ABSTRACT
Pharmaceutical firms typically have a large number of validated SAS language programs. Many of these programs are run repeatedly with alterations to produce required output. A common improvement on this process is to pass parameters to validated autocall macros. Still, programs are often hand-edited, tested, then output is copied and pasted to distribute results. In many cases, multiple manual steps are required that are expensive, error-prone and, importantly in a regulated environment, hard to reproduce reliably.

The use of SAS language autocall macro libraries provides a method for effective reuse of SAS program logic. To make best use of available skilled workforce, firms need to fully control and automate this logic as well as expand software tooling and process automation to encompass Python, R, SQL, Perl, Powershell, Linux shell scripts, batch scripts and other programming technologies.

The proposal here is to fully automate distribution and invocation of SAS language programs along with Python and R by enabling any consumer (browser, Excel, middleware applications, web portals, programming environments etc.) to make web-based REST API calls with API arguments being passed through as macro parameters when executing program logic.

This paper describes a novel, easy-to-implement architecture and workflow upgrading validated autocall macro libraries and exposing them as API calls passing parameters to auto-generated API endpoints. A related goal is to describe a scalable software architecture based on the OpenAPI Initiative (OAI) and its OpenAPI Specification (OAS) enabling these capabilities.

INTRODUCTION
Application developers will learn how easily and quickly to publish and track API endpoints integrating various programming tools. Application consumers benefit by having a single and consistent method to acquire output and results. Consistent use of REST APIs enforces rigorous abstraction and artifact publishing and management aids effective, controlled application development.

We describe a workflow enabling SAS language programs to be callable as REST APIs:

1. Optionally validate the program is host-independent by executing it on Windows and Linux.
2. Push the program to a source code repository such as Git
3. Publish the program together with any supporting files or data assets as an artifact to a Hub. Hub software architecture is described in detail below.
4. Deploy the program to enable client programs to call the program though auto-generated URLs with defined, typed parameters (numbers, dates, strings, choice lists, etc.) required for invocation of the API.
5. Define a pipeline to chain together deployed programs. A pipeline uses a flow control graph based on success/failure of pipeline steps, integrating Powershell, Linux Shell scripts, Windows Batch scripts and control nodes Wait, Any, and All.
6. Consume the program using browsers, Postman, R scripts, and SAS language PROC HTTP syntax that call the published REST API endpoint with a URL including parameters to execute the program’s logic.

REST APIS
This section provides a brief description on the fundamentals of REST APIs, how URL strings are passed from the consumer (browsers, etc.) and how servers fulfill these requests. Using a REST API layer, SAS
language programs are abstracted from the consumer through a “contract”. This allows different consumers to make the same API calls without needing to know the implementation or infrastructure used to fulfill requests, or even that SAS language logic was used. This allows, for example, an R or Python user to consume a SAS language program without any knowledge of SAS language syntax, and vice versa.

Representational State Transfer (REST) is a standardized method using web-based protocols (HTTP or HTTPS) to present remote resources to a local client. A server provisions an ‘endpoint’, a URL that services HTTP requests such as POST, GET, DELETE, UPDATE, etc. to accept client instructions to the server to operate on resources – application logic, data etc. - according to the request.

For example, a SAS language user wanting access to an already-written program may use a %INCLUDE statement with the path to a file containing a SAS language program in order to use the included program’s logic. The SAS language user must know details like where to find the program, what macro parameters etc. A common challenge is knowing if that path points to the most up-to-date version of the program. If the desired program is located on a remote file system, then they may need to use FTP or similar to copy the program locally, or use the SAS language FILENAME FTP access method, requiring the user to have detailed network knowledge and credentials.

Using a SAS language autocall macro library is a step towards abstracting file system and networking details allowing SAS language programmers easily to call program logic without knowing such details. However, SAS language autocall macro libraries provide abstraction only to SAS language programmers. In today’s multi-language (polyglot) world that includes R and Python and potentially many more emerging languages such as Julia, LUA and Golang, organizations will benefit from SAS language assets and resources being readily consumable beyond the SAS language user community. By mapping critical SAS language programs to REST APIs, organizations can make SAS language code re-usability for any consumer throughout an organization.

Later in this paper, we illustrate R programs calling a REST API to execute a SAS language program by importing the R library httr.

**HUB SOFTWARE ARCHITECTURE**

SLC Hub is the name of Altair’s enterprise analytics management software. The software provides a secure scalable platform that offers a range of services to support enterprise analytics processes including user security and management, user and data access provisioning, API definition, publishing and deployment, on-demand API execution, workload pipelines and scheduling, workload balancing for interactive and production workloads across a grid or cluster of worker nodes, auditing and logging. The software delivers a simple, modern, software stack that is easy to deploy, manage and control. The Hub software can be deployed on-premises, in the cloud or in a hybrid environment. The Hub software can run on Linux or Windows.
The SLC Hub architecture is built to be an elastic and scalable platform. It is composed of a Hub node as the control plane, and a collection of worker or compute nodes managed by the Hub(s) to which workloads are distributed by the Hub node for execution according to workload profile rules. Results are returned to the Hub and returned to end-users and machine-to-machine clients (browsers, SLC Excel plug-in, business applications making REST API requests).

The software has five major sub-components described below.

1. **Eclipsed-based Analytics Workbench** is a rich GUI/IDE used to develop analytics applications both through visual drag-and-drop Workflows and coding. Programmers compose SAS language programs, SQL, Python, and R. Non-programmers use a palette abstracting coding to visually design workflows.

   During development, programs and workflows can be executed interactively both locally and remotely via the Hub. The Workbench has a visual interface to wrap programs in a parameterized package ready for testing and deployment into a Hub. The parameterized package can be executed and tested locally in the Workbench without needing to be published to a Hub. Once a package is ready for publishing, it can be transmitted to a Hub as a Hub package (described next) using a simple point-and-click wizard in the Workbench.

   A Hub administrator defines profiles that map resources to specific worker nodes. For example, requests to invoke a program that reads data from a particular Oracle database may require a profile defined for worker node instances 1 and 2 running Linux. Access to SQL/Server requirements is fulfilled by worker node 3 running Windows etc.

2. **Hub Console**. This management portal is accessed by a web-based GUI providing services to manage services such as auth domains, library definitions, execution profiles (fine-grain control for
multiple SAS language program execution configurations), control of repositories, building workload pipelines (chained programs), job scheduling, REST API deployment and publishing.

All configuration management and change control activities are performed through the Hub Console.

3. **User Portal.** This portal is accessed in a similar way to the Hub Console, but is intended for end users to run deployed programs and access results.

4. **Workers:** Deployed programs are orchestrated for execution by the Hub node onto one of multiple worker nodes using Hashicorp’s open-source Nomad workload management infrastructure. Multiple, different teams of users can have workloads specified to run on specific worker nodes according to security, performance and platform requirement profiles. Workloads can include both production scheduled processes as well as interactive user processes. Nomad provides many facilities that are common in a Kubernetes containerized environment but using conventional VMs instead of containers. A Containerized version of Hub may be offered in future, but the Nomad/VM version was found to be easier to manage for many organizations at this time.

5. **Opensource components.**
   a. Hashicorp’s open-source Vault for back-end encryption for at-rest and in-flight passwords, database credentials, and secrets. This eliminates storing credentials in clear text when accessing resources.
   b. Minio: Worker nodes need a central location to store execution artifacts. For non-AWS deployments Minio provides an S3-like access method for shared resources created among Workers. This obviates deployment overhead for shared filesystems like IBM’s SpectrumScale, FSx for Lustre (AWS), FileStore (GCP), or Azure Files.
   c. Nomad for orchestration. Nomad provides a flexible and scalable platform for managing applications across a wide range of infrastructure. Nomad is an open-source cluster manager and scheduler that automates the deployment, scaling, and management of containerized and non-containerized workload.
   d. PostgreSQL is used as a backing store for configuration and related Hub application data. PostgreSQL stores resources used to control Hub components. These include configuration files, namespace definitions, TLS certificates and so on.
   e. NGINX is used as a webserver to facilitate communication between Hub components that are architected as micro services. No user-intervention is needed at install and configuration time. NGINX also serves HTML pages presented by the Hub’s portal interface.

### API PRODUCTION PROCESS

This section includes a step-by-step description of the workflow process described in the abstract. We start with a simple use case to illustrate an R script consuming a SAS language REST API:

1. Write an SLC program that outputs a range of dates. The start of the date range is a parameter (SAS macro variable &start_date) to be passed as part of the REST API call.
2. Test/validate this program by executing it using SLC Workbench to supply program attributes to determine how output is presented through the API call. For R, we chose to stream JSON output. Alternatively, we could have chosen to output an R DataFrame. Once the program is validated, we create an API package.
3. From SLC Workbench, upload the validated API package to the Hub storing it in a Hub-managed repository we named Pharmasug2023.
4. Using the Hub Console, we create a ‘deployment package’ to publish the program as a REST API.
5. Test the API call using Postman using Basic Auth and call the API with an arbitrary value for the &start_date parameter (SAS macro variable). This test ensure the API call is producing the expected output.

6. Write an R script using the httr library to call the API by passing basic authorization parameters and the value of the parameter representing the start_date. Load the jsonlite library to “flatten” the returned array and create the R DataFrame.

Create the SLC program called generate_dates

```plaintext
data dates;

do date = "&start_date" to "31Dec2023"d;
   output;
end;

format date YYMMDd10.; run;
```

In this form, the program would fail if executed in a “stand-alone” mode. For the API call we are using &start_date as a parameter whose value is determined at the time the rest API is called. SLC Workbench provisions configuration menus to determine the behavior of this parameter and related parameters to define the overall behavior of the REST API. See Figure 2. Hub Configuration Values for generate_dates.
For Results Detail, we select Dataset for the Result type. For Results format we select JSON. These choices “extend” our original ‘generate_dates.sas’ program by calling PROC JSON to write output in JSON format.

For Parameter Style, we select “Macro variables”. Alternatively, if we had a larger number of parameters to manage as part of the API design, we would select “Dataset”. For Type we select Date from the following choices:

---

Figure 2. Hub Configuration Values for generate_dates
For Categories, we chose the name Pharmasug2023. Within the Hub Console, we can search programs and deployment packages using this string as a search value. Once we save the Hub Configuration values, the program is marked as an API package shown in Figure 4. The API package contains the source code for the program as well as the selected Hub parameter values.

**TEST AND VALIDATE THE ‘GENERATE_DATES’ PROGRAM**

As we mentioned earlier, our ‘generate_dates’ program needs the appropriate execution context. Once the Hub parameters in Step 1 are defined, we test our program before uploading the API package to SLC Hub.
UPLOAD VALIDATED API PACKAGE TO SLC HUB

Once the API package is defined and validated it is uploaded to the SLC Hub instance we instantiated in AWS. SLC Hub deployments are cloud-agnostic. See Figure 5.

![Upload Hub Artefact](https://via.placeholder.com/150)

*Figure 5. Upload API Package to SLC Hub in Step 3*

The target deployment for this validated API package is the Pharmasug2023 repository that is a Hub managed repo we created previously. After clicking OK, the API package is uploaded to the Hub as an artifact.

CREATE A ‘DEPLOYMENT PACKAGE’ TO PUBLISH PROGRAM AS A REST API

With the API package uploaded to the Hub, we use the Hub Console to create a Hub Deployment package which allows the program to be called through a REST API. The Hub portal is designed to bring governance and control to all SLC assets.
In this step, we concentrate on the Deployment Services. We confirm the artifact "generate_dates" is available in the Hub environment.

Next, we create a Deployment package. Select the Deployment type, in our case a Hub-managed repository and the Execution profile. See Figure 8.
The Hub administrator defines Execution profiles which map resources to specific worker nodes. For example, requests for access to read data from Oracle is a configuration defined for worker node instances 1 and 2 running Linux. Access to SQL/Server requirements are met by worker node 3 running Windows, and access to DB2 are met by worker node 4.

Here we select the Standard Execution profile. The repo where “generate_dates” defined in Step 1, and the Version number. See Figure 9.

Next, we select a unique deployment path (a portion of the URL string) clients use to call the API. We select the Create API endpoint check-box. See Figure 10.
The final action for this step is to select the category name. The category name was defined in Step 1, when defining the Hub attributes for the API package in the Workbench. This action provides a convenient filtering mechanism to easily locate artifacts belonging together.

An artifact can be a member of multiple, different categories. See Figure 11.

Defining the “generate_dates” Deployment package is complete and available for any client to consume. The Hub Directory services returns the API end-point. See Figure 12.
In our case, the generated API endpoint is:

https://xxxxxxxx-xxxx.xxxx.altair.com/run/generate_dates/RunGenerateDates

We also see “start_date” is a required parameter. Selecting the Execute tab on this page permits us to call the defined REST API. See Figure 13.

In the Results pane we see the JSON array returned when the start_date parameter value is set to 30/12/2023. Recall the date values are formatted in our initial program with the SLC-supplied YYMMDDd10. date format.

**TEST THE REST API CALL USING POSTMAN USING BASIC AUTH**

Once an API endpoint is defined, we describe a validation process to certify calls into the API are working and produce intended results. This is accomplished by using different clients; Microsoft Excel, a browser, and a language of SAS program calling the API endpoint and so on.

An indispensable tool for testing REST API endpoint is Postman. Postman is an API platform for building and using APIs. Postman simplifies each step of the API lifecycle and streamlines collaboration.
Access to Hub’s APIs require client-side authentication. The Hub API endpoints support Bearer Tokens as illustrated in the Postman screenshots. See Figure 14.

The API endpoint passed is:

https://xxxxxxx.xxxxxxxx.altair.com/run/generate_dates/RunGenerateDates?start_date=2023-12-28

Notice how the start_date parameter value is 2023-12-28 returning 4 dates as a JSON array in the response body. In addition, we confirm with Postman that the SLC Hub endpoint will accept Basic Auth which is simpler to implement in subsequent endpoint API calls than using Bearer tokens.

Figure 14. Postman calling SLC Hub REST API endpoint using Bearer Auth token

As an example of client-side Basic Auth for authentication to access the API endpoint, consider the use of a browser such as Chrome.

Figure 15. Chrome browser authentication to access SLC REST API endpoint

The user must authenticate with their Hub credentials which can be managed internally by SLC Hub, or though Vault, or by establishing an auth domain to a credentials management services like Active Directory. Creating an auth domain within the Hub offers the ability for organizations to support Single Sign-On (SSO).
Once user credentials are validated, then the request is completed. In our case the five dates are returned as a JSON array.

Figure 16. Chrome browser calling REST API after authentication to SLC Hub

WRITE AN R SCRIPT USING THE HTTR LIBRARY TO CALL THE REST API

By understanding how authentication is handled by the Hub, we can complete our use-case. Write an R script that calls the defined API.

```r
# ChatGPT3 Prompt: # write an R program to illustrate basic auth for httr library

# Set up the authentication details
auth_user <- "rbetancourt"
auth_pass <- "<HUB Password String>"
auth <- paste(auth_user, auth_pass, sep = ":")

# Encode the authentication details in base64 format
auth_enc <- enc2utf8(base64_enc(auth))

# Create a request with the authentication header
req <- GET("https://smartworkshub-xxxxxxxxxxx.xxxxxxxxxx.xxxxxxxxxx.altair.com/run/generate_dates/RunGenerateDates?start_date=2023-12-01",
  add_headers("Authorization" = paste("Basic", auth_enc))
)

# Extract the JSON content from the response
json_content <- content(req, "text")

# Convert the JSON content to a DataFrame
df <- fromJSON(json_content)

# Print the DataFrame
print(df)
```
The output from the R script executed in VS Studio Code is shown in Figure 17.

**Figure 17. R DataFrame from calling SLC Hub’s REST API Endpoint**

---

**OPEN API STANDARD**

In designing SLC Hub’s ability to publish API endpoints, Altair engineers relied on the Open API Initiative. The OpenAPI Initiative (OAI) is a collaborative open-source project that aims to create, maintain and evolve a standard specification for building and documenting RESTful APIs. The initiative was created in 2015 by a group of industry experts and companies including Google, IBM, Microsoft, and Apigee, and is now governed by the Linux Foundation.

The OpenAPI Specification (OAS), formerly known as Swagger Specification, is the foundation of the OAI. It is a machine-readable format for describing APIs that can be used to generate documentation, code, and other resources automatically. The OAS provides a common language for describing API operations, parameters, responses, authentication mechanisms, and other aspects of a RESTful API.

The OAI aims to promote interoperability and standardization among API providers and consumers. By using a common specification, developers can create and consume APIs in a more consistent, efficient, and scalable way, regardless of the programming language, framework, or platform they are using. The OAI also provides tools, resources, and best practices to help developers design, test, and deploy APIs that conform to the OAS.

To illustrate the standardization offered by OAS, we wrote a SAS language program using PROC HTTP as the method for returning metadata about the SLC Hub example programs deployed at Hub configuration time. This program illustrates the handling of Bearer Tokens as the authentication method.

This program calls the auth API endpoint using PROC HTTP by passing a request body containing username and password key/value pairs. The returned auth string is included as part of the response header in a subsequent API call (all calls for resources must be authenticated) to return metadata about the composition of the Hub’s repositories. See Appendix A: PROC HTTP Calls to Hub’s Open APIs.

**CONCLUSION**

With the large scale use of the SAS language in the pharma and life sciences industry, program reusability is essential. The utilization of SAS Autocall macros as a common method for code re-usability among SAS users and programmers has not changed over the past 30 years. Further, SAS Autocall libraries are not accessible to non-SAS language applications.
In today’s world of multiple languages for reporting and analysis, REST APIs for critical SAS language macros and code templates extend code reusability beyond SAS programmers to include R, Python, Lua and other programming languages. The innovation behind exposing SAS language programs as REST APIs translates directly to increased staff productivity, less redundancy, and less re-work.

REFERENCES
2. Introduction to SLC Hub Administration: https://hubdoc.worldprogramming.com/5EA-2.2.0.108/use/introduction/#hub-administration
8. Introducing ChatGPT. https://openai.com/blog/chatgpt

CONTACT INFORMATION
Your comments and questions are valued and encouraged. Contact the authors at:

Randy Betancourt
Altair
rbetancourt@altair.com
www.pythonforsasusers.com

Oliver Robinson
Altair
orobinson@altair.com
APPENDIX A: PROC HTTP CALLS TO HUB’S OPEN APIs

/* Program variables */
%let url = http://xx.xx.xx.xx:9090/api/v2/auth/login;
/* https://hubdoc.worldprogramming.com/5EA-2.2.0.108/api/auth/restapi.html#operation/passwordLogin */

filename in TEMP;
filename out TEMP lrecl=32767 recfm=v;
filename headerin TEMP lrecl=32767 recfm=v;;

/* Macro to print FILE output to log */
%macro prntfile(file);
option nonotes;
data _null_
  infile &file end=eof;
  input;
  put _infile_
  if eof then put /;
run;
option notes;
%mend;

/* Create parameters to form Request Body in JSON format */
data _null_
  file in;
  put '{
"username": "rbetancourt",
"password": "XXXXXXXXXXXXX"
}';
r

NOTE: The file in is:
Filename='C:\Users\RBETAN-1\AppData\Local\Temp\WPS Temporary Data\TD19812\#LN00002',
  Owner Name=PROG\rbetancourt,
  File size (bytes)=0,
  Create Time=13:27:19 Jan 02 2023,
  Last Accessed=13:27:19 Jan 02 2023,
  Last Modified=13:27:19 Jan 02 2023,
  Lrecl=256, Recfm=V

NOTE: 4 records were written to file in
  The minimum record length was 1
  The maximum record length was 28

/* Create Request Header */
data _null_
  file Headerin;
  put 'Content-Type: application/json, text/plain' /
      'Accept: application/json, text/plain' /
      'Accept-Language: en-US';
r
run;
NOTE: The file header in is:
Filename='C:\Users\RBETAN~1\AppData\Local\Temp\WPS Temporary
Data\TD19812\#LN00004',
  Owner Name=PROG\rbetancourt,
  File size (bytes)=0,
  Create Time=13:27:19 Jan 02 2023,
  Last Accessed=13:27:19 Jan 02 2023,
  Last Modified=13:27:19 Jan 02 2023,
  Lrecl=32767, Recfm=V

NOTE: 3 records were written to file header in
The minimum record length was 22
The maximum record length was 42

42
43    /* Request (POST) for Auth Bearer Token string */
44    proc http
45       url = "&url"
46       method = "post"
47       ct="application/json, text/plain"
48       in = in
49       out = out
50       headerin = headerin;
51    run;


52
53    /* Reformat Auth Bearer Token for passing to next HTTP request */
54    data _null_
55       infile out;
56       length line $ 1024;
57       input line:$1024.;
58
59       token_x = substr(line,11);
60       token = substr(token_x,1,index(token_x,\'\'')-1);
61
62       call symput('bearer_token', token);
63    run;

NOTE: The infile out is:
Filename='C:\Users\RBETAN~1\AppData\Local\Temp\WPS Temporary
Data\TD19812\#LN00003',
  Owner Name=PROG\rbetancourt,
  File size (bytes)=995,
  Create Time=13:27:20 Jan 02 2023,
  Last Accessed=13:27:20 Jan 02 2023,
  Last Modified=13:27:20 Jan 02 2023,
  Lrecl=32767, Recfm=V

NOTE: 2 records were read from file out
The minimum record length was 0
The maximum record length was 992

64
65    /* Update Request Header file to include properly formatted Auth
Bearer Token */
66    data _null_;
file headerin mod;
length hdr_rqst $ 1024;

%let prefix = %nrbquote(Authorization: Bearer);
%let auth_str = &prefix %trim(&bearer_token);

hdr_rqst = symget('auth_str');
put hdr_rqst;
run;

NOTE: The file headerin is:
Filename='C:\Users\RBETAN-1\AppData\Local\Temp\WPS Temporary Data\TD19812\#LN00004',
Owner Name=PROG\rbetancourt,
File size (bytes)=106,
Create Time=13:27:19 Jan 02 2023,
Last Accessed=13:27:20 Jan 02 2023,
Last Modified=13:27:19 Jan 02 2023,
Lrecl=32767, Recfm=V

NOTE: 1 record was written to file headerin
The minimum record length was 985
The maximum record length was 985

%let url =
http://xx.xxx.xx.XXX:9090/api/v2/ondemand/artifactrepos;

/* https://hubdoc.worldprogramming.com/5EA-2.2.0.108/api/ondemand/restapi.html#operation/queryArtifactRepos */

proc http
    url = "&url"
    method = "get"
    ct="application/json, text/plain"
    out = out
    headerin = headerin;
run;

libname get_repo json fileref=out;

libname get_repo json fileref=out;
NOTE: Library get_repo assigned as follows:
Engine:        JSON
Physical Name: C:\Users\RBETAN-1\AppData\Local\Temp\WPS Temporary Data\TD19812\#LN00003

data _null_;  
file log;
set get_repo.alldata(where=(P2 ne ""));
put P2 " = " @32 Value;
run;
_created = 2022-12-14T17:58:33.086846Z
_id = c0c57ded-a17b-4a19-80e0-6059a8720e64
_modified = 2022-12-28T21:43:15.472878Z
allowPrereleaseVersions = true
allowReleaseVersions = true
description = Altair SmartWorks Hub Example Programs
name = Altair SmartWorks Hub Example Programs
_created = 2022-12-29T21:31:15.108314Z
_id = ebc9bce6-5ad1-4e92-a583-9599a62c53bb
_modified = 2022-12-29T21:35:13.085284Z
allowPrereleaseVersions = true
allowReleaseVersions = true
description = Added on 30Dec2022
name = SLCRepo

NOTE: 14 observations were read from "GET_REPO.alldata"