Gustav Bernard
Gustav Bernard is an Associate Director at IQVIA who has been with the company for 18 years. His work focuses on the implementation of CDISC Standards (SDTM, ADaM and Define-XML) within IQVIA. He is currently Leading ADaM Innovation @iqvia.

He has also created the Define-XML 2.0 automation process within IQVIA, the ADaM Spec Generator application and the ADaM Designer application.

Gustav earned a Bachelor of Business in computer science from the University of the Orange Free State in South Africa in 2004.
ADaM Automation Roadblocks

NC PharmaSUG 2022

Gustav Bernard, IQVIA, Associate Director
ADaM Automation Roadblocks

- Time
- Mindset
- Expertise

- Connecting Processes
- Change Management
- Innovation
Making Sense of a Data Universe in Biostatistics

- **Digitize**
  Transforming ‘analog’ content into digital building blocks for creation and consumption

- **Centralize**
  Aggregating ‘data’ into a single organized, governed single source of truth

- **Standardize**
  Establish reusable, connected content with supporting processes to drive compliance

- **Automate**
  Identify opportunities to drive efficiency and quality across manual and repetitive tasks

Clinical Data Universe
A robust BIOS solution that meets the new demands of the market

**Better BIOS Solutions**

**Centralize**
Aggregation of operational clinical data enriched through BIOS processes

**Standardize**
Application of clinical data standards, mapping libraries, documents, processes, etc.

**Connect**
Connected lineage of data assets based on metadata-driven derivations

**Automate**
Automation tools and platforms taking advantage of the data flow ecosystem

**Integrated Processes:**
- Self-Serve Data Availability
- Statistical Programming & Analysis
- Clinical Reporting
- Data Issue Management

---

**100% Data Ingestion Automation**

**100% Data Consumption Automation**

**Real-Time Data Flow & Access**

**Data-Driven Standardized Process**
Connecting Processes

Web Based with Single backend Database (ADaM Focus)

- SAP Metadata
- Shells Metadata
- SDTM Metadata
- ADaM Standards
- NCI CT

ADaM Metadata

TLF Metadata
Change Management

Continues Evolution of processes

- Excel Template
- Standard in PDF document
- VBA Application
- CDISC Standards Template with Customized Metadata
- NCI CT Terminology
- SDTM Metadata
  - Domain Level
  - Variable Level
  - VLM Level
  - Variable and VLM Terminology
- Web Based Version
- API Link To CDISC Library
- CDISC Standards Template with Customized Metadata
- API link To NCI CT Terminology
- SDTM Metadata
  - Domain Level
  - Variable Level
  - VLM Level
  - Variable and VLM Terminology
Change Management

Continues Evolution of processes
# Traditional Spec

<table>
<thead>
<tr>
<th>Final Variable Order in Dataset 2</th>
<th>Variable Name</th>
<th>Variable Label</th>
<th>Type</th>
<th>Length</th>
<th>Controlled Terms, Codelist or Format</th>
<th>Origin</th>
<th>Computational Algorithm or Method (CDISC only)</th>
<th>Details of derivation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 STUDIOID</td>
<td>Study Identifier</td>
<td>Char</td>
<td>20</td>
<td></td>
<td>SDTM.CM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 SUBJID</td>
<td>Unique Subject Identifier</td>
<td>Char</td>
<td>40</td>
<td></td>
<td>SDTM.CM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 SUBJID</td>
<td>Subject Identifier for the Study</td>
<td>Char</td>
<td>20</td>
<td></td>
<td>SDTM.CM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 SITEID</td>
<td>Study Site Identifier</td>
<td>Char</td>
<td>4</td>
<td></td>
<td>SDTM.CM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 BIRTHDT</td>
<td>Date of Birth</td>
<td>Char</td>
<td>10</td>
<td>ISO8601</td>
<td>SDTM.CM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 BIRTHDTN</td>
<td>Date of Birth (N)</td>
<td>Num</td>
<td>8</td>
<td>DATE9</td>
<td>Derived</td>
<td></td>
<td>convert BIRTHDT to a numeric value</td>
<td></td>
</tr>
<tr>
<td>7 AGEDT</td>
<td>Age</td>
<td>Num</td>
<td>8</td>
<td></td>
<td>SDTM.CM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 AGEU</td>
<td>Age Units</td>
<td>Char</td>
<td>10</td>
<td>AGEU</td>
<td>SDTM.CM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 SEX</td>
<td>Sex</td>
<td>Char</td>
<td>2</td>
<td>SEX</td>
<td>SDTM.CM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 SEXN</td>
<td>Sex (N)</td>
<td>Num</td>
<td>8</td>
<td>SEXN</td>
<td>Derived</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11 WBCF</td>
<td>Childbearing Potential</td>
<td>Char</td>
<td>SDTM.SUPPDM</td>
<td></td>
<td>SUPPORT QVAL where SUPPDM.QNM = &quot;CMGWCPNT&quot;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 GWCFN</td>
<td>Childbearing Potential (N)</td>
<td>Num</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13 RACE</td>
<td>Race</td>
<td>Char</td>
<td>20</td>
<td>RACE</td>
<td>SDTM.CM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 RACEN</td>
<td>Race (N)</td>
<td>Num</td>
<td>8</td>
<td>RACEN</td>
<td>Derived</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 RACEOTH</td>
<td>Race Other – Specify</td>
<td>Char</td>
<td>100</td>
<td>SDTM.SUPPDM</td>
<td>SUPPORT QVAL where SUPPDM.QNM = &quot;CMRACE9T&quot;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 ETHNIC</td>
<td>Ethnicity</td>
<td>Char</td>
<td>22</td>
<td>ETHNIC</td>
<td>SDTM.CM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17 ETHNICN</td>
<td>Ethnicity (N)</td>
<td>Num</td>
<td>8</td>
<td>ETHNIC</td>
<td>Derived</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 ETHNICAB</td>
<td>Ethnic Abbreviated</td>
<td>Char</td>
<td>4 ETHNICAB</td>
<td>Derived</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19 ICOT</td>
<td>Date of Informed Consent</td>
<td>Num</td>
<td>8</td>
<td>DATE9</td>
<td>Derived</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Edit Variable Attributes

#### Variable Categories

<table>
<thead>
<tr>
<th>Variable</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>STUDYID</td>
<td>Study Identifier</td>
</tr>
<tr>
<td>USURID</td>
<td>Unique Subject Identifier</td>
</tr>
<tr>
<td>SUSERID</td>
<td>Subject Identifier for the Study</td>
</tr>
<tr>
<td>SITEID</td>
<td>Study Site Identifier</td>
</tr>
<tr>
<td>SEX</td>
<td>Sex</td>
</tr>
<tr>
<td>RACE</td>
<td>Race</td>
</tr>
<tr>
<td>COUNTRY</td>
<td>Country</td>
</tr>
<tr>
<td>ETHNIC</td>
<td>Ethnicity</td>
</tr>
<tr>
<td>AGE</td>
<td>Age</td>
</tr>
<tr>
<td>AGEDU</td>
<td>Age Units</td>
</tr>
<tr>
<td>AGEGR1</td>
<td>Pooled Age Group 1</td>
</tr>
<tr>
<td>AGEGR2</td>
<td>Pooled Age Group 2</td>
</tr>
<tr>
<td>AGEGR3</td>
<td>Pooled Age Group 3</td>
</tr>
</tbody>
</table>

#### Variable Edit

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Variable Label</th>
<th>Pooled Age Group 1</th>
<th>Pooled Age Group 2</th>
<th>Pooled Age Group 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGEGR1</td>
<td>Pooled Age Group 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGEGR2</td>
<td>Pooled Age Group 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGEGR3</td>
<td>Pooled Age Group 3</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Description

Character description of a grouping or pooling of the subject’s age for analysis purposes. For example, AGEGR1 might have values of “<18”, “18-45”, and “>45”. AGEGR2 might have values of “Less than 50 y old” and “At least 50 y old.”
Why have a Tool \ Application

**Re-usable**
- Building a searchable database
- Sponsor-specific template availability

**Web-based**
- Improved performance
- 24/7 IT support
- Use wherever you can access the web

**Digitized**
- Structured documents
- Built from smaller components
- Easier to re-use
- Easier to manage
ADaM Automation Keeping Track

- Keeping Track
- What has been created
- What is being developed in current release
- What to develop in next release
- What to develop for the following release
- What functionality needs improvements
- What functionality is not working as expected
- What functionality needs to be investigated
- What linking with other applications needs to happen when
Automation Implementation

Tool vs Process vs Macro

• Tool

• Specification Metadata

• Should be setup to be able to handle SAS \ R \ Python

• Process

• For Example, Macro that Runs over all SDTM datasets to create SDTM Evaluation that feeds into Tool.

• Macro that runs over SDTM Evaluation and Create Excel Output with recommended assignment of Parameters from SDTM → ADaM
  - Output Read in by SAS programs to create Parameter variables
  - Output Read by tool to create VLM Metadata

• Macros

• Macro that uses Metadata from either Tools or Processes to create variables or add additional rows of data.
Innovation

Macro Management

- Global ADaM Macros
- Project\Client Specific Macros
- Global ADaM Processes
- Project\Client Specific Processes
Innovation - Setting up a Macro in a Global Macro Registry

ADaM Macros

• # of Parameters
• Name and Description for each Parameter
• Set allowed contents for parameter (Restricted vs not)
  - Restricted when Set of expected values allowed
    › Add Restricted list of allowed Values (e.g., 0 or 1)
  - Free Text
  - Selection from Metadata \ Where Clause
  - Default Values
• Variable(s) to be derived using Macro
  - What variables should be auto assigned (All Predecessor variable handled by one macro)
• Expected Single call or Multiple call
  - Datetime macro that will create DTN, TM, DTM and DY variables only needs to called once.
• Expected Code Level / Should always be called

• Code Levels
  - Initialize
  - Customize
  - Formalize \ Standardize
  - Finalize
• Each macro should have adequate defensive programming for each macro parameter with clear messaging if parameter used incorrectly.
• Version Control of each Macro
• Creating software dependent macros that are controlled via parameter calls
Code Management (Code Creation and Maintenance)

**Within a controlled system?**
- Code maintenance becomes a concern and can lead to rework.
- Highly controlled

**Outside of a system?**
- Tried and tested works but uncontrolled

**Best of Both Worlds**
- System Controls Certain code levels
  - Initialize
  - Formalize \ Standardize
  - Finalize
- Output each level per dataset
- User calls each Level to execute in program and can add Customized Level.
Innovation

Decision Tree

Variable Level

• Variable Derivation Sequence
• Variable Dependencies
• Automated until User takes control
• Code Level and Variable Priority

VLM Level

• Test Level Derivation Sequence
• Test Level Dependencies
• Automated until User takes control
• Code Level and Test Priority
Traditional vs Automated

**Traditional (AOCC01FL)**

- **Initial Creation**
  - Explain Derivation in Specification
  - Code Derivation in Production Program
  - Code Derivation in QC Program

- **Maintenance (Rework)**
  - Update Derivation in Specification
  - Update Derivation in Production Program
  - Update Derivation in QC Program

**Automated (AOCC01FL)**

- **Initial Creation**
  - Explain Derivation in Specification
  - Macro **Extracts** Derivation in Prod Prog
  - QC Macro **Extracts** Derivation in QC Prog

- **Maintenance (Rework)**
  - Update Derivation in Specification
  - Rerun Production Program
  - Rerun QC Program
Question Based Approach to include Occurrence Variable
Variable populated with Text Using Keywords
Derivation = Variable Calculation

Programming Derivation
Set to Y for: First Within: [First]; Where: [Where]; Using Sort Order: [Sort].

Define-XML Comment
The first occurrence of any [First] within [Where]

Programming Derivation
Set to Y for: First Within: AEBODSYS; Where: TRTEMFL="Y"; Using Sort Order: USUBJID AEBODSYS.

Define-XML Comment
The first occurrence of any AEBODSYS within TRTEMFL="Y"

Utilize Spec Metadata
Call %adam_initilaize

Call %adam_flags

Identify all Flag variables
Identify if automated syntax present

Extract Metadata: variable, first, where, sort
Calculate Variable
Library Usage Numbers

- Number of **Studies** Processed: 264
- Total Amount of ADaM **Datasets**: 3692
- Average Amount of **Datasets Per Study**: 13
- Average Amount of **Variables Per Datasets**: 83
# Class Level Analysis

## Trend \ Usage Analysis

<table>
<thead>
<tr>
<th>CLASS</th>
<th>Nr</th>
<th>Ave Variables</th>
<th>ADSL Variable Ave</th>
<th>Custom Variables Ave</th>
<th>SDTM Variables Ave</th>
<th>ADaM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject Level Analysis Dataset</td>
<td>265</td>
<td>91</td>
<td>39.41</td>
<td>7.83</td>
<td>43.76</td>
<td></td>
</tr>
<tr>
<td>Occurrence Data Structure</td>
<td>1099</td>
<td>87</td>
<td>43.31</td>
<td>5.43</td>
<td>22.31</td>
<td>15.95</td>
</tr>
<tr>
<td>Basic Data Structure</td>
<td>2268</td>
<td>81</td>
<td>41.68</td>
<td>3.42</td>
<td>13.05</td>
<td>22.85</td>
</tr>
<tr>
<td>Basic Data Structure - TTE</td>
<td>48</td>
<td>57</td>
<td>41.77</td>
<td>1.54</td>
<td>0.17</td>
<td>13.52</td>
</tr>
</tbody>
</table>
# Dataset Level Analysis

<table>
<thead>
<tr>
<th>Dataset Name</th>
<th>Nr of Studies with Dataset</th>
<th>Source = Assigned</th>
<th>Source = Derived</th>
<th>Source = Predecessor</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADSL</td>
<td>265</td>
<td>8</td>
<td>57</td>
<td>25</td>
</tr>
<tr>
<td>ADAE</td>
<td>257</td>
<td>5</td>
<td>31</td>
<td>80</td>
</tr>
<tr>
<td>ADCM</td>
<td>237</td>
<td>2</td>
<td>14</td>
<td>72</td>
</tr>
<tr>
<td>ADVS</td>
<td>237</td>
<td>8</td>
<td>17</td>
<td>55</td>
</tr>
<tr>
<td>ADLB</td>
<td>233</td>
<td>12</td>
<td>27</td>
<td>63</td>
</tr>
<tr>
<td>ADMH</td>
<td>228</td>
<td>1</td>
<td>5</td>
<td>60</td>
</tr>
<tr>
<td>ADEG</td>
<td>197</td>
<td>9</td>
<td>19</td>
<td>56</td>
</tr>
</tbody>
</table>
## Individual Dataset Level Analysis

### Trend \ Usage Analysis

<table>
<thead>
<tr>
<th>Dataset Name</th>
<th>Amount of Studies with Domain</th>
<th>% of Studies with Domain</th>
<th>Overall Ave Variables Per Domain</th>
<th>Source_Assigned</th>
<th>Source_Derived</th>
<th>Source_Predecessor</th>
<th>ADLB</th>
<th>ADSL Variable</th>
<th>Custom Variables</th>
<th>SDTM Variables</th>
<th>Timing and Visit Variables</th>
<th>Parameter Variables</th>
<th>Baseline and Change from Baseline Variables</th>
<th>Record-level Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>233</td>
<td>ADLB</td>
<td>40.61</td>
<td>5.52</td>
<td>21.51</td>
<td>6.61</td>
<td>.81</td>
<td>3.52</td>
</tr>
</tbody>
</table>

### Analysis Criteria and Indicators

<table>
<thead>
<tr>
<th>Analysis Criteria and Indicators</th>
<th>Dataset Initialize</th>
<th>Other Analysis Indicators</th>
<th>Custom Specified - SDTM Input</th>
<th>Custom Specified - ADaM Availible</th>
<th>Toxicity and Range Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5.61</td>
<td>.81</td>
<td>.04</td>
<td>.79</td>
<td>5.30</td>
</tr>
</tbody>
</table>
Study Library

• Each Completed Study becomes a Library Project that can be Reused by any other Study.
• Client Specific Libraries
• Safety Libraries
• Therapeutic Area Libraries

• Using Libraries to startup code
• After having enough connected metadata (500+ studies) can start using AI \ ML to help identify best fit to startup new studies
Building the COVID19 Library

SDTM VLM Metadata + TLF VLM Metadata → Classify BDS ADaM Datasets

**Inputs for Domain**

a) Parameters Assigned from SDTM Only
b) Parameters Assigned from SDTM and New Derived Tests
c) New Derived Parameters Only

**Baseline Variables Needed**

a) Yes
b) No

c) New Derived Parameters Only

**Criteria or CTCAE Variables Needed**

a) Yes
b) No
Process to Create Parameter Variables (PARAMCD, PARAMN, PARAM, PARCAT1, PARCAT1N, PARCAT2, PARCAT2N, PARCAT3 and PARCAT3N) using a Data Driven approach along with Algorithms to ensure Consistency.
Parameter Process

- VLM Summary
- TESTCD
- TEST
- STRESU
- ORRESU
- CAT
- SCAT
- LOC SPEC POS METHOD
- Findings Domains
- Evaluate SDTM
- SDTM
- PARAMCD
- PARAM
- PARCAT1
- PARCAT2
- PARCAT3
- PARCAT1N
- PARCAT2N
- PARCAT3N
- Algorithms
- Parameters
  - By Findings Domains
  - SDTM Merging Variables By Domain
  - Parameter Variables
Automation Processes

CRIT Process

Process to Decrease the amount of Files that Store and Maintain Criteria information.
ADaM Automation Roadblocks

Common Considerations

• CDISC Define-XML Standard not very set
• CDSIC ADaM Standards multiple ways to do the same thing
• FDA requesting sas \ r code
• When Automating SAS code
  - Need to still be able to submit to FDA
  - Spec needs to still be in adequate for
    › Sponsor Review
    › Create Define-xml
    › Programming Needs
    › Code Automation
Q&A

• Gustav.Bernard@iqvia.com