A PARADIGM SHIFT IN CLINICAL DATA PREPARATION

THE POWER OF GRAPHICAL DATA FLOW

Vineet Jain
Nimble Clinical Research
INTRODUCTION

- **Traditional Statistical Programming**: Tiedious, & error-prone

- **Graphical Data Flow**: A visual approach to data integration and transformation that is intuitive and efficient

- **The Imperative**: With the changing technology landscape, and the demand for real-time insights, there’s an opportunity for change.
CURRENT OPTIONS

- **Tableau Prep**: Interactive platform for data cleansing and preparation with a drag-and-drop interface.
- **Alteryx Designer**: End-to-end platform offering data blending, analytics, and visualization capabilities.
- **KNIME**: Open-source analytics platform, harnessing data through modular workflows.
- **Talend**: Cloud-based tool focusing on big data integration and management through a visual approach.
- **SAS Based Solutions**:  
  - Data Integration Studio
  - SAS Enterprise Guide
  - SAS Studio
CHALLENGES

Complex Logic & Custom Code
• Graphical interfaces might not capture the nuances or allow for highly customized code.
• May not support SAS or R

Lack of Targeted tools
• Lack of tools targeting CDISC standards and Clinical Data needs

Performance & Scalability
• For large datasets and complex operations, concerns about the performance and scalability

Interoperability & Integration
• Integration with existing workflow may be challenging.
### Evolving Technology Landscape
Shift towards R & cloud, with integration capabilities with powerful open-source technology solutions making graphical data flow viable.

### Industry Pressures for Efficiency
With increasing pressures for faster drug development and data analysis, efficiencies in data management become critical.

### Collaboration
As pharma becomes more interdisciplinary, tools that can bridge the knowledge gap and provide self-service capabilities in demand.

### Demand for Real-time Analysis
The need for instant insights and real-time data analysis drives the adoption of more intuitive tools.

### Integration of AI/ML Capabilities
Web based platforms increasingly offer AI/ML integration
OUR JOURNEY: EMBRACING THE GRAPHICAL APPROACH

Identifying the Needs
- Interactivity
- Simplicity
- Low-code Env.
- Excel like Specs in the web

Flexible Programming Backbone
- Choose R
- Open-source benefits
- Meet complex needs

Design a modern Web-App
- Scalable in cloud
- Programmable in browser
- Responsive Design

Programming Flexibility
- Functions
- Standard Nodes
- Custom Nodes
- Custom Scripts

Productivity/Ease of Use*
- Automations, validations
- AI integration
- Define.xml and compliance

*To be ready by end of 2023
USE CASE: SDTM

Data In
Read Source Data

Mapper
Map variable, test-level, & value-level data in an excel like structure, embedded with R-code snippets for granular mapping control

Finalize
Split, names, labels, sort, order variables, trim, basic checks per CDISC expectation

Data Out
Save datasets to target data libraries
## USE CASE: SDTM SPECIFICATIONS

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Source</th>
<th>Target</th>
<th>Destination</th>
<th>ID Variable</th>
<th>Function</th>
<th>Params</th>
<th>Label</th>
<th>Type</th>
<th>Map Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>VS1</td>
<td>CNO</td>
<td>CNO</td>
<td>SUPP</td>
<td>VISITNUM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VS1</td>
<td>F_STATUS</td>
<td>STATUS</td>
<td>SUPP</td>
<td>VSDTC</td>
<td></td>
<td></td>
<td>Status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VS1</td>
<td>PAGE</td>
<td>PAGE</td>
<td>CO</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VS1</td>
<td>PATNO</td>
<td>PATNO</td>
<td>NA</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VS1</td>
<td>PNO</td>
<td>PNO</td>
<td>NA</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VS1</td>
<td>REC_ID</td>
<td>REC_ID</td>
<td>NA</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VS1</td>
<td>R_DRUG</td>
<td>R_DRUG</td>
<td>NA</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VS1</td>
<td>SCTR</td>
<td>SCTR</td>
<td>NA</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VS1</td>
<td>TAREA</td>
<td>TAREA</td>
<td>NA</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VS1</td>
<td>VSDAT</td>
<td>VSDTC</td>
<td>conv2DTC</td>
<td>&quot;sasdt&quot;: &quot;df$VSDAT&quot;,&quot;sastm&quot;: &quot;df$VSTIM&quot;</td>
<td></td>
<td></td>
<td>Date/Time of Measurements</td>
<td>C</td>
<td>ISO 8801</td>
</tr>
<tr>
<td>VS1</td>
<td>EVENT_ID</td>
<td>VISITNUM</td>
<td>unschVisit</td>
<td>&quot;date&quot;: &quot;df$VSDTC&quot;</td>
<td></td>
<td></td>
<td>Visit Number</td>
<td>N</td>
<td>VISITNUM</td>
</tr>
<tr>
<td>VS1</td>
<td>VSDY</td>
<td>VSDY</td>
<td>NA</td>
<td>NA</td>
<td></td>
<td></td>
<td>Study Day of Vital Signs</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>VS1</td>
<td>VSTIM</td>
<td>VSTIM</td>
<td>NA</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VS1</td>
<td>VSYN</td>
<td>VSYN</td>
<td>NA</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VS1</td>
<td>STUDYID</td>
<td>STUDYID</td>
<td>Value</td>
<td>&quot;code&quot;: &quot;NIM-01&quot;</td>
<td></td>
<td></td>
<td>Study Identifier</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>VS1</td>
<td>DOMAIN</td>
<td>DOMAIN</td>
<td>Default</td>
<td>&quot;code&quot;: &quot;paste('NIM-01', df$SCNO, df$PATNO, sep =&quot;-&quot;&quot;)&quot;</td>
<td></td>
<td></td>
<td>Domain Abbreviation</td>
<td>C</td>
<td>DOMAIN</td>
</tr>
<tr>
<td>VS1</td>
<td>USUBJID</td>
<td>USUBJID</td>
<td>Value</td>
<td>&quot;code&quot;: &quot;paste('NIM-01', df$SCNO, df$PATNO, sep =&quot;-&quot;&quot;)&quot;</td>
<td></td>
<td></td>
<td>Unique Subject Identifier</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>VS1</td>
<td>VSEQ</td>
<td>VSEQ</td>
<td>Default</td>
<td>&quot;code&quot;: &quot;paste('NIM-01', df$SCNO, df$PATNO, sep =&quot;-&quot;&quot;)&quot;</td>
<td></td>
<td></td>
<td>Sequence Number</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>VS1</td>
<td>VTEST</td>
<td>VTEST</td>
<td>Default</td>
<td>&quot;code&quot;: &quot;paste('NIM-01', df$SCNO, df$PATNO, sep =&quot;-&quot;&quot;)&quot;</td>
<td></td>
<td></td>
<td>Vital Signs Test Name</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>VS1</td>
<td>VTESTCD</td>
<td>VTESTCD</td>
<td>NA</td>
<td>Default</td>
<td>&quot;code&quot;: &quot;paste('NIM-01', df$SCNO, df$PATNO, sep =&quot;-&quot;&quot;)&quot;</td>
<td></td>
<td></td>
<td>Result or Finding in Origin</td>
<td>C</td>
</tr>
<tr>
<td>VS1</td>
<td>VSORRES</td>
<td>VSORRES</td>
<td>Default</td>
<td>&quot;code&quot;: &quot;paste('NIM-01', df$SCNO, df$PATNO, sep =&quot;-&quot;&quot;)&quot;</td>
<td></td>
<td></td>
<td>Result or Finding in Origin</td>
<td>C</td>
<td></td>
</tr>
</tbody>
</table>
USE CASE: GRAPHICAL PATIENT PROFILE

- **Plug-n-play**: Plug a SDTM dataset to feed into graphical patient profiles
- No programming background needed
- The output dataset, feeds into graphical patient profile
USE CASE: GRAPHICAL PATIENT PROFILE

Subject ID: 100014, Age: 66, Sex: F, Race: White

Milestones:
- Visit 1
- Visit 2
- Visit 3
- Visit 10

AE:
- Diarrhoea
- Haemorrhoids
- Headache
- Vomiting
- Blisters

CHEMISTRY:
- Bilirubin
- Blood Urea Nitrogen
- Glucose
- Vitamin B12
- Vitamin B9
- LANSOPRAZOLE
- MIDAZOLAM
- PANTOPRAZOLE SODIUM
- PROCHLORPERAZINE EDISYLATE

CONCOMITANT MEDICATIONS:
- QTcB - Bazett's Correction
- QTcF - Fridericia's Correction
- Summary (Mean) PR Duration
- Summary (Mean) QRS Duration

Vitamin B9
- Date: Oct 6, 2003
- 52.8 nmol/L HIGH

H/L/Abnormal
USE CASE: RBQM

- Data-In squeezed into * icon on top of node
- Plug-n-play KRIs design
- No programming background needed
USE CASE: RBQM
USE CASE: ADAM CREATION

- ADaM programming harder to standardize
- For now, all the ADAE programming in a single node
- For efficiency, code can be created offline and pasted in the web
CONCLUSION

One will not turn to graphical data flow for show, but for the promise of better workflow!

• Lower-level nodes in other applications such as join, sort and so on… are not convenient to meet complex clinical data analysis needs

• Flexible Scripting node needed for complex custom programming

• Sophisticated high-level nodes can facilitate:
  • Low code environment
  • Plug-n-play solutions
  • Self-service environment
  • Rapid Prototyping
THANK YOU

Q & A