

## Paper 175-28

## SAS® Metadata, Authorization and Management Services – Working Together for You

Michelle Ryals, SAS Institute Inc., Cary, North Carolina

### ABSTRACT

How does your company manage resources? Are you able to restrict users from accessing certain information defined as confidential? Are you accessing your data directly every time you need to find information? SAS is working together for you to make your business as successful as possible. We have designed solutions to provide ease of use when configuring your environment, users, permissions, and overall metadata management.

One of the key enablers of intelligence that gives you information critical to succeed in understanding and managing your data is metadata. Is your company using metadata? Are you aware of the advantages metadata can bring to your organization? If you are not, you should be. As your business grows, so does the data contained within. Metadata is becoming more essential than ever for those who wish to stay ahead of their data, and thus, ahead of their competition. Speed is of the essence and metadata is the driver.

This paper will give you valuable information needed to ensure you are using every resource in your organization to its fullest potential. We will provide a sample scenario that leads you through understanding the importance of metadata, to setting up your environment via one central application, to establishing permissions that provide protection of confidential information. SAS has been designed to provide customers with ease of use and guidance in making better business decisions regarding organization, management and usage of your resources.

### METADATA 101

What is metadata? There are a number of definitions, but it can be described in the most basic terms as “data about data”. Metadata is information surrounding the data resources in an organization. In addition, it contains descriptive information about the structure and content of data and the applications that process and manipulate that data. If SAS allows you *The Power to Know*<sup>™</sup>, then metadata allows you *The Power to Understand*.

Metadata gives you faster, more focused search and retrieval ability. If you store information about your data you will have easier access and be able to get answers to direct questions quickly. Metadata can be viewed as an essential tool that helps you to maintain information which will in turn allow you to concentrate more on your business, instead of your data.

### BREAK IT DOWN

Metadata is typically broken down into two levels:

- Technical
- Business

Technical metadata supports the development, maintenance and management of an Information Technology Environment. This type of metadata is concrete and normally answers the where and how. Some examples of technical metadata include physical storage structures, server systems, installed applications, and data manipulation processes. Technical metadata will allow you to answer questions such as:

- Where are my servers located?
- How are the servers configured?
- How many servers are available?

- How many server types have been defined?
- How are my data libraries defined?

Business metadata makes the data and services in the environment easier to understand and use. Though it is less concrete than technical, it allows the business analyst to make sound decisions based on the data and normally answers the why. Some examples of business metadata can include data classification, presentation definitions, business meaning and business usage. Business metadata will allow you to answer questions such as:

- Why are sales slipping in the third quarter?
- How are the sales defined?
- Why do we need this data?
- Who is using the data?
- How is a change in data captured?
- Why can't I view this report?

### WHY SHOULD I CARE ABOUT METADATA?

One thing we can all agree on is the fact that information is not beginning to slow down and level out. It seems the definition of advancement involves the ability to digest information at a faster rate. This means the amount of data in organizations is growing at a rapid pace, as well as the number of systems producing this data. Likewise, the number of users who need to access and understand the data are growing. This increase can only create more complexity and less understanding. How useful is your data if it is not understood? It is not to your advantage to have a wealth of data without a wealth of knowledge to help interpret and understand it.

The data resources in your organization can be one of your most valuable assets. Without metadata, these assets can go underutilized because they cannot be found, accessed, or understood. Metadata is the answer to this problem and will allow you to fully utilize these resources. One way this is accomplished is by providing integration in a world where many data sources are talking different languages. Communication is the key. Figure A illustrates how many different data stores could use a common metadata model, which many clients could then query.

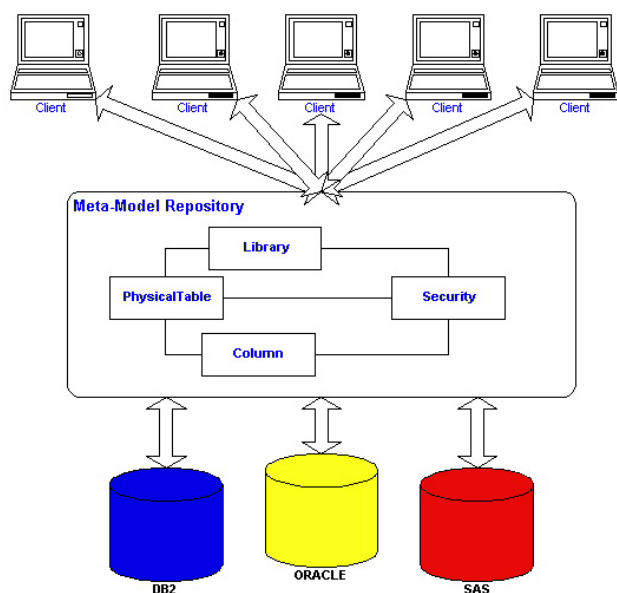


Figure A

Having the ability to query your data through the means of metadata will allow you to answer more direct questions quickly. You will have the ability to know where your reports are stored, when the most recent changes occurred and how they were calculated. You will be able to facilitate impact analysis, which helps explain how data relates with other data throughout the system.

Metadata can offer a common, consistent and secure representation of your data. It is easily accessible and compatible with existing and evolving standards, such as Object Management Group's (OMG) Common Warehouse Metamodel (CWM).

SAS V9.1 is taking advantage of compatibility and consistency ideas and putting them to practice with our own applications in order to provide the best solutions for our customers. Allowing our products to "speak the same language" by utilizing a common metadata model will allow solutions packaging to target your company's needs. No longer are you required to export information from one application and import it into another in order to make sense of it. The SASV9.1 solution removes these complicated steps and allows you to concentrate on solving customer needs.

### SAS WORKING TOGETHER FOR YOU

Most will agree that getting started is usually the hardest part and where our customers spend the majority of their time. SAS has addressed these concerns in V9.1.

How does a company begin to create an environment that contains metadata used to define the environment (i.e. what servers exist, what libraries exist, what users exist, etc.). Likewise, how does a company define rules to help in authorizing who can see what reports, and in establishing basic configuration issues? The answer can be found in the SAS Metadata Architecture, SAS Management Console, and basic guidelines for configuration. Making all of these work together for you so your business can get to business quicker and easier.

### THE SAS METADATA ARCHITECTURE

The SAS Metadata Architecture is a metadata management facility that provides common metadata services to SAS and other applications. This facility is key to allowing SAS applications to integrate, which provides a common, consistent

and secure representation of metadata. Allowing SAS applications to share metadata allows for better solutions.

The architecture itself consists of interface, model, supervisor, and repository engine layers.

- **The Metadata Interface**, or Application Programming Interface, provides utilities for maintenance of metadata from a client application. XML is used as the transport mechanism to pass information between the client and server. Support for the XML transport format and industry standard metadata models increases the likelihood of compatibility between SAS applications and other software applications.
- **The Metadata Model** defines the common semantics for the types of metadata, the attributes and how the individual pieces of metadata relate. A definition of common metadata types allows integration between SAS products. These metadata types are created to store information about your data in a repository.
- **The Metadata Supervisor** provides runtime services and offers authorization to the metadata. The supervisor is taking advantage of a multi-user, multi-threaded design using an in-memory cache and allows for surfacing of metadata from one or more repositories.
- **The Repository Engines** layer provides interfaces to relational databases and other repositories.

Figure B defines the SAS Metadata Architecture. You will notice the Metadata Manager, which allows you to manage your resources, metadata, and other critical information. The Metadata Manager is available through the SAS Management Console and will be discussed in detail later.

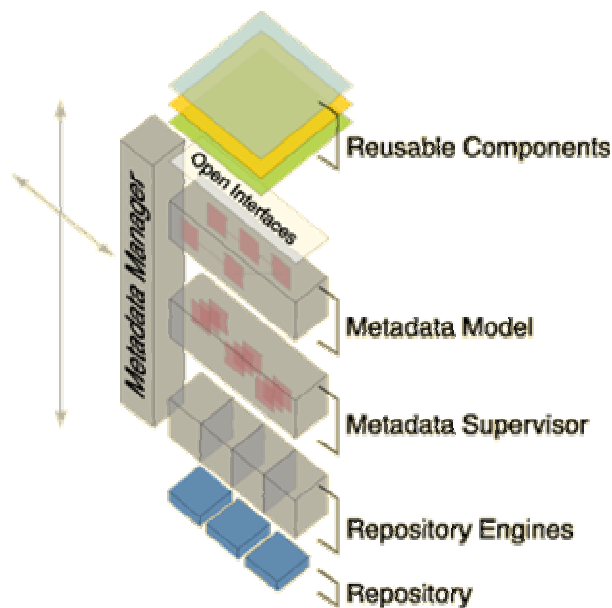


Figure B

Again, because we now have a foundation in SAS allowing our applications to "talk to one another in the same language" we are able to surface metadata to the customer more consistently and faster. Our customers can now reuse metadata created by other applications, import metadata from external sources (i.e. CWM Common Warehouse Metamodel), facilitate impact analysis to determine which programs are dependent on which reports, and identify resources and reports.

## MANAGEABILITY WITH SAS

In the past you probably determined the number of items in your environment by determining the number of configurations you had to deal with in order to get up and running. SAS has attempted to help with this situation by providing the SAS Management Console. The SAS Management Console is a new Java framework that provides a standardized interface for managing various SAS resources – most importantly metadata. It is this metadata that is used to manage these resources. Specific administrative tasks are supported via plug-ins to the SAS Management Console.

There are multiple roles using the SAS Management Console. Namely, a Systems Administrator or a Project Leader. The Systems Administrator is typically responsible for installing SAS software and SAS Servers, defining the needed servers, connections and then maintaining them, and creating user and group definitions. The Project Leader is typically a database administrator or project leader who understands key concepts for OLAP, ETL Studio, etc. They will typically be maintaining and managing cubes, warehouses, etc. for others to use in reports. In addition, the Project Leader would be the one authorizing user/group access to resources. Figure C helps explain SAS Management Console Usage.

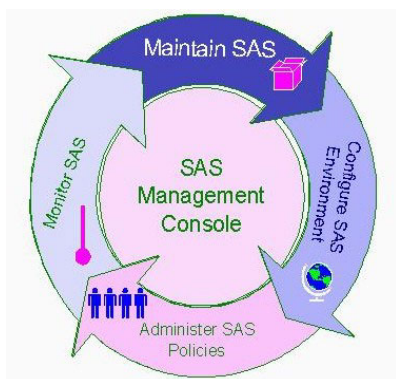


Figure C

Having “one stop shopping” for metadata management can add many benefits to your company. Providing an enterprise class single point of control for SAS administrative tasks will cut down on configuration headaches. In addition, a central place for defining the SAS environment while using an extensible framework which allows every SAS product to expand the console to administer its applications. Not only are SAS plug-ins supported, user and third party written plug-ins are available. Therefore, if you have a specific environment that requires administrative tasks, the SAS Management Console can plug-in that capability.

Figure D highlights different SAS Management Console capabilities. We will explain many of these, and others, in detail. Additional SAS Management Console plug-ins can be delivered within SAS products based on individual product needs.

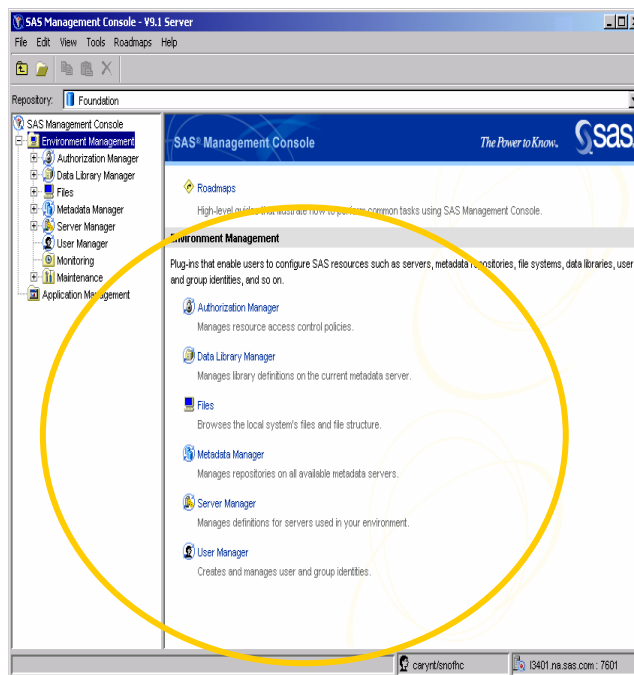


Figure D

**The Metadata Manager** portion of the SAS Management Console is responsible for creating and managing repositories and servers. You have the ability to administer a metadata server (i.e. stop, pause, etc.), as well as create the repositories residing on that server. Once the repositories are created, there is still a level of management that can take place.

The metadata manager also provides change management and import/export interfaces. The change management interface allows users to make changes to metadata, while maintaining a history of these changes. Information regarding any changes to metadata is persisted in the metadata itself for future reference and reporting capabilities. The import/export interfaces allow metadata to be processed from different definitions (i.e. Common Warehouse Metamodel).

With the Metadata Manager, users have the ability to perform replication and promotion of metadata between metadata servers.

**The Server Manager** allows you to define various types of servers, as well as information necessary to bind to those servers. The Server Manager can be particularly useful in an environment where there are multiple servers in use.

**The User Manager** allows you to define and manage users and groups of users. When we create authorization rules on specific items of metadata (e.g. servers, library, objects) we are establishing privileges for users that are defined via the User Manager.

**The Authorization Manager** allows you to create permissions, access control templates and define resource authorizations. The permissions defined here are provided to the users and/or groups of users defined via the User Manager.

**The Library Manager** allows you to manage your SAS Library definitions, as well as database schema definitions.

**The License Manager** allows you to store metadata that describes licensed SAS products. This can be helpful when determining what products/solutions are installed in your



environment, where they are installed, and when your setinit expires.

Figure E will help illustrate the many ways SAS Management Console can help with administrative tasks in every organization.

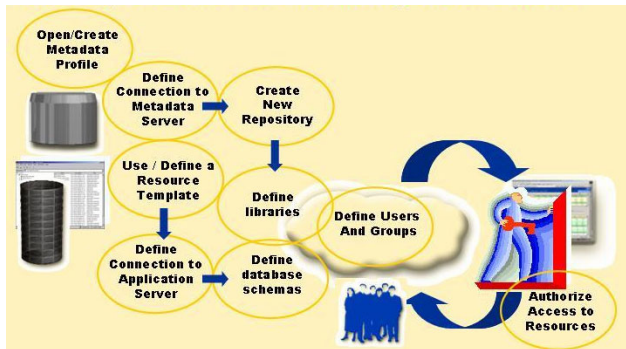


Figure E

### PULLING IT ALL TOGETHER

Let's take what we have learned and walk through an example to show how we would put this information to use. We will start at the end result and work our way backwards.

#### A SAMPLE BUSINESS SCENARIO

Do not categorize metadata as only useful in the form of Information Technology. Metadata is excellent at empowering the best and most accurate business decisions. Any detective can solve the case if metadata has been defined properly. Therefore, let us begin an investigation.

We can start with an executive of a successful company with regional offices in a number of cities throughout the United States. In order to stay on top of things our executive, Bob Billion, always begins his day reviewing the latest sales reports. One day in particular Bob notices what appears to be a huge error in sales reporting for the southeast region. Figures F and G are snapshots taken from SAS Report Studio, which is just one of the many products in SAS V9.1 using the SAS Metadata Architecture in order to share metadata across SAS applications.

Figure F demonstrates sales for the different region, while Figure G concentrates on sales for the southeast region. In particular Figure G shows sales for a period of one year, where the most current month of the current year is illustrated first, followed by months from the previous year.

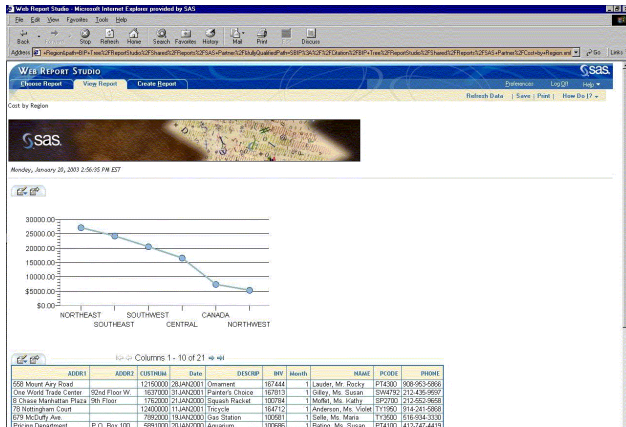


Figure F

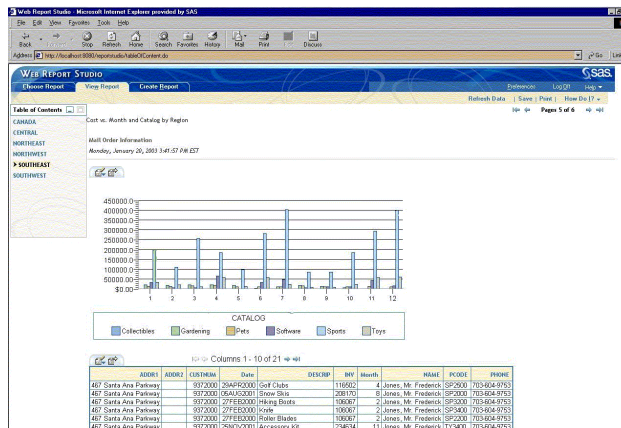


Figure G

Notice in Figure G the amount of sales falling from \$40,000 in December to \$20,000 in January of 2003. This unexpected result leads Bob to ask: "Can this be true?" What could have happened between last month and this month to cause a dramatic decrease in sales for this region? Metadata will allow you to answer this question. Not only can metadata give you the answer, but it can also help you make better business decisions by allowing you to see the big picture. This is the true advantage of being able to understand and manage your data.

#### BEHIND THE SCENES WITH METADATA

There are a number of metadata objects that contain information relevant to this investigation. Figure H is an example taken from the SAS Management Console, which includes metadata tables defined for a particular library called Sales. The library itself is part of an Analytics server, also defined via SAS Management Console.

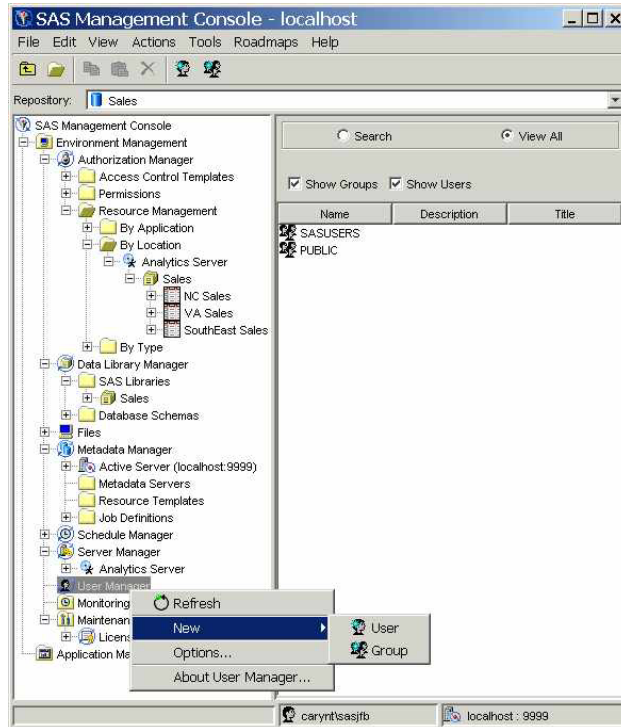


Figure H

One of the first steps the administrator will take in configuring an enterprise involves defining servers and repositories. Each server can contain multiple repositories. Each repository is usually defined for a specific purpose. In our example we have

one repository named *Sales*. We could choose to define other repositories as they relate to other areas of our organization (e.g. personnel, inventory, etc.). Figure I is a screen shot taken from the Metadata Manager plug-in of the SAS Management Console. SAS has made the creation of repositories easy by guiding you through questions to help the server know where the repository is located, what it is named, and what SAS engine it is using to access metadata.

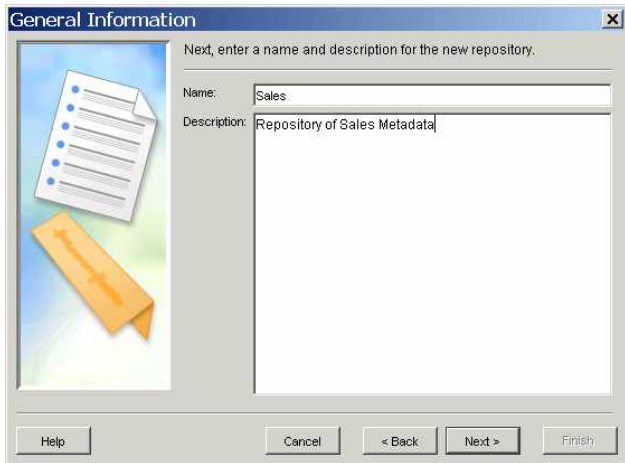


Figure I

The metadata tables defined in the repository for our sample scenario (*NC Sales, VA Sales, SouthEast Sales*) corresponds to the metadata types defined in the metadata model, which we explained earlier is part of the SAS Metadata Architecture.

**BEHIND THE SCENES WITH AUTHORIZATION**

We also want to help you understand the management and authorization side of metadata in this sample scenario. We know this report is generated from metadata, but a few questions that you may want to ask include:

- How did the report get created?
- Who is able to view the report?

Again, the metadata is consistent with the metadata model described in the SAS Metadata Architecture, but the actual authorization pieces are defined through permissions established for the users and/or groups of users via the SAS Management Console. You will notice in Figure H we highlighted the User Manager area where a new User or Group of Users can be defined.

**USERS AND GROUPS**

In our sample scenario we have an executive, an administrator and an administrative assistant. We have the ability to define different permissions for each user and each group of users. Figure J shows our definitions for our purposes in this scenario.

Name	Description	Title
SASUSERS		
PUBLIC		
Executives		
Administrators		
Bob Billion	Big Dog	Executive Chief
Joe Admin		Administrator
Mark Brown		Executive Admin Assistant

Figure J

It is important to understand the flexibility available when creating these permissions. There are two different types of definitions available, Groups and Users. A group can contain definitions of individual Users and/or other groups. A User is defined as a single individual. Permissions are available to Users and Groups.

In this example, we have Users *Bob Billion, Joe Admin* and *Mark Brown*. The *Executives* group, depicted by a "group" of heads, includes Bob Billion and the *Administrators* group includes Joe Admin. Mark Brown is not quite ready to be included in either of these groups and currently only has read permissions.

Client applications have the ability to restrict access to reports depending on permissions. For example, our Sales report in Figure G is only available to the executive and/or group of executives. The metadata tables themselves are available to our administrator and/or group of administrators. When we are creating our users and groups we must take this type of information into account. Who should be authorized to see what reports? Who will be required to make changes to the actual metadata tables for management purposes?

As the authorization setup begins, we start with a default that allows all users, *PUBLIC*, to read and write metadata. We can easily deny public any access and strictly define individual users with targeted permissions. The next series of figures helps demonstrate this idea.

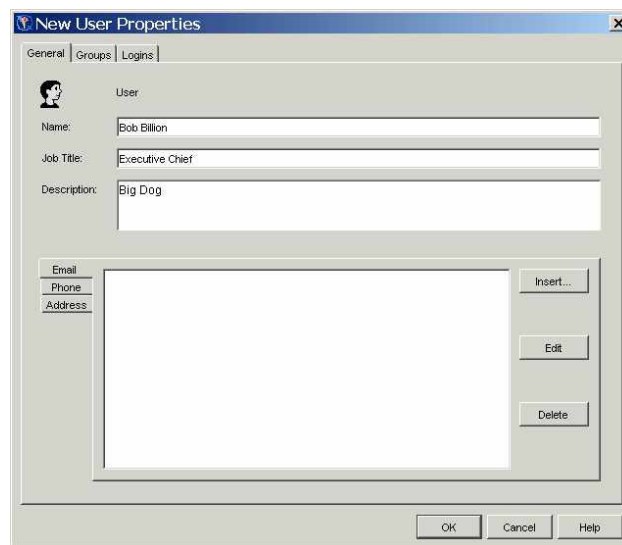


Figure K

We can see the actual creation of the user Bob Billion in Figure K. In addition, Figure L identifies Bob's login credentials. As the administrator is creating users and groups, he will be asked for login properties for each user. Once the individual users are defined, they may then be added to a group of users. Bob Billion is one of the users currently included in the *Executives* group.

**LOGIN PROCESSING**

The SAS Metadata Server depends on the host operating system to authenticate users when they connect to the server. When an authorization request is made, the authorization facility gets the host-authenticated domain and user ID of the connection making the request and searches the *UserID* attribute of all of the Login objects on the server for one with a matching domain and user ID. When a match is found, the identity object (Person or IdentityGroup) that owns the Login object is used as the primary identity on which an authorization decision is made.

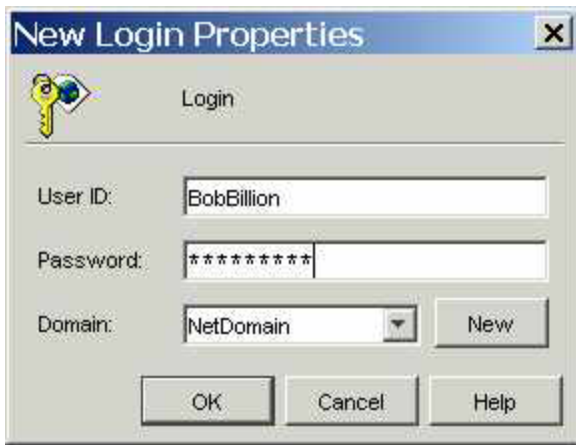


Figure L

The user ID stored in the *UserID* attribute of a given Login object should exactly match the host user ID. For example, on Windows, if the user ID is domain-qualified, then you would enter a domain-qualified user ID in the *UserID* attribute of the Login object, in the form "domain\userid" or "userid@domain", in addition to specifying the domain in the *Domain* attribute. If an identity has access to multiple domains, then a separate Login object should be created representing each domain.

If multiple people need to use the same user ID to perform a particular task, then an IdentityGroup should be created and the login to be shared should be owned by the IdentityGroup. All members of the group will be able to access the Login object, but only those with WriteMetadata permission on the IdentityGroup itself will be able to update it.

**PERMISSIONS**

Next we want to determine who has the ability to see what. We use SAS Management Console to create permissions. As we view the properties of our Sales library, shown in Figure M, we'll be able to explore the permissions currently defined.

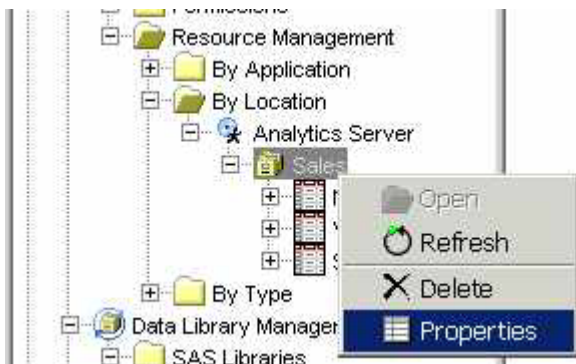


Figure M

Permissions can be set on a variety of levels. The administrator can set permissions at the server level, all the way down to the object level. This provides the maximum level of authorization for all client applications. Initially the *PUBLIC* group has the ability to read and write metadata. The *PUBLIC* Group contains all those individuals not defined in Users and/or Groups. For our sample scenario, see Figure N, we have denied access to all permissions for the *PUBLIC* group. Again, this is discretionary, but we decided we would like to define permissions for specific Users and Groups.

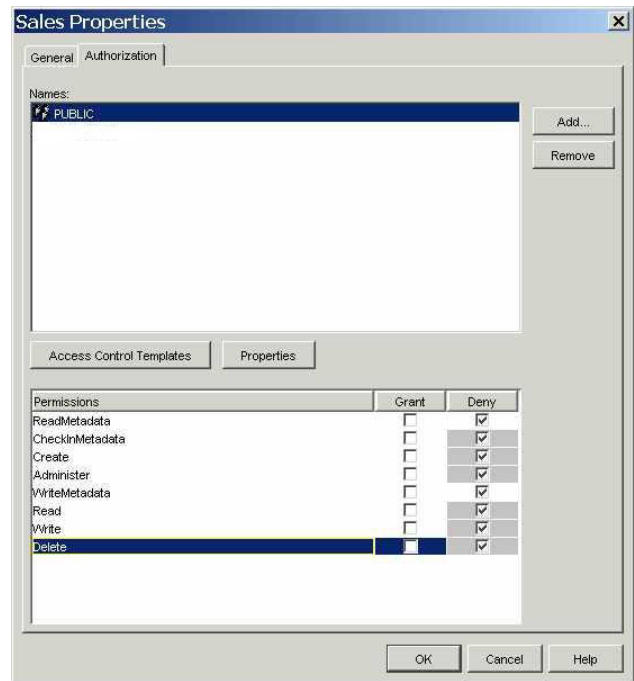


Figure N

Therefore, as Figure O illustrates, we have opened a variety of permissions to the *Executive* group of users. You will notice however, these users do not have the ability to administer or delete metadata. Though we have no problem with the executive folks viewing and creating, we don't want them to have the ability to destroy.

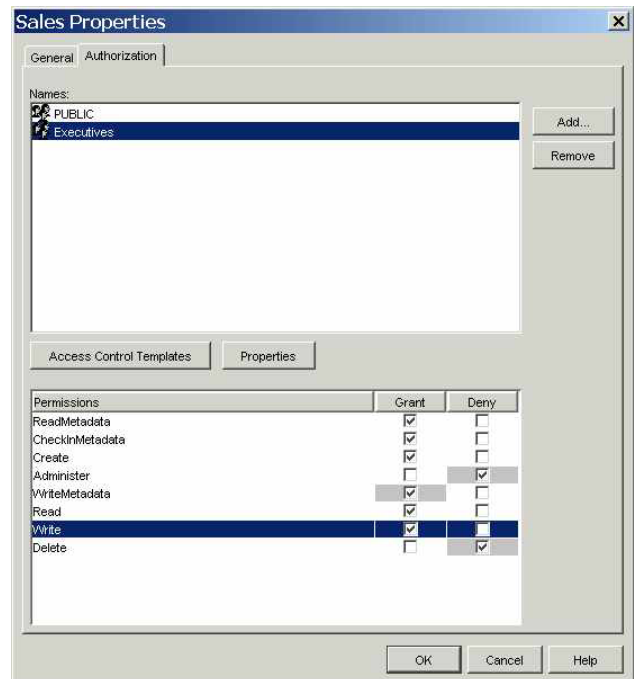


Figure O



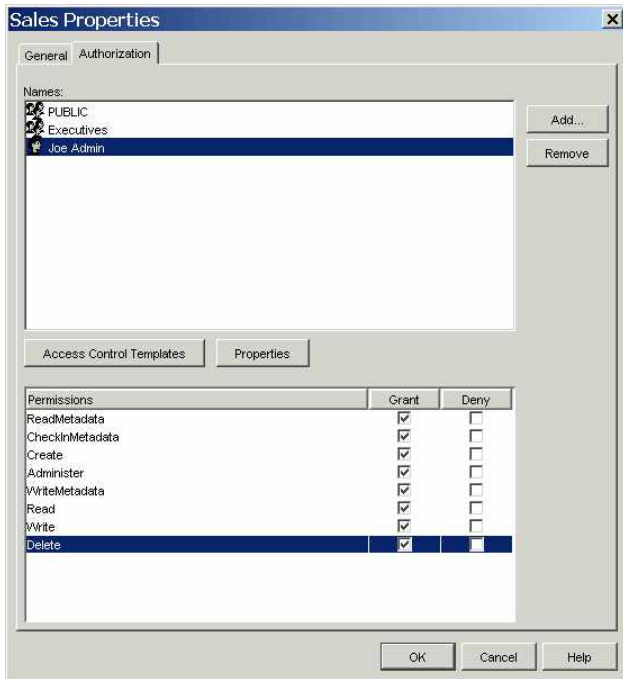


Figure P

We also want to show in Figure P what permissions are available to Joe Admin. Since Joe is in charge of these reports and ensuring they are running correctly, he must have full access.

### BEHIND THE SCENES WITH BUSINESS DECISIONS

Metadata has been carefully defined in the SAS Metadata Model to provide the smartest information, which will allow you to make the most out of understanding your data. For example, the *DateCreated* and *DateUpdated* attributes exist on each metadata type within our model and were established to answer when reports were defined and when they were last generated. Identifiers are established to create uniqueness throughout the metadata, and descriptions are included to define data items such as units of measure.

As we return to our sample scenario, there are two questions Bob decides to ask to determine the accuracy of this report: "When was the report created?" and "Who is responsible for creating the report?" Perhaps the report was generated several days or weeks ago. Depending on the reporting schedule, this information could be obsolete. Perhaps the data lineage information reveals that more than one table is involved and an error has occurred in one of these tables. There is also a chance that an error occurred during the actual generation of the report.

Understanding the data lineage can be the key to surfacing and solving business problems. Figure Q illustrates the data lineage for the *SouthEast Sales* metadata table.

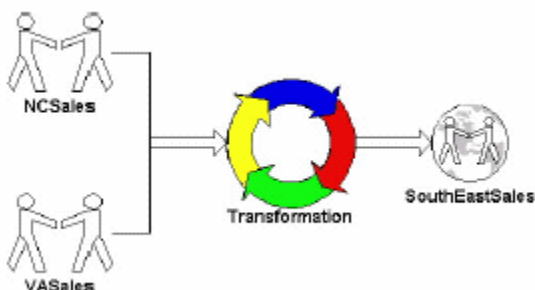


Figure Q

The transformation itself is a formula which defines what steps need to be taken to create the *SouthEast Sales* metadata table. This transformation is stored as metadata. Therefore, the *SouthEast Sales* report is created by reading the transformation information from metadata, which in turn generates the code necessary to generate the report.

As the metadata is studied in detail, Bob determines the report was generated within the last few days. Bob was able to determine this from the *DateCreated* and *DateUpdated* attributes in the *SouthEast Sales* metadata table. Retrieving this information consists of issuing a simple query to the metadata. Since Bob cannot determine any obvious problems, he attempts to answer the second question required to determine the accuracy of the report: "Who is responsible for creating this report?" Again, we can use the metadata to determine that Joe Admin should be given a call.

### THE INVESTIGATION CONTINUES

To determine if the sales are really as bad as they look on paper, the first question Joe decides to ask: "Is the report accurate?" The investigation so far has proven the numbers to be recent, but that does not prove accuracy.

Joe decides there are two questions he should ask in determining accuracy: "How was the value of *Sales* in the *SouthEast Sales* report calculated?" and "Have there been any changes to the report?" It is important to know how the value of *Sales* was calculated. What tables did the report retrieve information from? Is there an obvious problem with one of the tables?

It could be that one of the reports changed which in turn lead to an inaccurate calculation. This could occur with a change in a unit of measure or the definition of the unit of measure. For example, if a "package" definition has changed from one dozen to one half dozen and we continue to sell the same amount then our "quantity sold" would double. Does this mean our revenue doubles? No, and in fact it will remain the same, but this may not be obvious by glancing at the report. When these definitions change your report can take on a completely different face, which could lead to inaccurate assumptions. Defining descriptions is another example of where metadata can help you make informed decisions.

Metadata will allow Joe to conduct quick queries to determine how these numbers were calculated. We go back to our transformation for this type of information. Joe was able to determine that the *SouthEast* region consists of two sales areas, *NCSales* and *VASales*.

### CULPRIT REVEALED

The transformation metadata has allowed Joe to successfully answer his question regarding data lineage: "How was the value of *Sales* in the *SouthEast Sales* report calculated?" His query result included the *NCSales* and *VASales* metadata tables.

A closer examination of these metadata tables will answer his second question in determining accuracy: "Have there been any changes to the report?" After comparing the results of sales information, let's say he discovers that sales from both states were cut in half. The sales numbers are cut in half, but there appears to be no apparent changes to the report that would indicate a problem with calculations. Does this mean the calculations are accurate? Well, technically yes, but there still must be a reason the numbers are lower.

If we take a look at a different view of the data lineage, we will have the ability to dig into different areas of the "big picture". The business view of the data lineage in Figure R will help Joe focus on metadata queries touching all areas surrounding the *SouthEast Sales* report.

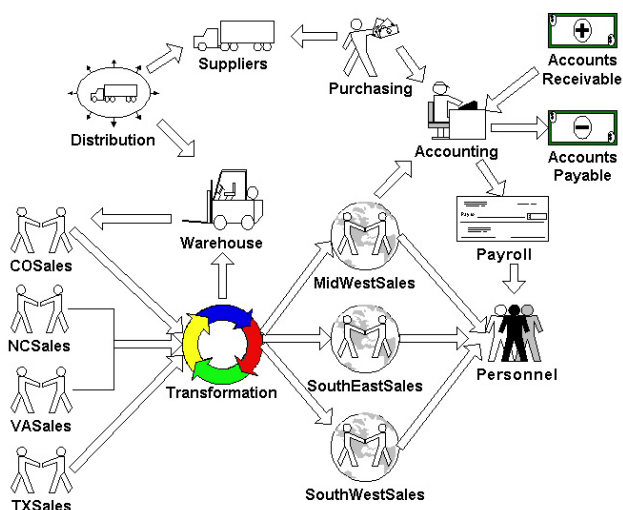


Figure R

There are a variety of spokes defined within the larger scope of metadata. This will provide Joe with the information needed to trace the business problem directly as opposed to combing through tedious data. There is no doubt that Joe will have a much easier time making sense of this using the metadata that has been defined.

### THE VERDICT

Through a series of metadata queries, Joe is able to understand why the SouthEast Sales report shows a decline in sales. It turns out that the shipments to the stores in the North Carolina and Virginia areas were coming from the wrong warehouse. The company had recently purchased a warehouse in the Florida area, but shipments to the North Carolina and Virginia stores were coming from a warehouse in Colorado. What Joe discovered was a potential logistics problem and was able to discover this through the use of metadata. Since shipments were coming from Colorado, they were taking much longer and supplies were running out. This didn't mean the sales were down, but rather the supplies were down. The metadata allowed Joe to make technical as well as business analysis, which in turn allowed Bob to make some decisions regarding their shipping policies. Changing the shipping to come from the Florida warehouse was a logical choice. The end result was a savings in costs and sales.

### CONCLUSION

Many companies today are faced with large "upfront" costs. The time to set up an environment which involves configuration, authorization, organization, and manageability issues can be extensive. Once the environment is set up, there are additional maintenance costs associated with running your business. SAS addresses these customer pains and more.

The solutions available with SAS include the integration of SAS products to allow for sharing of information with ease. The SAS Metadata Architecture is the infrastructure for this solution and can be described as a general purpose metadata management facility that provides common metadata services to SAS and other applications. This facility allows SAS applications to integrate, which provides a common, consistent and secure representation of metadata.

One aspect of this integration of SAS Products is that there is no longer a requirement to set up each individual SAS product with a different set of rules. SAS provides the SAS Management Console to allow for a single point of interaction for all SAS

products. This solution provides an enterprise class single point of control for SAS administrative tasks, which includes definitions of authorization rules and overall manageability of your enterprise resources. This central place for defining the SAS environment uses an extensible framework which allows every SAS product to expand in order to administer its applications.

By defining authorization rules within the SAS Management Console, client applications have the ability to restrict access to reports depending on permissions. Permissions can be set on a variety of levels ranging from the server level, all the way down to the object level. This range will provide the maximum level of authorization for all SAS applications.

SAS is working together for you by providing integration. Integration provides a common, consistent and secure representation for all SAS products, as well as the power to reuse information. Our business is to provide integrated solutions which allows you to have more time to concentrate on your business.

### CONTACT INFORMATION

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

Michelle Ryals  
 SAS Institute Inc.  
 100 SAS Campus Drive  
 Cary, North Carolina 27513  
 (919) 531-5671  
[michelle.ryals@sas.com](mailto:michelle.ryals@sas.com)