Implementing the CDISC Library API in software applications: first experiences

Jozef Aerts
XML4Pharma
Phuse Connect Amsterdam
November 2019
What is the CDISC Library API?

- It is a huge metadata repository (MDR) ultimately containing almost all of the CDISC standards

- Machine-readable

- Queryable using RESTful web services
What are RESTful Web Services? What is an API?

• RESTful web services are services running over the internet (usually HTTP/HTTPS)

• GET / POST / PUT / POST / DELETE

• Return structured information formatted as XML, JSON, RDF-Turtle, …

• Can usually be tested in the browser, but are meant to be used by applications

• The API describes the available methods and how to run them
Example in browser


```xml
<datasetVariable>
    <ordinal>5</ordinal>
    <name>RFSTDTC</name>
    <label>Subject Reference Start Date/Time</label>
    <description>
        Reference Start Date/time for the subject in ISO 8601 character format. Usually equivalent to date/time when subject was first exposed to study treatment. See Assumption 9 for additional detail on when RFSTDTC may be null.
    </description>
    <role>Record Qualifier</role>
    <simpleDatatype>Char</simpleDatatype>
    <core>Exp</core>
    <describedValueDomain>ISO 8601</describedValueDomain>
</datasetVariable>
```
HATEAOS principles

- Response contains links to parent and children and related information
  Example: LB - LBTESTCD (SDTM-IG 3.3)
HATEAOS principles
LBTESTCD Codelist

- CodeLists have … versions
HATEAOS principles
LBTESTCD Codelist 2018-12-21

• CodeLists versions have ... terms

```xml
<term>
  <conceptId>C100461</conceptId>
  <submissionValue>ALMICG</submissionValue>
  <definition>
    A measurement of the alpha-1 microglobulin in a biological specimen.
  </definition>
  <preferredTerm>Alpha-1 Microglobulin Measurement</preferredTerm>
  <synonyms>Alpha-1 Microglobulin; Protein HC</synonyms>
</term>
```

• And terms have a
  • Submission value
  • Definition
  • Preferred term
  • Synonyms ...

```xml
<term>
  <conceptId>C100429</conceptId>
  <submissionValue>ALAGLP</submissionValue>
  <definition>
    A measurement of the alpha-1 acid glycoprotein in a biological specimen.
  </definition>
  <preferredTerm>Alpha-1 Acid Glycoprotein Measurement</preferredTerm>
  <synonyms>Alpha-1 Acid Glycoprotein</synonyms>
</term>
```

```xml
<term>
  <conceptId>C80167</conceptId>
  <submissionValue>A1ANTRYP</submissionValue>
  <definition>
    A measurement of the alpha-1 antitrypsin in a biological specimen.
  </definition>
  <preferredTerm>Alpha-1 Antitrypsin Measurement</preferredTerm>
</term>
```
What is currently covered?

• Controlled Terminology
• CDASH
• CDASH-IG
• SDTM
• SDTM-IG
• SEND-IG
• ADaM

Coming: TAUGs (Therapeutic Area User Guides), Code tables (ValueLists), Questionnaires, Conformance rules …
Implementation in Software

• RESTful web services are extremely easy to implement in modern software

• "How to" examples for different computer languages on the Wiki: https://wiki.cdisc.org/display/LIBSUPRT/Getting+Started%3A+Programmatically+connect+to+CDISC+Library+API
  - Python
  - SAS & R
  - Java
  - Coming: XSLT, XQuery
Implementation in Software: Java
(Using Jersey-2 libraries)

- CDISC Library requires "Basic Authentication"
  - This needs to be taken into account in applications

```java
private String base = "https://library.cdisc.org/api"; // the base of the service
private String username = "my.name@mycompany.com"; // your username
private String pass = "xxxxxxxxxxxx"; // your passkey

private Client client;

public BasicCDISCLibraryClient() {
  ClientConfig clientConfig = new ClientConfig();
  // see e.g. https://jersey.github.io/documentation/latest/logging_chapter.html for how
  // clientConfig.property(LoggingFeature.LOGGINGFEATURE_VERBOSITY_CLIENT, LoggingFeature
  client = ClientBuilder.newClient(clientConfig);
  // Basic Authentication
  HttpAuthenticationFeature feature = HttpAuthenticationFeature.basic(username, pass);
  client.register(feature);
}
```
Implementation in Software: Java (Using Jersey-2 libraries)

- Example: get SDTM Variable Properties, with parameters "IG-version", "domain", "variable"

```java
public String getSDTMVariableProperties(String igVersion, String domain, String variable) {
    WebTarget webTarget = client.target(base).path("mdr").path("sdtmig").path(igVersion).path("datasets").path(domain).path("variables").path(variable);
    Invocation.Builder invocationBuilder = webTarget.request(MediaType.APPLICATION_XML);
    // Response requires javax.annotation-api-1.2.jar, hk2-api-2.5.0-b42.jar, hk2-locator-2.5.0-b42.jar, hk2-utils-2.5.0-b42.jar
    // and javax.inject-1.jar, javax.inject-2.5.0-b42.jar
    // and jersey-media-json-binding.jar
    // from the Jersey-2 library - see https://jersey.github.io/
    Response response = invocationBuilder.get();
    System.out.println("HTTP Response Status = " + response.getStatus());
    String xml = response.readEntity(String.class);
    //System.out.println(xml);
    return xml;
}
```
Implementation in Open Source Software: Java example: "Smart Submission Dataset Viewer"

<table>
<thead>
<tr>
<th>VSTESTCD</th>
<th>VSTEST</th>
<th>VSPOS</th>
<th>VSORRES</th>
<th>VSORRESU</th>
<th>VSSTRESN</th>
<th>VSSTRESU</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIABP</td>
<td>Diastolic BP...</td>
<td>SUPINE</td>
<td>64</td>
<td>mmHg</td>
<td>64</td>
<td>mmHg</td>
</tr>
<tr>
<td>DIABP</td>
<td>Diastolic BP...</td>
<td>STANDING</td>
<td>83</td>
<td>mmHg</td>
<td>64</td>
<td>mmHg</td>
</tr>
<tr>
<td>DIABP</td>
<td>Diastolic BP...</td>
<td>STANDING</td>
<td>57</td>
<td>mmHg</td>
<td>64</td>
<td>mmHg</td>
</tr>
<tr>
<td>DIABP</td>
<td>Diastolic BP...</td>
<td>STANDING</td>
<td>68</td>
<td>mmHg</td>
<td>64</td>
<td>mmHg</td>
</tr>
<tr>
<td>DIABP</td>
<td>Diastolic BP...</td>
<td>STANDING</td>
<td>59</td>
<td>mmHg</td>
<td>64</td>
<td>mmHg</td>
</tr>
<tr>
<td>DIABP</td>
<td>Diastolic BP...</td>
<td>STANDING</td>
<td>71</td>
<td>mmHg</td>
<td>64</td>
<td>mmHg</td>
</tr>
<tr>
<td>DIABP</td>
<td>Diastolic BP...</td>
<td>SUPINE</td>
<td>56</td>
<td>mmHg</td>
<td>64</td>
<td>mmHg</td>
</tr>
<tr>
<td>DIABP</td>
<td>Diastolic BP...</td>
<td>STANDING</td>
<td>61</td>
<td>mmHg</td>
<td>64</td>
<td>mmHg</td>
</tr>
<tr>
<td>DIABP</td>
<td>Diastolic BP...</td>
<td>STANDING</td>
<td>67</td>
<td>mmHg</td>
<td>64</td>
<td>mmHg</td>
</tr>
<tr>
<td>DIABP</td>
<td>Diastolic BP...</td>
<td>SUPINE</td>
<td>56</td>
<td>mmHg</td>
<td>64</td>
<td>mmHg</td>
</tr>
<tr>
<td>DIABP</td>
<td>Diastolic BP...</td>
<td>STANDING</td>
<td>61</td>
<td>mmHg</td>
<td>64</td>
<td>mmHg</td>
</tr>
<tr>
<td>DIABP</td>
<td>Diastolic BP...</td>
<td>STANDING</td>
<td>65</td>
<td>mmHg</td>
<td>64</td>
<td>mmHg</td>
</tr>
<tr>
<td>DIABP</td>
<td>Diastolic BP...</td>
<td>STANDING</td>
<td>51</td>
<td>mmHg</td>
<td>64</td>
<td>mmHg</td>
</tr>
<tr>
<td>DIABP</td>
<td>Diastolic BP...</td>
<td>STANDING</td>
<td>61</td>
<td>mmHg</td>
<td>64</td>
<td>mmHg</td>
</tr>
<tr>
<td>DIABP</td>
<td>Diastolic BP...</td>
<td>SUPINE</td>
<td>67</td>
<td>mmHg</td>
<td>64</td>
<td>mmHg</td>
</tr>
<tr>
<td>DIABP</td>
<td>Diastolic BP...</td>
<td>STANDING</td>
<td>50</td>
<td>mmHg</td>
<td>64</td>
<td>mmHg</td>
</tr>
<tr>
<td>DIABP</td>
<td>Diastolic BP...</td>
<td>STANDING</td>
<td>65</td>
<td>mmHg</td>
<td>64</td>
<td>mmHg</td>
</tr>
<tr>
<td>DIABP</td>
<td>Diastolic BP...</td>
<td>STANDING</td>
<td>64</td>
<td>mmHg</td>
<td>64</td>
<td>mmHg</td>
</tr>
<tr>
<td>DIABP</td>
<td>Diastolic BP...</td>
<td>STANDING</td>
<td>55</td>
<td>mmHg</td>
<td>64</td>
<td>mmHg</td>
</tr>
<tr>
<td>DIABP</td>
<td>Diastolic BP...</td>
<td>STANDING</td>
<td>53</td>
<td>mmHg</td>
<td>64</td>
<td>mmHg</td>
</tr>
</tbody>
</table>

**Define-XML metadata:**
- **Name:** VSORRESU
- **Label:** Original Units
- **Mandatory:** No
- **Datatype:** text
- **Length:** 200
- **CodeList:** VSUNIT
- **CodeList NCI Code:** C71620

**CDISC Library information:**
- **Name:** VSORRESU
- **Label:** Original Units
- **Core:** Exp
- **XPT simple data type:** Char
- **CodeList NCI Code:** C66770

**Discrepancies with CDISC Library:**
- Incorrect CDISC/NCI codelist assigned
- Expected: C66770

[UNIT] codelist was assigned instead of [VSRESU] codelist
Implementation in the "Smart Submission DatasetViewer"

Is the following OK?

<table>
<thead>
<tr>
<th>VSTESTCD</th>
<th>VSTEST</th>
<th>VSPOS</th>
<th>VSORRES</th>
<th>VSORRESU</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIABP</td>
<td>Diastolic Bl...</td>
<td>SUPINE</td>
<td>67</td>
<td>mmHg</td>
</tr>
<tr>
<td>DIABP</td>
<td>Diastolic Bl...</td>
<td>STANDING</td>
<td>61</td>
<td>mmHg</td>
</tr>
<tr>
<td>DIABP</td>
<td>Diastolic Bl...</td>
<td>RUNNING</td>
<td>65</td>
<td>mmHg</td>
</tr>
<tr>
<td>DIABP</td>
<td>Diastolic Bl...</td>
<td>SUPINE</td>
<td>56</td>
<td>mmHg</td>
</tr>
</tbody>
</table>

"RUNNING" was not defined as an "extended value" in the define.xml

Depends on what the define.xml says …

Define-XML metadata:
- Name: VSPOS
- Label: Vital Signs Position of Subject
- Mandatory: No
- Datatype: text
- Length: 8
- CodeList: VSPOS
- CodeList NCI Code: C71148

CDISC Library information:
- Name: VSPOS
- Label: Vital Signs Position of Subject
- Core: Perm

Discrepancies with CDISC Library:
- Define.xml codeList C71148 contains the value: [RUNNING]
  that could not be found in the CDISC Library CodeList and that was not defined as an extended value in the define.xml
Implementation in the "Smart Submission DatasetViewer"

CDISC CodeList Details for CodeList with NCI code C66770

Short name: VSRESU
Long Name: Units for Vital Signs Results
Extensible: true
Definition: The unit used to record and describe the result of a test investigating a vital sign.
Preferred term: CDISC SDTM Unit for Vital Sign Result Terminology
Synonyms: Units for Vital Signs Results

Discrepancies with CDISC Library

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Domain</th>
<th>Variable</th>
<th>Description</th>
<th>Expected Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>DM</td>
<td>DM</td>
<td>DTHFL</td>
<td>No CDISC/NCI codeList assigned</td>
<td>C66742</td>
</tr>
<tr>
<td>DM</td>
<td>DM</td>
<td>ARMACD</td>
<td>Incorrect casing of Label 'Planned arm code'</td>
<td>Planned Arm Code</td>
</tr>
<tr>
<td>DM</td>
<td>DM</td>
<td>DMDTC</td>
<td>Incorrect Label 'Date of Collection'</td>
<td>Date/Time of Collection</td>
</tr>
<tr>
<td>VS</td>
<td>VS</td>
<td>VSSTESTCD</td>
<td>No CDISC/NCI codeList assigned</td>
<td>C66744</td>
</tr>
<tr>
<td>VS</td>
<td>VS</td>
<td>VSSTEST</td>
<td>No CDISC/NCI codeList assigned</td>
<td>C67133</td>
</tr>
</tbody>
</table>
| VS      | VS     | VSPOS    | Define xml codeList C71148 contains the value [RUNNING] that could not be found in the CDISC Library CodeList and that was not defined as an extended value in the define xml
| VS      | VS     | VSORRESU | Incorrect CDISC/NCI codeList assigned | C66720          |
| VS      | VS     | VSORRES  | Define xml codeList C71620 contains values [BEATS/MIN. IN] that could not be found in the CDISC Library CodeList and that were not defined as extended values in the define xml

CDISC Library CodeList Details

<table>
<thead>
<tr>
<th>STUDYID</th>
<th>DOMAIN</th>
<th>USUBJID</th>
<th>VSSEQ</th>
<th>VSSTESTCD</th>
<th>VSSTEST</th>
<th>VSPOS</th>
<th>VSORRES</th>
<th>VSORRESU</th>
<th>VSSTRESN</th>
<th>VSSTRESU</th>
<th>VSTEST</th>
<th>VSLOC</th>
<th>VSTLFL</th>
<th>VSNUM</th>
<th>VS</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDISCPII</td>
<td>DM</td>
<td>VS</td>
<td>D10-2</td>
<td>1001-1615</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4001</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>CDISCPII</td>
<td>DM</td>
<td>VS</td>
<td>D10-2</td>
<td>1001-1615</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4001</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>CDISCPII</td>
<td>DM</td>
<td>VS</td>
<td>D10-2</td>
<td>1001-1615</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4001</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

CDISC Library Discrepancies Report
How validation SHOULD work

The sponsor's metadata truth

The CDISC metadata truth

Sponsor Submission Data → Validate against → Sponsor define.xml → Validate against → CDISC Library
Example: A GET Request to the CDISC Library That Uses the HEADERS Statement
Adjusted by: Marcelina Hungria
Date: 2019-04-03
Notes: Requires SAS 9.4M3 or later

filename response TEMP;

proc http
   url="https://library.cdisc.org/api/mdr/products/Terminology"
   AUTH_BASIC WEBUSERNAME="user" WEBPASSWORD="pass"
   out=response;
   headers
      /* Use "application/xml" instead of "application/json" to get the response in xml format */
      "Accept"="application/xml"
   run;

data _null_;
infile response;
input;
put _infile_;
run;
"Search" function in CDISC Library API

• New is a "search" function in the CDISC Library API
  • https://wiki.cdisc.org/display/LIBSUPRT/Search

• This functionality is extremely powerful, but a bit more complicated to use

• Example: "Search for LBTESTCD" in the scope of the "LB" domain and the IG-version "3.3":
  https://library.cdisc.org/api/mdr/search?q=LBTESTCD&domain=LB&product=SDTMIG%20v3.3

• Currently only a JSON implementation is available
Implementation in XSLT

- Mapping Software "SDTM-ETL" uses XSLT to execute the mappings

CDISC ODM (e.g. from EDC) → Mapping Instructions (XSLT) → SDTM / SEND (CDISC Dataset-XML) → SDTM / SEND (outdated XPT) → Mapping Instructions (propriety language)
Implementation in Software: XSLT/XQuery

• Problem:

• XSLT/XQuery very well supports RESTful web services

• But does not well support "Basic Authentication"

• Most used Java-Library for XSLT "Saxon" however allows to write XSLT extensions so that "Basic Authentication" can be applied

• We are currently implementing this into the "SDTM-ETL" mapping software
Planned further Implementations

• The "Open Rules for CDISC Standards" initiative has all CDISC / FDA / PMDA validation rules in a machine- AND human-readable format, and can be executed by ANY modern software

• "Open Rules for CDISC Standards" already uses RESTful Web Services for LOINC, UNII, SNOMED-CT and MedDRA lookups

• As the CDISC Library is "the CDISC single source of truth", we will implement the RESTful web services in the "Open Rules for CDISC Standards"
Planned further Implementations

• The "Define-XML Designer" is an extremely user-friendly software tool to develop and generate define.xml files for SDTM, SEND and ADaM submissions

• The define.xml is "the sponsor's truth" about the submission

• But this "truth" also has to comply with CDISC / FDA / PMDA "truth"
  • CDISC "truth": use CDISC Library API
  • FDA/PMDA "truth": use "Open Rules for CDISC Standards"
CDISC Library API
Summary of first experiences

• Very easy to use and implement in software
• HATEAOS makes "chaining" very easy
• The API is very well documented
• Support over "product inquiry" website is excellent
  • Usually good answers within hours
• Greatest advantages will become eminent with new versions of Standards, Implementation Guides and Controlled Terminology
  • No more "copy and paste" from PDFs or website pages
  • No more futzing around with Excel files
The CDISC Library
The CDISC Single Source of Truth