A Guide to Deploying Subversion for Version Control of SAS Programs in the Pharmaceutical Industry

PhUSE 2013

Tim Williams
Senior Application Manager
Global Data Sciences
UCB BioSciences, Inc.
Subversion Environment

- **Client:** TSVN 1.6, 64 bit client on Windows Server 2008
- **Repository:** SVN. 1.6, Linux
- **Work Area:** NetApp, *shared*
What is SVN?

SVN is:

- Source Control Management (SCM) tool
- version control for programs
- open source

SVN does not provide:

- traceability between data, programs and outputs

SVN is not ideal for:

- binary files (SAS datasets, outputs)
- use in a shared-folder structure environment
Getting Started

IT

- Servers, OS
- Existing deployments of SVN
- Skill sets: PERL? Python?

Programmers

User Requirements

- Development, Validation, Production?
- Incremental moves to Production?
- How is code identified as Production? Archived?
- Separate data from programs and outputs?
- Shared versus private work areas?
- Workshop proof of concept.
Reasons to Validate a Computerized System

4. Pass audits and inspections

2. Confirm reliable handling and protection of data

3. Confirm works as expected with network, other systems, infrastructure

1. Ensure the system performs the activities users need

- Dr. Teri Stokes
Validation Requirements

Define with Quality Assurance & Compliance

Validation Plan

Business Case (formal or informal)

Systems Development Life Cycle (SDLC)
- User Requirements
- Traceability matrix
- Supplier Risk Assessment, Qualification
- Impact Assessment
- Risk Analysis
- Installation Qualification
- Functional Specification
- Design Specification
- Installation Qualification Protocol/Execution/Report
- Operational Qualification Protocol/Execution/Report
- Performance Qualification Protocol/Execution/Report
- Validation Summary Report
- Periodic Review

Responsibilities Matrix
- Create
- Review
- Approve

combine : FSDS

OQ
PQ
Deployment: Complexity versus Simplicity

“The architect should strive continually to simplify...”

Frank Lloyd Wright 1869-1959
American Architect

“Simplify, man!!

“The Hippie Guy”
The Simpsons
Configuration Overview

- Repository locations
- Authentication Method
- Authorization File

Repository

- access_log
- error_log

Collabnet subversion httpd.conf

- access_products
- access_standards
- access_svnconfig

Hooks
- pre-commit

Setup scripts
- setupAnalysis.pl
Access = Authentication & Authorization

**Authentication** – *Who are you?*

- Active Directory
  - same username
  - password expiration, length and complexity

**Authorization** – *May you access this item?*

- access file
- path-based permissions
- Highly configurable
Path-based security in SVN

# $HeadURL: http://*******************/config/prod_server/svn_access_products $
# $Rev: 1619 $
# $Date: 2013-08-02 09:54:48 -0400 (Fri, 02 Aug 2013) $
# $Author: twilliams $ Author of Last Commit

[groups]
ucb_allssas = twilliams, jsasman, jillsweetcoder
ext_croSDG = fred, velma, scoobydoo, shaggy, daphne

#---END SCRIPT-BUILT GROUPS SECTION. DO NOT EDIT SECTION ABOVE ------#

# Custom groups
svnmadmin = twilliams, ucbbadmin1
svnmsetups = e140253,e140122a

[Product1:/Study1]
@svnmsetups = rw
@ucb_allssas = rw
@ext_croSDG= rw

[Product2:/Study1]
@svnmsetups = rw
@ucb_allssas = rw
velma = rw
fred = r
scoobydoo =
svn:externals for Validated Macros

Work area

svn:externals http://<validated repo path>/Release/2.0 validated

Authorized

access_products

(groups)
ext_croSDG = scoobydoo, shaggy...

[product1 /Study1]
@ext_croSDG = rw

Product 1 Repository

Authenticated

Active Directory

Standards Repository

[validation macros:/Macros/Release]
@ucb_standards = rw
@ext_croSDG = r

X Not Authorized

access_standards

Authenticated

Active Directory
SVN Usage and Administration
Blame – A Perception Problem

How some programmers see “Blame”
Commit – A Pervasive Problem
Integration with other tools

UltraEdit

TortoiseProc.exe /command:commit /path:"%f" /closeonend:0

SAS Enhanced Editor

“How to add TortoiseSVN-Functionality to SAS® Enhanced Editor”

- Oliver Wirtz, UCB Biosciences
Breaking Old Habits: Backup Files and Folders

SVN Solutions
- history
- audit trail
- svn tags
Git versus SVN

Git
- Highly configurable
- Performance
- Distributed teams
- Flexible merging

SVN
- Strict governance
- History cannot be altered or changed
- Larger binaries
- Familiar terminology
- Ease of deployment and validation
- Intuitive structures in repository
- Better merging in 1.8
Recent improvements in SVN

- Better merging
- Properties wizard
  - Custom properties
- Performance
- Single `.svn` folder at root of checkout
Conclusion

- A top choice for SAS program code
- SVN does not provide end-to-end traceability between data, programs and outputs. It is a Source Control Management (SCM) system.

Key elements of a successful deployment

- Users
- Process and Work Culture
- Training
- Validation
- Follow up
Q & A