

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

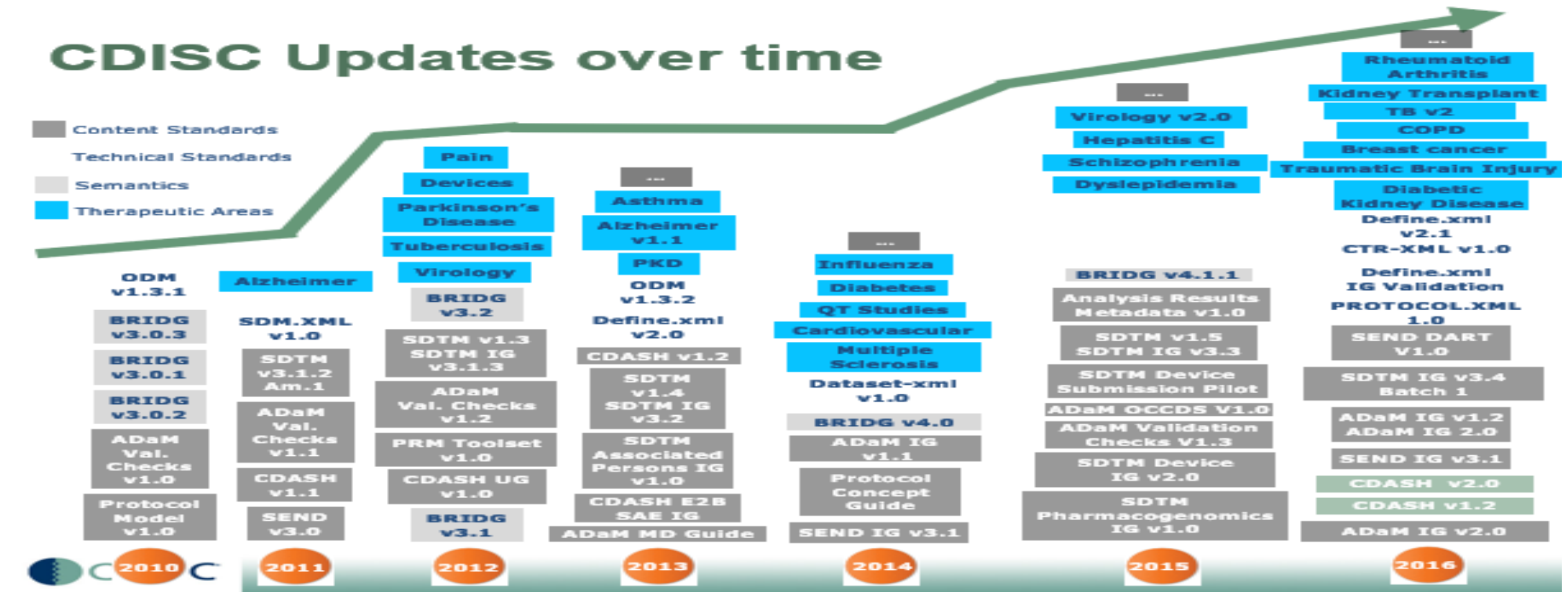
Jim LaPointe - Managing Director, Life Sciences & Healthcare

PhUSE US Connect 2019
2019-Feb-25

Agenda

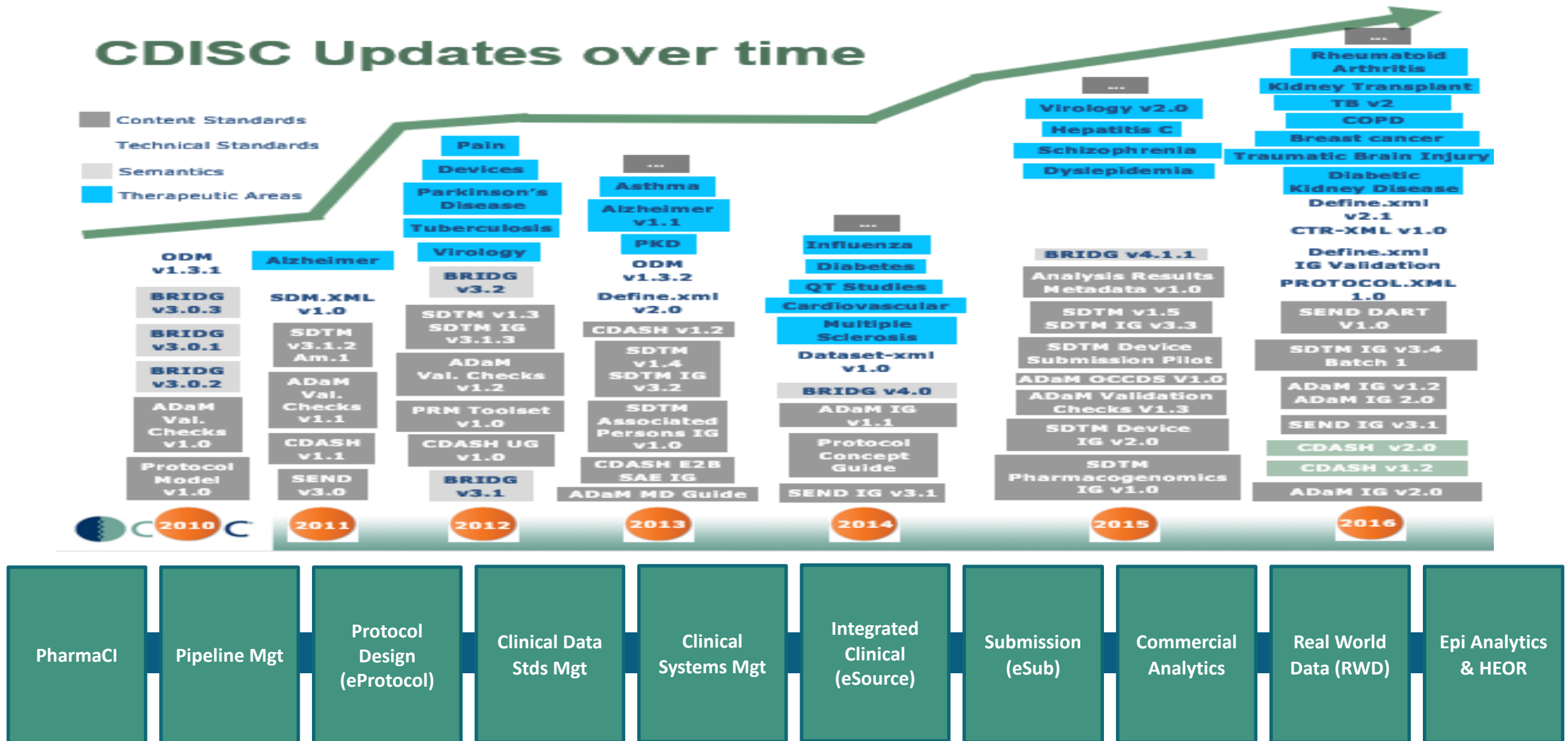
- Clinical Data Management Challenges
- A Graph-based, Ontology-driven Data Management Technology
 - Basics
 - CDISC Implementation
- Ontology-driven Use Cases
 - Highlights
 - Lessons
- Summary

Challenge 1 – Clinical Data Standards Evolve Over Time



Source: CDISC International Interchange Presentation, 2017

Challenge 2 – Consistency Across Disparate Clinical Systems



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

CDISC Standards in RDF Reference Guide



CDISC Standards in RDF Reference Guide
Version 1.0 Final

Prepared by
The PhUSE CS Semantic Technology Working Group

Notes to Readers

This document is the specification of the RDF vocabulary and the RDF data sets that represent the current foundational CDISC standards in a W3C endorsed semantic format.

Revision History

Date	Version	Summary of Changes
2014-12-12	Draft 1.0	Version 1.0 for public comment.
2015-06-18	Final 1.0	Version 1.0 Final

© 2015 Clinical Data Interchange Standards Consortium, Inc. All rights reserved. Page 1
Version 1.0 Final June 18, 2015

RDF Defines Data Context in a Machine-readable Format

Jack London
American novelist



John Griffith London was an American activist. A pioneer in the world of commercial fiction, he was one of the first writers to become a world fortune from writing. Wikipedia

Born: January 12, 1876, San Francisco, CA
Died: November 22, 1916, Glen Ellen, CA
Short stories: To Build a Fire, The Law of Life, The White Silence, MORE
Education: University of California, Berkeley (1896–1897), Oakland High School (1896)

Quotes [View 7+ more](#)

The proper function of man is to live, not to exist. I shall not waste my days in trying to prolong them. I shall use my time.

You can't wait for inspiration. You have to go after it with a club.

I would rather be a superb meteor, every atom of me in magnificent glow, than a sleepy and permanent planet.

Books [View 45+ more](#)



People also search for [View 15+ more](#)

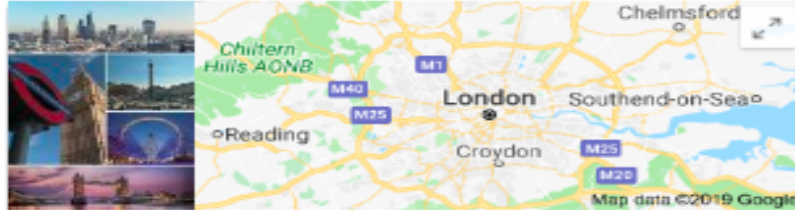


Mark Twain Robert Louis Stevenson Ernest Hemingway Charmian London Charles Dickens

<http://schema.org/familyName>

London
?

<http://schema.org/City>



London
Capital of England

London, the capital of England and the United Kingdom, is a 21st-century city with history stretching back to Roman times. At its centre stand the imposing Houses of Parliament, the iconic 'Big Ben' clock tower and Westminster Abbey, site of British monarch coronations. Across the Thames River, the London Eye observation wheel provides panoramic views of the South Bank cultural complex, and the entire city.

Area: 607 mi²
Elevation: 36'
Population: 8.136 million (2011) United Nations
Local time: Wednesday 8:13 PM
Weather: 46°F (8°C), Wind S at 6 mph (10 km/h), 81% Humidity

Plan a trip

[London travel guide](#)

3-star hotel averaging \$105, 5-star averaging \$293

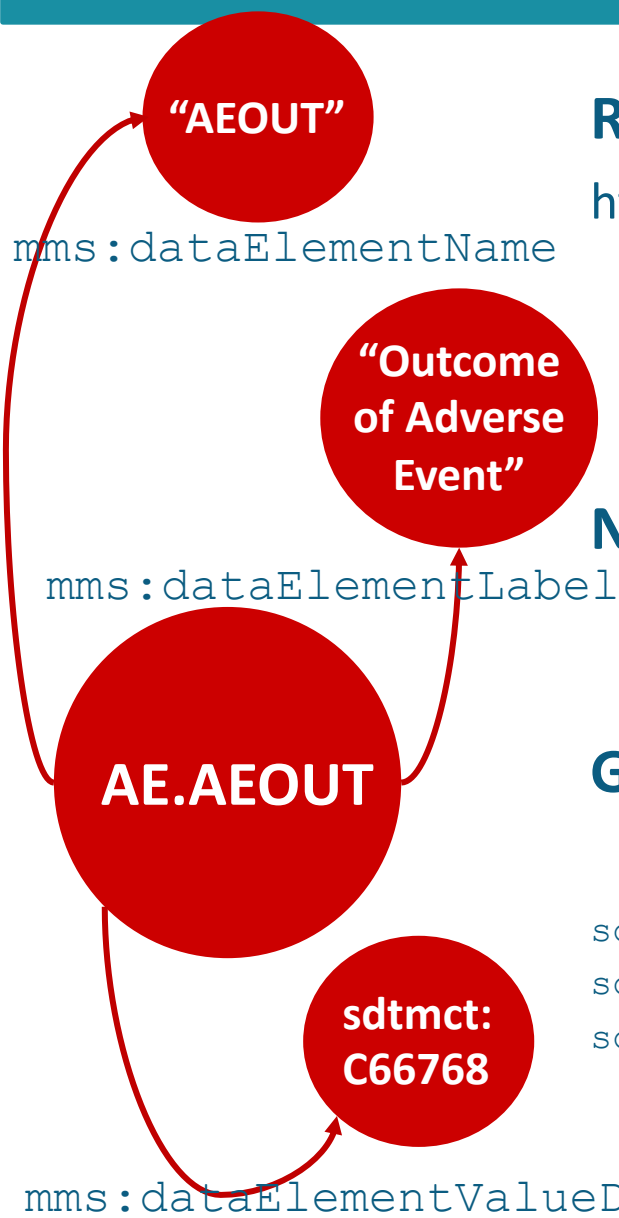
[College London, MORE](#)
2nd-most-populous European city by population (25,001). wikipedia.org

People also search for [View 15+ more](#)



England Paris United Kingdom New York City Berlin

CDISC RDF Basics ...



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`

standard internet communications protocol
authoritative source path name (IRI or URI)
resource fragment identifier

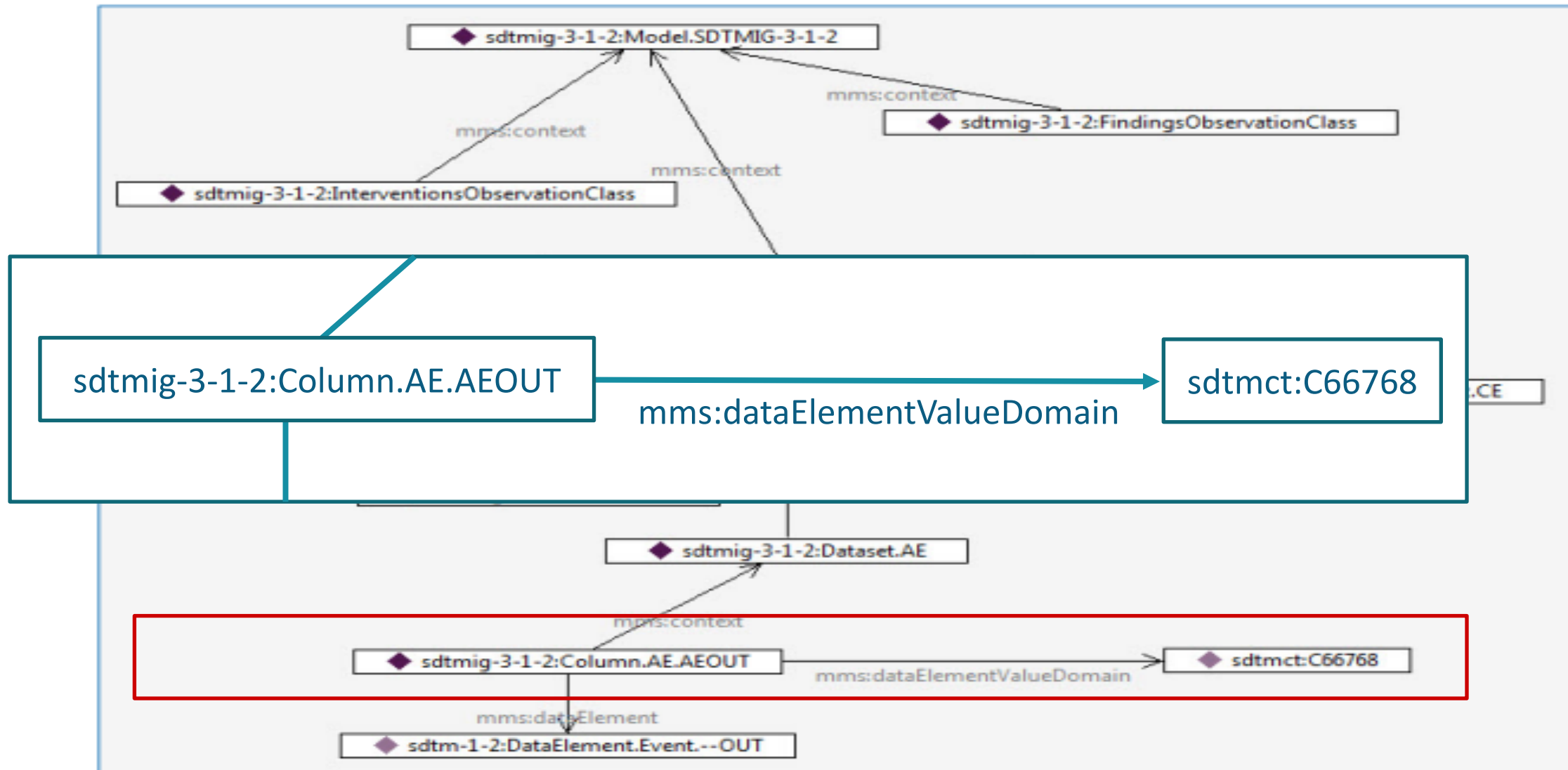
Namespace Identifier (shorthand notation)

- `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)

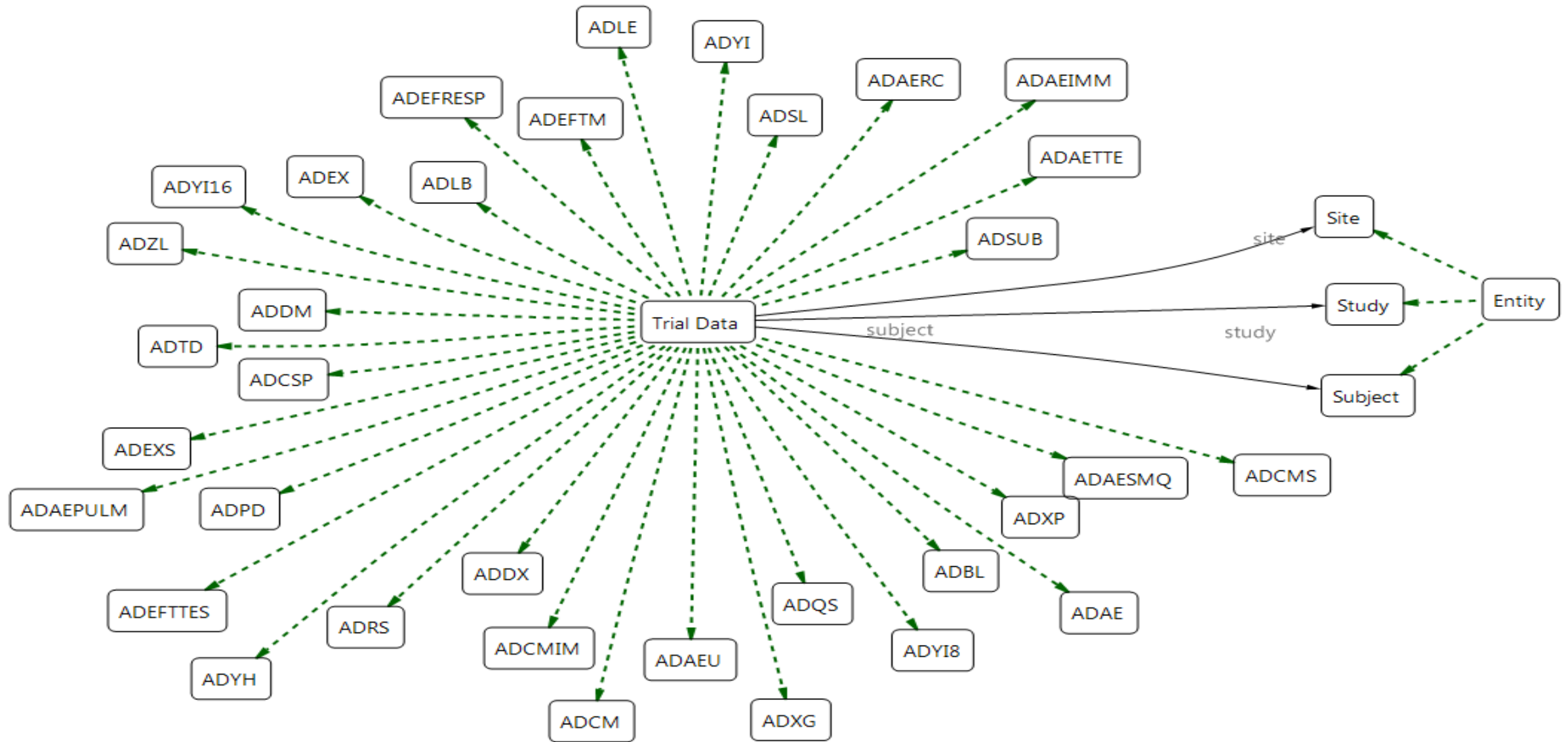
Subject	→ Predicate	→ Object (value)
<code>sdtmig-3-1-2:Column.AE.AEOUT</code>	<code>mms:dataElementName</code>	<code>"AEOUT"</code>
<code>sdtmig-3-1-2:Column.AE.AEOUT</code>	<code>mms:dataElementLabel</code>	<code>"Outcome of Adverse Event"</code>
<code>sdtmig-3-1-2:Column.AE.AEOUT</code>	<code>mms:dataElementValueDomain</code>	<code>sdtmct:C66768</code>

RDF Defines Data Context in a Machine-readable Format

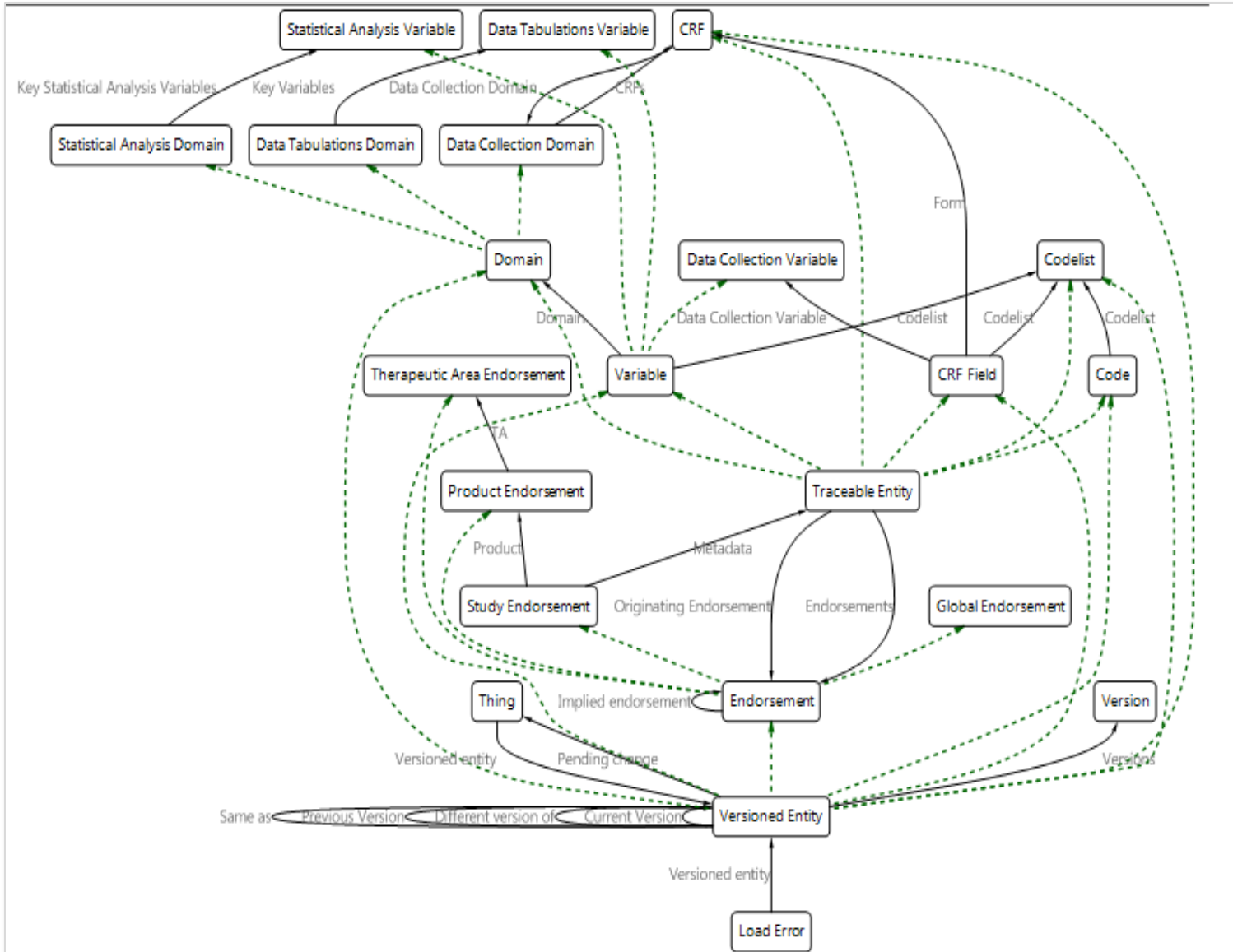


Source: CDISC Standards in RDF Reference Guide, Version 1.0 Final

RDF & Ontology Driven Use Cases



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

ANZOQ Metadata Management Logout

Filter menu items

Home

Consultation Requests >

Metadata Maintenance >

DMRACE - Form Item

Scope: 0.0 v1 Production Wed Oct 31 2018

Metadata Comments(0)

Select Another Form Item

Item RefName (Required)	Question Text
DMRACE	Race
Field Type (Required)	Field Format Length
Radio Button	
Mandatory/Optional	Required/Not Required
Mandatory	Required
Display Override	SDV Required
	true
SDV Critical	DB Value - CODELISTREF eVOC Dictionary
false	race (CDISC RACE)

V4.0.5.1201809281436 ANZOQ

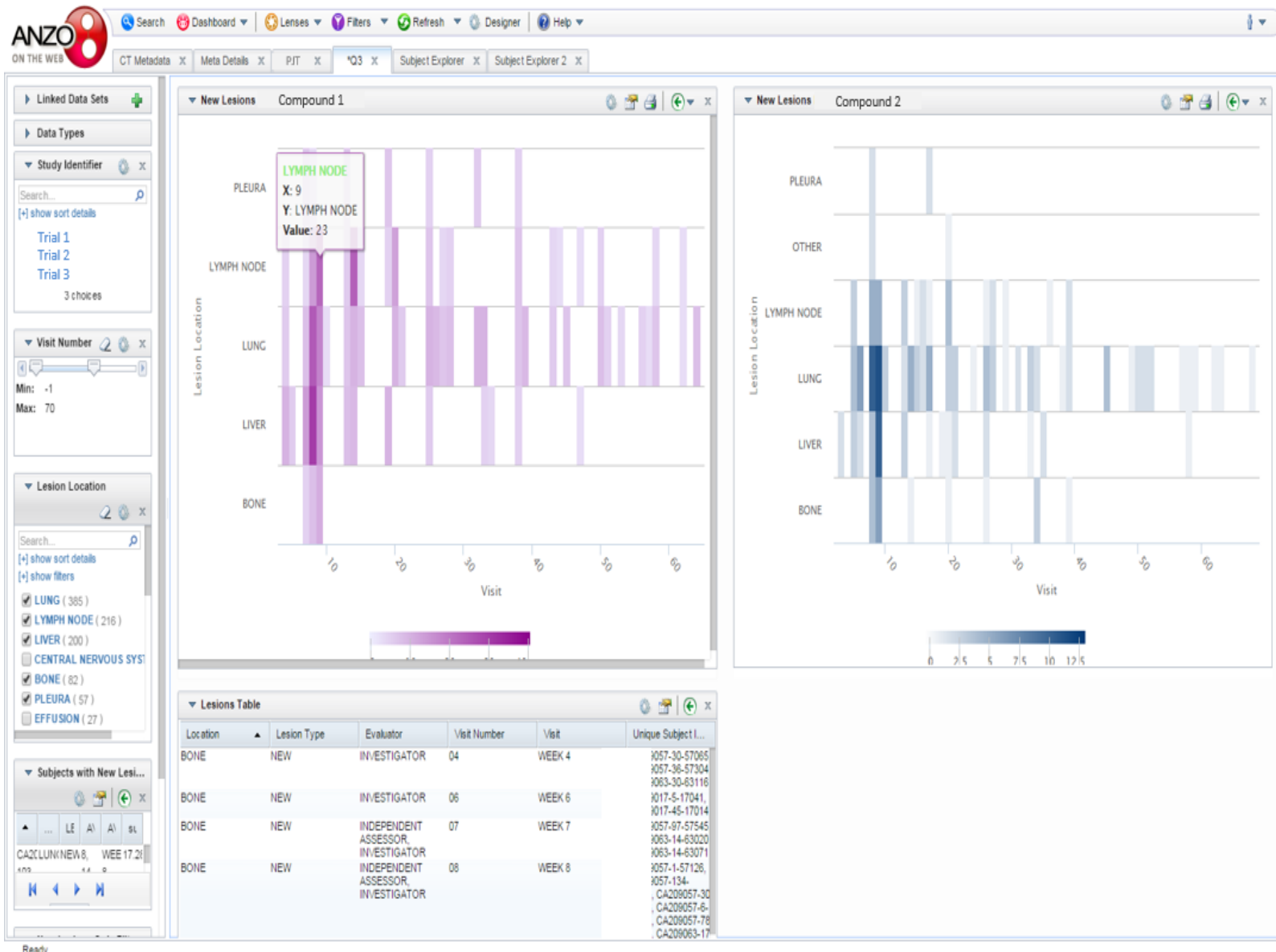
Highlights

- 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)

Activity	Visit																													
	01 Screening Visit	02 Practice Tests1	03 Practice Tests2	04 Practice Tests3	05 Randomization Visit	06 Treatment Visit 1	07 Treatment Visit 2	08 Treatment Visit 3	09 Treatment Visit 4	10 Treatment Visit 5	11 Treatment Visit 6	12 Treatment Visit 7	13 Treatment Visit 8	14 Treatment Visit 9	15 Treatment Visit 10	16 Treatment Visit 11	17 Treatment Visit 12	18 Treatment Visit 13	19 Treatment Visit 14	20 Treatment Visit 15	21 Treatment Visit 16	22 Treatment Visit 17	23 Treatment Visit 18	24 Treatment Visit 19	25 Treatment Visit 20	26 Treatment Visit 21	27 Treatment Visit 22	28 Treatment Visit 23	29 End of Treatment Visit 24	30 End of Study Visit 25
Vital Signs	x				x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Visual Function Test		x	x	x	x			x			x			x			x			x			x			x				x
Visual Analog Scale for Global Impression of Well-being					x			x			x			x			x			x			x			x				x
Urine Pregnancy Test					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Urinalysis	x				x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
SF-36 Health Survey					x						x						x													x
Serum Pregnancy Test	x																													
RNA Analysis					x			x									x						x							x
Randomization					x																									
Physical Examination	x				x						x						x						x							x
Expanded Physical Examination	x				x						x						x						x							x
Expanded Sample					x																									
Expanded Physical Examination					x												x													x
Expanded Physical Examination		x	x	x	x			x			x			x			x			x			x			x				x
MRI Scan	x										x						x													x
Expanded MRI scan	x										x						x													x
Medical History	x																													
Lipid Profile	x				x	x	x	x	x	x	x			x			x			x			x			x				x
Informed Consent	x										x			x			x			x			x			x				
Hematology	x				x	x	x	x			x			x			x			x			x			x				x
Expanded Disability Status Scale	x				x			x						x			x			x			x			x				x
European Quality of Life, 5-dimensions Questionnaire					x						x						x													x
ECG (12 – lead)	x				x	x	x	x			x			x			x			x			x			x				x
Dispense Study Drug					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Dispense drug					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Complete Physical Exam	x				x																									x
Blood Chemistry	x				x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

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

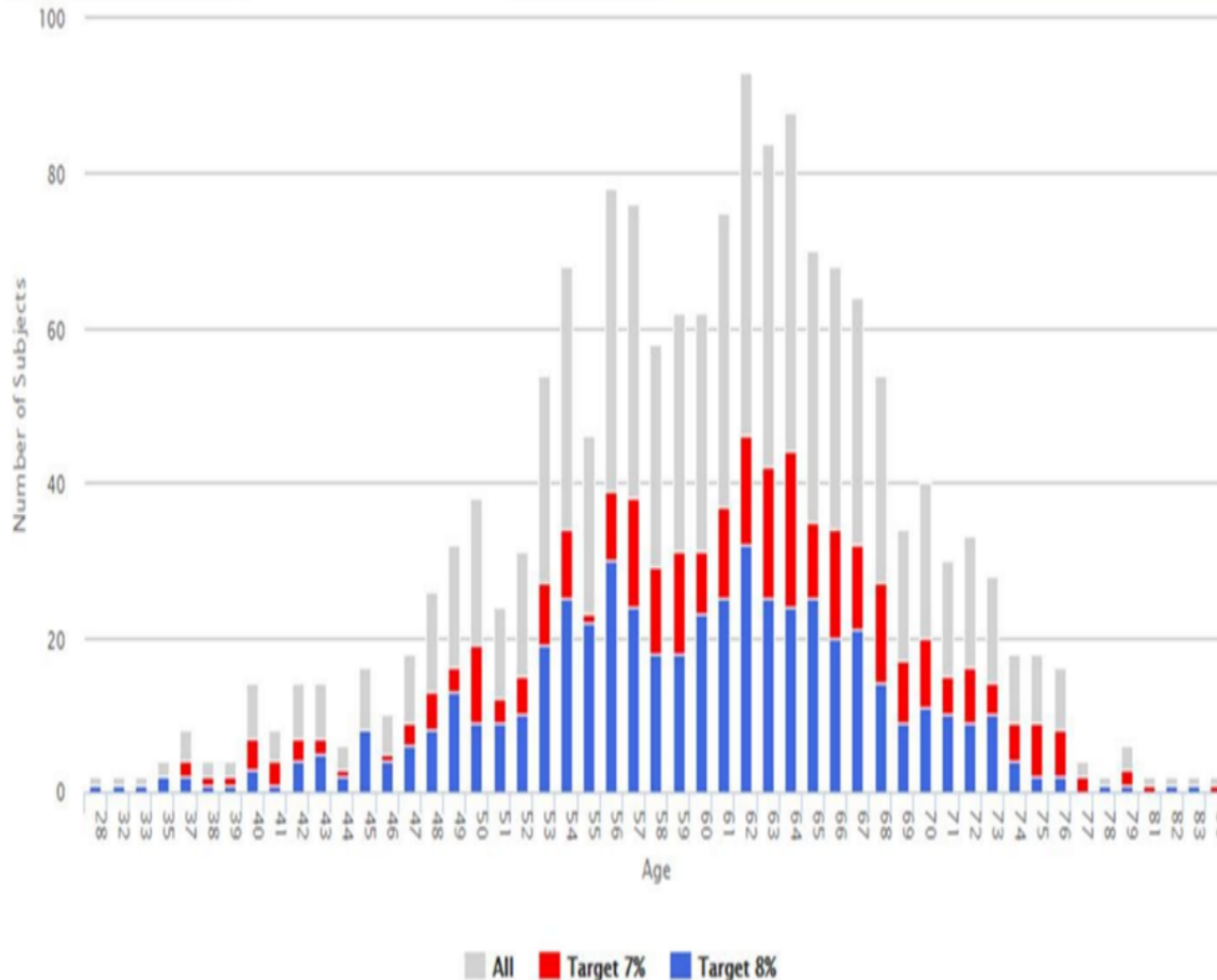
[illegible]

- 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

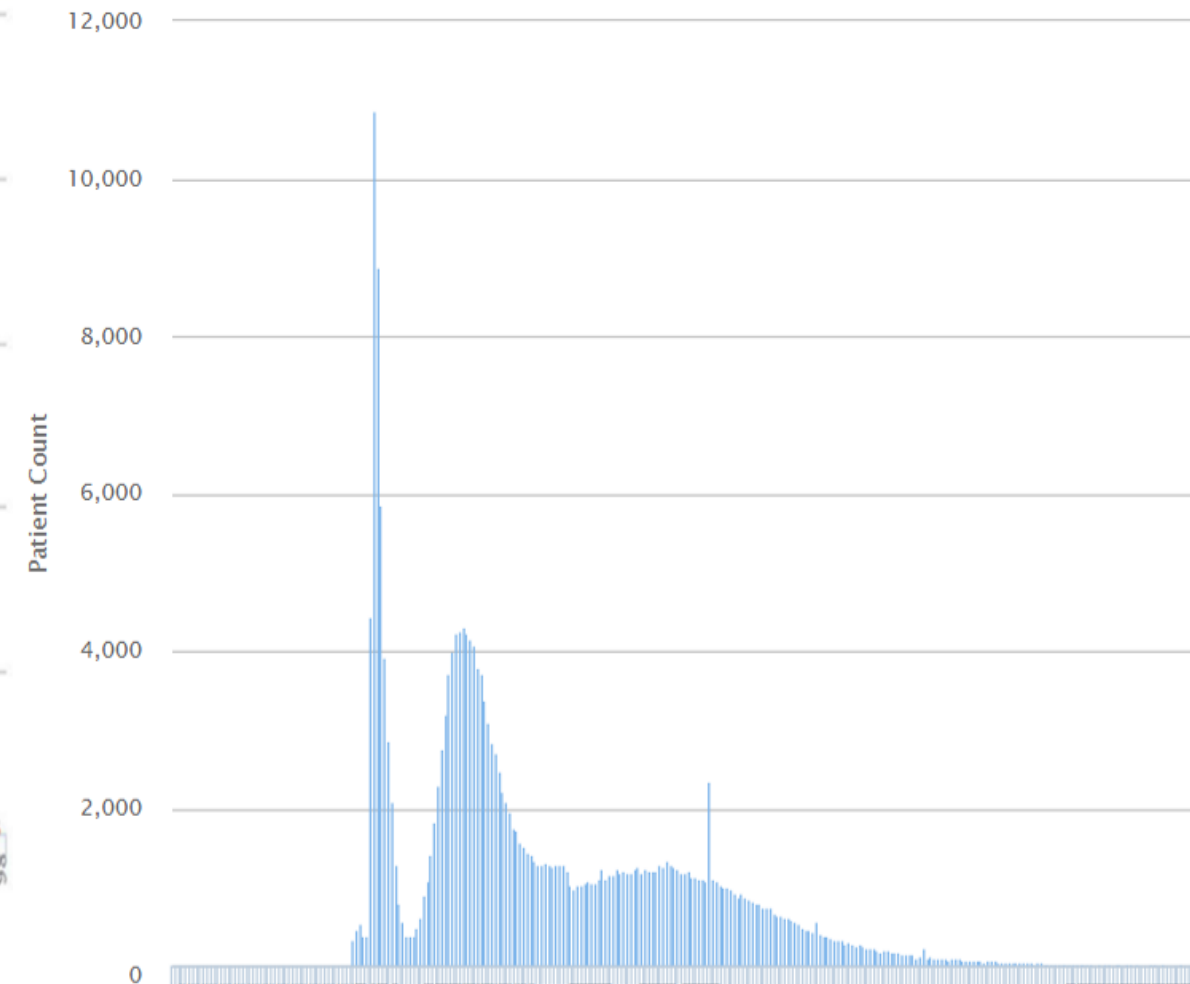
- AWS infrastructure for scale
- Loose integration for 'big data' scale

Use Case 5 – RDF-based Real World Data (RWD) Analytics

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

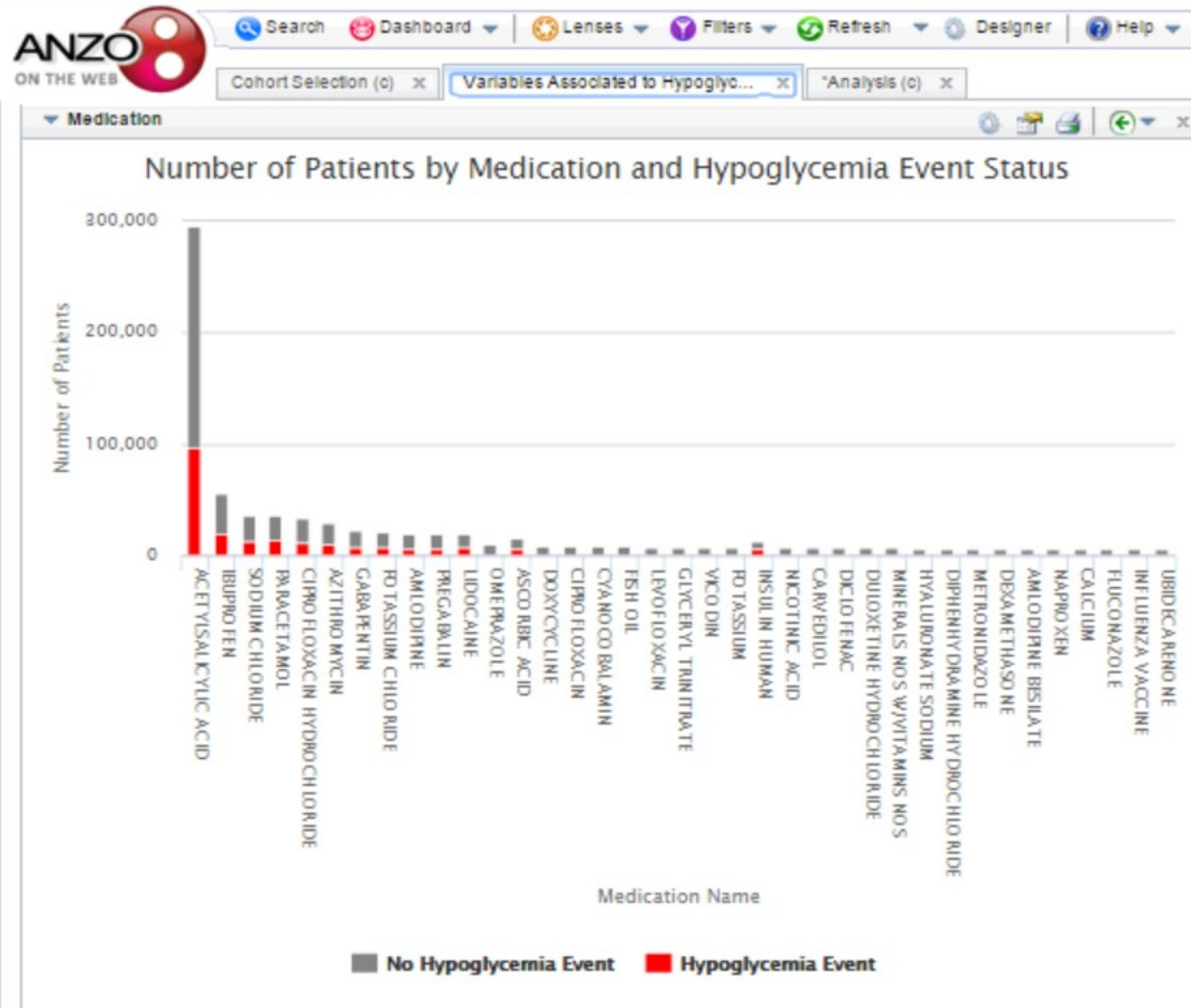


EMRs: N= 3,611,202, HbA1c=420,087 labs

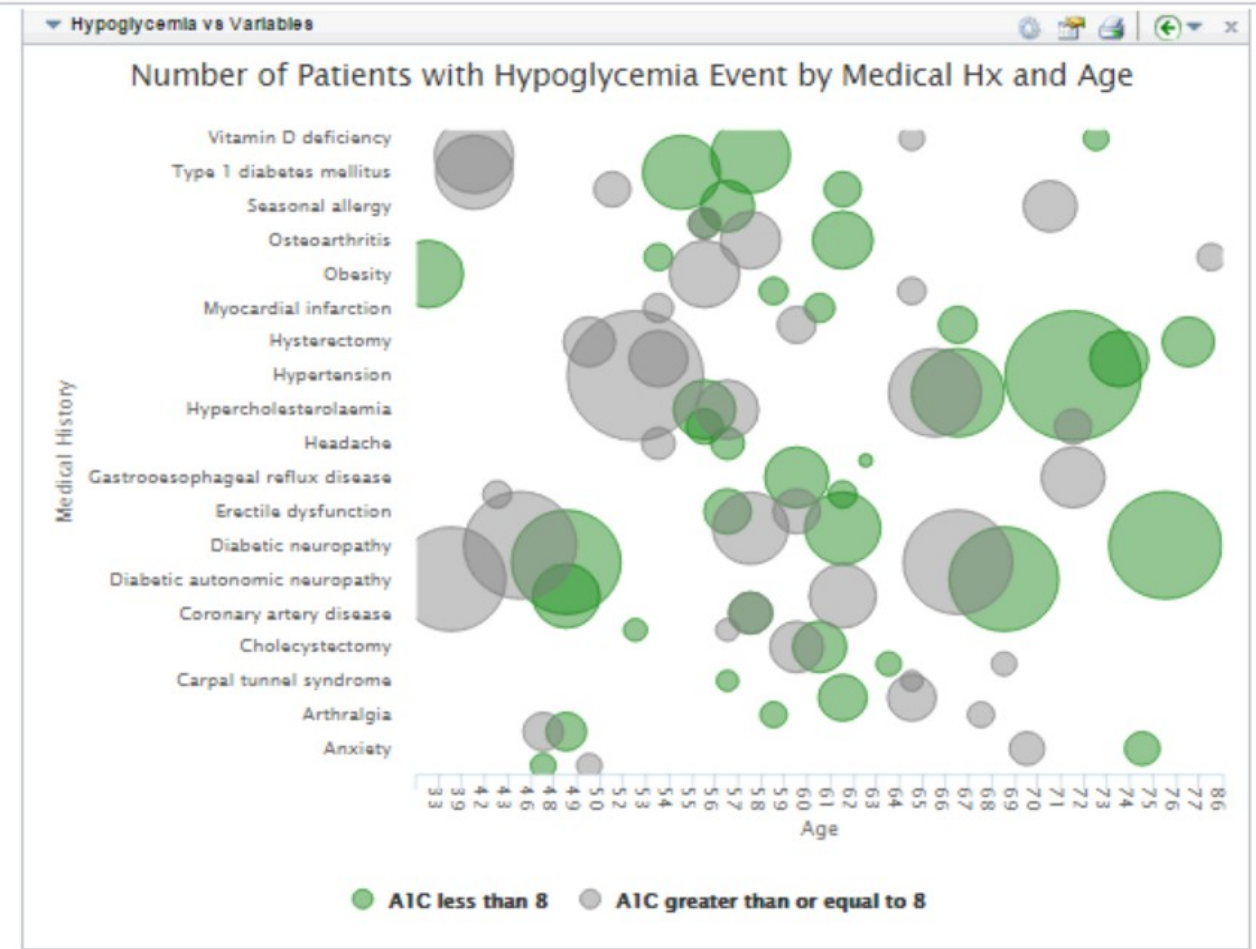


Use Case 5 – RDF-based Real World Data (RWD) Analytics

Hypoglycemia by Concomitant Medication



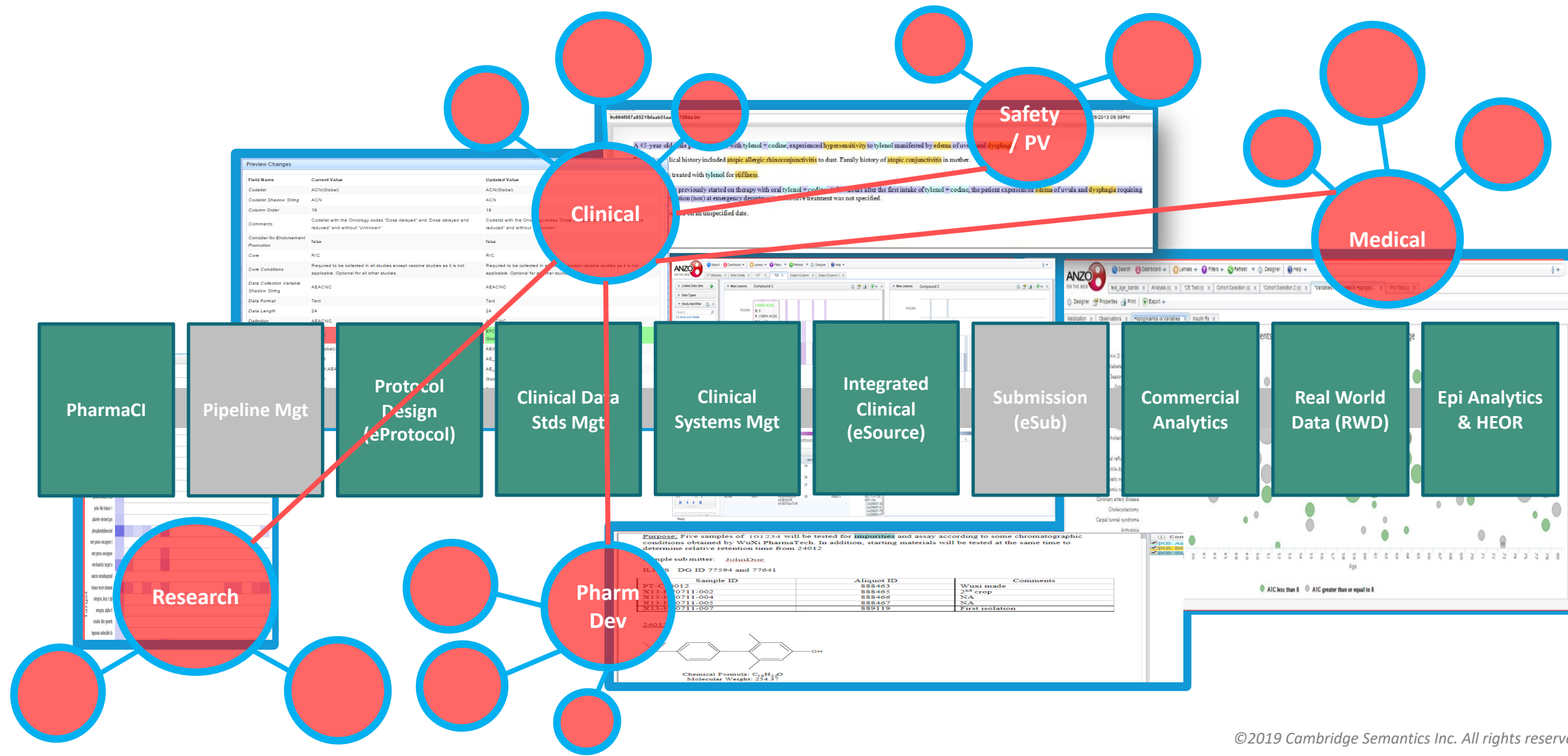
Hypoglycemia by Age & Medical History



Next Generation Clinical Data Management – Linked Data



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

Google

london

All Maps News Images Videos More Settings Tools

About 4,760,000,000 results (1.22 seconds)

Things to do in London



Big Ben
London's iconic national timepiece



Buckingham Palace
Home of the British Queen & State



Coca-Cola London Eye
Iconic riverside observation wheel



Tower of London
Medieval castle housing the Crown

[London travel guide](#)

Top stories



London knife crime: 'I don't know how many people I've stabbed'

BBC
1 hour ago



Failed London Garden Bridge project cost £53m

BBC
7 hours ago



Waterloo delays: 'Utter pandemonium' at major London station after trespasser on tracks...

Evening Standard
2 hours ago

[More for london](#)

London - Wikipedia

<https://en.wikipedia.org/wiki/London>

London is the capital and largest city of both England and the United Kingdom. Standing on the River Thames in the south-east of England, at the head of its ...

History of London · City of London · London (disambiguation) · Greater London

Welcome to London - visitlondon.com

<https://www.visitlondon.com/>

Welcome to Visit London, your official city guide to London, England. Find things to do in London, days out in London, London attractions and sightseeing, ...

[Tickets & offers](#) · [Things to do](#) · [Accommodation](#) · [Traveller information](#)

London

Image

Map

Name (EN)

administrativeArea

Description

Area

Elevation

Population

London, the capital of England and the United Kingdom, is a 21st-century city with history stretching back to Roman times. At its centre stand the imposing Houses of Parliament, the iconic 'Big Ben' clock tower and Westminster Abbey, site of British monarch coronations. Across the Thames River, the London Eye observation wheel offers panoramic views of the South Bank cultural complex, and the city's skyline.

Area: 607 mi²
Elevation: 36'
Population: 8.136 million (2011)
Local time: Wednesday 10:52 PM
Weather: 41°F (5°C), Wind S at 4 mph

Plan a trip

[London travel guide](#)

3-star hotel averaging \$105, 5-star averaging \$293

6 h 20 min flight, from \$490

Colleges and Universities: Imperial College London, MORE

Did you know: London is the third-most-populous European city by population within city limits (8,825,001). [wikipedia.org](https://www.wikipedia.org)

People also search for

[England](#) [Paris](#) [United Kingdom](#) [New York City](#) [Berlin](#)

[View 15+ more](#)

[Feedback](#)

Anzo Platform®

The leading platform for building an Ontology-based Data Fabric



End-to-end



Open Standards



Enterprise Scale

Contact: Jim LaPointe

Email: jtlapointe@cambridgesemantics.com