

Paper 199-2007

Best Practices for SAS[®]9 Metadata Administration

Diane Hatcher, SAS Institute Inc., Cary, NC

Ken Matz, SAS Institute Inc., Cary, NC

ABSTRACT

SAS[®]9 provides several tools to help with metadata administration, but you must implement internal processes to optimize the usage of these tools. This paper provides some process best practices for backing up your repositories and managing changes in the SAS[®]9 environment and discusses how you can leverage SAS[®]9 tools as an integral part of the process. A brief overview of the road map for future enhancements is also given.

INTRODUCTION

Before you deploy your SAS[®]9 environment, you must determine a plan for how the environment will be used and maintained. Your organization's expected usage patterns and processes will help to determine the appropriate configuration for your environment. This is very similar to things you consider when you are building a new house—you need to think about your needs and how you expect to live in the house. Designing a blueprint upfront allows you to ensure that you get the house right on the first try. You do not really have the luxury of building it again if it is not quite right. The same principle applies to the metadata server; you must establish a blueprint before you even start to put up the foundation.

When you design a house, there are things that are very basic; for example, you know you want a kitchen and at least one bathroom. Other basic features include the number of bedrooms and bathrooms, whether you want a dining room, living room, and/or great room. Extras could include a media room, basement, and study. If you think about it, rooms are just containers for furniture (content for your house). With the SAS[®] Metadata Server, you have to go through similar exercises to determine what type of folder structure (rooms) you need to house your content (furniture).

Other things you need to consider for a new house include planning how each room will be accessed and who should have access to each room. Should the master bedroom be on the first floor or on the second floor? Should the garage lead into the laundry room or the great room? Do I want the study next to the media room? How do I keep the children away from the antique china cabinet? Answers to these questions determine how you will structure your house. Before you deploy the metadata server, you must ask yourself similar types of questions. Who will access the environment? How will users get to the content? How do I ensure that users have access only to what they are allowed to see?

The SAS[®] Enterprise Intelligence Platform includes components (building tools) in the following categories:

- **Data Integration and ETL**
The data integration and extract, transform, and load (ETL) components enable you to consolidate and manage enterprise data from a variety of source systems, applications, and technologies. Components are provided to help you cleanse, migrate, synchronize, replicate, and promote your data. Metadata for all your intelligence resources is stored centrally and controlled through a single management interface.
- **Business Intelligence**
The business intelligence components enable users with various needs and skill levels to create, produce, and share their own reports and analyses. Easy-to-use interfaces enable users to obtain answers to their business questions while system administrators can still retain control over the quality and consistency of the data.
- **Analytics**
SAS offers the richest and widest portfolio of analytic capabilities in the software industry. The portfolio includes solutions for statistical data analysis, data and text mining, forecasting, econometrics, quality improvement, and operations research. You can use any combination of these tools with the SAS Enterprise Intelligence Platform to add extraordinary precision and insight to your reports and analyses.
- **Intelligence Storage**
The intelligence storage options are optimized for analytical processing. This enables you to quickly retrieve and report on large volumes of data. The options include simple relational databases, a threaded multidimensional database that supports online analytical processing (OLAP), and relational storage with a threaded multiple input/output (I/O) subsystem for intensive use by focused applications.

The outputs of these components come together in the SAS Metadata Server. Therefore, it is very important to consider how the platform will be leveraged to determine the best way to configure the metadata. This paper puts forward a methodology to help you identify those questions you need to consider before deploying the metadata server in your environment. Every organization will have different detailed requirements, but the basic areas to consider are shown in Figure 1:

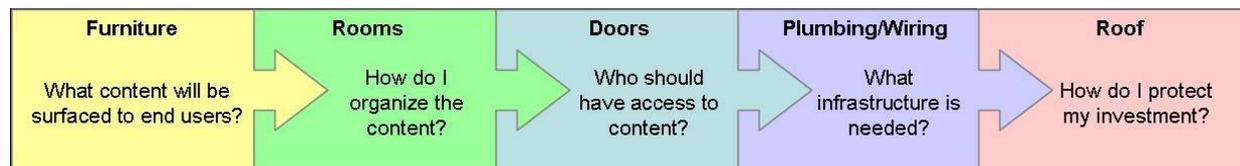


Figure 1. Planning a New House or Planning for SAS Metadata Deployment

Once all of these areas have been considered, you will be ready to build your house and deploy the metadata server.

FURNITURE: WHAT CONTENT WILL BE SURFACED TO END USERS?

As is the case for many projects, planning should start with the *end* in mind. Even when designing a new house, you often start with what you know—what furniture will have to fit in the house, for example. Particularly, if you are moving from an existing house to a new house, you have a set of “legacy” furniture that you want to keep. You might also know that you will have new furniture, like a 50-inch flat-screen plasma HDTV for which you need to design a proper place. In other words, you need to know the inventory of all the items that need to fit as a factor in designing the house.

The same thing applies to planning for SAS Metadata Server deployment. You need to determine an inventory of content to be surfaced to end users from the SAS environment. What reports will be generated? What data sources will be required? Because SAS® has an integrated platform, you need to consider both Business Intelligence (BI) content and Data Integration (DI) content.

BUSINESS INTELLIGENCE CONTENT

BI content consists of a variety of objects that provide specific information to one or more end users. Types of BI content include:

- Web reports
- stored processes
- information maps
- data explorations
- publication channels and archived result packages
- analytical models

DATA INTEGRATION CONTENT

DI content consists of metadata related to the physical data sources that are available in support of BI content. This includes both the data structures themselves, as well as the processes required to source the data. Types of DI content include:

- relational data sources
- OLAP cubes
- DI jobs used to extract, transform, and load the physical data structures
- transformation definitions

ROOMS: HOW DO I ORGANIZE THE CONTENT?

Let us return to planning for the new house. Once you have an inventory of the furniture, you can start to group the pieces together to fit into specific rooms. Bedroom suites would stay together, and dining room tables and chairs would be grouped together. Groupings would be based on context, relevance, and usage. In addition, it is also important to think about how you will maintain the rooms. How should the furniture be set up in the room to facilitate cleaning and maintenance? It is important to design the rooms carefully in the blueprint, as it will be difficult to change them after the house has been built.

You should consider the same things for organizing content in the metadata server. It is important to set up the folder structure (rooms), as you need them, the first time since it will be difficult to adjust them once your end users start adding their own content.

USAGE

The starting place to group content is to establish an initial folder structure based on separation of business concerns. Separation of business concerns implies patterns of security that can be maintained to allow similar access to one or more business groups to content within a single folder or multiple folders that is relevant to those groups. Proper separation of business concerns means that the content is meaningful and properly accessible to everyone in those particular groups. For example, content that is accessible by the Finance department could be in one or more folders. Those folders should not necessarily be accessible by the Sales or Human Resources departments. Likewise, sensitive data such as employee salaries should probably be restricted to access only by the Human Resources Department and restricted from the Finance and Sales departments.

It is also a good idea to allocate non-public folders (rooms) for content (furniture) that will be consumed by only a few specific users as well as folders that are more public for content available for a wider audience.

Consider Figure 2. This image depicts a simple folder structure where corporate content is split out into three folders: **Corporate**, **Dept A**, and **Dept B**. The content in the **Corporate** folder would be relevant to everyone, whereas the content in the folders **Dept A** and **Dept B** would be relevant only to the respective departments.

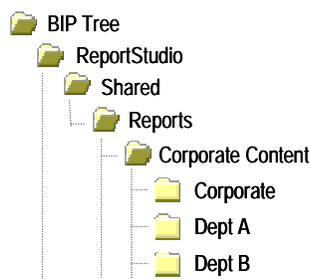


Figure 2. Example Metadata Organization Showing Separation of Business Concerns

At this point, the folder structure does not explicitly dictate security, as this is just a means for segregating content. However, the closer the folder structure can conform to security requirements, the easier it is to establish and maintain access controls. *Security* is discussed below, under the section “Doors: Who Should Have Access to Content?”

For Data Integration-related content, you should also consider whether the content is in support of data staging or for surfacing information in reports. As part of the ETL process, it is common to have physical data structures and jobs that are used in the process to generate end-user consumable information. This content typically does not need to be surfaced to end users directly and should be segregated into folders separate from more relevant data sources.

CONTEXT AND RELEVANCE

Once the initial folder structure is in place, you can now group specific content together based on context and relevance. *Context* refers to the subject matter, such that all objects in a single folder discuss the same business question. *Relevance* means that the content is meaningful to everyone in a particular audience. For example, content relevant for **Dept A** could be in one folder, and content relevant for **Dept B** could be in a different folder.

Context and relevance lead to usage. How will objects be used in conjunction with each other? Complementary usage is another angle that can determine how to group objects together. One example is grouping a report and its associated information map in the same folder. The two items are always used together to generate the appropriate output, so locating them together in a single folder makes both logical as well as business sense. Another example is having a report on actual sales for the current time period that is usually used with a report on sales targets. It would make sense to group these reports in the same folder.

In Figure 3, we can now extend the metadata folder structure from the example above.

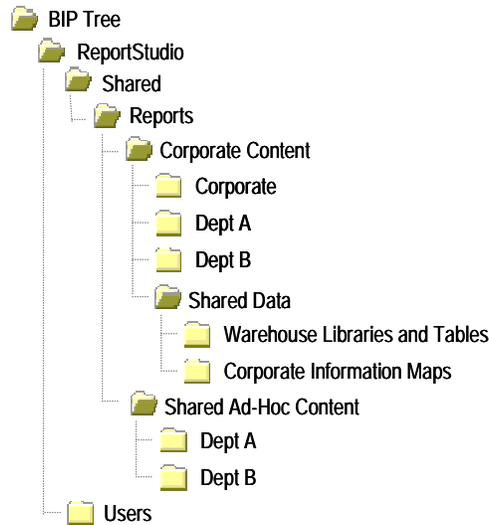


Figure 3. Example Metadata Organization to Group Content

Starting with SAS® 9.1.3 Service Pack 4 (Ship Event 26 or higher), both DI and BI content can be stored in the same metadata tree and the content can be managed together. For new metadata environments, you can configure the client applications to share the same directory structures. For existing metadata environments, you can move content from SAS® Data Integration Studio into the `ReportStudio/Shared` directory. (In Figure 3, the folder `Shared Data` is intended to contain metadata for libraries, tables, and information maps that are available as corporate-wide data sources.)

MAINTENANCE

Historically, production environments were primarily the domain of IT departments. IT controlled the corporate content that was published. With the SAS® Enterprise Intelligence Platform, a number of tools are provided that empower business users to create and share content without the help of IT. However, this means that the production environment now needs to handle the creation and management of ad-hoc content.

To control the impact of changes from multiple constituencies, it is imperative that the metadata is organized to segregate corporate content from ad-hoc content. Changes to corporate content should only occur via established change control processes. Business users should not be able to change corporate content. Ad-hoc content should be separated such that ongoing changes do not affect any corporate content. As an example, the folder structure that was displayed earlier provides separation of corporate content from ad-hoc content (see Figure 4).

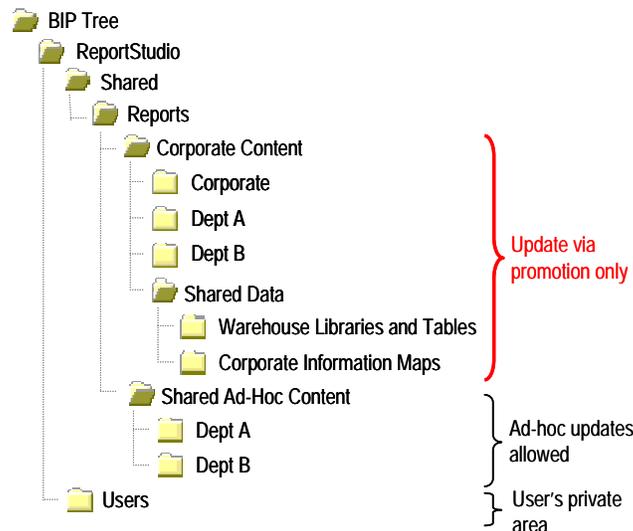


Figure 4. Example Metadata Organization to Group Content

DOORS: WHO SHOULD HAVE ACCESS TO CONTENT?

In planning for the new house, once you have an inventory of the furniture in the house and you have grouped that furniture together to fit into specific rooms, it is time to plan for doors (and locks) to safeguard those rooms of furniture. For example, access to the entertainment room for listening to the stereo and viewing the latest videos on the big-screen theater should not be allowed for small children who can easily break that equipment. Doors and “keyed” access to those doors should be based on separation of concerns for the furniture in each room. You should consider the same things for securing content in the metadata server.

SEPARATION OF BUSINESS CONCERNS

Above, we talked about creating folders (rooms) that group content based on separation of business concerns. You can use metadata permissions to restrict the ability and the level of access to corporate content. The specific settings can vary among the DEV, TEST, and PROD environments, depending on your promotion process, but all environments will need to restrict unauthorized users from write and update permissions to the metadata content. For more details around setting metadata authorizations, please consult the *SAS® 9.1.3 Intelligence Platform: Security Administration Guide* (SAS 2006).

USAGE PATTERNS

Client applications create objects in the metadata repository and leverage metadata authorization control on those content items. Some applications have specific repository locations for content metadata; other applications let the SAS administrator create and manage the locations. SAS administrators must manage groups of users. Each group requires secure access to specific sets of content. The administrator creates subfolders in the metadata root folder and applies access permissions so that PUBLIC is denied access while specific group access is granted. The administrator should also grant access to an Administrators group for content management. For more information about applying access controls, see the online help for the Authorization Manager plug-in to SAS® Management Console. Access to content is controlled by two permissions: ReadMetadata and WriteMetadata. These permissions are enforced by the server based on the identity of the connecting client.

- ReadMetadata permission is needed to navigate and read content. If a user does not have ReadMetadata for a content item, the item is not found in a search and is not viewable in a metadata browse client.
- WriteMetadata permission is needed to create new objects such as trees (folders) or content objects, such as Stored Processes, Reports, and Information Maps.

ACCESS CONTROLS AND THE REPOSITORY DEFAULT ACT

You can use access control templates (ACTs) to apply sets of access controls to metadata objects. The Default ACT has a unique role in that it applies (through inheritance) to all objects in the repository. Whenever a new object is created, the access controls of the Default ACT govern the initial creation of the object. This means that the identity that creates objects in the repository must be granted WriteMetadata in the Default ACT.

TWO APPROACHES TO DEFINING GROUP ACCESS CONTROLS

It is a best practice to grant permissions through groups. Using this approach, you can modify the Default ACT as follows:

- **Identity-based groups**—with this configuration, users who have identities in metadata have permissions to access the Portal and other Web Applications. Administrators, including the Portal Admins group (SAS Web Administrator, SAS Trusted User), have permissions to manage Web Application security and operations. All Web Application users must have an identity in the SAS Metadata Server because PUBLIC users have been denied permissions. Most content folders will be accessible; we will describe permissions for securing content later in this document.
- **Application user groups**—in Option 2, you have the same user permission pattern as in Option 1, but now you are creating specific application user groups (for example, Web application users) and granting permissions to the group instead of the broader permission grant to SASUSERS.

Note: You will need to create a group for each application user group in your system and grant each group ReadMetadata at a minimum and WriteMetadata for certain activities.

BEST PRACTICES FOR A STRONGER METADATA-BASED SECURITY POLICY

The following are some best practices to use when employing a strong metadata security policy:

- Limit content locations to a small number of metadata content root folders, preferably only in the **BIP Tree** root folder.
- Deny WriteMetadata access whenever possible. This prevents users from writing content in unsecured locations.

- Generally, the ReadMetadata permission must be broadly granted to metadata objects for navigation and searching, and can be denied at the lowest subfolders that must be secured to the owning user or group.
- When applying a direct ACE or ACT to deny ReadMetadata permission for an identity, always deny WriteMetadata as well for the identity.

PLUMBING AND WIRING: WHAT INFRASTRUCTURE IS NEEDED?

As a best practice, it is recommended that you set up at least three separate functional metadata environments to support your SAS[®]9 deployment: Development (DEV), Test/QA (TEST), and Production (PROD) (Figure 5). Each environment would reflect a complete install of the SAS[®]9 platform, including backend application servers and mid-tier components. There is no need to have multiple installs of the desktop client applications to access these environments because a single client instance can access multiple functional environments.

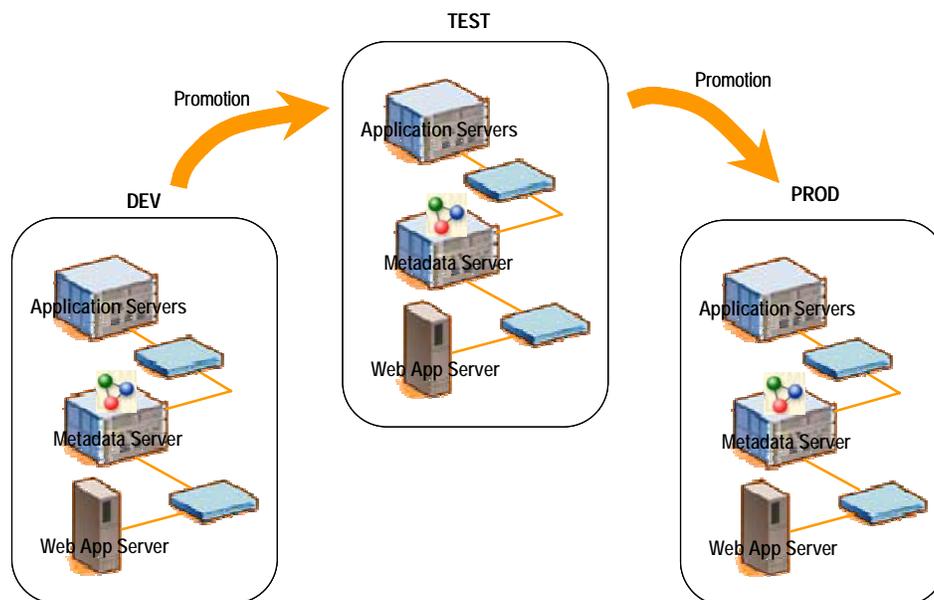


Figure 5. Physically Separate Environments for DEV/TEST/PROD

With this setup, you can completely isolate changes made to one of the environments. Staging each environment is also straightforward because full repository promotion is a viable option as a staging mechanism. If your environment includes the SAS[®] Information Delivery Portal, you can create an initial portal deployment in the DEV environment. Then, with full repository promotion, you can duplicate that baseline environment into TEST and PROD.

In this scenario, changes are isolated to individual environments, and processes can be put in place to manage when promotions occur. Each environment can be on separate sets of servers, or partitioned on a large server (or servers) with multiple CPUs. The distinction is that each environment is completely isolated from each other.

An alternative scenario would be to combine DEV and TEST in a single environment and keep PROD in its own environment (Figure 6). The DEV/TEST content would be segregated via metadata on the same machines—which means using separate folder structures within a metadata repository tree.

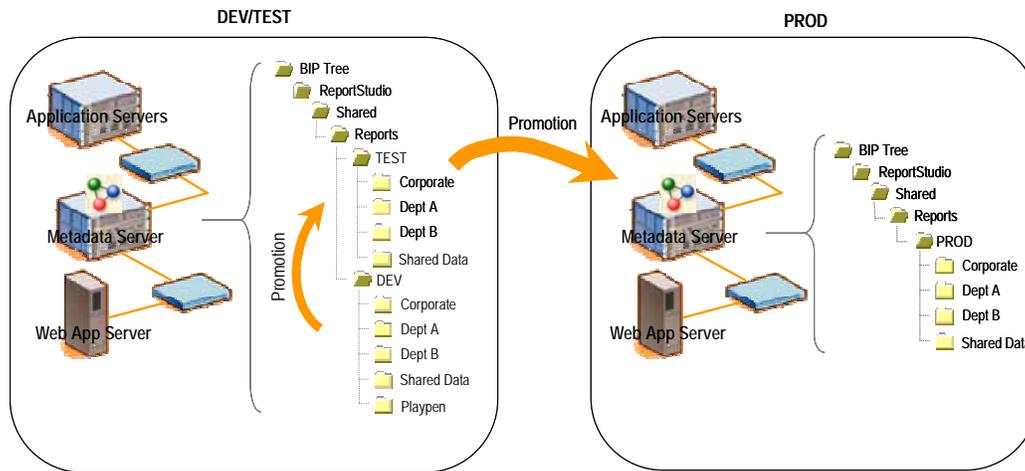


Figure 6. Combine DEV/TEST and Keep Separately from PROD

In this setup, you do not need as much server hardware to support, as you can leverage a single machine for multiple environments. Partial promotion techniques work best here, as they support copying of content both within a repository as well as across repositories. Note that the folder structures for DEV, TEST, and PROD are virtually identical. This makes the partial promotion process much easier to manage and control. The BI Manager plug-in within SAS Management Console provides partial promotion support, and it assumes that the target metadata repository will have the same metadata folder structure as the source metadata repository.

It is a best practice to have the PROD environment physically separated from your DEV and TEST environments—to isolate production content from ongoing changes. It is possible to extend the single environment to support a production environment, if you do not have resources to support multiple physical environments, but much care should be taken to set up an appropriate physical and metadata tree structure and apply authorizations appropriately.

ROOF: HOW DO I PROTECT MY INVESTMENT?

The components in the SAS Enterprise Intelligence Platform all leverage a common set of servers to generate data, analyses, and reports. The term *server* refers to a program or programs that wait for and fulfill requests from client programs for data or services. Each server is connected to content stores that need to be factored into the backup strategy for your SAS environment. Whenever you back up the SAS Metadata Server, you must also back up these content stores. (See Figure 7.)

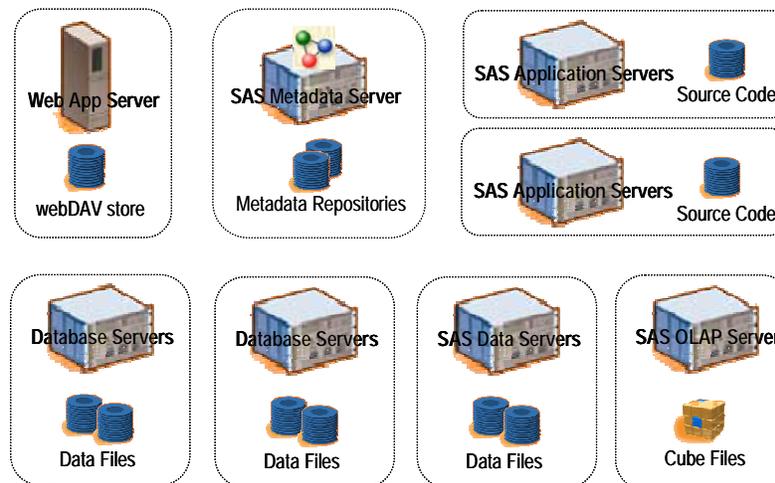


Figure 7. Content Storage in the SAS Enterprise Intelligence Platform

Web App Server—Web-based clients, such as SAS® Web Report Studio and SAS Information Delivery Portal, leverage a WebDAV (Web-Based Distributed Authoring and Versioning) store to house content that is surfaced via their interfaces. The recommended WebDAV store in the SAS platform is Xythos WFS.

Metadata Server—The SAS Metadata Server facilitates the sharing of content across SAS applications to provide a consistent representation of the business to all constituents in your organization. All SAS applications leverage the SAS Metadata Server to store information about their content in the metadata repositories.

Application Servers—The SAS application servers execute SAS analytical and reporting processes for distributed clients. The source code for these processes is stored in a physical location that is accessible to the server. Types of SAS application servers include the SAS Workspace Server and the SAS Stored Process Server.

Data Servers—Enterprise data includes SAS data (such as data sets, OLAP cubes, and SAS® Scalable Performance Data Server® tables), relational database management systems, or Enterprise Resource Planning (ERP) system tables. These represent the actual physical data stores that feed the SAS platform.

ESTABLISH A FORMAL BACKUP PROCESS

A formal backup process should be established and managed specifically as a key IT process. Management oversight should be provided to ensure that resources are available to support the backup process and to be responsible for corrective measures if a backup fails. There should be clear documentation regarding the steps required for completing and archiving backups. The following information should be included in the documentation:

- stopping, starting, and pausing software server processes
- network locations of content to be backed up
- locations, media, and restore facilities where backup copies will be archived
- utilities and other software tools required to perform the backups
- schedule of when backups should occur and related service levels appropriate for archive, restore, and scale requirements
- amelioration and recovery processes for situations where backups fail or take longer than service levels allow

BACKING UP THE SAS®9 ENVIRONMENT

The SAS®9 environment creates both metadata and physical content that must remain in synch. As a result, it is critical to incorporate the three key components into the content backup strategy: metadata, WebDAV content, and physical files (such as source code files and data files). The natural inclination is to back up each component separately, but because of the integrated nature of the SAS environment, you must back up all components within the same backup process. For SAS metadata and metadata-related content, you should always perform a full backup versus an incremental backup. This will ensure that all dependencies and cross-references stay in synch. Incremental backups can be effectively leveraged for databases, such as relational databases or SAS Scalable Performance Data Server tables.

To back up the content in the SAS®9 environment, follow these general steps:

1. Stop or pause the SAS Metadata Server.
2. Lock down Write access to physical content locations, such as source code directories and database tables.
3. Back up metadata repositories.
4. Resume the SAS Metadata Server with its repositories in read-only state, so that applications can read information but not create or edit it.
5. Back up WebDAV locations and physical content, such as source code files and SAS data files.
6. Resume the SAS Metadata Server to active status when all backups are completed.
7. Restore access to physical content locations.

BACKING UP METADATA

For regular maintenance, you should always back up the entire metadata server. Full backups include the Foundation metadata repository and all custom and project metadata repositories that are registered on the metadata server. By performing full backups, you ensure that dependencies and cross-references between metadata repositories are correctly maintained and that the SAS Metadata Server can be restored to a consistent state.

The metadata repositories that underlie the SAS Metadata Server are built on a set of SAS data sets. When the SAS Metadata Server is active, these data sets are in a “locked” state; they are essentially inaccessible for copying. In order to back up those SAS data sets, you must unlock access to them by pausing the SAS Metadata Server

process. Although you can manually back up the SAS Metadata Server, we recommend automating the process by using the %OMABAKUP macro. The %OMABAKUP macro is a SAS batch interface that enables you to perform a complete backup of the SAS Metadata Server. The macro is part of the SAS autocall libraries and runs in a dedicated SAS session.

Because the SAS Metadata Server must be paused, SAS applications cannot access metadata during the backup process. To minimize the down time of the metadata server, you can consider using a third-party tool that can take a quick snapshot of the metadata repositories instead of using the %OMABAKUP macro. The steps to follow are:

1. Use the Metadata Manager plug-in in SAS Management Console or PROC METAOPERATE to pause the repositories to an offline state.
2. Snapshot the metadata repositories.
3. Use the Metadata Manager plug-in in SAS Management Console or PROC METAOPERATE to resume the metadata repositories.
4. Back up the snapshot using the %OMABAKUP macro.

By making a file system that contains only the SAS Metadata Server backup files, you can quickly take a snapshot of this file system for archiving and maintain the maximum up time for the SAS Metadata Server. For storage area network (SAN) users, making a snapshot of the data to a spare area on the SAN is a way to ensure that the backup process happens quickly and minimizes the down time of the SAS Metadata Server. The backup of the data can then be done after the SAS Metadata Server is restarted. The tradeoff, if you use this technique, is that you need enough extra disk space to support the second copy of the SAS data store.

If the server is not stopped or paused, the resulting backup might be unusable. If no backup exists or the backup was not performed correctly, completely rebuilding the metadata repository might be required. For detailed backup instructions, see "Backing Up and Restoring the SAS Metadata Server" in the SAS[®] 9.1.3 *Intelligence Platform: System Administration Guide* (SAS 2006).

RESTORING METADATA

There are three methods to consider when restoring metadata from a backup:

- Restore the full repository.
To restore the repository, you can use the Metadata Manager plug-in in SAS Management Console or create a SAS program using the provided macros. This approach restores all metadata, but it does not restore any associated physical files. You need to update those manually, if needed.
- Update in place using a backup file from the partial promotion process.
If only a specific location in metadata needs to be restored, importing the backup SPK file (created during the promotion process) directly to the same target folder location can "reverse" any changes to objects that existed before the import. This process essentially overwrites objects with their previous state. However, it does not delete any new objects that were created as part of the import. You can manually delete any objects after the fact.
- Delete and re-import using a backup file from the partial promotion process
If only a specific location in metadata needs to be restored, another method is to delete the target metadata folder in its entirety and re-import from the backup SPK file (created during the promotion process). This method returns the target metadata folder to its former state and includes deleting any new objects. Obviously, having a good backup is a prerequisite for this approach.

The restore method that you choose to use should be dependent on the situation. The delete and re-import method is more risky because if there is a problem with the backup file, you cannot recover the content. In either case, there should be an established process for requesting a reversal and gaining approval.

BACKING UP WEBDAV

A WebDAV store contains content that is used in Web applications. SAS Web Report Studio and SAS Information Delivery Portal use a WebDAV server to store content. You should back up the WebDAV store whenever you back up the metadata server to keep the content in synch. The SAS[®]9 platform supports two WebDAV providers: Apache and Xythos. The Apache WebDAV store is straightforward; it uses standard file system support. You can use system tools to back up the files. SAS can also provide a WebDAV dump utility (similar to the WFSDump utility discussed later). If you want access to this utility, contact SAS Technical Support.

The Xythos WFS WebDAV architecture uses a relational database, and the file system and both components must be backed up. Your backup and recovery strategy will vary depending on the number of Document Stores being used as well as whether External Storage is enabled. For a discussion of backup and recovery considerations, see the "Database Recovery Guide" that is bundled in the installation:

- WFS 4.0: <xythos_home>wfs-4.0.x/doc/wfs_dbrecovery.html
- WFS 4.1: <xythos_home>wfs-4.1.x/doc/en/wfs_admin.html#_DB_Recovery
- WFS 4.2: <xythos_home>wfs-4.2.x/doc/en/wfs_admin_file_system.html#_backup_restore

SAS also provides a WFSDump utility, which can be found in the Xythos install directory (<xythos_home>/custom/bin). This utility requires that the SAS Metadata Server be available to respond to requests. You can pause the metadata repositories to a read-only state (as discussed earlier) to prevent any content from being modified or added during the backup process. The WFSDump utility is primarily used to migrate content from one WebDAV server to another, but you can use the same utility to create WebDAV backups.

BACKING UP PHYSICAL CONTENT

Because SAS also leverages physical content that is stored on the file system, your backup strategy should include the following information:

- source code files for stored processes and stored process results that are stored in permanent packages
- packages that have been published to a channel's persistent store location (if using the file system or FTP location)
- data sources, including the following:
 - SAS data files, such as SAS data sets, SAS catalogs, and OLAP structure files
 - SAS Scalable Performance Data Server data
 - relational database sets
- XML files and XML maps that are contained in libraries
- custom code, processes, and post-processes that are associated with SAS Data Integration Studio jobs
- source files for SAS Data Integration Studio jobs that have been scheduled or deployed for scheduling, which could include the work area for Platform Computing Process Manager software.
- generated and user-written SAS code, reports, sample and work data sets, and project properties for SAS Enterprise Miner models

These physical content files are stored as standard operating system files. You can use operating-system commands or third-party tools to back up these files. You should ensure that all in-memory changes have been flushed to disk before performing backups. You can do this by issuing a "sync" command which is supported by both Windows and UNIX environments.

BACKING UP SAS DATA FILES

SAS data files, such as SAS data sets and OLAP data files, are locked if there is an active server process that is accessing them. In order to back up these files, you must stop the SAS processes. This might require stopping all application servers before starting the backup process, which includes workspace servers, stored process servers, and OLAP servers. Once these files are available, they can be backed up using standard backup tools.

BACKING UP SAS SCALABLE PERFORMANCE DATA SERVER

The SAS Scalable Performance Data Server (SAS SPD Server) tables can be backed up using standard backup tools and integrated into existing backup processes. Full backups are recommended if the SAS SPD Server data is partitioned, as restoration time would be fairly quick. However, if the data tables are large, an incremental backup strategy might be more efficient.

SAS SPD Server also provides backup and restore utilities to alleviate resource problems that are associated with backing up large data tables. Instead of backing up the entire table, the utility backs up only the records that changed after the previous table backup date.

For more details about backing up your SAS SPD Server tables, see the *SAS® Scalable Performance Data Server: Administrator's Guide*, Second Edition (SAS 2006).

HOME IMPROVEMENTS: HOW DO I PLAN FOR CUSTOMIZATIONS?

Now that you have built your house, you will want to make changes and adjustments to deal with actual usage patterns and growth of your family and assets. Changing needs might require you to change furniture or add and expand rooms. The same is true for the metadata server. Organizational needs will constantly change, and you will want to make changes to the content.

METADATA PROMOTION FOR CHANGE CONTROL

Change control is the process to coordinate changes to content in a given environment in a controlled and organized manner. Typically, organizations set up three environments: a playpen environment (Development), a quality assurance environment (Test), and a published environment (Production). Changes can be made in an ad-hoc fashion in Development (DEV) and promoted to Test (TEST) for quality assurance. After the content is validated, the changes can be promoted to Production (PROD) for broad end-user consumption. Promoted content is usually not editable by end users, and any changes must go back through the DEV/TEST/PROD change control process.

SAS PARTIAL PROMOTION TOOLS

With SAS 9.1.3 Service Pack 4, partial promotion can be handled either via the BI Manager plug-in in SAS Management Console or via SAS Data Integration Studio 3.4.

The SAS[®]9 framework provides the ability to export selected metadata content from one metadata repository and import it into another (or the same repository), as a way to facilitate partial promotion. The export results in a SAS package file (similar to a ZIP file) that contains the selected content from the source metadata repository as well as any associated physical files. There is a user interface for re-mapping properties to match the properties in the target repository in order to account for different paths, library names, server names, and so on.

The BI Manager provides partial promotion for the following metadata objects:

<u>BI-related content</u>	<u>DI-related content</u>	<u>Other content</u>
Reports	Data Integration Jobs	Mining results
Stored Processes	Data Libraries	Metadata Tree folders
Information Maps	Data Tables	
Data Explorations	Generated Transformations	
	Notes	
	Documents	
	External Files	

CHANGE CONTROL PROCESS

The goals of an effective change control process for SAS metadata are straightforward:

- to minimize the impact of changes on metadata integrity and service levels
- to allow coordination and planning of changes to provide a stable production environment
- to maximize productivity of the personnel involved in the promotion process

As business intelligence content is made increasingly available to end users, there is a growing demand to support both production and ad-hoc reporting requests. Corporate level reports will still be centrally managed, but business end users need the ability to ask ad-hoc questions to perform their day-to-day jobs. This presents a challenge for SAS system administrators to control where changes are made. A robust change control process should not allow ad-hoc changes in the production repository that might affect the integrity of standard management reports and their consumption.

1) ESTABLISH PROCESS FOR CHANGE CONTROL

In a survey of 40 corporate IT managers, 60% admitted that their change control processes are not effective in communicating and coordinating changes in their production environment. Gaps in the process included lack of communication, lack of thorough testing, and ineffective approval processes (Stickel 2005). Therefore, as a best practice, it is important to establish a consistent change control process that addresses these gaps. This process should include the following elements:

- Request for change—includes communication to stakeholders.
- Assess scope of the change—includes testing the requested changes.
- Approve appropriate changes.

The request for change should be the first step of the process. This starts the change control process by identifying what needs to be changed (that is, promoted) and communicating the request to all stakeholders. The request should clearly state who is requesting the change, what objects are to be changed, when the changes should occur, and justification for the change. All stakeholders impacted by the proposed change should receive notification of the request. Gather as much detail upfront as possible, so expectations can be measured against the results of the promotion.

When requesting a change using the SAS platform, you might consider requiring that the content to be promoted has already been consolidated into a SAS package file (by using the BI Manager plug-in) and that package file is included as part of the request. The scope of the change should be assessed to determine if the change could be contained within a single parent folder. If the change crosses multiple peer folders without a clear parent (before the root of the metadata tree), then justification must be provided on the need for a wide-scope change. This should send up a warning that a change of this magnitude could affect the entire repository, so it must be reviewed carefully before approval.

Approvals from all stakeholders should be required before the promotion can occur. This ensures that the request has been communicated appropriately and that everyone has agreed on the expected outcome. Approvals should be documented for auditing and traceability. Systematic tracking of these approvals is not provided today by the SAS platform. You should implement a tracking process based on your organization's best practices.

2) EXPORT APPROVED CHANGES

After necessary approval has been acquired, the next step is to execute the approved changes. This means actually promoting the revised content from DEV to TEST or from TEST to PROD. Essentially, the act of partial promotion can be emulated by exporting content from the source metadata repository and importing that same content into the target metadata repository.

As a best practice, you should strive to promote a complete folder structure, instead of selecting individual objects. If you need to promote objects from multiple folders, you should promote the parent folder that contains all relevant content (even though this might include additional metadata that is not changing). By promoting the parent folder, you provide a well-defined container in which the changes occur. If something should go wrong with the promotion action, you can easily isolate the impacts to the container. The person who owns the request for change should be responsible for the content in the export package file—by either creating the file himself or working with a Promotion Administrator to create the file.

3) DOCUMENT CHANGES

As part of your change control process, you should maintain a history of changes. Capturing what changed, who changed it, and when and why the changes occurred provides a complete record that can be used for auditing and tracing. Using a source code control system can provide revision history, release management control, and audit capabilities. Although the SAS package file is not source code, you can still use a source code control system to help manage and document promotion activity.

After the SAS package file is created, you can check the file into your source code control system. Relevant information should be captured when this occurs, such as the owner ID, the date, and a description of the contents. In addition, there should be documentation about how to execute the change, including the following information:

- the name of the source metadata repository
- the name of the parent metadata folder from which the content was exported
- the name of the target metadata repository
- the name of the parent metadata folder to import into
- any reconnection information that might be required for the target repository
- any dependency requirements regarding specific content that must already exist in the target repository for successful promotion

Changes to the package file itself should result in a new version of the complete package file being checked in. Only package files that have been checked into the source code control system and documented appropriately should be available for import. When a change request is made, the specific version of the package file should be included.

4) IMPORT CONTENT

Promotions should be managed on a scheduled basis in order to control when changes can occur. This will aid greatly in any troubleshooting that needs to be done during the promotion process.

The import process should follow an established set of steps to ensure that all precautions have been taken for a successful import.

- Ensure that the documented dependencies exist in the target repository.
- Back up the target location and associated physical content.
- Launch the import wizard, and refer to the documentation for reconnection information, if needed.
- Review the log files for successes and/or failures.

5) VALIDATE CHANGES

The owner of the change request should verify that the changes were imported appropriately and that the new content is functioning as expected. If there are problems found, it might be necessary to back out the changes from the target system. Depending on the scope of changes made, you should consider either restoring the repository from a backup or reloading from a backup file created during the partial promotion process.

Before starting the restore process, it is highly recommended that you back up the existing state of the repository using full repository backups. This will provide you with a checkpoint state that you can return to if you run into problems during the restore.

6) DOCUMENT RESULTS

Whatever the results of the validation, they should be documented as part of the change control process. All stakeholders should be informed of the outcome. Any additional changes that might be required should go back through the process in order to provide the essential documentation that would be necessary to understand what activity is taking place in the metadata repositories.

After the promotion is completed, you should have complete documentation for what happened. This would include:

1. a copy of the SPK file that was created
2. the log file from the export process
3. the log file from the import process
4. the complete record of the change request, including the person who requested the change, the person who made the change, the documentation about how to execute the change, reconnection parameters, and existing dependencies

FUTURE UPGRADES: WHAT IS COMING THAT I SHOULD BE AWARE OF?

With each new release of SAS, the individual tools and components will continue to evolve with changing market needs. No two metadata server deployments will be exactly the same, so you need the flexibility to build what your organization needs. The SAS roadmap includes plans for additional capabilities within the administrative interfaces of the SAS Intelligence Platform to provide that flexibility while streamlining routine maintenance tasks.

ENHANCEMENTS FOR PARTIAL PROMOTION

Partial promotion will become integrated within the interface of SAS Management Console. The BI Manager as a plug-in will be removed, and all of its functionality will be absorbed by the metadata folder tree. We will continue to add support for more metadata item types for partial promotion of content between Development, Test, and Production environments. In the next release of SAS, planned content to be supported include:

- OLAP Cube definitions
- scheduled jobs and job flows
- prompt definitions

REPOSITORY ARCHITECTURE CHANGES

In the upcoming release of the SAS Intelligence Platform, there will be more flexibility for how you can organize your content. Several of the restrictions that are in place today will be removed, so you have to maintain only a single metadata folder structure that supports secure segregation of content. If you still require using physically separate repositories, custom repositories will still be available, and all SAS applications will support them. Custom repositories would behave just like another folder in the metadata folder tree, so there would be no restrictions on accessing content in other repositories. Of course, security authorizations would still be honored.

METADATA UTILITIES: REPLICATION VERSUS PROMOTION

The Metadata Manager plug-in in SAS Management Console is being modified to consolidate Promotion and Replication jobs into a single activity. Metadata Utilities will surface only a Replication wizard, but you will still have the option to specify substitution values for metadata objects, similar to the way Promotion jobs work today.

CONCLUSION

Careful planning and design when building a house can go a long way to making sure that the investment in the house is well spent and well protected. Planning for the proper number and size of rooms, the inventory of furniture and proper placement within those rooms, the proper type of doors to ensure safe and secure access to those rooms, the appropriate plumbing and wiring as well as the right roof to cover the house are all essential to doing it right. There is nothing more frustrating than having to spend more money to account for changes or mistakes in the house's design or to have to buy a new house because the first one that you built did not fit your needs. One can never be 100% sure in terms of meeting a family's needs forever, but sound planning and design can help a lot in terms of reducing unnecessary maintenance or additional investment down the road.

The same careful planning and design are essential to deploying a robust SAS metadata environment to meet your organization's needs. By combining best practices for metadata management within your SAS®9 environment, you can establish an effective change control system for metadata-based content.

REFERENCES

SAS Institute Inc. 2006. "Backing Up SAS Content in Your SAS®9 Enterprise Intelligence Platform." Available at support.sas.com/documentation/whitepaper/technical/contentbackup.pdf.

SAS Institute Inc. 2006. "Best Practices for SAS®9 Metadata Server Change Control." Available at support.sas.com/documentation/whitepaper/technical/MetadataServerchngmgmt.pdf.

SAS Institute Inc. 2006. *SAS® 9.1.3 Intelligence Platform: System Administration Guide*, Fifth Edition. Available at support.sas.com/documentation/configuration/bisag.pdf.

SAS Institute Inc. 2006. *SAS® 9.1.3 Intelligence Platform: Security Administration Guide*. Available at support.sas.com/documentation/configuration/bisecag.pdf.

Stickel, Edward. "Change Control vs. Change Management: Moving Beyond IT." March 14, 2005. Available at www.itsmwatch.com/itil/article.php/3367151.

RESOURCES

Amadeus International Inc. 2006. "Change Control Best Practices." Available at www.amadeussolutions.com/english/practices/bp_change_control.htm.

Kaplan, J. "Change Management Best Practices." February 2003. Available at auditnet.org/docs/CMbp.pdf.

Leopoldi, Rick, RL Consulting. "IT Service Management: Change Management Methods and Implementation Best Practices." June 19, 2002. Available at www.itsm.info/ITSM%20Change%20Management%20Best%20Practices.pdf.

ACKNOWLEDGMENTS

Special thanks go to Scott Sweetland for providing input for the paper. And, of course, we must thank Kathy Walch for doing her usual superb editing job!

CONTACT INFORMATION

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

Diane Hatcher, Worldwide Product Management
SAS Institute Inc.
SAS Campus Drive
Cary, NC, 27513
(919) 531-0503
diane.hatcher@sas.com

Ken Matz, Senior Solutions Architect
SAS Institute Inc.
SAS Campus Drive
Cary, NC, 27513
(919) 531-0363
ken.matz@sas.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.