



Medidata RAVE and Veeva CDMS Data Extraction Macros (SM13)

Craig Mistal & Nicola Tambascia

PHUSE EU 2024 - 13-Nov-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

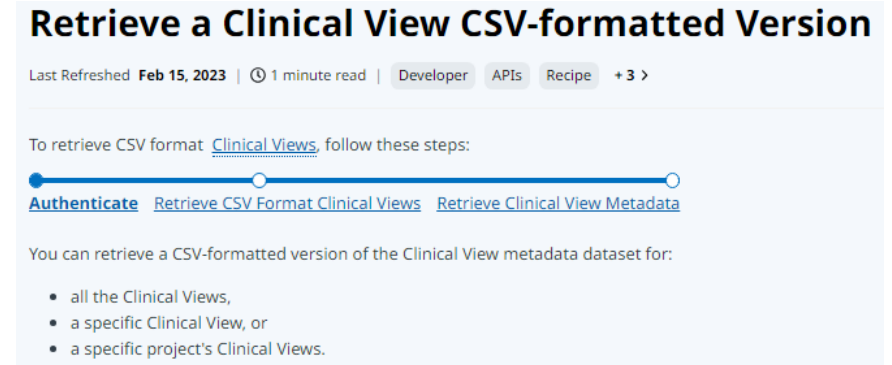
Overview & History

- 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 Pages 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/24.2/#cdms-api-overview>
- Data extraction: <https://developer-cdms.veevavault.com/api/24.2/#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/24.2/#user-name-amp-password>

The screenshot shows 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 of the endpoint and a list of system datasets that can be extracted. The left sidebar contains a navigation menu with options like "Jobs", "Subject Progress Listing", "Event Progress Listing", "Form Progress Listing", "Query Detail Listing", "Core Listings", "Data and Definition Export", "Audit Trail by Study", "Audit Trail by Site", "Audit Trail by Subject", "Data Change Report", "Cancel Job", "Retrieve Job Status", "Retrieve Job Output File", "Retrieve Job Log", "Users", "Study Design", "Vault CDMS Help", and "Veeva Support". The main content area on the right lists the 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. A "Request - Start SDE" section shows a cURL command and a JSON request body for starting the SDE job.

Study Data Extract (SDE)

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:

- SYS_SITE: Lists all Sites.
- SYS_SUB: Lists all Subjects.
- SYS_EVT: Lists all Events (visits) with the Event Date and Status.
- SYS_FORM: Lists all Forms, including any Forms marked as intentionally left blank, with the Review Status, Freeze Status, Lock Status, and some additional metrics.
- SYS_Q: Lists all Queries, along with some query metrics. Only the first query message is listed in this dataset. The full list of query messages (first message, answers, etc.) is included in the SYS_QT dataset.
- SYS_QT: Lists all Query Messages.
- SYS_ILB: Lists all Items that are marked as intentionally left blank.
- SYS_LINKS: Lists all Forms that are part of Form to Form linking (or Form to Item). Each link is a pair of records in the dataset.
- SYS_ASM: Lists all Assessments that exist in the Study, if Assessments are configured in the study design.
- SYS_ASMR: Lists all Questions and Answers for all Assessments, if Assessments are configured in the study design.
- SYS_PD: Lists all Protocol Deviations, if this feature is enabled in the study.
- SYS_RAND: Lists randomization details, if Randomization is enabled for the study.
- SYS_LABRANGES: Lists Normal Ranges, if Local Labs is enabled for the study.
- SYS_LABLOC: Lists Lab Locations, if Local Labs is enabled for the study.
- SYS_ANALYTES: Lists Analytes, if Local Labs is enabled for the study.

See [CDMS Help](#) for a detailed list of all columns in the Study Data Extract output.

Request - Start SDE

```
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_type": ["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

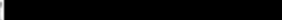
Veeva > extract >

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
 apmhp.d.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. Clinical Views Metadata issue
 - Use Production Metadata for Live data
 - Non-Production Metadata for all other data
 - Validate against SAS-OnDemand extract
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);
```

Implementation - Veeva

Macro Parameters

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

Implementation - Veeva

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

Listing of Common Variables with Differing Attributes

Variable	Dataset	Type	Length	Format	Informat	Label
Folder	WORK.AEYN	Char	150	\$150.	\$150.	Folder OID
	WORK.CMP_AEYN	Char	50	\$150.	\$50.	Folder OID
StudyEnvSiteNumber	WORK.AEYN	Char	60	\$60.	\$60.	StudyEnvSiteNumber
	WORK.CMP_AEYN	Char	20	\$60.	\$20.	StudyEnvSiteNumber

- RAVE differences we noticed
 - Sorting
 - Multiply length by 3 except: Folder, StudyEnvSiteNumber
 - Line feeds removed
- 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
 - Rave Audit Records
 - Veeva Query Detail Listing
 - Veeva Audit Trail
 - Rave Admin Data



ICONplc.com

© 2024 ICON. All rights reserved.