can be retrieved stating who made the most recent updates to the program and when; who set the program to validated status and when; and when was the program last used to produce the final output. All these produced automatically means the system saves time, and delivers consistency, efficiency and can allow programmers to perform their tasks seamlessly with each other.

References

- Microsoft Excel Specifications and Limits:
<https://support.office.com/en-us/article/Excel-specifications-and-limits-1672b34d-7043-467e-8e27-269d656771c3>

Contact Information

Shafi Chowdhury
Managing Director
Shafi Consultancy Ltd.
Email: shafi@shaficonsultancy.com
www.shaficonsultancy.com

Md. Zobair
Statistical Programmer
Shafi Consultancy Bangladesh
Email: zobair@shaficonsultancy.com
www.shaficonsultancy.com