Paper 013-2013

Knowing When to Start, Where You Are, and How Far You Need to Go: Customized Software Tracks Project Workflow, Deliverables, and Communication

Eric Vandervort, Rho®, Chapel Hill, NC

ABSTRACT

In a clinical trials environment, projects can have multiple statisticians and statistical programmers working on tables, listings, and figures, or "displays", for project deliverables. Communication between the various team members regarding when to program, validate, review, or revise these displays is vital to the success of a project. This paper describes a custom web-based application that stores relevant data about displays, tracks programming and reviewing workflow, and provides a tool for project-level management overview.

INTRODUCTION

The application has a web interface that grants users access to a database where project information is stored; this allows for the creation and maintenance of multiple concurrent projects. With this information stored in a central location, project managers can view the status of a single deliverable or the overall project. Project teams can customize many functions, as the database stores project-specific information about the system's behavior for individual studies. Project-specific titles, footnotes, and other data are retrieved from the database to be used in SAS® programs. This metadata-driven design gives the application the flexibility to accommodate the varying needs of individual projects and clients. While this application was initially created to track statistical deliveries, the metadata-based design allows it to track workflow of any kind.

APPLICATION TERMINOLOGY, WORKFLOW, AND LAYOUT

TERMINOLOGY

The following is a list of common terminology used in this application:

Item Type

Items are grouped based on a category, or Item Type. A generic project layout consists of the item types Table, Listing, Figure, Macro, and Analysis Dataset. Additional item types may appear in certain projects if additional units of work are tracked.

Items

The main unit of organization is called an Item. It is a single unit and the end-product of a work process. An item's status goes through different stages as its tasks' statuses change. If an item's tasks are all complete, then the item is complete. A single table, for example, DM_TXA_01, would be 1 item.

Fields

Metadata stored along with an item are called Fields. Example fields may include Title, Population, Program Name, Validation Program, and Project Directory. Fields are defined using metadata, and customized for each item type.

Tasks

Tasks are discrete work activities required to produce an item. Tasks are acquired, assigned, started, canceled, completed, or put on hold. The collective task statuses for an item produce the item's status. For example, if at least one task is started, the item's status is "Started".

Deliveries

A delivery date is the date when the associated items are due to be completed. Deliveries are a set of items sent collectively to the client, or a set of items due internally. Items are organized into one or more deliveries.

WORKFLOW FOR A STATISTICAL DELIVERY1

A statistical delivery typically begins with a set of mock tables, which specify how the final tables should look. Mock tables contain information about formatting, what rows and columns should be presented, special instructions as needed, etc.

Clinical trial data are collected on a Case Report Form, or CRF, and stored electronically in datasets. The raw CRF or "clinical" datasets may not be well suited for analysis purposes due to inconvenient structure, the need for derived variables, etc. Therefore, using the mock tables as a roadmap, a statistician will write specifications for transforming the CRF datasets into analysis datasets ("analysis dataset specifications"). Specific instructions for how to use these analysis datasets in generating the displays (what dataset or subset of data to use, what variable belongs in each column, etc.), referred to as "annotations", are often "written" directly on the mock tables.

Once the specifications are ready, Biostatistics and Statistical Programming work together to create the analysis datasets, after which Statistical Programming will program and validate the displays.

The displays are delivered to the client for review and revisions are made as necessary. The process of delivering and revising is repeated until the tables are approved by the client.

CREATE DISPLAYS: LIFECYCLE OF A TABLE¹

The process of creating a table begins by adding the table to the application as a new item. Once the table is added, various pieces of metadata must be provided (e.g., title, population, number, etc.). These metadata, in conjunction with the analysis datasets and annotations, are used by Statistical Programming to create two programs:

- one program to generate the table itself
- one program (created by a second independent programmer) to validate the values produced by the first program

Once both of these programs are created, the programmers check to see whether they produce identical output. If they do not produce identical output, the programmers will revise their programs and check them again. This iterative process is repeated until the programs produce identical output.

Once the statistics are validated, the next step in the table-generation process is a cosmetic review. Elements checked at this stage include titles, row and column labels, formatting, and footnotes. This step does not typically require any special knowledge about statistics or the study. If the cosmetics are not as specified in the mock tables, the programmers are asked to make revisions. The process then moves back one step to the statistical validation, before once again moving ahead to the cosmetic validation.

Once the cosmetics are validated, the final step in the process is Quality Control (QC). Whereas cosmetic validation concentrates on comparing the table to its mock, QC focuses more on the questions: "Do the numbers make sense?" and "Are the numbers internally consistent with other tables?" This step is typically performed by the study statistician. If issues are found during QC, the programmers are asked to make revisions. The process then moves back two steps to the statistical validation before once again moving ahead to the cosmetic validation and QC.

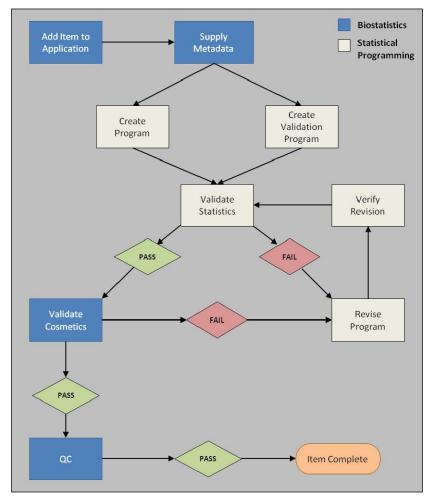


Figure 1. Lifecycle of a Table

EXAMPLE TABLE DISPLAY

A typical table display contains headers, footnotes, a title, population, and the table data. The table name, title, population, and footnotes are entered and stored in the back-end Oracle database. These data are usually entered using the web interface, and then accessed directly via a SAS program.

S	ummary of Demographi	Table: DM TX c and Baseline Cha Population: S	racteristics by Tr	eatment Group		
Characteristics	Treatment A N=141	Treatment B N=169	Treatment C N=143	Treatment D N=147	All Treatments N=600	
Sex - n (%) Male Female Missing	68 (48.2) 73 (51.8) 0 (0.0)	87 (51.5) 82 (48.5) 0 (0.0)	63 (44.1) 80 (55.9) 0 (0.0)	79 (53.7) 68 (46.3) 0 (0.0)	297 (49.5) 303 (50.5) 0 (0.0)	
Age (years) n Mean SD Median Range (Min, Max)	141 42.9 13.94 44.4 (18, 64)	169 40.5 13.87 39.2 (18, 64)	143 40.2 13.88 39.3 (18, 65)	147 39.9 13.41 40.9 (18, 65)	600 40.8 13.79 40.4 (18, 65)	
Race - n (%) Caucasian Black Asian Native American Hispanic Other Missing	41 (29.1) 58 (41.1) 4 (2.8) 0 (0.0) 38 (27.0) 0 (0.0) 0 (0.0)	57 (33.7) 55 (32.5) 1 (0.6) 0 (0.0) 52 (30.8) 4 (2.4) 0 (0.0)	50 (35.0) 45 (31.5) 2 (1.4) 0 (0.0) 45 (31.5) 1 (0.7) 0 (0.0)	42 (28.6) 43 (29.3) 2 (1.4) 0 (0.0) 58 (39.5) 2 (1.4) 0 (0.0)	190 (31.7) 201 (33.5) 9 (1.5) 0 (0.0) 193 (32.2) 7 (1.2) 0 (0.0)	
Baseline Weight (kg) n Mean SD Median Range (Min, Max)	141 71.03 10.892 70.58 (41.6, 89.7)	169 70.10 10.772 69.97 (41.0, 89.8)	143 69.03 12.183 68.42 (40.8, 89.9)	147 69.56 11.797 70.54 (44.0, 89.8)	600 69.93 11.397 69.89 (40.8, 89.9)	
entages are based on the numbe	er of subjects (N) in	n a given treatmen	t group for the pop	oulation analyzed.		
s calculated as BMI=Weight(kg	g) / [Height (m) 2.	25JUL2008 1				re

Figure 2. Sample Table Display for Table DM_TXA_01 Containing Headers, Footnotes, Titles, and Table Data

STUDY SETUP

For any given project, a set of item types are defined. Each item type contains data that describe the fields and tasks that are used to define and complete an item. For each study that is created, a series of questions are asked to determine the metadata used to create the study.

The project setup questions are as follows:

1. What categories or item types should be tracked?

The generic item types for a project setup are Tables, Listings, Figures, Analysis Datasets, SDTM Datasets, Macros, and Study Documents.

2. What data should be recorded and stored for each item?

Each item type has its own set of fields that will contain data particular to an item. These fields could have defined data types such as text, text area, select list, yes/no, and date.

3. What steps are involved in completing each item type's item?

A set of tasks or actions that defines the completion of an item. These are created when an item is added to the application.

4. Will there be revisions performed for these items? If so, what steps are performed for the revision?

The revision tasks can be the exact same tasks that are created when an item is created, or they can be a completely different set of tasks.

5. Will any item types have permission to be rerun?

Item types that potentially need future re-runs need not go through the normal revision process; instead, tasks specifically for rerunning will be created.

Step		Value(s	5)	
1) It	em Type	•	Table	
2) F	ields	•	Description (Text) Population (Text) Program Name (Text) Validation Program (Text) List File (Text) Group (Text) Sort Order (Text) Item Number (Text) Validation Method (Select List) Options Double Programming Code Review Manual Computation Manual p-value Validation? (Yes/No)	
3) T	asks	•	 Create Program Create Validation Program Validate Cosmetics Validate Statistics QC 	
4) R	evision Tasks	•	Revise Program Verify Revision QC	
5) R	erun Tasks	•	Rerun Verify Rerun	

Table 1. Sample Study Setup Values for the Table Item Type

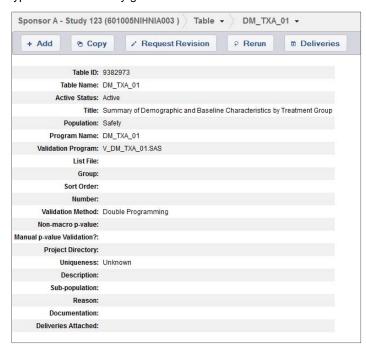
When the project specifications are completed, the default PL/SQL script is then edited to match the specifications and executed, creating the layout of the project in the database.

ADDING THE DATA

When a project is created, data pertaining to it can then be added. The first pieces of data that are added are items and their relevant metadata (e.g., Title, Population). If any of the display items contain footnotes, the footnotes are then added and attached to the items using them. Deliveries are created to track the progress of items that are due on a particular delivery date. Items can be associated with or attached to one or more footnote and/or delivery. Footnotes and deliveries can have multiple items associated with them.

ITEMS

Items can be added one of two ways: manually via the user interface or uploaded from a comma-delimited file. The latter assists in adding several items at once. When an item is added, all of the fields and tasks defined for its item type are automatically generated. The metadata are then entered into the application using the web interface.



Display 1. Result of Adding a New Table DM_TXA_01 and Related Metadata

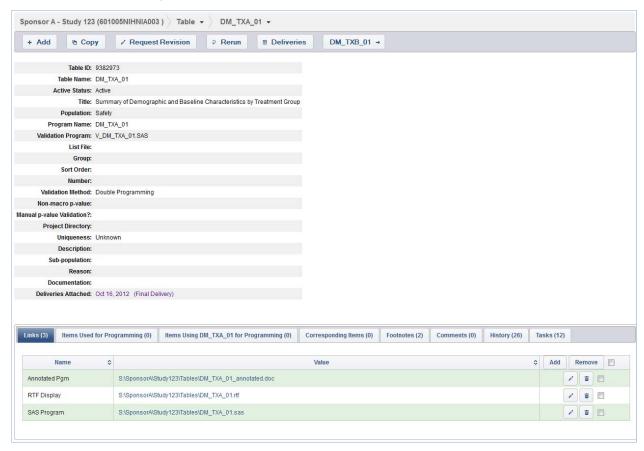
A main items page provides a list of all items grouped by their item type. This view provides an overview of the items and important metadata. Not all fields may display; field display settings are controlled in the project settings. Items can be printed, exported, or copied to another project.



Display 2. Items Overview: Listing of Table Items

ITEM LINKS

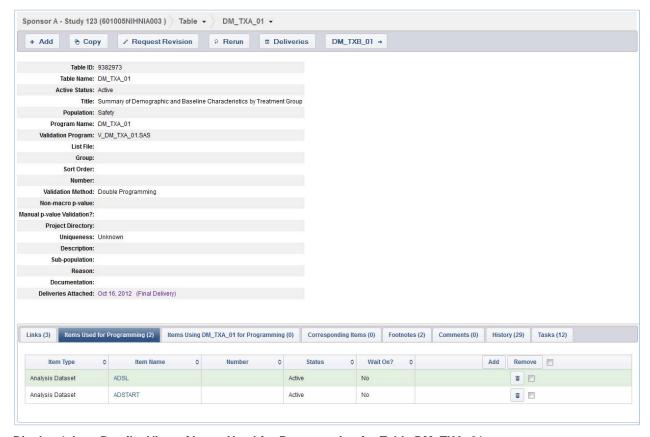
Items can contain one or more links. Links can refer to internal documentation or external websites. This allows a user to view any additional documentation that may exist for an item, such as Mock Tables, Annotated Programs, RTF Displays, and SAS Programs.



Display 3. Item Details: View of Links Section for Table DM_TXA_01

ITEM RELATIONSHIPS

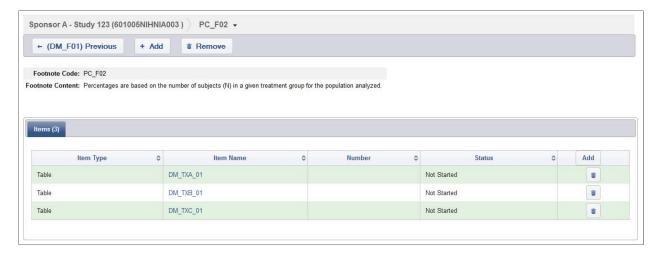
Items can be related to one or more other items in order to show dependencies or cross-references between them. Item relationships are grouped into different categories: Items Used for Programming, Corresponding Items (Displays), and Items Used for Validation.



Display 4. Item Details: View of Items Used for Programming for Table DM_TXA_01

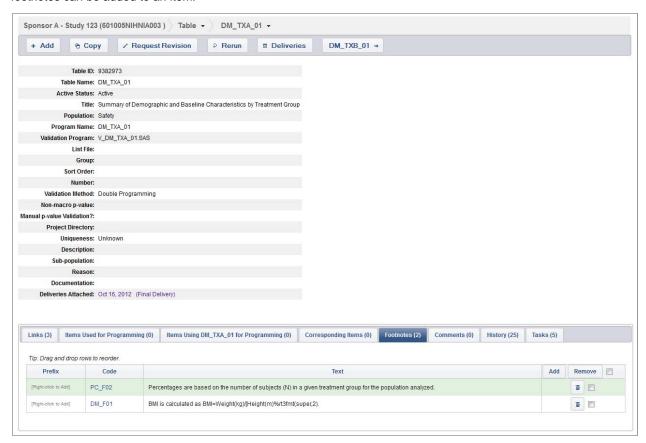
FOOTNOTES

Footnotes that will be used for available item types are added to the application. Once added, items and their related footnotes are then attached to each other.



Display 5. Footnote Details: Footnote PC_F02 and the Table Items that are using PC_F02

Footnotes and items can be attached in one of two ways. Multiple items can be added to a footnote, or multiple footnotes can be added to an item.



Display 6. Item Details: View of Table Item DM_TXA_01 with Attached Footnotes

DELIVERIES

A delivery is a set of items that are due on a specific date. Deliveries are added to the application, and items are then attached to specific deliveries. Items can belong to more than one delivery.



Display 7. Delivery Details: View of Delivery Oct 16, 2012 and Attached Items

THE LIFECYCLE OF AN ITEM

Work begins for an item once it has been added, its tasks generated, and metadata populated. One or more tasks are acquired by a user or assigned to a user by another person. If a user acquires a task, the task is started and a start date is automatically recorded. If the task was assigned to a user, the start date is recorded when the user acknowledges that they have started the task. Tasks can be dependent on other tasks, which means users may not be able to start or complete a task if it is waiting on completion of another task.

For Table DM_TXA_01, the statistical programmers will acquire and start creating the program and validation program. The biostatisticians cannot start the Validate Statistics or QC tasks until the statistical programmers complete their work. Once the programmers sign off, or complete, their tasks, the statisticians can start their work.



Display 8. Tasks: View of Right-Clicking a DM_TXA_01 Create Program Task to View Available Actions

The start date is automatically recorded when a user starts a task. When the work for a specific task is done, that task is signed off on, or completed, by the user. The task is then marked as completed and the completed date is automatically recorded.



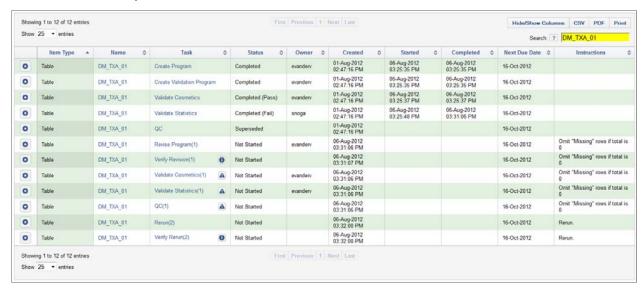
Display 9. Tasks: View of Task Progress for Table DM_TXA_01

REVISIONS

A revision can be created at any point during the lifecycle of an item. There are 3 scenarios in which a revision can be created for an item:

- 1. Manually by a user.
- 2. Automatically upon failure of a task.
- 3. Manually via rerunning.

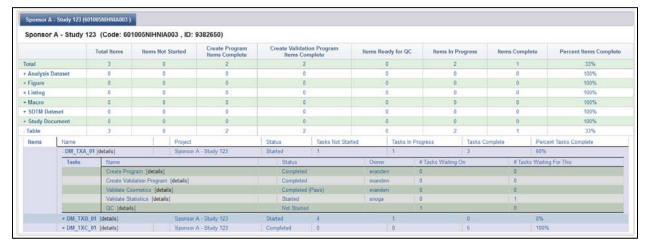
When revisions are created, an auto-generated sequence number is assigned for that revision's tasks. The revision instructions appear for each task for that single revision. Reruns and revisions both use this sequence number that shows the iteration lifecycle of an item.



Display 10. Tasks: View of Task Progress for Table DM_TXA_01 Including Revisions and Reruns

MONITORING PROGRESS

Progress for a project can be monitored for the study as a whole or for one or more deliverables. The project summary page shows the status of the project and is organized by item type. The item types can be drilled down to the item and task level for a more detailed view.



Display 11. Project Summary: Drilled Down Overview of Project

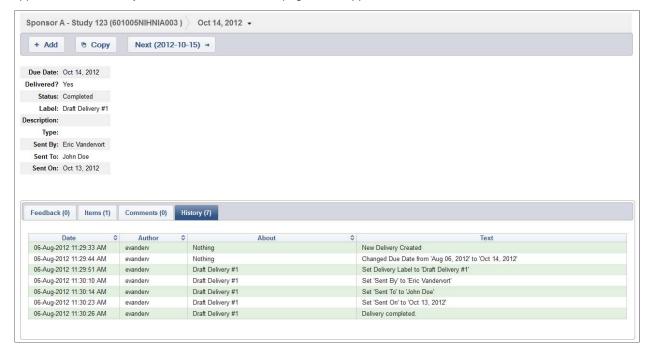
Deliveries are monitored much the same way as the entire project. Each delivery is organized by its delivery date, and then by item type. Deliveries can be drilled down by item type and item for a more detailed view.



Display 12. Deliveries View: Drilled Down View of a Delivery

AUDIT TRAIL

For each item, task, footnote, and delivery, a strict audit trail, or history, is kept of every change that is made. This appears under a history tab located on each detail page in the application.



Display 13. Delivery Detail: View of History tab for Oct 14, 2012 Delivery

METADATA-DRIVEN DESIGN

Projects, item types, fields, item relationships, display fields, and tasks are all defined by metadata. Each entity can be customized for a study simply by making changes to the metadata. With this design, any work processes can be added to the system and tracked, even down to simple checklists.

CONCLUSION

This application was initially created to circumvent problems with multiple people concurrently using a shared Excel file, which lacked in functionality and often led to crashing. How do you know who changed the file last? If you need to make changes, do you just edit the cells or copy and paste only to leave redundant data? Having a web-based application that tracks who did what, and when, allows for precise tracking of where a project or deliverable stands.

REFERENCES

¹Project Tracker Training Manual: Tee Bahnson, Blair McCauley, Laurie McLennan, Shane Rosenbalm, Eric Vandervort, Rho, Inc., Chapel Hill, NC, May 2009

ACKNOWLEDGMENTS

Russ Barnes

Allison Myers

Steve Noga

Brandon Welch

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Eric Vandervort Enterprise: Rho

Address: 6330 Quadrangle Drive, Suite #500 City, State, Zip: Chapel Hill, NC 27517

Phone: (919) 595-6309 Fax: (919) 408-0999

E-mail: Eric_Vandervort@RhoWorld.com

Web: www.rhoworld.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.