PharmaSUG Japan: CDISC Library Update

Presented by CDISC Data Science Team
October 24, 2019
Agenda

1. CDISC Library High-Level Technical Overview
2. CDISC Library Demo
CDISC Library High-Level Technical Overview
CDISC Library, an ISO 11179 applied Metadata Repository, built on semantic architecture using the Resource Description Framework (RDF). It uses linked data and a REST API to deliver CDISC standards metadata to software applications that automate standards-based processes. CDISC Library provides access to a substantially increased number of versioned CDISC standards and controlled terminology packages, including new relationships between the standards. CDISC Library is the single, trusted, authoritative source of CDISC standards metadata and represents a new way of creating, maintaining, and publishing this metadata.

In non-technical terms, it means that CDISC Library will remain flexible and adaptive for the future. It will continue to provide the key metadata described to support your organizational needs.
Accessibility Possibilities
API System Diagram

- MDR
- BRIDG
- NCI
- Others

RDF Models

RDF Triple Store

API Config.

API Engine

API Endpoint

Git Repository

RDF & JSON Git Client

API Client
API Access Overview

System/Application #1

System/Application #2

System/Application #...

API Engine

Query Engine

SPARQL Endpoint

RDF Triple Store
API Access Overview (cont.)

End User → Authentication → Process Request → CDISC Library

I want to get SDTMIG v3.2 standards metadata

Request confirmed, getting SDTMIG v3.2 standards metadata
# SDTM Implementation Guide (SDTMIG)

<table>
<thead>
<tr>
<th>Method</th>
<th>Path</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GET</td>
<td>/mdr/sdtmig/{version}</td>
<td>Get SDTMIG Product</td>
</tr>
<tr>
<td>GET</td>
<td>/mdr/sdtmig/{version}/classes</td>
<td>Get SDTMIG Class List</td>
</tr>
<tr>
<td>GET</td>
<td>/mdr/sdtmig/{version}/classes/{class}</td>
<td>Get SDTMIG Class</td>
</tr>
<tr>
<td>GET</td>
<td>/mdr/sdtmig/{version}/classes/{class}/datasets</td>
<td>Get SDTMIG Class Dataset List</td>
</tr>
<tr>
<td>GET</td>
<td>/mdr/sdtmig/{version}/datasets</td>
<td>Get SDTMIG Dataset List</td>
</tr>
<tr>
<td>GET</td>
<td>/mdr/sdtmig/{version}/datasets/{dataset}</td>
<td>Get SDTMIG Dataset</td>
</tr>
<tr>
<td>GET</td>
<td>/mdr/sdtmig/{version}/datasets/{dataset}/variables</td>
<td>Get SDTMIG Dataset Variable List</td>
</tr>
<tr>
<td>GET</td>
<td>/mdr/sdtmig/{version}/datasets/{dataset}/variables/{var}</td>
<td>Get SDTMIG Dataset Variable</td>
</tr>
<tr>
<td>GET</td>
<td>/mdr/root/sdtmig/datasets/{dataset}/variables/{var}</td>
<td>Get Root SDTMIG Dataset Variable</td>
</tr>
</tbody>
</table>

---

**Term**
- **operation**: Method to be performed. GET is used for retrieving information.
- **scheme**: The transfer protocol of the API.
- **host**: The host (name or IP) serving the API.
- **basePath**: The base path on which the API is served, which is relative to the host.
- **endpoint**: The available paths and operations for the API.
- **parameter**: Defines the specific content to render.
API Taxonomy: Top Level

CDISC Library API
library.cdisc.org/api

Search
/mdr/search

List Searchable Fields
/mdr/search/scopes

Scope Value List
/mdr/search/scopes/(scope)

Search Query
/mdr/search?q=(valid search syntax)

Search Query with Scopes
/mdr/search?q=(valid search syntax)&(scope1)=(value1)&(scopeN)=(valueN)

Standards Products
/mdr/products

Data Collection
/mdr/products/DataCollection

- CDASH Model
  /mdr/cdash/(version)
- CDASHIG
  /mdr/cdashig/(version)

Data Tabulation
/mdr/products/DataTabulation

- SDTM Model
  /mdr/sdtm/(version)
- SDTMIG
  /mdr/sdtmig/(version)
- SENDIG
  /mdr/sendig/(version)

Data Analysis
/mdr/products/DataAnalysis

- ADeM
  /mdr/adem/(product)

Terminology
/mdr/products/Terminology

- Controlled Terminology
  /mdr/ct/packages

Controlled Terminology
/mdr/ct/packages
API Taxonomy: SDTM/SDTMIG
API Taxonomy: CT

CDISC Library API
library.cdisc.org/api

Controlled Terminology Codelists Version List (Root)
/mdr/root/ct/{package-type}/codelists/{codelist}

Controlled Terminology Packages
/mdr/ct/packages

Controlled Terminology Package
/mdr/ct/packages/{package}

Controlled Terminology Package Codelists
/mdr/ct/packages/{package}/codelists

Controlled Terminology Package Codelist Details
/mdr/ct/packages/{package}/codelists/{codelist}

Controlled Terminology Package Codelists Term List
/mdr/ct/packages/{package}/codelists/{codelist}/terms

Controlled Terminology Package Codelists Term Details
/mdr/ct/packages/{package}/codelists/{codelist}/terms/{term}
What ways can CDISC Library API be accessed?

- **API Tools:** Postman, Insomnia, SoapUI, Katalon, etc.
- **Systems:** MDRs, CDMS, CTMS, EDCs, etc.
- **Applications:** Validators, Editors, ETLs, etc.
- **Macros/Code:** SAS, Python, Julia, R, Go, Rust, XML, etc.
- **Web Browser:** FireFox, Chrome, IE, Opera, Brave, Safari, etc.

And possibly others not described above.
ElasticSearch

Search Request

Indexes

SDTM → ....
SDTMIG → ....
SENDIG → ....
CDASH → ....
CDASHIG → ....
ADaM → ....
CT → ....

Search Results
ElasticSearch (cont.)

- **Words**: (race AND ethnicity) OR age
- **Phrases**: “One record” OR “per subject”
- **Specify Field**: Label=“Study Identifier”
- **Wildcard**: AGE* OR *DTC
- **RegEx**: Description: /s?dtmig(ath[en]if)/
- **Proximity (Edit Distance of 3)**: “supplemental”\(^3\)
- **Date Range**: date:[2019-09-09 TO 2020-01-09]
- **Date Before**: date:{* TO 2020-01-09}
- **Number Range**: ordinal:[6 TO 10]
- **Number Range Exclusive**: ordinal:[1 TO 5]
- **Number Ranges Unbounded**: ordinal:>8

And more…
Hypermedia

Richardson Maturity Model

Level 0: Single URI & Single Verb
Level 1: Multiple URI Resources & Single Verbs
Level 2: Multiple URI Resources & Verbs
Level 3: HATEOAS

Lowest Maturity | Highest Maturity

Most Common

Slide Note: *Reference: https://restfulapi.net/richardson-maturity-model/
Hypermedia (cont.)

```
{  "ordinal": "28",  "name": "AEACN",  "label": "Action Taken with Study Treatment",  "description": "Describes changes to the study treatment as a result of the event. AEACN is for actions unrelated to dose adjustments of study treatment. Examples of AEACN include:
- DOSE REDUCED, DOSE INCREASED, DOSE NOT CHANGED, UNKNOWN or NOT APPLICABLE.",  "role": "Record Qualifier",  "simpleDatatype": "Char",  "core": "Exp",  "_links": {    "self": "#/mdr/sdtmig/3-2/datasets/AE/variables/AEACN",    "title": "Action Taken with Study Treatment",    "type": "SDTM Dataset Variable"  },  "codelist": [  ]},  "type": "Class",  "priorVersion": {    "href": "/mdr/sdtmig/3-1-3/datasets/AE",    "title": "Adverse Events",    "type": "SDTM Dataset"
}
```
Computer science

From Wikipedia, the free encyclopedia

"Computer sciences" redirects here. For the American corporation, see Computer Sciences Corporation.

Computer science is the study of processes that interact with data and that can be represented as data in the form of programs. It enables the use of algorithms to manipulate, store, and communicate digital information. A computer scientist studies the theory of computation and the practice of designing software systems.\(^1\)

Its fields can be divided into theoretical and practical disciplines. Computational complexity theory is highly abstract, while computer graphics emphasizes real-world applications. Programming language theory considers approaches to the description of computational processes, while computer programming itself involves the use of programming languages and complex systems. Human–computer interaction considers the challenges in making computers useful, usable, and accessible.

Not to be confused with information science.

Data science is a multi-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from data in various forms, both structured and unstructured,\(^12\) similar to data mining.

Data science is a “concept to unify statistics, data analysis, machine learning and their related methods” in order to “understand and analyze actual phenomena” with data.\(^9\) It employs techniques and theories drawn from many fields, including mathematics, statistics, information science, and computer science.

Turing award winner Jim Gray imagined data science as a “fourth paradigm” of science (empirical, theoretical, computational and now data-driven) and asserted that “everything about science is changing because of the impact of information technology” and the data deluge.\(^16\)

In 2012, when Harvard Business Review called it “The Sexiest Job of the 21st Century,”\(^5\) the term “data science” became a buzzword. It is now often used interchangeably with earlier concepts like business analytics,\(^7\) business intelligence, predictive modeling, and statistics. Even the suggestion that data science is sexy was paraphrasing Hans Rosling, featured in a 2011 BBC documentary with the quote, “Statistics is now the sexiest subject around.”\(^4\) Nate Silver referred to data science as a sixed up term for statistics.\(^9\) In many cases, earlier approaches and solutions are now simply rebranded as “data science” to be more attractive, which can cause the term to become “illiterate beyond usefulness.”\(^10\) While many university programs now offer a data science degree, there exists no consensus on a definition or suitable curriculum contents.\(^7\) To its discredit, however, many data-science and big-data projects fail to deliver useful results, often as a result of poor management and utilization of resources.\(^11\)
Evolution of Media Types

What future media types can we expect?

- **ODM-XML** (Operational Data Model as XML)
- **YAML** (YAML Ain't Markup Language)
- **JSON-LD** (JavaScript Object Notation for Linked Data)
- **TOML** (Tom’s Obvious, Minimal Language)
- **CSON** (CoffeeScript Object Notation)

Maybe others….
CDISC Library Demo

• Sample RDF Graph in CDISC Library
• Hypermedia Representation from REST API
• Navigation within Data Standards Browser
Access CDISC Library now!

CDISC Membership Inquiry: https://www.cdisc.org/register
CDISC Library: https://www.cdisc.org/cdisc-library
Thank You!