

Automation of a Web-based Metadata Repository for SAS Datasets with Selenium Technology

Eugene Tsykalov

eugene.2.tsykalov@gsk.com

PHUSE US Connect,
14th – 18th June 2021

Clinical Trial Metadata and Depository

☐ Clinical Trial Metadata

☐ ADaM Datasets

- ☐ Datasets Specs, Controlled Terms, Computational Methods, Value Level Metadata

☐ SDTM Datasets

- ☐ Datasets Specs, Controlled Terms

☐ Data Displays (Titles, Footnotes, Program Names)

☐ Web Based Metadata Depository

☐ Advantages and problems

☐ User Workflow for ADaM datasets Metadata

- ☐ Creating Metadata in Excel File
 - ☐ Manual input of Metadata to Web Based Application
-

Agenda



Web Automation Technologies

- ❑ Introduction to Web Automation Technologies
 - ❑ Automation of Web Applications Testing
 - ❑ Web Data Scraping
 - ❑ Commercial Web and Desktop Applications testing
 - ❑ Quick Test Professional (QTP), Hewlett-Packard
 - ❑ Unified Functional Testing (UTF), Micro Focus
 - ❑ Open Source Selenium Technology
 - ❑ Main Functionalities
 - ❑ Supporting Browsers
 - ❑ Supporting Languages
-

☐ ADaM Metadata Handling with Selenium VBA

☐ User Workflow for Metadata with Selenium VBA

☐ Initial Creation of ADaM Metadata in Excel

- ☐ Manual input of Metadata in Excel

- ☐ Copying Existing Metadata from Excel with reference dataset

- ☐ Downloading Existing Metadata from Web to Excel with Selenium

☐ Metadata processing with Selenium VBA

- ☐ Downloading Metadata from Web Depository to Excel

- ☐ User Editing Metadata in Excel

- ☐ Uploading Metadata from Excel to Web

Agenda



Selenium – Universal Method for Web Automation

- ❑ Metadata Automation in Clinical Trials and Beyond
 - ❑ Automation of Metadata for ADaM, SDTM, TLFs
 - ❑ Downloading and Uploading Data to any Web Based Application
 - ❑ Data Mining (Web Scrapping)
 - ❑ Control of Web Based Applications
-

Clinical Trial Metadata and Depository



Clinical Trial Metadata, p1

Clinical Trial Metadata required for FDA submission consist of a large scope of metadata types and numerous attributes, per CDISC specs

Examples of Metadata in Clinical Trials at GSK:

ADaM Datasets Variables Specs (17 attributes):

(Name, Label, Type, Length, Sort Order, Format, Origin, Source, CT*, CM*, VLM*, Comments, etc.)

Controlled Terminology (7 attributes)

Computational Methods (5 attributes)

Value Level Metadata (12 attributes)

SDTM Datasets Variables Specs (17 attributes)

(Name, Label, Type, Length, Sort Order, Format, Origin, Source, Role, etc.)

Controlled Terminology (7 attributes)

Clinical Trial Metadata, p2

Data Displays Metadata (12 attributes):

(Titles, Footnotes, Type, Category, Program Name, etc.)

Let's estimate the number of Metadata Attributes (Objects) for a reporting example:

Number of ADaM Datasets, Variables, Objects: Nadam=15, Nvars1=100, Nobj1=17

Number of SDTM Datasets , Variables, Objects: Nstdm=30, Nvars1=50, Nobj2=17

Number of Data Displays: Ndispl=200, Nobj3=12

Number of Metadata Objects:

$N_{meta_rep} = N_{adam} * N_{vars1} * N_{obj1} + N_{stdm} * N_{vars2} * N_{obj2} + N_{displ} * N_{obj3}$

$N_{meta_rep} = 15 * 100 * 17 + 30 * 50 * 17 + 300 * 12 = 25,500 + 25,500 + 2,400 = 53,400 \text{ metadata objects!}$

Depending on number of datasets, the Nmeta could be up to 100,000 metadata objects in just one reporting (Dry Run, Interim, DSUR, ISS, ISE, etc.)

Clinical Trial Metadata and Depository



Web Based Metadata Depository, p1

Estimate of Metadata Amount for a big Pharma Company example:

Number of Reporting Efforts in a study:	Nrept=10 (Dry Run, DSUR, Data Look, etc.)
Number of Studies for a Compound:	Nstud=10 (Phase 2, Phase 3, Integrated, etc.)
Number of Compounds in a TA*:	Ncomp=5
Number of Therapeutic Areas:	Nta=5

Number of Metadata Objects in a Company Depository:

$$N_{\text{meta_company}} = N_{\text{meta}} * N_{\text{rept}} * N_{\text{stud}} * N_{\text{comp}} * N_{\text{ta}}$$

$$N_{\text{meta_company}} \sim (50,000-100,000) * 10 * 10 * 5 * 5 \sim (125-250) * 10^6 \text{ metadata objects!}$$

All this metadata is stored in a Data Depository providing:

- centralized user access
 - traceability
-

Clinical Trial Metadata and Depository



Web Based Metadata Depository, p2

Common Solution for Clin Trial Metadata storage: *Web Based Depository*

Main advantages of Web Based Depository on a Server:

- Multi-user centralized concurrent access
- Interactivity for user actions (looking at the data and outputs, running programs, etc.)
- Traceability in a secure production environment

Problems:

- Difficult to display large amount of metadata in a limited Web pages space
- Difficult to view/review multiple metadata objects at Web pages
- Tedious, time consuming and error prone *manual input and editing* of metadata at Web pages
- Difficult and time consuming adding new functionalities to existing Web Based Application

Common User Workflow for ADaM datasets Metadata

- Creating and Editing Metadata in Excel file
 - Manual input/update of Metadata to Web Based Application
-

Web Automation Technologies



Introduction to Web Automation Technologies

Automation Technologies were created to do:

- Automation of Desktop and Web Applications Testing
(emulating user mouse and keyboard actions in Apps)
- Web Data Scraping (Data Mining)
(collecting data from Web sites using automatic bots)

Available Automation Solutions:

Commercial Programs for Web and Desktop Application testing:

- Quick Test Professional (QTP), Hewlett-Paccard
- Unified Functional Testing (UTF), Micro Focus

Open Source Web Automation Framework:

- Selenium Technology
-

Web Automation Technologies



Open Source Selenium Technology

Selenium is matured, powerful, well developed and supported Web Automation technology

Main Functionalities:

- emulates most (if not all) user actions at Web pages (mouse moves and clicks, keyboard entries, drag and drop, etc.)
- does not require any changes in Web Application
- does not break security of Web Applications
- much faster than user manual actions
- improves accuracy, quality and consistency of metadata loading
- all complexities of metadata handling are programmed inside Excel macros

Supporting Browsers:

- Microsoft IE, Edge; Google Chrome, Firefox

Supporting Languages:

- R, Python, VBA, Java, Ruby
-

Web Automation with Selenium VBA



ADaM Metadata Handling with Selenium VBA

Selenium VBA is a Selenium Library, ported to VBA language, it works as macros in Excel

User Workflow for ADaM Metadata with Selenium VBA:

Initial Creation of Metadata in Excel:

- by manual input of metadata in Excel from scratch
- by copying existing metadata from Excel with reference dataset
- by downloading existing metadata from Web depository to Excel with Selenium

Metadata Processing with Selenium in Excel:

- Automatic Downloading of Metadata from Web Depository to Excel
 - User Editing and Maintaining of Metadata in Excel
 - Automatic Uploading of Metadata from Excel to Web Depository
 - Automatic built-in data checks and alerts
 - Automatic comparison of Metadata in User Excel vs Web Depository
-

Selenium – Universal Method for Web Automation



Metadata Automation in Clinical Trials and Beyond

Selenium is a universal technology, can be used for Automation of any Web Based Application

Automation in Clinical Trials:

Metadata processing for ADaM and SDTM datasets and TLFs at Web Based Depository

Other Applications of Selenium Technology:

Downloading and Uploading data to any Web Based Application:

- VeeVa Vault, Documentum, etc.

Data Mining (Web Scrapping)

Control of Web Based Application

- running programs, checking results, creating reports
-

Three overlapping, rounded, organic shapes in shades of orange and yellow, creating a layered effect on the left side of the slide.

Thank You