

Medidata RAVE and Veeva CDMS Data Extraction Macros

Nicola Tambascia, ICON plc, Mannheim, Germany

Craig Mistal, ICON plc, Canada

ABSTRACT

Web-based study data extractions allow for the immediate availability of clinical data for review, transformation, and analysis by Data Managers, Programmers, and Biostatisticians, respectively. The paper shows a technical overview and background of the SAS macros we created and validated to extract live EDC data from Medidata RAVE via the Rave Web Services (RWS) and Veeva CDMS via the Veeva CDMS API. The process around the APIs (e.g. access needed in the systems), and the advantages of using these methods of data extraction over procedures which utilize other scheduling methods, and requiring subsequent retrieval of extracted data from FTP sites is also detailed.

INTRODUCTION

Medidata's Biostat Adapter is a Rave Web Services solution which provides CSV or ODM XML extracts of Clinical Views to be used as source for users to create datasets for themselves. This paper describes how to programmatically access Medidata's Biostat Adapter solution to download the Clinical View Metadata, the CSV Format Clinical Views and the Data Dictionaries Dataset and how to use the metadata to intelligently build import programs to convert CSV Format Clinical Views to SAS datasets that should contain the same data as that received through Medidata's SAS on Demand solution.

Similar to the Medidata's Biostat Adapter, Veeva CDMS offers an API to download clinical data. It allows to download data as SAS® files, SAS® XPT files, CSV and other formats.

Macros were developed and validated to connect to the APIs, and download the data, so that the ICON Biostatistics and Programming Group can programmatically extract electronic data capture (EDC) data and use it for their work.

THE API

MEDIDATA RAVE

Medidata's Biostat Adapter is a Rave Web Services solution which provides CSV or ODM XML extracts of Clinical Views to be used as source for users to create datasets for themselves.

The API that is available for the Rave Web Services is documented online under

<https://medidatarwsnet.readthedocs.io/en/latest/>.

To connect via the API Users have to identify with their Username and Password assigned to their Rave Web Services Account. After that they can reach the clinical views by building the following URLs:

The Metadata of the Clinical Views in a study can be reached via the environment independent URL below:

`https://<client>-
<host>.mdsol.com/RaveWebServices/datasets/ClinicalViewMetadata.csv?ProjectName=<ProjectName>`

The clinical views can be reached via the URL:

`https://<client>-
<host>.mdsol.com/RaveWebServices/studies/<ProjectName>(<environment>)/datasets/regular/<dataset>.csv`

The SDTM Data Dictionaries in a study can be reached via:

`https:// <client>-
<host>.mdsol.com/RaveWebServices/datasets/SDTMDDataDictionaries.csv?studyid=<ProjectName>(<environment>)`

The Medidata RAVE API also provides a lot of additional data via the API, for example Audit Trail data can be downloaded using the API. For more details please refer to the documentation.

VEEVA CDMS

Per [Veeva Jobs – Study Data Extract \(SDE\)](#) Documentation: "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 system datasets".

To connect via the API Users have to identify with their Veeva CDMS Username and Password. After successful identification jobs can be started via the API and the clinical data will be downloaded in the requested file format and zipped as ZIP files. The session will need to be kept alive until the download is finished.

The Veeva CDMS API also provides a lot of additional data via the API, for example Audit Trail data can be downloaded using the API. For more details please refer to the documentation.

IMPLEMENTATION

MEDIDATA RAVE

A user calls the macro through a SAS stored-compiled macro. The stored-compiled macro utilizes user's credentials acquired from a custom compiled macro catalog for each user in order to access Rave Web Services in such a manner that user's credentials are never exposed in a program or a log. A user is not able to utilize access credentials stored in another user's compiled macro catalog to access Rave Web Services.

The macro then proceeds with the following major functions:

Accesses Metadata's Biostat Adapter solution to download and convert the Clinical View Metadata and Data Dictionaries Dataset for a given study based on the required macro parameters.

The g_CaE_extractRave macro utilizes the information in Clinical View Metadata to:

- Get a list of Clinical Views ('Regular' ViewType) for the specified environment
- Build an appropriate Biostat Adapter URL for each Clinical View, and download the CSV file for each Clinical View to the Output Location
- Confirm the 'end-of-file' marker in each downloaded CSV file to validate the whole dataset is received
- Apply variable name, label, type, length, format and informat information in Clinical View Metadata to convert CSV Format Clinical Views to SAS datasets. The solution also applies the same character variable length multiplier as applied in Medidata's SAS on Demand solution to accommodate multi-byte characters.

DIFFICULTIES

While developing the macro we faced multiple challenges. We had to find out how to use the correct set of metadata (production vs non-production) from Clinical View Metadata to build an appropriate Biostat Adapter URLs for all applicable Clinical Views, and apply variable name, label, type, length and format information in Clinical View Metadata to convert CSV Format Clinical Views to SAS datasets. Production Clinical View Metadata will be used if the user specifies that Live data (PROD data) should be extracted. For all other cases non-production Clinical View Metadata will be used.

To ensure all data was received for a file, check for the 'end of file' marker to validate if the whole dataset is received, then remove the 'end of file' marker during the conversion to SAS datasets.

We also had to remove embedded line feed characters inside double quotes, before converting the CSV file to a SAS dataset.

VEEVA CDMS

A user is able to call the macro through a SAS stored-compiled macro. The stored compiled macro utilizes users' credentials acquired from a custom compiled macro catalog for each user in order to access the Veeva CDMS in such a manner that user's credentials are never exposed in a program or a log.

A user is not able to utilize access credentials stored in another user's compiled macro catalog to access the Veeva Vault CDMS API.

The macro will then proceeds with the following major functions:

Retrieve Veeva Vault CDMS API SessionID

Macro utilizes User Name & Password Authentication (post User Name & Password to Veeva Vault CDMS API Authentication URL) and, on success, retrieves a SessionID granted by the Veeva Vault CDMS API which is then used for Bearer Token Authorization in subsequent API calls (e.g. Study Data Extract, Session Keep Alive, Retrieve Job Status, and Retrieve Job Output File).

Trigger a Study Data Extract in the Veeva Vault CDMS API

The Macro builds the appropriate study data extract request (JSON string), posts the request to the Veeva Vault CDMS API Study Data Extract (SDE) job URL, and retrieves the response from the Veeva Vault CDMS API.

Confirm Study Data Extract job has Completed

A successful Study Data Extract (SDE) job request results in a JSON response from the Veeva Vault CDMS API, which includes an assigned job_id. The Macro retrieves the job_id that is then used in subsequent API calls (e.g. Retrieve Job Status and Retrieve Job Output File).

The Macro uses the Retrieve Job Status URL to query the job status (e.g. Queued, In-progress, or Completed) at one-minute intervals until the job status indicates the job is Completed, or a until configurable maximum timeout period has been reached (e.g. 30 minutes).

The Macro will also use Session Keep Alive functionality to ensure the SessionID granted for the Study Data Extract does not expire during the time taken for the job to reach Completed status.

Download Study Data Extract ZIP file to fileserver

On confirmation the specific job_id has reached Completed status, the macro uses the Retrieve Job Output File URL to download the Study Data Extract ZIP file to a folder location on the fileserver that is specified by the user.

Unzip Study Data Extract ZIP file contents on fileserver

The Macro unzips SAS, XPT, and/or CSV file contents in the specific Study Data Extract ZIP file to the folder location on the fileserver that is specified by the user.

DIFFICULTIES

While developing the macro we faced multiple challenges. The authentication via the Veeva CDMS requires credentials (user name/mail address) separate from the ones used for Single Sign On in Veeva CDMS, as otherwise Veeva tries to use Single Sign On in the API and fails to do so.

The macro also needed to ensure that the session stays alive during extract from large studies due to the long waiting time until the download finishes. The macro uses Keep Session Alive functionality to ensure the Session ID acquired during the authentication step remains valid.

New versions of the API sometimes include new options or new implementations of existing options (for example JSON strings verses arrays), requiring a robust method to generate increasingly complex JSON strings, based on macro parameters.

VALIDATION & TESTING

The validation of both macros was done according to ICONs SOP for macro validation. The Programmer had to create User Requirements. A validation Lead had to create a Validation Plan and Test Cases, while the Testers had to test the Macros against the different User Requirements, based on the system (RAVE and Veeva).

For the testing we selected 3 studies per system in different stages of a clinical trial, locked, ongoing and a study that recently underwent CRF migration. For the tests the Tester had to manually extract data and compare it against the data extracted via the new macros

For Medidata RAVE extracts the Testers noted differences in the data, that were documented and deemed acceptable:

1. The data from the macro extracts were sorted differently
2. The multiplication for the variable length to handle multi-byte character is not done on the variables Folder and StudyEnvSiteNumber in SAS® on Demand extracts. As the macro extract multiplied the length, it was on the safe side.
3. For the conversion to SAS® the macro removes Line feeds which are included in SAS® on Demand extract

While preparing and during testing we realised that getting access to API accounts was not easy, as administrators were reluctant to give these access rights.

For Medidata RAVE a Rave Web Services Account with Data Extract Role, All Sites is needed, which is separate from the usual account and often only used by other tools or scripts.

For Veeva CDMS an account with API Read/**Write** access is needed, which concerned study teams.

To standardise the process we met with the IT Department and ensured that Users could easily get the needed access, once the macros were rolled out.

OUTLOOK

The same technology was implemented to pull data into our Clinical Data Repository and to convert the CSV files coming from Medidata RAVE into SAS® files.

As study teams started using then macros, we have also identified updates that we want to implement, e.g. fixing an issue that causes batch runs to fail or update to newer Veeva CDMS API version.

Based on the technology other groups raised interests in developing macros to extract Audit Trail, Query Listings and Admin Data for reporting via the API. These macros are currently under development and validation.

CONCLUSION

The macros are being used by different study teams and provide them with key advantages:

1. Immediate and automated extraction and access to data. Particularly important where fast turnaround times are required after CDMS is confirmed ready for extraction on a delivery day, or if a CDMS correction is required.
2. Programmatic extraction prevents user-error associated with manually selecting incorrect, or insistent options while using on-demand interfaces.
3. Speed: Extractions take minutes.
4. Transparent traceability in the log directly from CDMS to production workflows
 - a. The log provides the user the extracted the data, the location of the extracted data, as well as the data and time of the extract
5. No intervening systems (e.g. scheduling extractions with subsequent downloads from an FTP site, or other repository systems requiring multi-step extractions)
6. Veeva CDMS API macro allows for simultaneous and automated posting of a study data extract to a client FTP site.
7. Automated extraction and run of subsequent tasks without manual intervention, which is very important for scheduled runs and transfers

While the study teams were using the macro, ICON also moved data centers, causing the macro to fail due to closed ports. We worked together with IT to get the ports reopened.

REFERENCES

Medidata.RWS.NET Documentation

<https://medidata.rwsnet.readthedocs.io/en/latest/>

Medidata Knowledge Hub

https://learn.medidata.com/en-US/bundle/rave-web-services/page/get_started_with_rave_web_services.html

Veeva Vault CDMS API Overview

<https://developer-cdms.veevavault.com/api/24.2/#cdms-api-overview>

Veeva Vault CDMS Study Data Extract (SDE)

<https://developer-cdms.veevavault.com/api/24.2/#study-data-extract-sde>

Veeva Vault CDMS User Name & Password

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

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Author Name: Nicola Tambascia

Company: ICON plc

Address: Gottlieb-Daimler-Straße 10, 68165 Mannheim

Work Phone: +49 621 8782529

Email: nicola.tambascia@iconplc.com

Website: <https://www.iconplc.com/>

Author Name: Craig Mistal

Company: ICON plc

Address: Canada

Work Phone: +1 267-481-9657

Email: craig.mistal@iconplc.com

Website: <https://www.iconplc.com/>

Brand and product names are trademarks of their respective companies.