MERCK METADATA REPOSITORY:

BUSINESS-BASED VALUE IN AN MDR

A. Brooke Hinkson
Manori Turmel
Karl Konrad

PhUSE Connect Conference, Raleigh NC, 4-6 June 2018
Keep your mind open to change all the time. Welcome it. Court it. It is only by examining and reexamining your opinions and ideas that you can progress.

— Dale Carnegie —
Business Problems to Address

Current information management practices are not sufficient or scalable

My Wish List:

**Technology-enabled information management business processes**

- Enable informed business decision making capabilities regarding the definition, use, maintenance and sharing of clinical info.

**Standards and flexibility**

- Leverage industry standards to implement well-defined data definitions that are harmonized across the clinical lifecycle
- Apply “fit for purpose” data definitions into a trial

**Model driven information management**

- Information is modeled, shared, and used to drive automation of processes

**A foundation on which future processes & technologies can be applied**

- Enabling a true collaboration between machines and subject matter experts
Consistently define, use, and reuse information globally through the use of standards to achieve process and scientific excellence by:

**Define**
- Ensuring information transparency, quality, compliance, & traceability
  - Quality by Design
  - Leverage industry standards
  - Reuse/consistency
  - Flexible use of standards
  - Drive process automation

**Use**
- Streamlining business processes associated with the definition, collection, integration, analysis and use of clinical data
  - Insight into actual study designs: where and how information was used
  - Traceability beginning-to-end across info. lifecycle and applications
  - Metrics on use/reuse & compliance

**Share**
- Facilitating the exchange of information with partners and customers
  - Merck internal stakeholders
  - External collaborators and partners
  - Health Authorities

**Maintain**
- Meeting existing regulatory guidelines and preparing for future regulatory, eHealth & life science industry standards
  - Contemporary industry versions
  - Maintain multiple versions of standards
  - Curation of standards
  - Change Control and impact assessment

Define → Use → Share → Maintain →
Transition from spreadsheets to managing data definitions within a semantic-based metadata repository (MDR)

“Master Specification” Context

– **Scope**: Basic data collection information to support EDC build, SDTM mappings (for multiple IG versions), and data definitions to support the generation of our Data Management Workbench (DMW) review models

– **Includes**:
  - Within a workbook: A minimum of 8 worksheets; number increases with each form version
  - Within a worksheet: Numerous columns (e.g., Columns A – CG) and rows (e.g., over 550)

– **The Master Specification represents one data definition spreadsheet**
  - One Master Specification for each unique eCRF
  - Over 20 standard components today
    – 11 spreadsheets representing different information have to be harmonized and fed into DMW
  - Will reduce the number of standard components as we transitions info. into the MDR
Benefits of a Semantics-based Approach

Enabling new business capabilities

• **Ability to rapidly ingest data / definitions from disparate sources**
  • Graph model that represents information & related items

• **Versioning capabilities**
  • Item level versioning

• **Linked items**
  • Impact assessment of proposed changes
  • Visual traceability of the information flow

• **Flexibility to easily extend / modify the model**
  • Change is a constant: Science is rapidly advancing, industry data standard are evolving, interfacing technologies are changing, and we’re still learning!
Implementation - Release Strategy

• Frequent, Incremental improvement with each release
• Allows the organization to grow with the features of the tool
• Allows the tool to adapt to the changing organization
Change Management - Stakeholders

Data Use
- Investigator Data Entry
- Transactional
- Safety Reporting
- Data Review
- SDTM Conversion
- Analysis Reporting

Data Model
- InForm Items
- Reporting Dataset
- Review Model 1
- Review Model 2
- Review Model 3
- SDTM

MDR Visual Navigator

Form
- Form DOV v1
- Form Section Date of Visit v1
- DOVPHASE v1

RD Dataset
- RD_DOVEM v1
- DOVPHASEIT v1
- DOVPHASEIT v1

RM1 Dataset
- RM_DOV v1
- DOV_PHASE v1
- DOV_PHASE v1

RM2 Dataset
- RM_DoV v1
- DOV_PHASE v1
- DOV_PHASE v1

RM3 Dataset
- RM3_DoV _SV v1
- PHASE v1

SDTM Dataset
- SDTM Dataset _SV v1 (3.1.1)
Change Management - Communication

- Process Walk-through
- Facilitated Workshop
- 1-1 SME consultation
- Mandatory Training
- PMO Office
- User Demo Forums
- Word of mouth
- How-to Videos
- Department Meetings
- Public Awareness Sessions
- Community SharePoint Space
- User Manuals and Guides
Visualize Information Flow: SDTM Up-versioning Example

- When comparing the SDTM definitions across two different SDTM versions, it’s not obvious to see what has changed
- Once information is loaded in an MDR, the data traceability is clear
  - This is represented across 4 worksheets in the Master Spec
- In the example below, the Date/Time of assessment is collected on the eCRF, goes through several data transformations, and in SDTM v3.1.1, it goes to the CF domain, while in SDTM v3.1.3, it goes to the FA domain
Visual Navigator - CMENDTCL

Start at SDTM 3.1.3 Item:

Start at RM3 Item:

Start at RM2 Item:

Start at Form Item:
Semantic Links: Library Migration “Integrated” (Legacy) to “PDAM” (Next-Gen)

**PDAM Data Flow for VS02TPT:**

<table>
<thead>
<tr>
<th>Library</th>
<th>Item Name</th>
<th>Question</th>
<th>Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integrated</td>
<td>VSTPT02ESDE</td>
<td>Scheduled Time</td>
<td>Predose 1 hour post dose 2 hours post dose 5 hours post dose 12 hours post dose</td>
</tr>
<tr>
<td>PDAM</td>
<td>VS02TPT</td>
<td>Scheduled Time</td>
<td>Predose 1 hour post dose 2 hours post dose 5 hours post dose 12 hours post dose</td>
</tr>
</tbody>
</table>

Semantically, VS02TPT and VSTPT02ESDE are equivalent. Thus, the data flow of the migrated Integrated variable is the same as the PDAM variable, and you can re-use the data flow to VSTPT and VSTPTNUM in SDTM.
Near-term Business Value

- **Cross-company collaborations**
  - Understand similarities, and be able to link related items between the two companies
  - Once a link is established, then the data can be better understood by using the existing relationships, programs, etc.

- **Metadata Access for All**
  - Utilize a Data connector to reach into the MDR from other tools

- **Coordinate metadata handoffs to downstream systems**
  - SDTM programming is not performed in the MDR
  - Mapping definitions are described in the MDR
  - Utilize the versioning and mapping instructions in MDR to select which standard programs to execute for SDTM programming

- **Coordinated Roadmap Discussions**
  - Seems to be an endless number of expectations for an MDR
  - By using smaller, more frequent releases, the MDR can adapt to the current business need
THANK YOU