Qualification Process for Standard Scripts in the Open Source Repository with Cloud Services

Hanming Tu, Accenture, Berwyn, PA, USA
Dante Di Tommaso, Roche, Basel, Switzerland
Dirk Spruck, Accovion GmbH, Germany
Christopher Hurley, MMS Holdings, Inc., Canton, MI, USA
Nancy Brucken, inVentiv Health, Ann Arbor, MI, USA

ABSTRACT

This paper describes the steps, platform and progress of initiating a qualification process for standard scripts hosted in a code repository with cloud services. The open source repository is used as a collaborative development platform for hosting the specialized programs to be used as analytic tools for clinical trial research, reporting, and analysis through cloud services. We will present how to access the repository, how to contribute to the repository and more importantly how to ensure the quality of the scripts being stored in the repository.

INTRODUCTION

The Pharmaceutical Users Software Exchange's (PhUSE) Computational Science Symposium (CSS) formed a "Standard Scripts for Analysis and Programming" group for two purposes: 1) to create white papers outlining recommendations for safety analysis and reporting for clinical trial study reports and integrated safety-related submission documents, and 2) to establish a platform for sharing and developing standard scripts collaboratively and for implementing the recommendations through cloud services. Since it was formed in year 2012, the working group persistently works on not just building the structure of the repository but also adding more contents in the repository. With more scripts are added into the repository, how we could provide easy access to the scripts and build the users' confidence on the scripts become the important tasks to the group. We will present how to access the repository, how to contribute to the repository and more importantly how to ensure the quality of the scripts being stored in the repository.

BUILDING THE REPOSITORY

In the past three years, the group (http://www.phusewiki.org/wiki/index.php?title=Standard_Scripts) has accomplished the following tasks:
- Selected and built a script repository:
  - Google Code (https://code.google.com/p/phuse-scripts/) was initially chosen as the repository after evaluating half dozen of repositories
  - MIT license was chosen after comparing different open source licenses
  - Tasks, duties, and responsibilities were evaluated and defined for each role
  - The script repository is being migrated to Github (https://github.com/phuse-org/phuse-scripts) since Google is closing its Google Code project.
- Established the basic structure and process for managing the repository:
  - Process guidelines were developed
  - Folder structure and name conventions were proposed
  - Process of tracking issues is recommended
  - Required metadata tags and programming style for scripts were recommended

- Added contents into the repository
  - o Three sets of scripts from an regulatory agency were added
  - o Two dozen scripts from Scriptathon events were hosted in the repository
  - o Some test data sets were added.
  - o Index page for easy accessing to the scripts was created.

HOW TO ACCESS THE REPOSITORY

- *Access to the scripts*: The team built a semi-automatic process to create an index page, so that users can easily access the scripts stored in the repository. The following screenshot shows the top part of the index page http://www.phusewiki.org/wiki/index.php?title=Standard_Script_Index:

Standard Script Index Page

Name	Title	Source	Type	Language	Keywords	Qualification
AE Scripts.zip 🔒	AE Analysis Panel	Contributed	table(AE)	SAS (9.2)	AE, Scripts	Contributed
Demographics_20140516.zip 🔒	Demographics Analysis Panel	Contributed 🔒	table(none)	SAS (9.3)	DM, table	Qualified
MedDRA at a Glance Scripts.zip 🔒	MedDRA at a Glance Script Package	Contributed	table(None)	SAS (9.2)	MedDRA at a Glance, Adverse events, AE, Scripts	Contributed
ae_common.sas 🔒	Table of Common AEs	White paper, Figure 7.1 Summary of Common Treatment-Emergent Adverse Events by Treatment 🔗	figure(figure)	SAS (9.4)	Adverse Events,Common, Treatment-Emergent	Contributed
ae_pref.sas 🔒	AE Table of Preferred Terms	White paper 📄	table(table)	SAS (9.4)	AE, Table, Preferred Terms	Contributed
ae_serious.sas 🔒	Table of Serious AEs	White paper 📄	table(plot)	SAS (9.4)	AE, Table, Serious	Contributed
box_chg_time.Rnw 🔒	Boxplot of change from baseline vs time	White paper 📄	figure(boxplot)	R (3.0.2)	Central, Boxplot, labs, vital signs, ECGs	Contributed
box_obs_time.Rnw 🔒	Boxplot of observations vs time	White paper 📄	figure(boxplot)	R (3.0.2)	Central, Boxplot, labs, vital signs, ECGs	Contributed
box_obs_time.sas 🔒	Boxplot of observations vs time	White paper 📄	figure(boxplot)	SAS (9.2.1)	Central, Boxplot, labs, vital signs, ECGs	Contributed
Box_Plot_Baseline.sas 🔒	Boxplot of change from baseline vs time	White paper 📄	figure(boxplot)	R (3.0.2)	Central, Boxplot, labs, vital signs, ECGs	Contributed
mean_time.sas 🔒	Mean time plot	White paper 📄	figure(plot)	SAS (9.2.1)	Central, Mean Time Plot, labs, vital signs, ECGs	Contributed
demo_summary.R 🔒	Demographic Summary	White paper 🔗	Table(Summary)	R (x.x.x)	Demographics, Table	Contributed
demo_summary.sas 🔒	Demographic Summary	White paper 🔗	Table(Summary)	SAS (9.2.1)	Demographics, Table	Contributed
discontinuation.sas 🔒	Reason for Discontinuation	White paper Demographics 🔒	table(listing)	SAS (9.2.1)	Central, reason, discontinuation	Contributed
disposition_a.sas 🔒	summary of discontinuation reasons	White paper 📄	table(summary)	SAS (9.2.1)	Central, discontinuation, reason, summary	Contributed
					Central,	

Display 1. Screenshot of the Standard Script Index Page

- *Access to the test data in the repository*: The following script sample SAS® code shows how to access the test date stored in the repository:

```
filename source url "http://phuse-
scripts.googlecode.com/svn/trunk/scriptathon2014/data/adpc.xpt" ;
libname source xport ;
data work.adpc ;
set source.adpc ;
keep usubjid ;
run ;

proc print data=work.adpc ;
title1 "Test Dataset Access from the PhUSE Code Repository" ;
```

```
run ;
```

- *Access the scripts through the cloud systems*: The following examples show how to use R and SAS® in the cloud system to access the scripts in the repository:

```
R example
source("http://phuse-
scripts.googlecode.com/svn/trunk/lang/R/report/test/src/adsl.R",echo=T)

SAS® example
options source2 ;
filename code url "http://phuse-
scripts.googlecode.com/svn/trunk/lang/SAS/report/test/src/summary.sas ;
%include code ;
```

HOW TO CONTRIBUTE TO THE REPOSITORY

The standard script working group currently has five active projects after some projects were finished and closed.
- Become a script evaluator by joining the Script Discovery and Acquisition (SDA) team to discover existing scripts and acquire permission from their authors to contribute their scripts.
- Become a script developer by joining the Repository Content and Delivery (RCD) team to develop and curate new scripts or become a script contributor and contribute your own scripts.
- Become a governor of the repository by joining the Repository Governance and Infrastructure (RGI) team to maintain the repository infrastructure and develop the process and structure of hosting the scripts.
- Join Analysis and Display White Papers (ADW) team to develop white papers and identify more targets for script development. White papers describe recommended analyses, tables, figures, and listings, which will be developed in the future.
- Become communicator and advocate of standard script working group by joining the Communication, Promotion and Education (CPE) team to develop communication plans, maintain outreach schedules, manage promotional events and develop educational programs.
- Or simply join our Scriptathon events to get hands-on experience on script development. Here is a list of Scriptathon events that we had so far:
  - 1st Scriptathon: PhUSE/CSS March 17, 2014, Silver Spring, MD, USA
  - 2nd Scriptathon: PharmaSUG, June 1-4, 2014, San Diego, CA, USA
  - 3rd Scriptathon: PhUSE, October 12-15, 2014, London, UK
  - 4th Scriptathon: PhUSE/CSS March 17, 2015, Silver Spring, MD, USA

QUALIFYING THE SCRIPTS

THE IMPORTANCE OF QUALIFYING THE SCRIPTS

It is very important to ensure the quality of the scripts stored in the repository and to build users' confidence by defining a sound qualification process and providing a mechanism that users can be involved in the process. We provided accessibility of the repository through the script index page while we need to build the usability of the scripts through qualification process.

QUALIFICATION PROCESS

The following diagram shows the qualification process and the roles and tasks at each stage. The whole qualification process is composed of four stages: Contributed, Developed, Reviewed and Qualified.
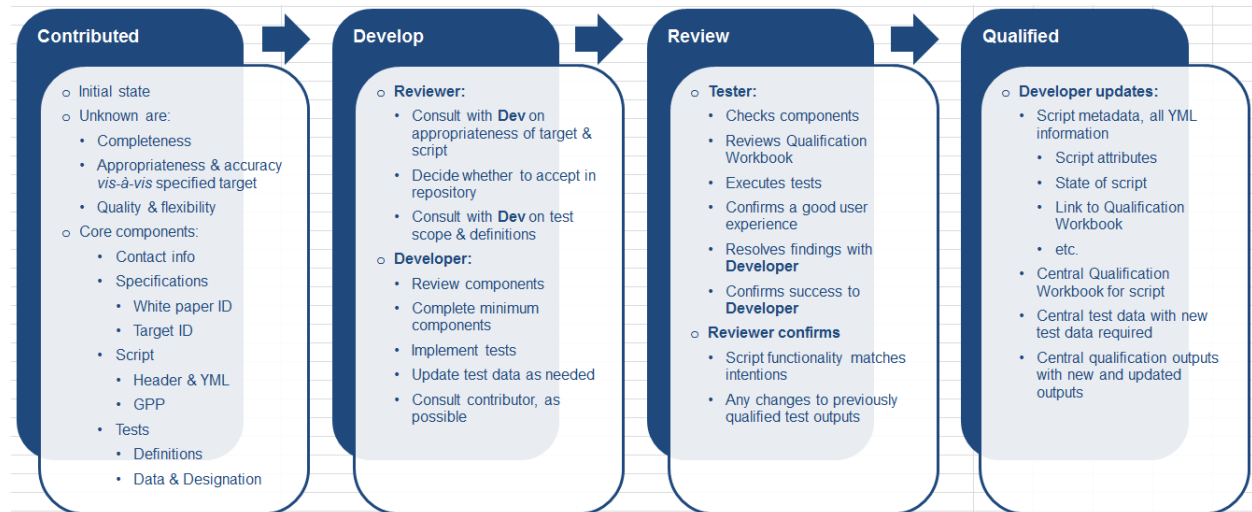

Figure 1. Standard Script Qualification Process

Here is the description of each stage:
- **Contributed**: A script is at contributed stage when it is received from any source in the community and initially stored in the repository.
- **Develop**: A script is under development when at least one developer agrees to make progress with the script.
- **Review**: A script is under review when at least one volunteer has agreed to review the script.
- **Qualified**: A script is qualified when it has passed review and all the required documentation is done.

There are many roles involving in the qualification process. Here is a brief description of each role:
- **Contributor**: Anyone with appropriate skills and interests contributes to any of tasks in the process.
- **Developer**: A volunteer in standard script working group who is familiar with the objectives of a script and has the skill to further develop the script.
- **Tester**: A volunteer in the working group who is familiar with the testing objectives and is willing to execute the script, resolve any finding and confirm success to developers.
- **Environment Tester**: Anyone in community who is able to set up automatic test replication in their work environment.
- **Reviewer**: Author of white papers, designers of script targets, familiar with script objectives.

QUALIFICATION PLAN

The working group relies on distributed volunteers to review, develop, test and qualify Standard Scripts for general use. Volunteers can easily find ways to contribute, either by improving our processes or working directly on the qualification of Standard Scripts. In 2014, we proposed the above process and associated roles and also chose YAML as the metadata format for describing a script. YAML is a recursive acronym for "YAML Ain't Markup Language" or "Yet Another Markup Language." It is a format that can be read by both machine and human. We have recommended metadata tags for the current existing scripts. The script metadata is used to generate the index page. Here is what we plan to do in the coming year:

- Exercise and refine this qualification process while delivering a "Central Tendencies" package based on the published CT white paper from P08 - Analysis and Display White papers.
- Work on the phuse-scripts repository in Github.
- Refine the folder structure to accommodate the needs of each qualification stage.
- Adopt a simpler approach that simply documents tests and checks in scripts in the qualification folder.

CONCLUSION

The goal of the standard script working group is to produce recommendations and establish a platform for the collaborative development of specialized programs to be used as analytical tools for clinical trial research, reporting, and analysis. From what we have experimented, we can conclude
- Open source repository such as Google Code or Github provides a scalable, reliable, and fast collaborative development environment for developing and sharing standard scripts and documents for data transformations and analyses.
- It is mutually beneficial to both the industry and the regulatory authorities if an open source repository is available for storing and developing standard scripts and known test data sets.
- It is a good practice to use the white papers defining the cross-industry analysis and reports based on CDISC standards such SDTM and ADaM.
- It is possible to implement the open source repository through cloud services, as the process for utilizing the standard scripts from the repository in the secure cloud systems has been tested in the Scriptathon events.
- It is very important to conduct the qualification process for each script stored in the repository. The process that we demonstrated covers the minimum requirements for testing the accessibility, usability and functionality.

Work together openly and collaboratively and we can achieve more than what we have hoped for!

ACKNOWLEDGMENTS

CONTACT INFORMATION

Name: Hanming Tu
Enterprise: Accenture
City, State ZIP: Wayne, PA 19087
Work Phone: 610-535-6500, x5564
E-mail: Hanming.H.Tu@accenture,com

Name: Dante Di Tommaso
Enterprise: F. Hoffimann-La Roche, Ltd.
City, State ZIP: CH-4052 Basel, Switzerland
Work Phone: +41 61 687-4314
E-mail: dante.di_tommaso@roche.com

Name: Dirk Spruck
Enterprise: Accovion GmbH

City, State ZIP: 35037 Marburg, Germany
Work Phone: + 49 6421 9484937
E-mail: dirk.spruck@accovion.com

Name: Christopher Hurley
Enterprise: MMS Holdings, Inc.
City, State ZIP: Canton, MI 48187
Work Phone (734) 245-0469
E-mail: churley@mmsholdings.com

Name: Nancy Brucken
Enterprise: inVentiv Health
City, State ZIP: Ann Arbor, MI 48108
Work Phone: 734-887-0255
E-mail: Nancy.Brucken@inventivhealth.com