Next Generation Data Management: Case Studies for Ontology-driven, Graph Databases

Jim LaPointe - Managing Director, Life Sciences & Healthcare

PhUSE US Connect 2019
2019-Feb-25
Agenda

• Clinical Data Management Challenges

• A Graph-based, Ontology-driven Data Management Technology
  – Basics
  – CDISC Implementation

• Ontology-driven Use Cases
  – Highlights
  – Lessons

• Summary
Challenge 1 – Clinical Data Standards Evolve Over Time

Source: CDISC International Interchange Presentation, 2017
Challenge 2 – Consistency Across Disparate Clinical Systems

CDISC Updates over time

- **Content Standards**
- **Technical Standards**
- **Semantics**
- **Therapeutic Areas**

- **Alzheimer**
- **Virology**
- **PKD**
- **Asthma**

- **Pain**
- **Devices**
- **Parkinson’s Disease**
- **Tuberculosis**

- **ODM v1.3.1**
- **Define.xmi v2.0**
- **BRIDG v3.1**
- **SDTM v1.3 IG v3.1.3**
- **PrM Toolkit v1.0**
- **COASH v1.1**
- **SEND v3.0**
- **BRIDG v3.1**

- **Rheumatoid Arthritis**
- **Kidney Transplant**
- **Hepatitis C**
- **Diabetic Kidney Disease**

- **Commercial Analytics**
- **Real World Data (RWD)**
- **Epi Analytics & HEOR**

©2019 Cambridge Semantics Inc. All rights reserved.
Addressing These Clinical Data Management Challenges

In 2015, PhUSE and CDISC introduced RDF-based Foundational Clinical Data Standards

- World Wide Web Consortium (W3C) standards technology, Feb. 2004
- RDF – Resource Description Framework
- OWL – Ontology Web Language

Requirements

- authoritative source of unambiguous clinical data standards
- flexibility to support sponsor-unique standards
- extensible to enable future standards to evolve
- machine-readable for interoperability
RDF Defines Data Context in a Machine-readable Format
Resource Definition

http://rdf.cdisc.org/std/sdtmig-3-1-2#Column.AE.AEOUT

- http://
- rdf.cdisc.org/std/sdtmig-3-1-2
- Column.AE.AEOUT

Namespace Identifier (shorthand notation)

- xmlns:sdtmig-3-1-2#="http://rdf.cdisc.org/std/sdtmig-3-1-2#
- sdtmig-3-1-2:Column.AE.AEOUT

Graph Structure, Resource - Relationship (Triple Notation)

```
<table>
<thead>
<tr>
<th>Subject</th>
<th>Predicate</th>
<th>Object (value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>sdtmig-3-1-2:Column.AE.AEOUT</td>
<td>mms:dataElementName</td>
<td>&quot;AEOUT&quot;</td>
</tr>
<tr>
<td>sdtmig-3-1-2:Column.AE.AEOUT</td>
<td>mms:dataElementLabel</td>
<td>&quot;Outcome of Adverse Event&quot;</td>
</tr>
<tr>
<td>sdtmig-3-1-2:Column.AE.AEOUT</td>
<td>mms:dataElementValueDomain</td>
<td>sdtmct:C66768</td>
</tr>
<tr>
<td>mms:dataElementName</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mms:dataElementLabel</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mms:dataElementValueDomain</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

mms: dataElementName

"AEOOUT"

"Outcome of Adverse Event"

sdtmct: C66768
RDF Defines Data Context in a Machine-readable Format

Source: CDISC Standards in RDF Reference Guide, Version 1.0 Final
Use Case 1 – Integrating Clinical Data Across Different Systems

Highlights

- Apply CDISC’s Operational Data Model (ODM) as the harmonizing ‘umbrella’ data model
- Extend the Quantum clinical MetaData Repository (MDR)
- Integrate to Oracle Central Designer, InForm Electronic Data Capture (EDC), Data Management Workbench (DMW) & SAS macros
- Conduct multi-system Impact Analyses for metadata changes

Lessons

- XML-based data structures led to complex ontologies & mappings
- Monitoring focus, not Governance
Use Case 2 - An RDF-based Clinical Metadata Manager

Highlights

• Migrate 1,000s of existing metadata specification worksheets
• Harmonize prior worksheets into a consistent & versioned metadata data model (ontology)
• Support a metadata curation & governance process
• Integrate to Oracle Central Designer, InForm & DMW
• Support Impact Analysis for metadata changes

Lessons

• Metadata specification quality issues
• Complex mappings across specifications and to CDISC standards
Use Case 3 - RDF-based Integrated Clinical Analysis

Highlights

• Implement a Therapeutic Area (TA) specific ontology
• Harmonize nearly 100 studies for cross-trial (meta-analysis) analytics
• Automate the ingestion and mapping of new studies into the TA ontology
• Integrate to SAS / R for complex derivations, e.g., Kaplan-Meier
• Develop key ‘screening’ dashboards & visualizations
• Integrate to Spotfire

Lessons

• AWS infrastructure for scale
• Complex conformance pipeline
Use Case 4 – An RDF-based Protocol Schedule of Activities (SoA)

**Highlights**

- Implement an early version of CDISC’s Protocol Representation Model (PRM)
- Extend the PRM model to capture SoA concepts and properties
- Define a simple visit entry mechanism
- Integrate to standard activity specification worksheets
- Capture visit – activity special conditions / constraints
- Investigate linking standard activities to standard electronic Case Report Forms (eCRF)

**Lessons**

- Protocol documents were too inconsistent to parse
- Investigate Common Protocol Template (CPT) potential
Use Case 5 – RDF-based Real World Data (RWD) Analytics

Highlights

• Harmonize 4 studies using an ADaM-based ontology for Diabetes
• Develop a non-standard Electronic Medical Record (EMR) ontology by Natural Language Processing (NLP) techniques from a data dictionary
• Create a canonical ontology for clinical & EMR data
• Develop key ‘screening’ dashboards & visualizations
• Integrate to Shiny R for simulation

Lessons

• AWS infrastructure for scale
• Loose integration for ‘big data’ scale
Use Case 5 – RDF-based Real World Data (RWD) Analytics

4 Clinical Trials: N=4,681, HbA1c=29,012 labs

EMRs: N=3,611,202, HbA1c=420,087 labs

©2019 Cambridge Semantics Inc. All rights reserved.
Use Case 5 – RDF-based Real World Data (RWD) Analytics

Hypoglycemia by Concomitant Medication

Hypoglycemia by Age & Medical History
Next Generation Clinical Data Management – Linked Data
Graphs & Ontologies Support Key Clinical Data Management Objectives

- Consistency with current CDISC foundational standards to promote data reuse
- Relate past CDISC foundational standards to current standards for maximum reuse and to support traceability
- A flexible, extensible clinical and healthcare data model based on harmonized data concepts and definitions
- Data governance process support for data model and data item changes
- Automated data transformation for interoperability amongst disparate clinical systems (vendor agnostic)
- Legacy clinical systems support while enabling future technology adoption
- User role security to protect confidential and sensitive data
- Scale to support global business volume
- Enable new use cases, such as impact analyses and enterprise-wide controlled terminology
- Support business transformation initiatives driven by data
Summary

Conclusions

• RDF and OWL has proven to be a foundational technology for numerous clinical solutions
• CDISC’s RDF-based data standards are key to accelerating clinical solution deployment
• New RDF-based data standards, e.g., Electronic Medical Records (EMR from HL7 FHIR) and scientific data (Allotrope Foundation), broaden the potential scope of sharing clinical data

Recommendation

• Continue sharing experiences of prior RDF-based & ontology-driven use cases and solutions
• Create a full lifecycle clinical data ontology based on RDF-OWL within a CDISC-HL7 collaboration
• Urge data management solution vendors to adopt RDF & OWL based data structures or create Application Programming Interfaces (API) to speed deployments and promote interoperability
Can RDF scale? - Google’s Linked Data Knowledge Graph

London

Image

Map

Name (EN)

administrativeArea

Area

Description

Elevation

Population
Anzo Platform®

The leading platform for building an Ontology-based Data Fabric

Contact: Jim LaPointe
Email: jtlapointe@cambridgesemantics.com