



Medidata RAVE and Veeva CDMS Data Extraction Macros

Craig Mistal & Nicola Tambascia

27-Jun-2024



Agenda

- Overview & History
- Medidata RAVE API
- Veeva CDMS API
- Implementation
 - Difficulties
- Validation and testing
 - The right access
 - IT Processes
- Usage
- Outlook

Overview & History

- Two macros to programmatically extract clinical data from CDMS systems developed
 - g_CaE_extractRave (v1.0.0)
 - g_CaE_extractVeeva (v1.0.0)
- Idea first came up in 2020
 - Paused due to issues faced during testing
- Restarted when Veeva API work was started in 2023

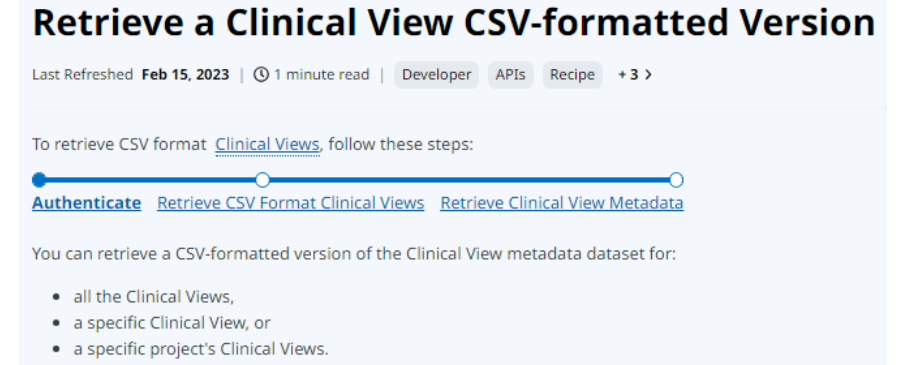
Overview & History

- Team:
 - Owner: Nicola Tambascia
 - **Developer: Craig Mistal**
 - Validation Team:
 - Validation Lead: Drikus Raath
 - Tester RAVE: Naveen Madugundu
 - Tester Veeva: Johan Niemann



Medidata RAVE API

- Documentation: <https://medidatarwsnet.readthedocs.io/en/latest/>
 - More detailed here: https://learn.medidata.com/en-US/bundle/rave-web-services/page/get_started_with_rave_web_services.html
 - Requires sign in
 - 400 page PDF also available
- Data extracts as CSV or ODM XML
 - Variable name, label, type, length and format information in Clinical View Metadata
- Study specific URL
 - `https://<client>-<host>.mdsol.com/RaveWebServices/studies/<ProjectName>(<environment>)/datasets/<details>`
 - **<details>**: ClinicalViewMetadata.csv, regular/<datasets>, SDTMDDataDictionaries.csv, SDTMPProtocolDeviations.csv, etc.
- URL, Username and Password needed to access Metadata's Biostat Adapter solution



Veeva CDMS API

- Documentation: <https://developer-cdms.veevavault.com/api/24.1/#cdms-api-overview>
- Data extraction: <https://developer-cdms.veevavault.com/api/24.1/#study-data-extract-sde>
 - Data extracts as SAS, XPT or CSV

- URL

- <https://<Vault Subdomain>/api/<SDE Version>/auth>

- Username and Password Authentication to create Session ID:
<https://developer-cdms.veevavault.com/api/24.1/#user-name-amp-password>

The screenshot displays the Veeva CDMS API documentation page for the 'Study Data Extract (SDE)' endpoint. The page is titled 'Study Data Extract (SDE)' and includes a description: 'The Study Data Extract (SDE) job allows the extract of study execution data from Vault, in CSV (UTF-8), and optional SAS/SAS XPT format. The job's output includes study datasets (AE, CM, etc.) as configured in the study design, plus the following system datasets:'. Below the description, there is a list of system datasets: SYS_SITE, SYS_SUB, SYS_EVT, SYS_FORM, SYS_Q, SYS_QT, SYS_ILB, SYS_LINKS, SYS_ASM, SYS_ASMR, SYS_PD, SYS_RAND, SYS_LABRANGES, SYS_LABLOC, and SYS_ANALYTES. Each dataset is followed by a brief description of its contents. The page also includes a 'Request - Start SDE' section with a cURL command for starting the SDE job.

```
curl -L -X POST 'https://my-vault.veevavault.com/api/v23.3/app/cdm/jobs/start_now' \
-H 'Accept: application/json' \
-H 'Content-Type: application/json' \
-H 'Authorization: {SESSION_ID}' \
-d '{
  "study_name": "ABCP-2022-01_DEV1",
  "request": {
    "job_type": "study_data_extract_v",
    "include_restricted_data": true,
    "include_study_design": true,
    "version": "v23.3",
    "export_file_types": ["CSV", "SAS"],
    "file_name": "ABCP-2022-01_DEV1_Study_Data_Extract",
    "use_external_ids": false,
    "include_formlib": true,
    "split_datetime": false,
    "exclude_blank_forms": false,
    "all_clinical_datasets": true,
    "all_system_datasets": true,
    "include_custom_objects": false
  }
}'
```

Implementation

- Authentication
 - Individual Username and Password stored in a compiled macro
 - Not saved in any programs
 - Not visible for others
 - Macro will only run if username matches with &sysuserid
- Steps in the program:
 1. Macros use proc http to connect to the APIs
 2. Download the data
 3. RAVE: convert the data to SAS
 4. Archiving
 5. Add data of extraction, location and Username running the program to log

Veeva > extract >

Name

 g_CaE_extractVeeva_11DEC2023_193552.log
 SAS_Sandbox_Test_Study_DEV1_Study_Data_Extract_11DEC2023_1935_CST.zip
 ae.sas7bdat
 aeyn.sas7bdat
 cm.sas7bdat
 cmyn.sas7bdat
 dm.sas7bdat
 ecg.sas7bdat
 eos.sas7bdat
 ex.sas7bdat
 ic.sas7bdat
 ie.sas7bdat
 lbc.sas7bdat
 mh.sas7bdat
 pa.sas7bdat
 pe.sas7bdat
 pg.sas7bdat
 pk.sas7bdat
 symp.sas7bdat
 sys_analytes.sas7bdat

Name

 clinicalviewmetadata.sas7bdat
 ae.sas7bdat
 aeisr.sas7bdat
 aeyn.sas7bdat
 apmhpds.sas7bdat
 cm.sas7bdat
 cmyn.sas7bdat
 cohort.sas7bdat
 datspect.sas7bdat
 dd.sas7bdat
 dm.sas7bdat

 filehash.sas7bdat
  2103_30NOV2023_165820.zip
 g_CaE_extractRave_30NOV2023_165820.log

Implementation – Difficulties RAVE

1. Metadata issue for CRF migrations

- Clinical View Metadata dataset is used to ‘drive’ the macro:
 - Contains both production and non-production metadata
- Production CSV Format Clinical Views may contain non-production variables
- Need to download production CSV Format Clinical Views, then keep production variables
 - Validate against SAS-OnDemand extract

B	C	D
viewname	ordinal	varname
prod.V_TVD_101_001H_AE	90	AEACNNST2_STD
prod.V_TVD_101_001H_AE	91	AERELNST3
prod.V_TVD_101_001H_AE	92	AERELNST3_STD
prod.V_TVD_101_001H_AE	93	AEACNNST3
prod.V_TVD_101_001H_AE	94	AEACNNST3_STD
V_TVD_101_001H_AE	90	AERELNST2_STD
V_TVD_101_001H_AE	91	AEACNNST2
V_TVD_101_001H_AE	92	AEACNNST2_STD
V_TVD_101_001H_AE	93	AERELNST3
V_TVD_101_001H_AE	94	AERELNST3_STD
V_TVD_101_001H_AE	95	AEACNNST3
V_TVD_101_001H_AE	96	AEACNNST3_STD

CO	CP	CQ	CR	CS
AERELNST3	AERELNST3_STD	AEACNNST3	AEACNNST3_STD	
Not Related	NOT RELATED	Dose Not Changed	DOSE NOT CHANGED	
Unlikely Related	UNLIKELY RELATED	Drug Withdrawn	DRUG WITHDRAWN	
Unlikely Related	UNLIKELY RELATED	Dose Not Changed	DOSE NOT CHANGED	
Not Related	NOT RELATED	Not Applicable	NOT APPLICABLE	
Not Related	NOT RELATED	Not Applicable	NOT APPLICABLE	
Not Related	NOT RELATED	Not Applicable	NOT APPLICABLE	
Not Related	NOT RELATED	Not Applicable	NOT APPLICABLE	
Not Related	NOT RELATED	Not Applicable	NOT APPLICABLE	
Not Related	NOT RELATED	Not Applicable	NOT APPLICABLE	
Not Related	NOT RELATED	Not Applicable	NOT APPLICABLE	
Not Related	NOT RELATED	Not Applicable	NOT APPLICABLE	
Not Related	NOT RELATED	Not Applicable	NOT APPLICABLE	

Count: 96

Implementation – Difficulties RAVE

2. Check for 'end of file' marker to validate if the whole dataset is received
 - Then remove 'end of file' marker during conversion to SAS datasets
3. Embedded line feed characters inside double quotes
 - Control characters presented challenges during initial validation

Implementation - RAVE

```
/* Sample call for %g_CaE_extractRave -Production data */  
%g_CaE_extractRave(Project=Sponsor1,  
    ProjectName=D1701-VrV101,  
    outloc=&output_location,  
    environment=PROD,  
    credloc=&credential_location);
```

```
/* Sample call for %g_CaE_extractRave -Mock data */  
%g_CaE_extractRave(Project= Sponsor1,  
    ProjectName=D1701-VrV101,  
    outloc=&output_location,  
    environment=dUAT,  
    credloc=&credential_location);
```

Implementation - RAVE

Macro Parameters

Parameter	Description
Project	Client-specific host address within Rave Web Services (e.g. <Group>-pra.mdsol.com)
ProjectName	Study-specific designation corresponding to ProjectName parameter in Rave Web Services URL (e.g. <a href="https://{host}/RaveWebServices/datasets/ClinicalViewMetadata.csv?ProjectName=<Study>">https://{host}/RaveWebServices/datasets/ClinicalViewMetadata.csv?ProjectName=<Study>)
outloc	Folder location on fileserver in which CSV files are downloaded, and corresponding SAS datasets are created
credloc	Folder location on fileserver in which custom user-specific, compiled-macro catalog containing g_anonymous_Rave macro, which contains the following: • Rave Web User: User-specific Rave Web Services account username • Rave Web Services password: User-specific Rave Web Services account password. Password can be encrypted with PROC PWENCODE. Compiled-macro catalog is generated using g_anonymous_Rave_template code.
environment	Environment: Corresponding host study environment in Rave Web Services URL
HostName	Corresponding host CRO in Rave Web Services URL e.g. <Group>-pra.mdsol.com, <Group>-icon.mdsol.com

Implementation – Difficulties Veeva

- Macro has to use Session Keep Alive functionality
 - i.e. post SessionID granted for the Study Data Extract to the Session Keep Alive URL
 - to ensure the SessionID granted for the Study Data Extract does not expire during the time taken for the job to reach Completed status.
- Job status timeout max 60 minutes
- Many parameters
- Versioned API

Implementation - Veeva

/* Sample call for basic study data extract job trigger */

```
%g_CaE_extractVeeva(credloc=&credential_location,  
    outloc=&output_location,  
    vault_subdomain=sb-prahs-cdms-cro.veevavault.com,  
    study_name=Sandbox Test Study_DEV1);
```

/* Sample call for specified clinical_datasets*/

```
%g_CAE_extractVeeva(credloc=&credential_location,  
    outloc=&output_location,  
    vault_subdomain=sb-prahs-cdms-cro.veevavault.com,  
    study_name=Sandbox Test Study_DEV1,  
    clinical_datasets=DM AE CM);
```

/* Sample call for specified true/false optional parameters and export_file_type=CSV */

```
%g_CAE_extractVeeva(credloc=&credential_location,  
    outloc=&output_location,  
    vault_subdomain=sb-prahs-cdms-cro.veevavault.com,  
    study_name=Sandbox Test Study_DEV1,  
    include_study_design=true,  
    use_external_ids=true,  
    include_formilb=true,  
    split_datetime=true,  
    exclude_blank_forms=true,  
    export_file_types=CSV);
```

Implementation - Veeva

Macro Parameters

Parameter - basic parameters required for full SDE	Description
vault_subdomain	vault_subdomain
study_name	name of study in Veeva CDMS
credloc	path to folder containing stored-compiled macro catalog with user's credential
outloc	path to output folder where study data extract is downloaded and unzipped, and log file saved
job_status_timeout	Maximum number of minutes macro queries Retrieve Job Status at 1 minute intervals until job status is Completed.

Optional Parameters - options in API for customized extract	Description	Expected Value Syntax / Example Values	Default Value
export_file_types	The type(s) of files of export	e.g. (any one or all off): SAS XPT CSV	SAS
external_connections	names of FTP connection names to push a copy of the SDE zip file	Must be pre-configured in Veeva CDMS .e.g.: "Veeva Vault SFTP"	
clinical_datasets	Names of specific clinical datasets to include in SDE (if not extracting all)	e.g. DM AE CM	
system_datasets	Names of specific system datasets to include in SDE (if not extracting all)	e.g. SYS_SUB SYS_FORM	
custom_objects	Names of custom object configurations to include in the export	e.g. CustomObj1_Docs CustomObj2_VisitTracking	
include_restricted_data	Whether the run of data should include data from restricted forms (or not). The API caller must have the permission for the restricted data when using true on this option.	true/false, false when omitted	false
include_study_design	Whether the run should include additional CSVs about the study design.	true/false, false when omitted	false
use_external_ids	Whether to use the Form / Item design External IDs for the naming of datasets and dataset columns. The default is to use Form / Item design name.	true/false, false when omitted	false
include_formilb	Whether the export run should include clinical dataset rows for forms in status of Intentionally Left Blank (ILB), or not.	true/false, false when omitted	false
split_datetime	Whether all datetime Item fields should be split into separate (additional) dataset columns for date only / time only.	true/false, false when omitted	false
exclude_blank_forms	Whether to exclude data from forms in status blank__v, or not.	true/false, false when omitted	false
include_rand_treatment	Only applicable if the study is using the Randomization module within CDMS, and for those with unmasked permissions in the module. The result (with parameter true) will be treatment values within the SYS_RAND dataset.	true/false, false when omitted	false
boolean_formatting	The format to use for boolean/checkbox values, y_n when omitted.	Possible values are: y_n, y_n_empty, t.f, t.f_empty, 1_0, 1_empty, 1_0_empty, true_false, true_false_empty, yes_no, yes_no_empty	y_n
item_display_label	The format to use for SAS labels.	Possible values are label and short_label. When omitted, the default is label.	label

Validation & Testing

- Validation and Testing according to our SOP
 - Create User Requirements (Programmer)
 - Create Validation Testing Plan (Validation Lead)
 - Create Validation Test Cases (Validation Lead)
 - Execute Test Cases (Tester)
 - Separate Tester for RAVE and Veeva
 - Repeat as needed
 - Create Macro Documentation
 - File everything

Validation & Testing

- 3 Studies per System selected
 - DEV and PROD data for each
 - Different stages of studies: locked, ongoing, CRF migration
- Testing was done by comparing to manual extracts
 - RAVE differences we noticed
 - Sorting
 - Multiple length by 3 except: Folder, StudyEnvSiteNumber
- Access for Testers to the studies needs to be planned well in advance
 - Study Teams reluctant to give access
 - Access roles needed were not clearly defined
 - IT supported us

Validation & Testing

- Access for Testers to the studies needs to be planned well in advance
 - RAVE:
 - Rave Web Services Account needed (different from the usual account)
 - Data Extract Role, All Sites.
 - Architect Project module access: EDC study Name, Role Read Only
 - Veeva:
 - Veeva Vault CDMS API account
 - with CDMS API Read **Write** access rights to the study
 - Credentials cannot be the same as those used for Single Sign On
- Before macro launch, we met with IT
 - to ensure people can raise the correct access requests
 - IT assigns the correct access

Usage

- Already in use by multiple study teams
 - In different departments
- During usage we moved data center
 - New firewall blocked URLs and Ports
- Allows to document the extraction date and time, as well as user extracting the data
- Allows automated data extraction **and** SDTM run
- CDR Team using same technology for RAVE
 - Planning to do the same for Veeva

Outlook

- Use technology to extract the data centrally from the CDR
 - Convert RAVE CSV files to SAS
- Updates already identified
 - Fix batch run issues
 - Update Veeva CDMS API version
- High interest from different groups to use same technology for other topics/macros
 - g_CaE_extractRaveAuditRecords
 - g_CaE_VeevaQueryDetailListing
 - g_CaE_extractVeevaAuditTrail
 - g_CaE_extractRaveAdminData



ICONplc.com

© 2024 ICON. All rights reserved.