AUTOMATED CREATION OF SUBMISSION-READY ARTIFACTS

SILAS MCKEE
AGENDA

1. Motivation
2. Automation Overview
3. Architecture
4. Validating the System
5. Pilot Study Results
6. Future State
WHY AUTOMATE?

Return (benefit) = ROI (Return on Investment)

\[
\frac{\text{Return (benefit)}}{\text{Investment (cost)}} = \text{ROI (Return on Investment)}
\]

- Reduce the per trial cost of submission artifact creation
- Reduce the cycle time for submission artifact creation
- Increase data quality

Source: https://xkcd.com/1319/
Moving from traditional programming to full automation

No Automation

- Organic Programming Methodology
  - Traditional programming
    - Smallest initial investment
    - Most Flexible
  - Low Efficiency
    ✓ Cycle times measured in Months

Full Automation

- Custom Macro Library
  - Traditional programming
    - Standard macro library for code re-use
  - Moderate Efficiency
    ✓ Cycle times measured in Weeks

- Automated Metadata Driven Processing (AMDP)
  - Metadata-driven code generating engine
  - High Efficiency
    ✓ Cycle times measured in Days
  - Repeatable outputs
  - Reduced cycle times
  - Reduced cost

Why Choose It

- Relative Efficiency
  - Most Flexible
  - Supports junior level resources
  - Enables code re-use

What It Is

- Traditional programming

AUTOMATION WITH METADATA

CDISC Metadata

1. CDISC Metadata
   - Provides structural metadata to inform dataset structure and variable creation statements
   - Ensures compliance with data standards

2. Operational Metadata
   - Provides instructions for join / merge operations
   - Provides instructions for macro calls including parameters values
   - Provides instructions on the algorithms used to transform, derive, or compute data.
### METADATA MODEL

#### Complexity vs. Expressiveness

There is a natural trade-off between the complexity of the metadata model and its expressiveness.

<table>
<thead>
<tr>
<th>Study Metadata</th>
<th>Table Metadata</th>
<th>Metadata Complexity</th>
<th>Expressiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA</td>
<td>Title: ‘Summary Table for Vital Signs’</td>
<td>Simple</td>
<td>Low</td>
</tr>
<tr>
<td>Alignment: Left, Font: Times New Roman, Text Size: 12</td>
<td>Title: ‘Summary Table for Vital Signs’</td>
<td>Medium</td>
<td>Medium</td>
</tr>
<tr>
<td>NA</td>
<td>Title: ‘Summary Table for Vital Signs’</td>
<td>Complex</td>
<td>High</td>
</tr>
</tbody>
</table>
AMDP ARCHITECTURE

Current State

METADATA SOURCES

- Standards
- Mapping Specs
- Controlled Terminology

METADATA SPECIFICATION

- CDISC Metadata
- Operational Metadata

SCE ENVIRONMENT

- Executable Programs
- Submission-Ready Artifacts

SAS SERVER

- SDTM
- ADaM
- TFLs

AMDP SAS Engine

- Generates
- Instructs

Utility Macros

Reads

AMDP SAS Library

Copyright © 2018 Accenture All rights reserved.
### VALIDATING THE SYSTEM

#### The Artifacts

1. Created a set of standard artifacts that included representation from the following categories across SDTM, ADaM, and TFLs

<table>
<thead>
<tr>
<th>SDTM</th>
<th>ADaM</th>
<th>TFL</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Observation Classes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Findings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Interventions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Events</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Special Purpose</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Special Purpose Relationship</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADaM Subject-Level Analysis Dataset (ADSL)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Occurrence Data (OCCDS)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basic Data Structure (BDS)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basic Listings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summary Tables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Change from Baseline Tables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shift Tables</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2. Validated the generation of these standard artifacts against data from two different studies
## VALIDATING THE SYSTEM

### Artifact Details

<table>
<thead>
<tr>
<th>Artifact Type</th>
<th>Template or Class</th>
<th>Artifact(s) Produced during Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDTM</td>
<td>Findings</td>
<td>DM, DS, EX, EG, VS, LB, IE, PE,</td>
</tr>
<tr>
<td></td>
<td>Interventions</td>
<td>CM,</td>
</tr>
<tr>
<td></td>
<td>Events</td>
<td>AE, HO</td>
</tr>
<tr>
<td></td>
<td>Special Purpose</td>
<td>DM</td>
</tr>
<tr>
<td>ADaM</td>
<td>ADSL</td>
<td>ADSL</td>
</tr>
<tr>
<td></td>
<td>OCCDS</td>
<td>ADCM</td>
</tr>
<tr>
<td></td>
<td>BDS</td>
<td>ADEG</td>
</tr>
<tr>
<td>TFL</td>
<td>Summary Table – Continuous and Categorical Variables Statistics</td>
<td>“Demography Summary Table”</td>
</tr>
<tr>
<td></td>
<td>Summary table – Frequency Count and Percent by Single or Double Level</td>
<td>“Concomitant Medications [by CMDECOD]”</td>
</tr>
<tr>
<td></td>
<td>Summary table – Frequency Count and Percent by Single or Double Level</td>
<td>“Concomitant Medications within Classes of Interest [CMCLAS*CMDECOD]”</td>
</tr>
<tr>
<td></td>
<td>Change from Baseline Table</td>
<td>“Summary of ECG Parameters and Change from Baseline by Time by Treatment”</td>
</tr>
<tr>
<td></td>
<td>Shift Table</td>
<td>“Summary of Post-Baseline Worst-Case ECG Parameters Shift from Baseline by Treatment”</td>
</tr>
<tr>
<td></td>
<td>Basic Listing</td>
<td>“Listing of Concomitant Medications”</td>
</tr>
</tbody>
</table>
PILOT STUDY RESULTS

Description

- The team received a proprietary global metadata library and source data for the study.
- The team then developed study specifications (both human and machine-readable) to define the conversion of the source data into 48 SDTM domains.
- Successfully completed complete SDTM conversion with high quality in 31 business days.

Quality Metrics

<table>
<thead>
<tr>
<th>Error Category</th>
<th>Error Count</th>
<th>Total Variables</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major</td>
<td>3</td>
<td>700</td>
<td>99.6%</td>
</tr>
<tr>
<td>Minor</td>
<td>6</td>
<td>700</td>
<td>99.1%</td>
</tr>
<tr>
<td>Total</td>
<td>9</td>
<td>700</td>
<td>98.7%</td>
</tr>
</tbody>
</table>

$Quality = 1 - \left( \frac{Error\ Count}{Total\ Variables} \right)$
## PILOT STUDY RESULTS

### Work Breakdown

<table>
<thead>
<tr>
<th>Work Type</th>
<th>Reduction Expected?</th>
<th>Relative Effort (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global Specification Interpretation</td>
<td>Yes</td>
<td>10%</td>
</tr>
<tr>
<td>Client-specific Team Training</td>
<td>Yes</td>
<td>5%</td>
</tr>
<tr>
<td>Human-readable Specification Development</td>
<td>Yes</td>
<td>15%</td>
</tr>
<tr>
<td>Machine-readable Specification Development</td>
<td>Yes</td>
<td>40%</td>
</tr>
<tr>
<td>AMDP Code Development</td>
<td>Yes</td>
<td>5%</td>
</tr>
<tr>
<td>QC and Compliance Reviews</td>
<td>No</td>
<td>15%</td>
</tr>
<tr>
<td>Client and Project Management</td>
<td>No</td>
<td>10%</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>100%</td>
</tr>
</tbody>
</table>

### Takeaways

- Time to perform most tasks expected to decrease with study volume
- Very little Code Development
- 55% of effort devoted to specification development
AMPD ARCHITECTURE

Future State

METADATA SOURCES

Standards
Mapping Specs
Controlled Terminology

METADATA SPECIFICATION

CDISC Metadata
Operational Metadata

Supervised Machine Learning

SCE ENVIRONMENT

Executable Programs

AMDP JAVA Engine
Generates

AMDP R/SAS Library
Utility Macros

R SERVER

Instructions

SAS SERVER

Submissions-Ready Artifacts
SDTM
ADaM
SEND
TFLs

Additional Artifacts

Table Ready Datasets

Populates

Automated Analysis Reports and Dashboard

Exploratory Analysis
Patient Profiles
Internal Reporting
KPIs

Copyright © 2018 Accenture All rights reserved.
RECAP

- Described an approach and system for the automation of submission-ready artifacts using a metadata-driven code generating engine
- Discussed the validation of the system
- Presented results from a pilot study to convert source data to 48 SDTM domains
- Proposed enhancements to the system
Questions?