Collaboration Calibration: How SAS® Data Scientists Collaborate Effectively Using SAS® Viya®

Weston McManus and Allison Becker, SAS Institute Inc.

ABSTRACT

SAS® Viya® seems great, but how can teams collaborate with each other to analyze big data? Who owns the data? What about security? And where do the reports live? Is there a best practice? **If you've found yourself asking these questions, you're not alone**—here at SAS, we experienced them first! In this paper, we walk through the steps of how we are establishing a collaborative environment so that your team can get the job done.

INTRODUCTION

The landscape of analytics is changing. Gone are the days of you, the data scientist/statistician, doing analysis in a siloed environment, exporting a graph in Microsoft Excel, and delivering a PowerPoint document. SAS has sought to change this landscape by leveling the three separate mountains of development, modeling, and delivery into one rolling hill. This requires asking questions that **we've never had the ability nor need**ed to previously ask as a data scientist.

Just like your company, SAS has data scientists who are on this journey asking these questions. Our **"Customer 0" d**ata scientists have integrated our workflow with SAS Viya, allowing us to evaluate some of the questions you might be facing.

In this paper we explore the questions of what it looks like to maintain a seamless, organized, collaborative development structure within and outside of our data scientist team in an enterprise environment. We also review the lessons **we've** learned to help quick start your journey toward a modernized analytics landscape.

FIRST THINGS FIRST: WHAT DOES WORKING WITH SAS VIYA MEAN?

WHAT IS SAS VIYA, ANYWAY?

It might **be helpful to first understand the "Wh**at is SAS Viya?**" before asking the "How within** SAS Viya?**". You've heard that** SAS Viya is the future of SAS. As a data scientist/statistician what exactly does this mean? What would life look like working with SAS Viya versus previous SAS technologies?

From a traditional data scientist perspective, SAS Viya is a new way to access the tools needed for exploring, mining, and modeling data. It brings a web-based platform to access similar tools that you might have familiarized yourself with over the years. (See SAS 9.x and SAS Viya Tool List in the Appendix). In addition, SAS Viya provides a seamless interface to SAS® Visual Analytics, which is the main reporting tool (though this has some modeling integrated with the product as well if you have the SAS® Visual Statistics package).

In other words, within SAS Viya, you have a code development environment, a modeling environment, a model managing environment, and a reporting environment all integrated within the same platform. The key to this integrated success is that there is nothing to install locally; user access to the SAS Viya system is as wide as your organization casts the net. The access to the entire SAS Viya suite of tools is potentially available to everyone, security dependent.

Note: If you are not familiar with SAS Viya, please refer to BASIC SAS VIYA TERMINOLOGY.

EVALUATING YOUR FOUNDATION OF COLLABORATION

Before addressing the collaborative benefits of SAS Viya, we need to confirm several questions:

- How is your team established?
- What is the **collaboration model you're currently working with**?
- What is the collaborative vision for your team?

All these questions come down to the overarching and most critical question: Why we do we care about collaboration?

If we can define our end delivery goals and roles, then we can understand our collaborative design and explore SAS Viya capabilities that support this along the way.

In other words, the questions in this section are extremely important for you, the data scientist, to address at this pivotal moment of your analytics landscape modernization. Instead of letting development drive workflow, this is your opportunity to empower your position to provide analytic and/or business decisions seamlessly through you. By becoming involved with the collaboration workflow in the beginning, you can ensure that your results (the what and the why) reach the people (the who) in the way you would want them (the when and the how).

WHAT IS BEING DELIVERED (AND TO WHOM IS IT DELIVERED)?

This is a question that will change as the analytic landscape changes. The expectation of a presentation and a graph is becoming a thing of the past. Consumers expect delivery of results in a consumable, palatable data visualization, often embedded within an interactive report. Delivering in SAS Visual Analytics also supports having one version of the truth and helps eliminate outdated numbers.

As "citizen" data scientists grow across the company, there might be more of an appetite for access to the model itself. There is a growing need for transparency and for some project sponsors to be more involved in understanding a model. For example, if there are other teams interested in seeing your work or other auditing entities that need to review the results of your work due to regulation, to some project sponsors, the collaborative capabilities of SAS Viya to share models can be just as important as the results.

SAS Visual Analytics has been a center piece of the business platform suite of tools for several years now. Since it is well known for its slick data visualization and results **reporting, we won't** address much about the reporting abilities of the product. The only **thing I'll note is that i**t looks better, faster, and stronger in HTML5, and that the security around reports has changed some in the SAS Viya **world. We'll touch on this later.**

As mentioned previously, anyone who has access to a project report has potential access to the rest of the SAS Viya system. In other words, if you plan to deliver your report to be accessed through SAS Visual Analytics, then the rest of SAS Viya can be configured such that the same people can access it. So, the question of what can be delivered expands well beyond the report of yesteryear.

SAS Viya In Action Example: Determining a Delivery Plan

For many **of my team's projects, the** project consumers are key business leaders and often include executives. There are three critical things to think about for us to consider for our environment:

- Can we handle one-time analysis and a continuous delivery cycle?

- Is our delivery method for business decisions flexible enough when data sensitivity becomes an issue?

- How will the business leaders digest the results, and can we provide additional details?

Through these questions, we made it a top priority for our collaboration environment to handle all these scenarios to efficiently get our data and the results of our projects into a business leader's hands on SAS Viya. Project artifacts are created in the SAS Viya environment so that additional content can be leveraged and the initial environmental decision on security, workflow, and delivery locations can be utilized in all future projects. Once you have this vision defined, we recommend that you work with your SAS Viya administration team to execute this vision in a sustainable way.

WHAT ARE THE ROLES AND RESPONSIBILITIES OF THOSE INVOLVED IN DEVELOPMENT?

Now that you have determined the what and to whom something is delivered, you can focus on defining the roles on the project team to make data available, creating models and reports for different projects. As a simple, but non-exhaustive list, the following roles might be involved in your current (or future) projects: you (data scientist), your core team (additional data scientists), your larger team (data prep developers, project manager), the tech support and/or SAS Viya admins, business SMEs, report designers, customer contacts, project liaisons, and project sponsors.

Of these people, how many will need direct access to SAS Viya? To your data? To your model? To your documentation? To your reports? Where do you draw the boundary between roles and how do you utilize security in the technology to support those boundaries? Some individuals should have their access restricted to the reporting layer only, while others need access to the models. Other people might need Read access only versus being able to edit the content.

SAS Viya can provide the ability for anyone connected to the active directory system to connect to any part of the system. It is up to you to decide where and how things should look.

SAS Viya In Action Example: Determining Roles

In our delivery model, we traditionally separate internal collaborators from external collaborators. For our internal roles, we have a small team that wears multiple hats of data scientist, data procurer, report creator, and project manager. This means we have both an open circle of trust and a bus factor (https://en.wikipedia.org/wiki/Bus_factor). As such, all members of our core team have full access to code, model, data, libraries, reports, and all other development artifacts/resources.

For work efforts that include sensitive data, anyone outside of our core team is treated as an external collaborator, including the SAS Viya admins. In other words, we need complete autonomous control over our resources after the initial environment set up that we discussed in the previous section. We worked very closely with the SAS Viya admins to

establish a SAS Viya environment in which the admins, themselves, cannot access certain resources once the data is made available in the environment. SAS Viya provides great flexibility from that standpoint (as discussed in the next section.)

For other external roles, we have established roles for project liaisons and project sponsors (usually executive sponsors). Project liaisons are those that might help with our reports and/or modeling but are prohibited from accessing code and have Read-Only access to the data. Project liaisons have more robust access to our data, models, and documentation. The end consumers have only simplified, streamlined access to reports and the underlying data.

WHAT MIGHT OUR SECURITY MODEL LOOK LIKE?

One of the biggest things to note is that there are two completely independent security models within the SAS Viya world: content and data. A simple example can highlight the major difference **and how it's handled by** SAS Viya.

The team just finished up the project and has created a report to present to the executives. Here are some security-related details:

- Alex has Read access to the project report but does not have Read access to the data. When Alex opens the report, he sees error messages on any content that is associated with this data.

- Betty has Read access to the data but does not have Read access to the folder that contains the text mining report. Betty cannot open the report, nor does she even know that the report exists.

In both instances, neither Alex nor Betty can read the contents of the report. Is this a good thing? Is it intended? Alex might be a design consultant with this project but was never allowed to see the actual data. Betty might be a data procurer who was never given access to the final report because of other sensitive comments contained in the report itself.

This simple example highlights the importance of how answering the previous two questions begins to help establish your security model. Antonio Gianni wrote an extremely informative and detailed paper for the 2018 SAS Global Forum that provides details about SAS Viya security model best practices. (There is a link to this paper in the [REFERENCES](#) section.) The paper is a much better resource than trying to elaborate here.

DAY-TO-DAY COLLABORATION

Wow, that was a lot of questions to answer before we even got into how we data scientists will work day to day. We know that asking the right questions is the best way to accomplish our goal. So, **let's** ask one more big question with a subtle stat joke:
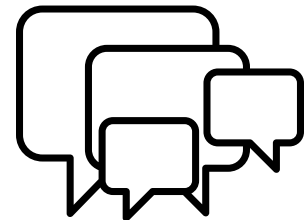
Assume SAS Viya. How do we want our day-to-day collaboration to look?

In other words, no one develops and delivers on an island. Most of us are collaborating with **others before, during, or after the "fun part" of modeling.** Even if you are by yourself, there is a good chance that other people need or want access to the project artifact**s that you've** created. Not only that, but can the same basic workflow apply to the different team members, over different projects, during different times?

For example**, let's assume that** your boss wants you to work on a new project that would take you months on your own, so the two of you need to work together. Or you need another coworker to review your model to make sure that your assumptions are correct. Or

you're tasked to work with Schmidt in Accounting on this project when you work in IT, under different security paradigms.

Collaboration is an undeniable part of the project workflow, and the SAS Viya framework **aims to overcome some of the previous collaboration barriers. Let's** investigate the details.

RESOURCE COLLABORATION

Perhaps one of the biggest pain-points over the years is the ability to provide a seamless location for code, documentation, models, model maintenance, and other resources. How can I share these resources to my team (or to an external team) easily, effectively, and securely? SAS Viya**'s integrative platform** allows us to do **this. I'll highlight two** major parts:

- Projects in SAS® Drive
- Code Development in SAS® Studio

Projects in SAS Drive

SAS Viya has a fully integrated option for general resource collaboration, *projects* Projectswith SAS Drive. Thought the projects idea is still being enhanced, the general purpose is to provide an aggregated spot outside of the folder security structure, for all project resources to reside. Think of projects as a primary space place to aggregate shortcuts to all project content.

Note: This paper discusses projects in a generic sense and SAS Drive projects. For differentiation purposes, from here forward, SAS Drive projects are referred to as *Projects*.

This means that within the Project, resources still maintain the original security model. Regardless of folder structure, and regardless of individual security concerns, a single Project can be shared across all team members. The visibility of the content within the Project is based on the security parameters of the resources themselves; the same Project will display differently based on the user roles.

Let's use the example in which there are three main resources that we have developed for our Project, Master Sales Project. This is shown in Table 1.

Resource	Location	Read Security on the Resource
MyDocumentation.doc	/Documentation/	Alex, Betty, Carl
MyExecutiveVAReport	/Reports/	Betty, Carl
MyModelManagerProject	IT/Models/	Betty, Dave
MyShinyCode.sas	/IT/Code	Alex, Betty
MyVDMML_Model	/IT/Models	Betty

Table 1. Master Sales Project Resources

Suppose Betty, as the team lead who can see everything, creates a new Project in SAS Drive, and shares it with Alex, a coworker, and Carl, an executive.

When Carl opens the Project, he only sees **"My Executive VA Report" and "My** Documentation.doc**",** based on the Read access.

When Alex opens the Project, he only sees the two pieces relevant to his work: **"My_shiny_code.sas"** and **"My Documentation.doc."**

Dave is a resource that is added to the Project later, and his job is to monitor model **performance. He sees the resource "MyModelManagerProject."**

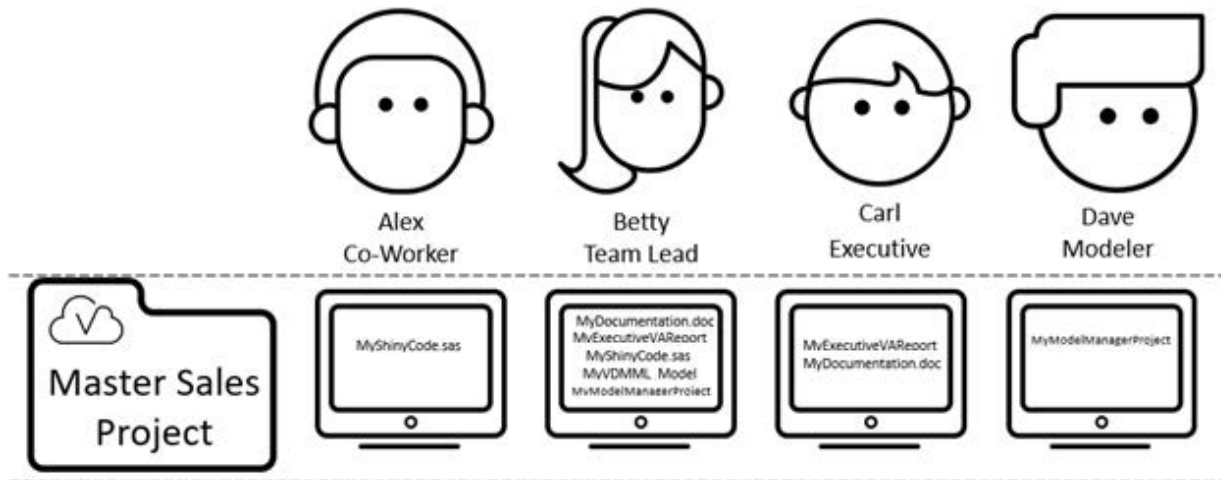These project resources are shown in Figure 1.

Figure 1. Project Resources

To highlight, Projects provide several things:

- One central location that points to all disparate artifacts of the development.

  - Collaborators do not have to search across the system to find associated resources.
  - From an archiving perspective, these Projects are readily retrieved.

- Elimination of a folder hierarchy structure in which all items must be neatly nested. For example, if code exists in multiple spots for different Projects, both can be referenced.

- Elimination of the need for complicated security models. Without having to nest all Project resources, access can be streamlined at a shallower folder depth.

In addition to the organization, Projects also provides the ability for collaborators to comment on the Project itself. If someone makes an update to one of the Project resources, it will appear in the history, with any comments made.

CODE DEVELOPMENT IN SAS STUDIO

Code development is another sticking point for a small team of data folks working on the same Project. SAS Studio for SAS Viya has a few means of collaboration. As mentioned above, there is the ability for code within Projects to be reached directly, just as you would through SAS Drive. However, there are two additional functionalities that are expected to be released toward the end of this year: Snippet/Task sharing and Git integration.

Snippet/Task sharing is the idea that there is a small piece of code, such as a macro or **useful routine, that I want to reuse. A snippet would be "How to assign libref to Sales caslib" or a macro "%macro loop_through_hr_data()." A task** might **be "textmine survey" that** requires input from the end user. When sharing modular code such as snippets and tasks becomes fully integrated, collaboration amongst your team will become much easier.

The Git integration in SAS Studio will integrate the ability for Git users to access/pull/push to the repositories through the SAS Studio interface. (SAS Studio for SAS 9.4 platforms made this available in this November 2018). This will empower users to collaborate using one of the most widely adopted code version control systems.

SAS Viya In Action Example: Resource Collaboration

I don't want to mince words: I love GitHub. **It's my favorite** code development collaboration tool, and **I'm very excited that SAS Studio for** SAS Viya will be integrating Git functionality. This is what I see as the future of code development internal to my team. We currently use GitHub on the SAS 9.4 architecture and utilize code on the SAS Viya architecture anticipating the migration of everything to SAS Viya when the Git integration is available later this year.

As of today, my team uses Project sharing through SAS Drive. We often collaborate with different departments that should not be able to access data from each other.

Our general Project flow is as follows:

1. Create a Project in SAS Drive, for example, **"HR Project 1."**

2. Create a folder under the tree for code to which only my team has access:

   /HR/Analytics Team/HR Project 1/

   Executives in IT do not have access to this folder structure, but my team does. Depending on the security of the Project, we might allow my entire group access or it might be privy to only a few people.

   a. Create subfolders for code, models, and so on.

   b. When you create these files, store them here.

      Note: Though as described, this is not 100% necessary, as files can be shared from many locations.

3. Create/Leverage a more public folder for documentation:

   /HR/Documentation for Analytics Team/

4. When the report is created, store it under the following location:

   /HR/Executive Reports/

5. Within SAS Drive, add all the resources to the Project itself.

Now let's repeat our general Project flow for a new P**roject, "IT Project 1."**

What is the result?

- **Both "IT Project 1" and "HR Project 1" are available in a single top**-level location, Projects.

- The folks in my data team can see both Projects, and the content within both Projects, by accessing the same top-level Project folder locations.

- IT executives **can see "IT Project 1" but cannot see the "HR Project 1"** Similarly, HR executives **can see "HR Project 1" but cannot see "IT project 1".**

- Each executive has access only to what they care about: the final report. **They don't** have to go digging for it.

  - These reports are located under a file structure, such that if the IT executives want to dig through the IT folder structure, they could. But it would be unnecessary.

DATA COLLABORATION AND SECURITY

So, what about the data? How do you set controls in place such that once it leaves the safe confines of your warehouse? For most SAS Viya products (outside of SAS Studio) to access your data, you need to have your project data and model results data loaded into a *caslib*.

Your source data must be accessible by SAS Viya, but ultimately the end consumers will not touch your source data. From a collaboration and security standpoint this is great because the security modeled around your source data can stay locked down to a more limited audience (typically ETL Developers). You can grant additional security to the data, be this access wider or narrower, within the SAS Viya framework and mostly independent of your warehouse security.

For example, suppose you have the database A that houses your data. Depending on your security model, SAS Viya gives you several options to share this data with your consumers. If users do not have (or even if they do have) access to the database A, you can grant permission to have a super user in SAS Viya connect directly to the server. You can then manage the permission at a caslib and/or table level.

SAS Viya In Action Example: Data Collaboration and Security

Continuing our "HR Project 1" example from above, our data generally goes into caslibs designated specifically for HR data. The HR database is locked down to only members of our team. We own and maintain this database, and there should be very little use for accessing it without coming through our team. However, there might be a use case in which HR executives want View access in order to create a report.
Through SAS Viya, we created caslib A that connects to our database. Caslib A is locked down to HR, Executives, and a small team of developers **allowing them "in the door."**

Thus, the database maintains the same security level, but the caslib (and/or tables within the caslib) can be surfaced to the users within SAS Viya. In addition, the individual tables and rows of those tables within caslib A are locked down based on user credentials.

ETL Developers   ETL Developers*
                          HR
                          Executives
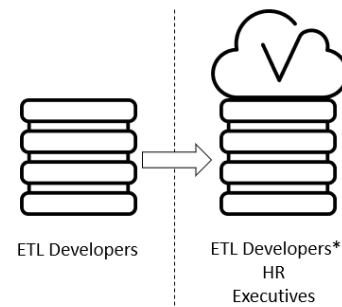
You might say to **yourself at this point "that sounds a bit tricky to maintain."** And "why should I care?" As data scientists, we need to be careful not to expose data within SAS Viya that individuals could not previously access. It can be cumbersome for sure, but SAS Viya makes this process easy. Our SAS Viya admins have set up a script that we use to maintain table-level and row-level security (based on the user IDs in a unified location). If you want to answer the question of whom has access to specific data, a simple SAS query can give you accurate information from one location.

MODELING COLLABORATION

Last, but not least, is the piece this is most relevant to a data scientist/statistician: the modeling piece. Without getting too much into the product specifics, there are two things to highlight: model sharing and The Exchange.

Model Sharing

One of the more game-changing collaborative aspects of SAS Viya is to be able to log on to **someone else's model and pick up where they left off.** This is accomplished by simply selecting the option to share and then selecting the applicable user name.

*Truly, i**t's that simple. Open** the browser to SAS® Visual Data Mining and Machine Learning, right-click your Project and share. Boom. Your coworker has access to your model.* This is shown in Figure 2.
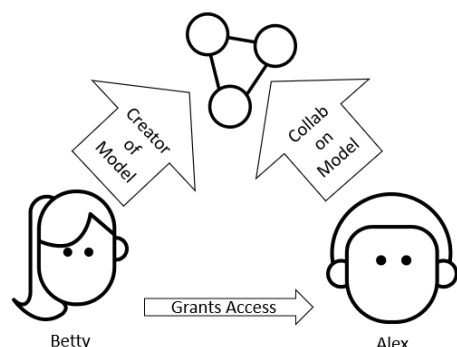

Figure 2. Model Collaboration

To highlight why this is so important, we can visit what this would take in SAS® Enterprise Miner™. It involves either:

- Creating a super user in metadata that analysts can share. All modeling work must be done under the super user ID.

- Cloning the workflow, which involves: exporting/importing a diagram, realigning the data, running the model, and comparing what they did. For anyone who has done this, **it's not straight forward and costs development time.**

At this point in its life cycle, SAS Enterprise Miner is still the champion in terms of a complete armory of analytic tools (though this gap is closing with each new release). However, from a collaborative standpoint, the SAS Viya suite of modeling products will be making a push to enable a future generation of data scientists and statisticians.

Similarly, SAS® Model Manager deploys the same concept as SAS Visual Data Mining and Machine Learning. Suppose that another data scientist, Dave, comes from the business audit team and would like to see the efficacy of the model. You, or anyone else on the team who has Edit access, can simply right-click **and "Share"** the model management Project with Dave. Dave can work on what he needs to, and you can continue your work with little interruption.

The Exchange

The Exchange, a new concept within SAS Viya, is a place to share project templates. What is a project template? **It's a piece of your model that you would like to reuse, share, or** both.

For example, suppose your team uses a process flow (called a pipeline) within SAS® Visual Text Analysis that is very specific to the type of survey that you text mine. If you have created the pipeline for an initial project, you can save this to The Exchange out as a pipeline template for the more junior members on your team to use.

A simpler example might be that you create a node within your SAS VDMML project that took time to tweak to specific settings. Instead of jotting down the settings to re-create later, you can simply save this node template to The Exchange.

SAS Viya in Action: Modeling

We perform a lot of text mining within our team. As any data scientist knows, text mining is **an art. It's complicated, it's tedious, and it's very subjective.** Suppose I create a new text model in SAS Viya and want to have someone else **make sure my art isn't** a fashion show gone wrong. Using SAS Viya, **I don't ask them to either a)** re-create the model or b) come sit in my office for hours to **look at what I've performed. I simply share the resource with** my teammate and send them an email to critique my work. Model collaboration is fantastic.

CONCLUSION

The ability to seamlessly create and share will change the dynamic of the relationships of the people involved in modeling projects. As the data scientist, it is very important to understand how our roles will change, especially as we explore the ability and vision of SAS Viya products. Through asking the right questions and exploring the SAS Viya capabilities, the data scientists of the team will guide their larger team to the correct answers, faster.

SAS Viya aims to make your job a little more integrated and little less complicated.

APPENDIX

SAS 9.X AND SAS VIYA TOOL LIST

The below table is a basic translation of the tools that SAS Supports in 9.X and it's natural Viya analogy. In no way does this reflect all of the capabilities; many of the Viya tools are under development to reach the same maturity achieved within the analogous 9.X tools.

SAS 9.x Tool	SAS Viya Tool
SAS® Enterprise Guide®	SAS Studio
SAS Enterprise Miner	SAS Visual Data Mining and Machine Learning
SAS® Text Miner / SAS® Content Categorization	SAS® Visual Text Analytics
SAS Model Manager	SAS Model Manager
SAS® Stored Process Server	SAS® Job Execution
SAS Forecast Server	SAS Visual Forecasting

Table 2. SAS 9.x and SAS Viya Tools List

BASIC SAS VIYA TERMS TO KNOW

Caslibs

A caslib is defined as "*an in-memory space to hold tables, access control lists, and data source information.*" You can think of it as a libref on steroids. A caslib contains the in-memory data tables that the SAS Viya architecture leverages to be successful.

Like a libref, a caslib can point to any data source to access the data. The CAS engine allows SAS Viya to then load these data sources into memory within the same caslib. So essentially, you have two copies of the data: one referenced, one in memory. SAS Viya products perform most of their work on the in-memory version of this data, utilizing the patented in-memory technology.

For example, assume I have a database with a schema that contains ten data sets of golf scores, MYDB.SCORES_1 – MYDB.SCORES_10. If I create a caslib, GOLFSCORES that references this schema, then I would have a caslib that has ten total files. To use any of these data sets within SAS Viya, I need to then load these files into memory. This effectively doubles the number of data sets within my caslib to 20: ten representing the actual data source tables and ten representing the in-memory tables.

Projects

A Project is a new concept introduced with SAS Drive in SAS Viya 3.3. The idea of a Project is to be able to collect and organize project resources across multiple folders that can be shared.

For example, assume you have multiple resources under /Folder/Structure/One and /Another/Folder/Structure. Both are paramount to the project that you are working on, so it makes sense to have them both be referenced in the same place. A Project can create a collection of shortcuts to point at both. It can also contain links to external resources.

REFERENCES

Gianni, A. and F. Qamar 2018. "Understanding Security for SAS Visual Analytics 8.2 on SAS Viya" SAS Global Forum 2018. Available at https://www.sas.com/content/dam/SAS/support/en/sas-global-forum-proceedings/2018/2130-2018.pdf.

Support for Databases in SAS Viya 3.3,Available at https://support.sas.com/en/documentation/third-party-software-reference/viya/33/support-for-databases.html.

CONTACT INFORMATION

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

Weston McManus
SAS Institute Inc.
Weston.McManus@sas.com

Allison Becker
SAS Institute Inc.
Allison.Becker@sas.com