You are Connected to the CDISC Library – What's Next?

Santosh Kumar Davuluri (sdavuluri@vitadatasciences.com) and Hanumanta Rao Kavuluru (rkavuluru@vitadatasciences.com); Vita Data Sciences, Waltham, MA

Abstract

CDISC Library is an authoritative source for the machine-readable CDISC standards and terminology metadata. It uses linked data and a REST API to deliver these standards to software applications that automate standards-based processes. With the availability of the CDISC Library, it is not only important for one to learn the use of various API options to connect to the library, but also to understand how to make use of the various metadata tables extracted during the process to automate generation of the CDISC deliverables. In this presentation, we will demonstrate use of SAS-based API to connect to the library and more importantly how we utilize this metadata to automate processes linked to the development of SDTM and ADaM datasets.

Sample SAS code

```
libname libl "libl address";
libname lib2 "lib2 address";
filename libl temp;

proc http
     url = "URL Construct"
     auth_basic
     webusername = "username"
     webpassword = "password"
     out = libl;
     headers
     "Accept"="application/json";
run;

libname libl json fileref = libl;

proc copy in=libl out=lib2;
run;
```

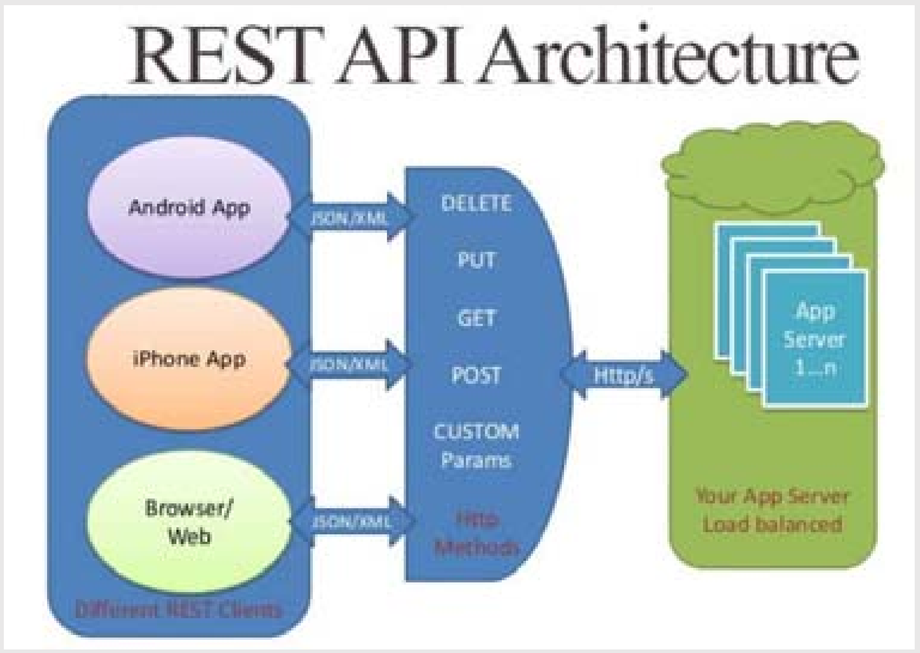
Metadata to generate SDTM specifications , SDTM, ADAM datasets.

SDTM specifications : The JSON engine used in the above SAS code helps us to transform JSON type of data extracted from CDISC library into SAS datasets. These SAS datasets can be used to automate the generation of SDTM specifications. First key step in this process is to find the relation between these SAS datasets. Once we make these SAS datasets relational then it would be easy to populate the code list, variables, value level, methods, and comments tabs in the pinnacle 21 specifications template by relating each of these SAS datasets through key variables.

SDTM/ADAM datasets :
Metadata design is the key aspect to automate SDTM/ADAM generation. This metadata design has three main components to it they are : **Source** , **Origin**, and **Target**.
Source : source data is the CDASH data which has to be mapped to SDTM.
Origin: Origin is how the target variable is obtained based on the source variable.
Target : the variables from SDTM to which CDASH variables are mapped to.
In case of ADAM datasets the source is SDTM instead of CDASH and target is ADAM instead of SDTM.

API and REST API

- API stands for **A**pplication **P**rogramming **I**nterface.
- API is a programmatic method to query and retrieve information between different applications and data bases.
- **R**epresentational **S**tate **T**ransfer (REST) is an architectural framework which uses **H**yper **T**ext **T**ransfer **P**rotocol (HTTP) to request and transfer data.
- Postman and cURL are 2 different tools which can be used for API development and testing purposes.
- Java Script Object notation (JSON) and XML are the generally used media type format.



URL construct and example

{scheme }://{host}/{basePath}/{endpoint}

- scheme: https
- host: library.cdisc.org
- basePath: /api
- endpoint: /mdr/ct/packages (example)

Example : https://library.cdisc.org/api//mdr/sdtmig/3-3

HTTP status codes

Code	Status	CDISC Library Description
200	OK	Request Processed successfully
400	Bad Request	Malformed URL or request
401	Unauthorized	Not authenticated
403	Forbidden	Not authorized
404	Not Found	Request resource not found
503	Service Unavailable	Service is not available, e.g., maintenance

Conclusion

Extraction of metadata from CDISC library using REST API is a major advancement towards automation. Nonetheless, it is important to have an understanding of how to use this metadata. Metadata extracted from CDISC library can be used to automate SDTM specifications generation by finding the key variables across the SAS datasets obtained through extraction. By mapping the metadata from CDISC library to company specific metadata can be done to automate the generation of SDTM/ ADAM and thereby TFL as well.