

Enterprise Data Governance across SAS® and Beyond

Jeffrey Stander, SAS Institute Inc.

ABSTRACT

As Data Management professionals, you have to comply with new regulations and controls. One such regulation is BCBS 239 in Financial Services. To respond to these new demands, you have to put processes and methods in place to automate metadata collection and analysis, and to provide rigorous documentation around your data flows. You also have to deal with many aspects of data management including data access, data manipulation (ETL and other), data quality, data usage, and data consumption often from a variety of toolsets that are not necessarily from a single vendor.

This paper shows you how to use SAS technologies to support data governance requirements, including:

- Business Data Glossary
- 3rd party Metadata collection, Data Lineage and Metadata Analysis
- Data Monitoring

Best practices will be highlighted in this paper including: - Implementing a Business Glossary - Data lineage across metadata silos - Collecting 3rd party metadata (using SAS Metabridge Relationship Loader and SAS® Metadata bridges) - Capturing metadata from SAS programs - Integrating Metadata using SAS® Relationship Service - Using SAS® Lineage Viewer, and SAS Relationship Reporter - Establishing controls for Monitoring Data. After attending this session, you will be familiar with the SAS tools used to meet the new requirements and having a more managed environment.

INTRODUCTION

Organizations are under increasing pressure to be able to link data in disparate systems at a logical level – that is to show how metadata is connected. At the same time, with the advent of Big Data systems and the concept of the “Data Lake”, it is ever more important from a practical, user-driven point of view, to have a system that tells data users where the data resides, and what data is used for what purposes.

Your data landscape is diverse including all types of data: big data, small data, operational data, analytical data, and external data. Your data resides in different places. You work alongside with many people managing different data, using different tools and techniques, and methods. Collaboration and communication with other users is key. Who actually knows what is in your data? Who can tell you what happens to your data before you work with it? How do you manage the metadata and associated jobs and tasks that process the data? As a data user, can you easily identify the best data resources available to you? Are you surprised by data quality issues appearing late in your processes?

In order to comply with new regulations and controls, Data Management professionals are putting into place new processes and methods to automate metadata collection and analysis, and to provide rigorous documentation around key data flows. This is particularly in response to regulation such as BCBS 239 regulation in the Financial Services industry.

Because these challenges have to be addressed sooner or later, there are a number of reasons to consider taking action and deploying a solution now. These issues can be addressed by using SAS technologies for data management, data quality and data governance.

OVERVIEW

Let's introduce what a solution typically includes, then look at how you can use SAS to meet the requirements. I will highlight best practices you will want to consider when using the SAS tools.

Typical requirements for an organization fall into 3 main areas: business data glossary, metadata collection and lineage, and data monitoring. As tightly linked applications, the three areas must be addressed to provide you with a view of your data. Typical users include business users and IT users.

Business users have the ability to better understand the data that is available to them, and often have more confidence in the data. IT users have the potential to bring order to data resources, and are often better prepared to communicate how data processes work and interact with each other.

With the capabilities in place, both business and IT users can access all kinds of metadata from a governed environment. You can search, define, create and edit metadata. You will be able to examine relationships, use workflows to manage change and ensure granular security. Plus you can enforce business and data rules, and set up alerts so you can be more proactive with data monitoring.

BUSINESS GLOSSARY

Imagine that you are a new employee in your organization. You want to get to work quickly and be as effective as possible at decision making and running the business.

One of your challenges is getting up to speed with (or learning about) the large number of data sources available to you. You also need to understand how data is used for decision-making.

As you consider how to accomplish this, you'll probably have several questions:

What data is available for me to use in my decision-making?

What form does the data have?

How is it collected?

How often/frequently is it updated?

What processes are used to update the data?

Who is the data owner?

How do I compare reports that come from different systems – what data feeds the reports? What fields are used in which reports?

When do you use one data source versus another?

USING SAS TO FULFILL THE BUSINESS GLOSSARY REQUIREMENTS

SAS® Business Data Network is the enabling technology used to provide a business data glossary. The glossary often contains business terms, but may have technical data dictionary metadata as well.

As a SAS user, you can provide consistent information to both IT and business users. You can empower users so they can accurately search for and identify all available information sources, discover attributes of each information source, and perform data lineage and other analysis on the data.

SAS® Business Data Network provides many capabilities including:

- Create, read, edit, and delete terms and definitions collaboratively
- Search for Terms
- Create hierarchical list of Business Terms
- Versioning, history & roll back of terms
- Lock & unlock definitions
- Email notifications
- Edit standard business term templates & attributes

The screenshot shows the SAS Business Data Network interface. On the left is a navigation sidebar with categories like All Terms, Notifications, My Drafts, Search, Assets, Basel III (which is selected and highlighted in green), CCAR, CCAR Program, CCAR Rules, General, SG, and Solvency II. Below this is a 'Filter' section. The main area has a toolbar with New Term..., Open, Tags..., and Lineage buttons. A table lists terms with columns for Term Name, Description, Type, Importance, Status, Workflow, and Last Modified. The 'Credit Risk' term is selected, showing its details in a modal. The 'Identification' tab is active, displaying the term's description ('The risk of a change in value due to actual credit losses deviating from expected credit losses due to the failure to meet contractual debt obligations.'), requirements ('Credit risk comprises default and settlement risk. Credit risk can arise on issuers of securities (in the company's investment portfolio), debtors (e.g. mortgagors), or counterparties (e.g. on reinsurance contracts, derivative contracts, or deposits) and intermediaries, to whom the company has an exposure.'), and details ('Type: Default', 'Status: Production', 'Importance: Medium'). Other tabs include Hierarchy, Associated Items, Notes and Contacts, and History. To the right of the term details are links to a Monitoring dashboard and Lineage, and a list of tags including Basel III and Solvency II.

Display 1. Main Interface for SAS® Business Data Network

A search facility enables you to search, locate and use the information that is relevant for you. It helps you answer questions about the data, like determining what data is available, whether it's suited for your purpose and where to go for more information.

You can search for terms directly in SAS® Business Data Network. You can also search for terms via a web page. If you are in your web browser, you can select text anywhere in the browser and click on the search bookmarklet, which will run the search and provide links into the glossary.

The screenshot shows a search results page for "debt" in the SAS Business Data Network. At the top, it says "14 results for 'debt': 14 Business Data items". The search bar contains "debt". Below, there is a list of 14 Business Data items:

- Credit Insurance**: A form of guarantee against loss resulting from default on the part of debtors.
- Credit Risk**: The risk of a change in value due to actual credit losses deviating from expected credit losses due to the failure to meet contractual debt obligations. The risk that a borrower (counterparty) will not repay a loan.
- Hybrid Capital**: Capital that has the form of a combination of two or more different financial structures or instruments. Examples are subordinated and deeply subordinated debt. Hybrid capital is used to provide cheaper funding than share capital.
- Insurance Obligation**: An insurer's contractual obligations/rights under an insurance contract.

At the bottom right of the search results, there is a link "SAS® Business Data Network Search extras".

Display 2. Searching from web browser for terms containing 'debt' in SAS® Business Data Network

From the search screen, you can drill down into the details for any term. You can view various kinds of attributes. The attributes include business attributes, and technical attributes if desired.

You can locate and use data that's relevant. You can answer questions about the data, like determining what data is available, whether it's suited for the purpose and where to go for more information.

If you are authorized, you can edit the term, and add your own information such as attributes, associated items, add notes, and associate contacts to the term. As you make changes a history is captured and you can view the history of changes, as well as rollback to a previous version. You can organize terms further by creating folders, arranging items in a hierarchy, and by applying tags to terms.

The screenshot shows a web-based application interface for managing business terms. At the top, there are tabs for 'View', 'Edit', and 'Lineage'. Below these are several navigation links: Identification, Hierarchy, Associated Items, Notes and Contacts, and History. The main content area is titled 'Description' and contains the text: 'Products such as futures, options and swaps, which derive their values from the movement in price of the underlying bank assets on which they are secured.' Under the 'Requirements' section, it says: 'Attribute has to be set. NULL or blank values are not allowed.' The 'Attributes' section is circled in red and lists the following data:

Risk Relationship:	Credit Risk
Escalation Method:	On detection of data quality issues, a problem solving track has to be opened in EGRC. (see also link) Procedures to solve DQ issues are outlined in EGRC. DQ issue correction is documented in EGRC
Validation Method:	Scheduled monitoring job on monthly basis to check data quality on a regular basis.
Regulatory requirements:	Relevant for BCBS 239
Source Applications:	YXC System Data Warehouse

Display 3. Viewing the definition and related attributes for a business term in SAS® Business Data Network

BEST PRACTICES

1. Populate your existing terms using the batch import facility. There is an “Import terms using XML file” – You can import existing data by using an XML import feature. To see the format required for the input XML file, you can export a term from the glossary, and see the structure of the file as an example. Then you use either SAS® Data Integration Studio or SAS® Data Management Studio to create a job that reads your data source and outputs an xml file.
2. Create custom term type to meet the specific requirements of your data. You may have both business terms, and technical data dictionary type of data. Each term type may have different attributes that you want to capture. By editing the term type, you can enforce certain attributes to be captured and displayed.
3. Associate items to your terms. Terms have an associated items attribute. You can associate SAS processes, jobs, data, and other assets like documents with terms.
4. Set up granular security using security controls and SAS® Metadata. User access to the glossary is managed by roles on the users and groups defined in the SAS® Management Console.

Show Users Show Groups Show Roles

User, Group, or Role /	Description
Business Data Network: Administration	Provides all functionality related to administrative activities.
Business Data Network: Business Approver	Provides all the functionality for approving terms.
Business Data Network: Business Editor	Provides all the functionality for creating and editing terms.
Business Data Network: Power Editor	Provides all the functionality for creating and editing terms.
Business Data Network: Technical Approver	Provides all the functionality for approving terms.
Business Data Network: Technical Editor	Provides all the functionality for creating and editing terms.
Comments: Administrator	
Data Management: Business Data Network	Provides default access to Business Data Network.

Display 4. SAS® Business Data Network Roles in SAS® Management Console

Effective permissions:

Add Identities... Remove ◆ = Direct Control

Identity	Read	Write	Create Childr...	Delete
Data Management Administrators	✓	✗	✓ ◆	✓ ◆
Data Management Business Approvers	✓	✗	✗	✗ ◆
Data Management Business Users	✓	✗	✗	✗ ◆
Data Management Power Users	✓	✗	✓ ◆	✓ ◆
Data Management Technical Approvers	✓	✗	✗	✗ ◆
SAS_AUTHENTICATED_USERS	✓	✗	✗	✗ ◆

Display 5. Term level permissions

5. Enforce update policies using SAS Workflows. Each term type has an optional workflow to manage the creation, update and delete process for terms. Use SAS® Workflow Studio to edit the workflow to align with your approval processes.

Settings **Attributes**

Type Identification

Name: Critical Data Item

Description: The Banks data items that describe transactions across reporting.

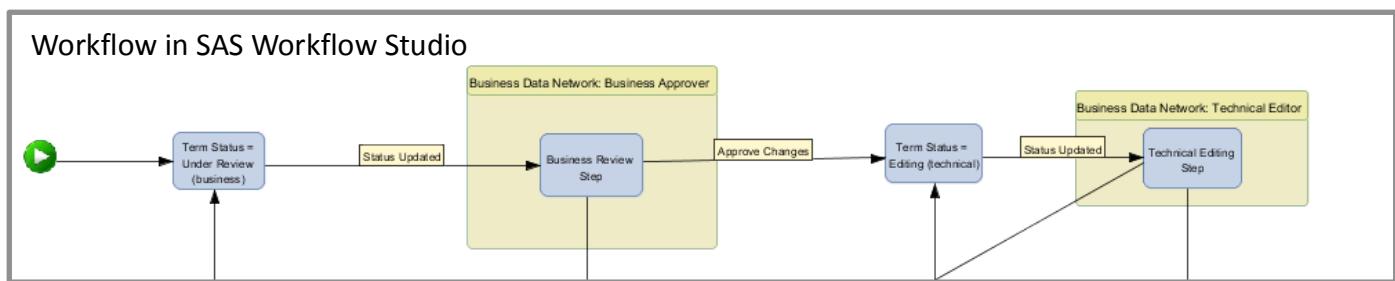
Workflows (optional)

Create: Extended Create Term

Edit: Extended Edit Term

Delete: Delete Term

Display 6. Setting a workflow on a Term Type



Display 7. A workflow in SAS® Workflow Studio

6. Make your business data glossary available to users at the start of projects, so they quickly locate what they are looking for, which can shorten the time it takes them to do projects. Add the URL link to SAS® Business Data Network as a favorite link your portals and web pages. Users will be delighted.

METADATA COLLECTION AND ANALYSIS

It's invaluable to have an automated system for metadata collection and analysis in any size or type of data management environment. Your information flows generate metadata during the end-to-end process of information delivery. SAS provides capabilities to collect that metadata, integrate the metadata and analyze the metadata.

Metadata can reveal to you which data is being used for which reports and which decisions.

You may want to understand how data is used for decision-making. In many cases, the underlying processes implemented for decision-making use different methods of defining terms, vocabulary and taxonomy. Sometimes different tools use different logic for calculating values and for extracting and massaging data. And they may use different data quality rules for manipulating data. To complicate matters, the technologies used for ETL, business intelligence and analytic processes store metadata in different repositories, often using different methods.

Using metadata, you can see who uses data sources and who will be affected by changes. You can analyze how information and metadata flows from data model design to database through ETL processes, BI tools and analytic environments. You can monitor data and detect changes that might be important to the information environment. Plus you can identify duplicate information and help resolve misunderstandings about the data.

A WORD ABOUT METADATA & CHALLENGES YOU MIGHT FACE

Metadata comes in many different formats.

Technical metadata describes data as it exists. It includes systems, databases, schemas, tables, columns, data types and other attributes. Technical metadata persists in a variety of formats, like data model, data dictionary, DBMS catalog and others.

Business metadata describes data as it's used in business processes. It can include objects such as terms, definitions, glossaries, dictionaries and taxonomies. Business metadata is often stored in different places. For example, spreadsheets, databases, emails, documents and in peoples' heads.

Process metadata is descriptive information about things that operate on your data. It may include information about data flows, how data is manipulated and how data is used in the organization. It has information about various ETL flows, reports consuming the data and analytic applications that transform the data into business information. It may also include business rules and decision processes.

Given these diverse types of metadata, you can see how challenging it can be to process it all. Each type of metadata presents unique challenges and opportunities.

Many software vendors store metadata in different file formats and repositories, and it's often proprietary to a particular vendor or tool. To complicate matters, vendors sometimes change the file structure between versions of the tools.

In some cases, you will need special administrative access permissions to access all levels of metadata in the repositories. Some metadata can be very large. Making sense of the metadata can be difficult. Knowing what to select and what is useful can require specific tool or domain expertise.

COLLECT METADATA

Using SAS, you can access and collect metadata from many different repositories.

SAS collects metadata and stores it in the SAS® Relationship Repository, which is part of the SAS mid-tier. The SAS Relationship Service manages this repository. SAS provides many ways to get data into and out of this repository. It depends upon the source of metadata. The repository is used to store and integrate the various metadata and is used by SAS® Lineage, SAS Relationship Reporter, and other tools.

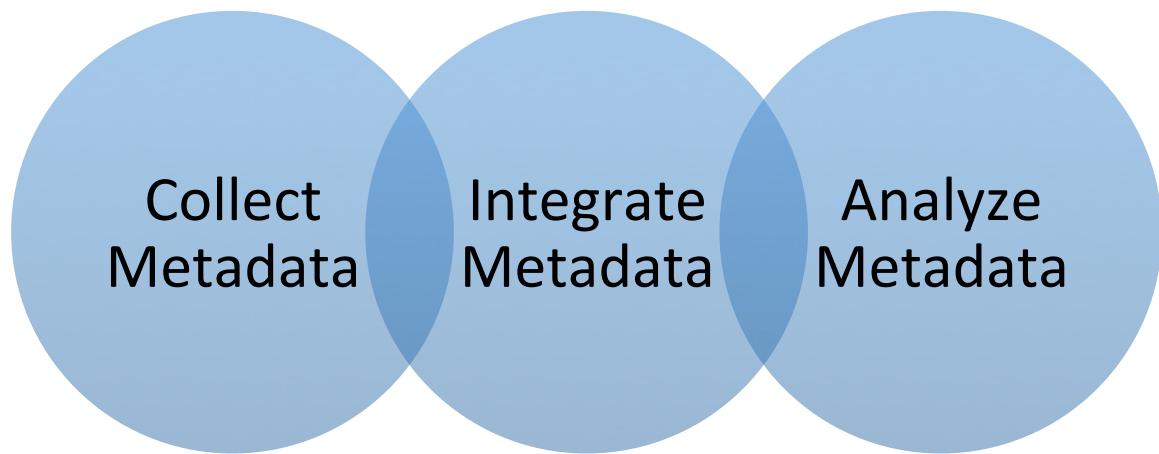


Figure 1. 3 groups of capabilities for metadata collection and analysis

COLLECT 3RD PARTY METADATA

SAS® Metadata Bridges help you access metadata from third-party formats. SAS® Metadata Bridges communicate with standard repositories and file formats that contain metadata. You can use the bridges to read metadata from a variety of enterprise tools including: Data Modeling, DBMS, Hadoop, ETL, Business Intelligence, Analytics tools and more. This allows you to capture metadata about the data sources and targets, and also the process metadata that the tool generates. The result is you can collect information on what ETL is being done, what stored procedures are in the databases, and what reports are being made in the BI tools from the various data sources. SAS® Metadata Bridges are powerful and can help you accelerate your task of metadata collection, especially from 3rd parties.

SAS® Metadata Bridges let you connect to both metadata repositories and files, select what you want to extract and then make the information available. There are also bridges for general industry standards, and for acquiring metadata from files such as .CSV and XML.

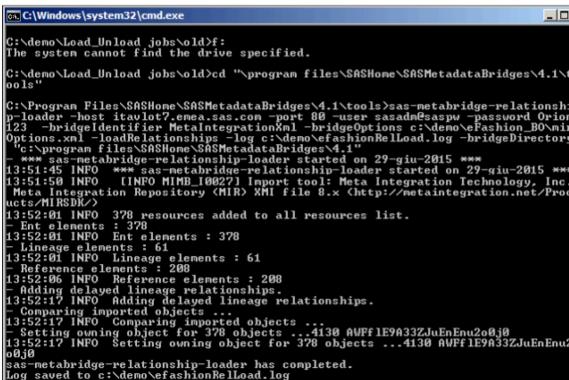
SAS® Metabridge Relationship Loader is a command line utility that comes with SAS® Metadata Bridges. Introduced with SAS® 9.4 M3, it automates the acquisition and loading of 3rd party metadata into the SAS® Relationship Repository. You can setup a command line batch script to periodically collect metadata from your enterprise metadata sources and populate the SAS® Relationship Repository.

For information on using the SAS Metabridge Relationship loader, see
<http://support.sas.com/documentation/cdl/en/bisag/68240/HTML/default/viewer.htm#p0pid3anvsyxkpn1wtj6xrmiiujp.htm>

Load Relationships using -loadRelationships

```
sas-metabridge-relationship-loader -host [host] -port 80 -user [user] -password [password] -bridgeIdentifier
JdbcOracle -bridgeOptions f:\demo\Oracle\OracleOptions.xml -loadRelationships -log f:\demo\OracleRelLoad.log
-brideDirectory "c:\program files\SASHome\SASMetadataBridges\4.1"
```

Output:



```
C:\>C:\Windows\system32\cmd.exe
C:\>demo\Load_Unload jobs\old>:
The system cannot find the drive specified.
C:\>demo\Load_Unload jobs\old>cd "c:\program files\SASHome\SASMetadataBridges\4.1\tools"
C:\>Program Files\SASHome\SASMetadataBridges\4.1\tools>sas-metabridge-relationship-loader -host itavilot7.emea.sas.com -port 80 -user sasadm@saspv -password Orion123 -bridgeIdentifier MetaIntegrationXMI -bridgeOptions c:\demo\efashion_BO\airOptions.xml -loadRelationships -log c:\demo\efashionRelLoad.log -bridgeDirectory c:\Program Files\SASHome\SASMetadataBridges\4.1
--> sas-metabridge-relationship-loader started on 29-jun-2015 ***
13:51:45 INFO *** sas-metabridge-relationship-loader started on 29-jun-2015 ***
13:51:50 INFO [INFO MIMB_10027] Import tool: Meta Integration Technology, Inc.
Meta Integration Repository (MIR) XMI file 0.x (http://metaintegration.net/Products/MIRSDK)
13:52:01 INFO 378 resources added to all resources list.
- Em elements : 378
13:52:01 INFO Ent elements : 378
Lineage Elements
13:52:01 INFO Lineage elements : 61
- Reference elements : 208
13:52:06 INFO Reference elements : 208
- Adding delayed lineage relationships.
13:52:17 INFO Adding delayed lineage relationships.
- Comparing imported objects ...
13:52:17 INFO Comparing imported objects ...4130 AUFF1E9A33ZJuEnEn2e0j0
13:52:17 INFO Setting owning object for 378 objects ...4130 AUFF1E9A33ZJuEnEn2e0j0
sas-metabridge-relationship-loader has completed.
Log saved to c:\demo\efashionRelLoad.log
```

Display 8. Using sas-metabridge-relationship-loader to collect metadata from a third party source

COLLECT METADATA FROM SAS® METADATA SERVER

SAS® Relationship Loader is a command line utility that automates the acquisition and loading of metadata from SAS® Open Metadata Repository (OMR) into the SAS® Relationship Repository. You can setup a command line batch script to periodically collect metadata from SAS and populate the SAS® Relationship Repository. The SAS Relationship Loader provides options to select and read all objects or to filter certain objects to load.

Below are two examples using the command line utility:

Load relationship data for all content objects in SAS OMR

```
sas-relationship-loader -host [hostname] -port 80 -user [user] -password
[password] -loadAll
```

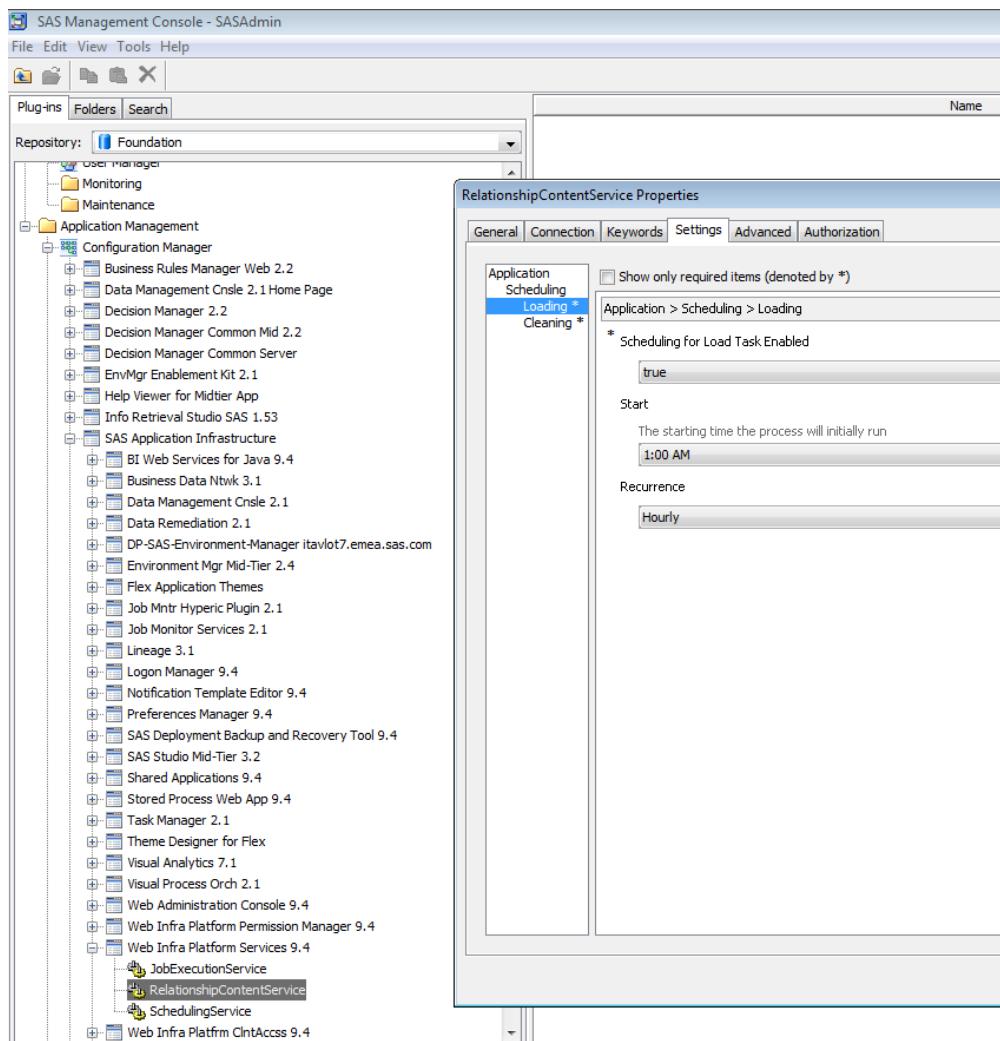
Load relationship data for all tables that are in the /Shared Data folder in SAS OMR

```
sas-relationship-loader -host [hostname] -port 80 -user [user] -password
[password] -types Table -foldertree '/Shared Data'
```

For information on using SAS Relationship loader, see

<http://support.sas.com/documentation/cdl/en/bisag/68240/HTML/default/viewer.htm#p12rspweeziuktn1dkxb24p8g9ez.htm>

In addition to the command line utility, you can use SAS® Management Console and set up the service to run in an automatic fashion. The RelationshipContentService can be scheduled to load and clean at certain intervals. See display 9 below.



Display 9. Configure the Relationship Content Service to run regularly using SAS® Management Console

COLLECT METADATA FROM SAS® DATA MANAGEMENT STUDIO

SAS® Data Management Studio maintains metadata about data quality processes and other monitoring tasks in its own repository. To collect the metadata from SAS® Data Management Studio, use the SAS provided job called: pjExportLineage. This job reads the metadata and loads it into the SAS® Relationship Repository.

COLLECT METADATA FROM SAS CODE

If you have SAS programs or SAS code, you can collect metadata from them as well. This occurs by leveraging the SAS Code Analyzer, otherwise known as the SCAPROC Procedure.

The SAS code analyzer watches as your SAS code executes. It logs information about every procedure and datastep. It logs things like what data it reads from and writes to. It captures descriptive information about the data sources as well. This information is not only logged, but can be used as input into SAS® Data Integration Studio to create a SAS DI job, in SAS Metadata. The output can also be used to create a SAS® Enterprise Guide project. Because the results are in SAS Metadata, the information can be

populated into the SAS® Relationship Repository as part of the automated load, using SAS Relationship Loader.

SAS CODE ANALYZER

- SAS procedure that analyzes SAS code as it is running
- Documents I/O, Proc and Datastep execution

```
proc scaproc;
  Record 'c:\temp\program1.sca' attr;
run;

libname data 'c:\temp';
data data.females;
  set sashelp.class;
  where sex='F';
run;

proc scaproc;
  write;
run;
```

Figure 2. SAS Code Analyzer

SAS CODE ANALYZER OUTPUT

```
/* JOBSPLIT: DATASET INPUT SEQ #C00001.CLASS.DATA */
/* JOBSPLIT: LIBNAME #C00001 V9 'C:\Program Files\SAS\SASFoundation\9.2\core\sashelp' */
/* JOBSPLIT: DATASET OUTPUT SEQ DATA.FEMALES.DATA */
/* JOBSPLIT: LIBNAME DATA 'c:\temp' */
/* JOBSPLIT: FILE OUTPUT c:\temp\program1.sca */
/* JOBSPLIT: ATTR #C00001.CLASS.DATA INPUT VARIABLE:Name TYPE:CHARACTER LENGTH:8 LABEL: FORMAT: INFORMAT: */
/* JOBSPLIT: ATTR #C00001.CLASS.DATA INPUT VARIABLE:Sex TYPE:CHARACTER LENGTH:1 LABEL: FORMAT: INFORMAT: */
/* JOBSPLIT: ATTR #C00001.CLASS.DATA INPUT VARIABLE:Age TYPE:NUMERIC LENGTH:8 LABEL: FORMAT: INFORMAT: */
/* JOBSPLIT: ATTR #C00001.CLASS.DATA INPUT VARIABLE:Height TYPE:NUMERIC LENGTH:8 LABEL: FORMAT: INFORMAT: */
/* JOBSPLIT: ATTR #C00001.CLASS.DATA INPUT VARIABLE:Weight TYPE:NUMERIC LENGTH:8 LABEL: FORMAT: INFORMAT: */
/* JOBSPLIT: ATTR DATA.FEMALES.DATA OUTPUT VARIABLE:Name TYPE:CHARACTER LENGTH:8 LABEL: FORMAT: INFORMAT: */
/* JOBSPLIT: ATTR DATA.FEMALES.DATA OUTPUT VARIABLE:Sex TYPE:CHARACTER LENGTH:1 LABEL: FORMAT: INFORMAT: */
/* JOBSPLIT: ATTR DATA.FEMALES.DATA OUTPUT VARIABLE:Age TYPE:NUMERIC LENGTH:8 LABEL: FORMAT: INFORMAT: */
/* JOBSPLIT: ATTR DATA.FEMALES.DATA OUTPUT VARIABLE:Height TYPE:NUMERIC LENGTH:8 LABEL: FORMAT: INFORMAT: */
/* JOBSPLIT: ATTR DATA.FEMALES.DATA OUTPUT VARIABLE:Weight TYPE:NUMERIC LENGTH:8 LABEL: FORMAT: INFORMAT: */
/* JOBSPLIT: SYMBOL GET SYS_IOUSEEE */
/* JOBSPLIT: ELAPSED 62 */
/* JOBSPLIT: PROCNAME DATASEP */
/* JOBSPLIT: STEP SOURCE FOLLOWS */

libname data 'c:\temp';
data data.females;
  set sashelp.class;
  where sex='F';
run;

/* JOBSPLIT: END */
```

Figure 3. SAS Code Analyzer – SCAPROC – output

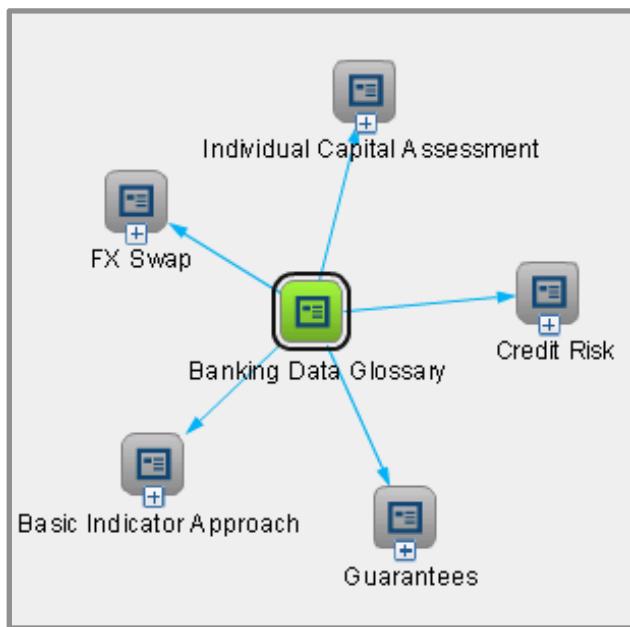
ANALYZE METADATA USING SAS® LINEAGE

In SAS® Business Data Network, you can click on Lineage to bring up SAS® Lineage on any object. This is a useful starting point for most users. You can search and navigate resources in the glossary, and when you find one of interest, you can drill into SAS® Lineage.

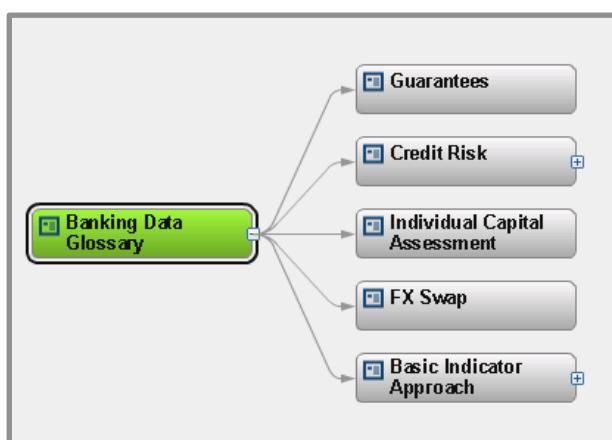
SAS® Lineage is the viewer that shows objects, relationships, and attributes that are stored in the SAS® Relationship Repository. SAS® Lineage shows all kinds of metadata and associated items and relationships including: business terms, SAS metadata including data integration jobs, reports, and analysis, database connections, business rules, projects, libraries, and 3rd party metadata including BI, ETL, and database and modeling tools for example.

You can view the lineage in different views including: All Relationships, Data Governance, and Impact Analysis. With each view, there is a details panel.

The details panel shows details about the selected object. When viewing for example, third party metadata, you can see things such as the SQL logic used to perform a transformation task.



Display 10. SAS® Lineage – All relationships view



Display 11. SAS® Lineage – Governance view



Display 12. SAS® Lineage – Impact analysis view

SAS RELATIONSHIP REPORTER

The SAS Relationship Reporter is a command line utility that queries the SAS® Relationship Service and produces reports about the relationships among objects in the relationship repository. It provides the following types of reports: lineage, impact, direct dependencies, and indirect dependencies. The report output can be .CSV or text file format. The sas-relationship-reporter utility is typically located in the following path: SAS-installation-directory/SASPlatformObjectFramework/9.4/tools.

Here is an example:

```
sas-relationship-reporter -host host01.example.com -port 80 -user
domain1\user123 -password pw123 -rel {-depth 5 -types D A} -types Report -
name "Sales by Product Group" -file Custom_Report
```

Below is the output from this command:

```
"/Shared Data/Orion Star Data/My Folder/Sales by Product Group.srx" (SAS
report)

--> Is dependent on: "/Shared Data/Orion Star Data/Product Orders"
(Information map (relational))

--- Is associated with: "SASApp" (Application server)

--> Is dependent on: "/Shared Data/Orion Star Data/ORDER_FACT"
(Table)

--- Is dependent on: "/Shared Data/Orion Star Data/orgold"
(Library)

--- Is associated with: "SASApp" (Application server)

--> Is dependent on: "/Shared Data/Orion Star Data/orgold" (Library)

--- Is associated with: "SASApp" (Application server)

--> Is dependent on: "/Shared Data/Orion Star Data/PRODUCT_DIM"
(Table)

--- Is dependent on: "/Shared Data/Orion Star Data/orgold"
(Library)

--- Is associated with: "SASApp" (Application server)

--> Is dependent on: "/Shared Data/Orion Star Data/Sample: Multiple
Output Formats" (Stored process)

--- Is associated with: "C:\Program
Files\SASHome\SASFoundation\9.4\inttech\sample" (Directory)

--- Is associated with: "SASApp - Logical Workspace Server"
(Logical workspace server) - in "SASApp" (Application server)
```

BEST PRACTICES

1. Metadata Integration (Stitching)

Metadata from SAS applications is mostly pre-integrated and views easily. Metadata coming from 3rd party tools and data stores is often in silos that the sources come from. For example, you can view the metadata from an ETL source, or a BI source. But there are no linkages between one metadata type and the next. You may want to stitch these together so you can achieve an integrated lineage view. The figure below shows lineage across metadata coming from different tools including ETL, BI, analytics, legacy system, databases, and data modeling tools. I annotated the diagram with different colored circles around the different types of metadata shown as an example.

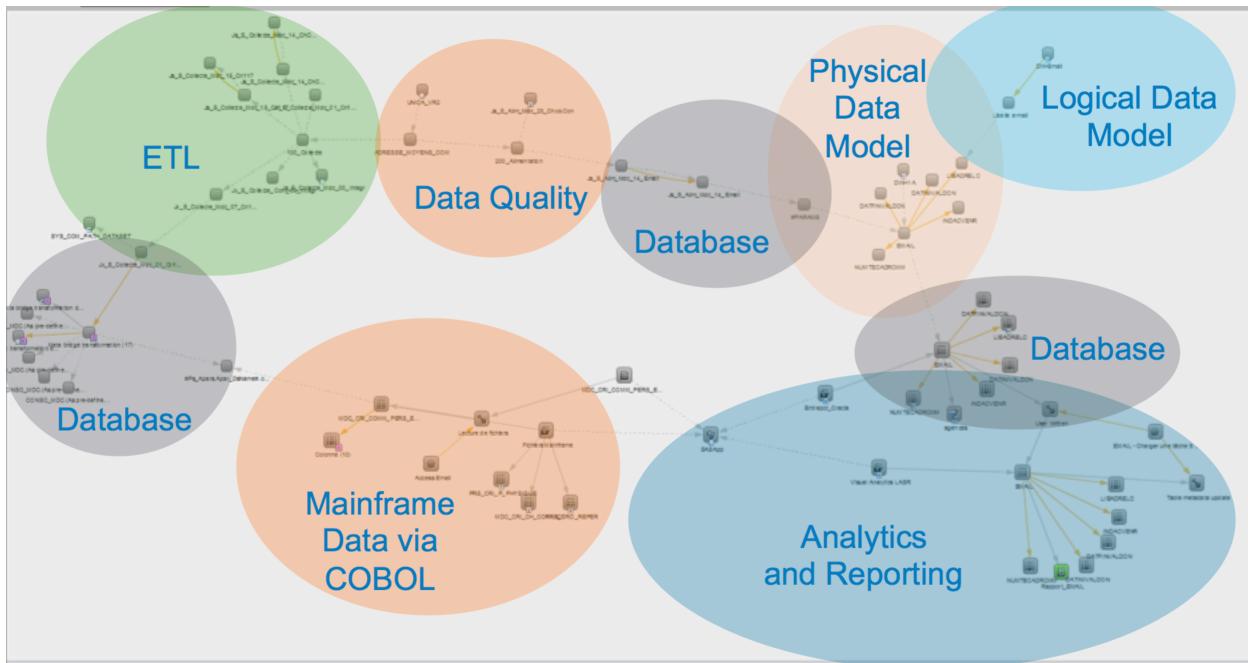


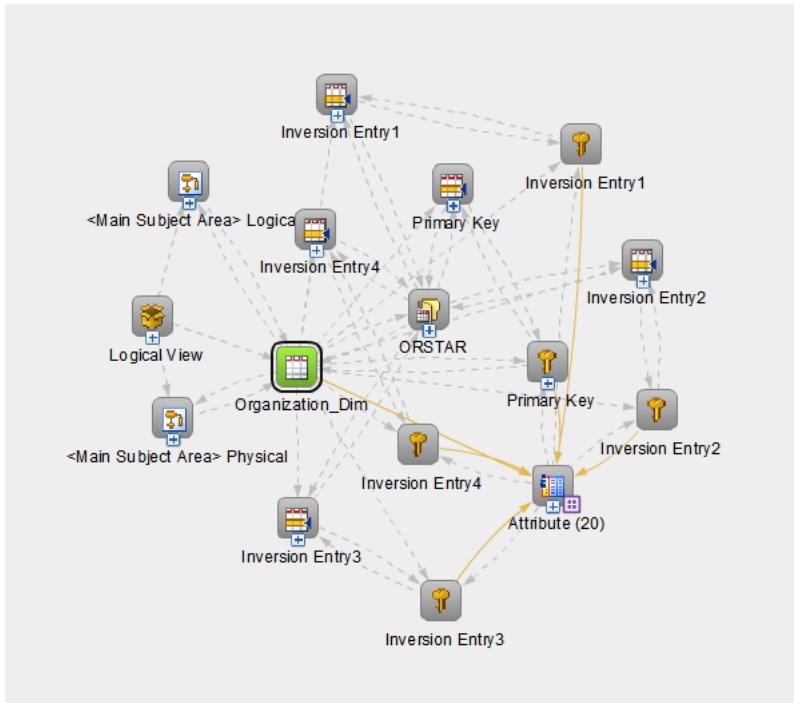
Figure 4. Annotated view of SAS® Lineage – All relationships with multiple metadata sources stitched together

Metadata stitching is an integration technique. It involves taking metadata from different sources that refer to the same elements and associating them together. One common technique for stitching metadata is to identify common elements, typically by querying the metadata, and identifying common data points. Once that query runs, we can then create additional relationships between the metadata imports.

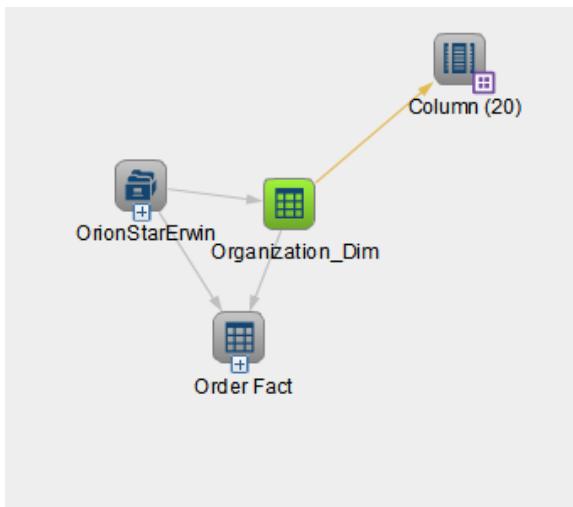
2. SAS® Lineage Viewer: Subset the types for faster search results. SAS® Lineage provides options in the search window for selecting what types of metadata to search for. When working with large volumes of metadata, it is often useful to filter your search for objects of a certain type you are interested in.
3. Metadata quality and quantity: Limit the objects you import to only what you need. Especially when working with 3rd party metadata sources, you should be aware that there is a lot of metadata that may not be useful for governance purposes. Some metadata is only useful for the tool itself, and is noise for the consumer of the information in SAS® Lineage.

SAS® Metadata Bridges provide options, in many cases, to reduce the amount of metadata imported. Below are two images of the same imported metadata. The first one contains

everything and the second image shows just a subset of the metadata. Find out what is required and limit your import to just that metadata, where possible.



Display 13. An example full import of third party metadata (note there are lots of details here that are not needed)



Display 14. An example import of third party metadata (note this is a subset, and more useful)

DATA MONITORING

Establishing some controls to monitor data is a key part of any data governance program. Data quality measures provide you with a key piece of information beyond the glossary and metadata. You want to know if there is an issue with your data, so that actions can be taken to remediate the issue.

Data monitoring helps you and the many people you work with. For example, it provides an additional safety check to developers that their ETL processes are correct. It gives the business users more confidence in the data. And it can provide an auditor a checkpoint to measure.

You can potentially reduce errors and rework of failed processes due to data errors. You can catch data errors earlier in the process, and know when there is a data issue sooner.

Using SAS, you can proactively measure and validate data according to standard business rules. In this section you will learn about SAS capabilities used in this space.

SAS® Data Management Studio provides many data monitoring capabilities including:

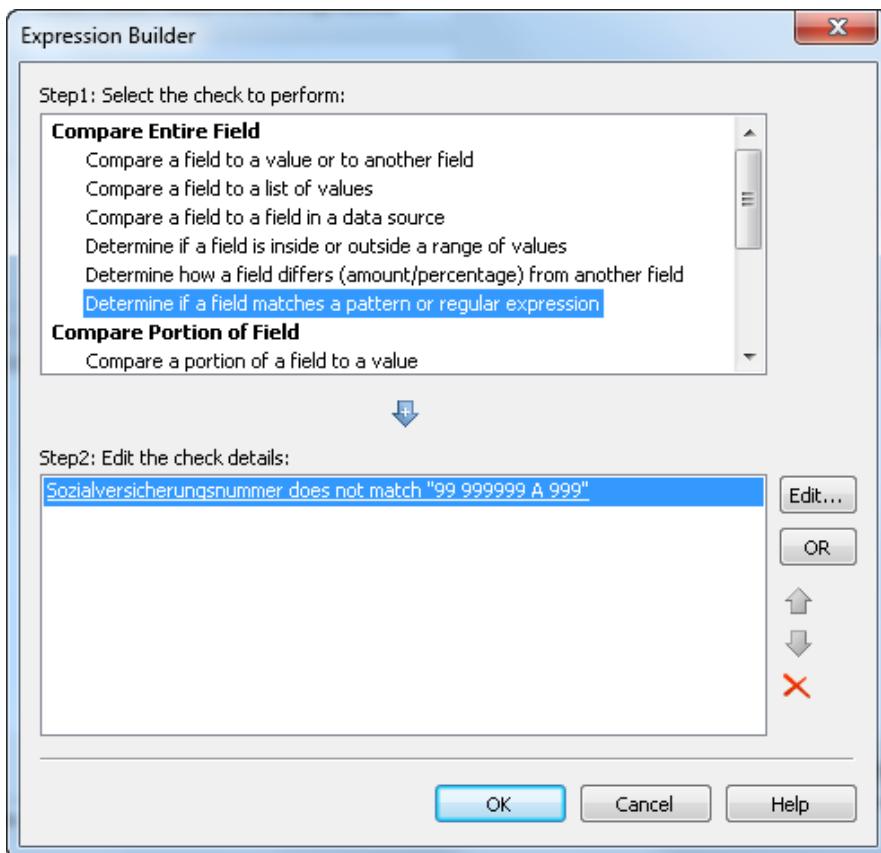
- Business rules editor
- Business rules repository
- Data Jobs to apply the business rules
- Monitoring results
- Historical reporting

You use SAS® Data Management Studio's business rules manager to specify data quality controls and rules. These rules can be technical, for example, that check to see if fields have certain properties. The rules can be logical and validation oriented, such checking to see if data is reasonable and logical based upon the business activity and business process. The rules can also reference historical data, such as check to see if there has been a significantly larger increase or decrease than observed over a period of time.

Business rules drive the monitoring process. The rules are designed to work independently of the data source. So you define the rules once, and then you can apply them to any data using field mappings.

Rules can also be grouped into rule sets. This allows you to build complex rules in a simple way. Rules drive actions. If a rule is triggered, you can set actions to take place. The actions can include notifying someone, logging an event, or triggering a process to run, for example.

You can assign different weights to rules based upon your prioritization or business needs. The weights are used to create a score along a dimension, so that you can report on various measures.



Display 15. SAS® Data Management Studio business rules manager – this rule checks to see if data in a field matches a particular pattern.

Display 16. SAS® Data Management Studio business rules manager – this rule validates data using a more complex expression using reference data and cross field validation

The screenshot shows a window titled "Task: _L Cashflows validieren". The task details are as follows:

- Task label: Cashflows validieren
- Description: Cashflow-Daten für LCR-Berechnung validieren
- Last execution time: March 03, 2013 12:34:21 PM
- Last execution source: (None specified)
- Last execution result: Completed successfully, 22691 triggers

Rules and events:

Rule	Events	Weight	Threshold	Dimension
_L Check CF CASH INFLOW Subt...	Log error to repository	1.00	Risiko	Consistency
_L Check CF CASH OUTFLOW Su...	Log error to repository	1.00	Risiko	Consistency
_L Check CF InstID	Log error to repository	1.00	Risiko	Completeness
_L Check CF InstID unique	Log error to repository	1.00	Risiko	Uniqueness

Fields:

- [ID](#)
- [ISSUE_KEY](#)
- [LIQUIDITY_SUBTYPE](#)
- [LIQUIDITY_TYPE](#)
- [PARENT3_CD](#)
- [PARENT4_CD](#)
- [PRODUCT_CD](#)

Display 17. SAS® Data Management Studio business rules manager – a task that logs records with errors to a repository when certain rules hit certain thresholds.

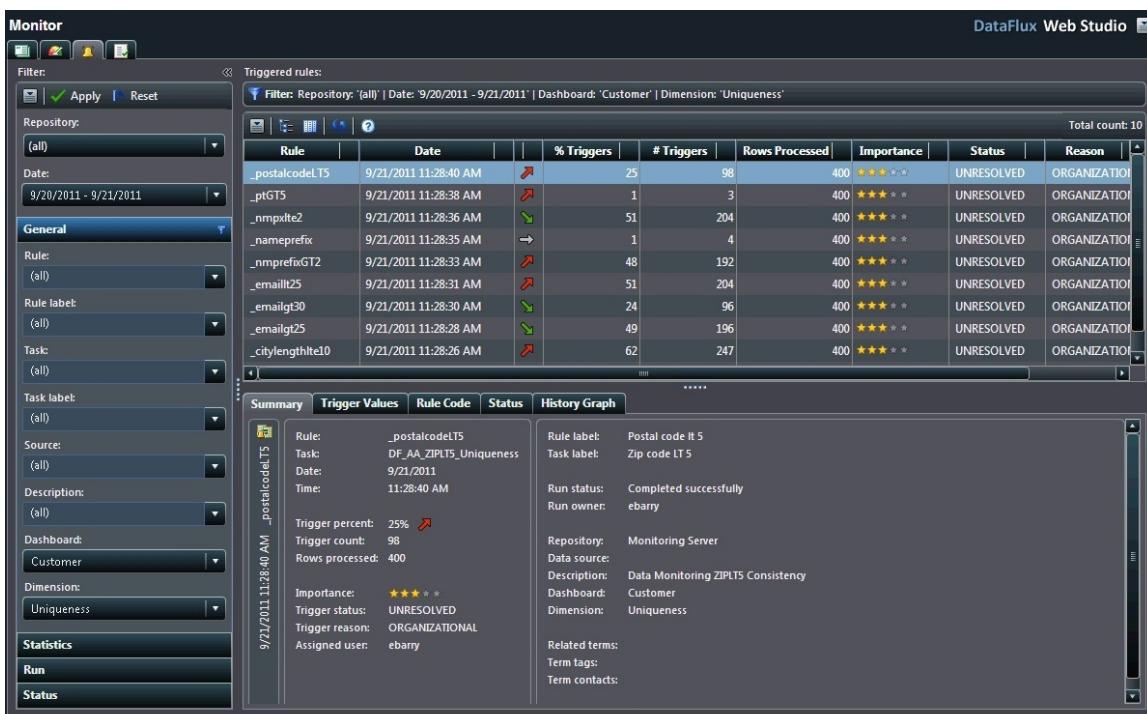
After you define the business rules, you apply them to data sources by using a data job. The data job typically is scheduled to run periodically. The data job reads the data source(s), applies the rules, and writes results to a monitor repository. This is available to independently report and analyze afterwards. You have several options for using the monitor results including:

- Alerts
- Dashboards and Monitoring
- Reports

DataFlux Web Studio provides out of the box reports such as the ones shown below.



Display 18. DataFlux® Web Studio dashboard showing rules that were triggered, along with a score, and information on the trend

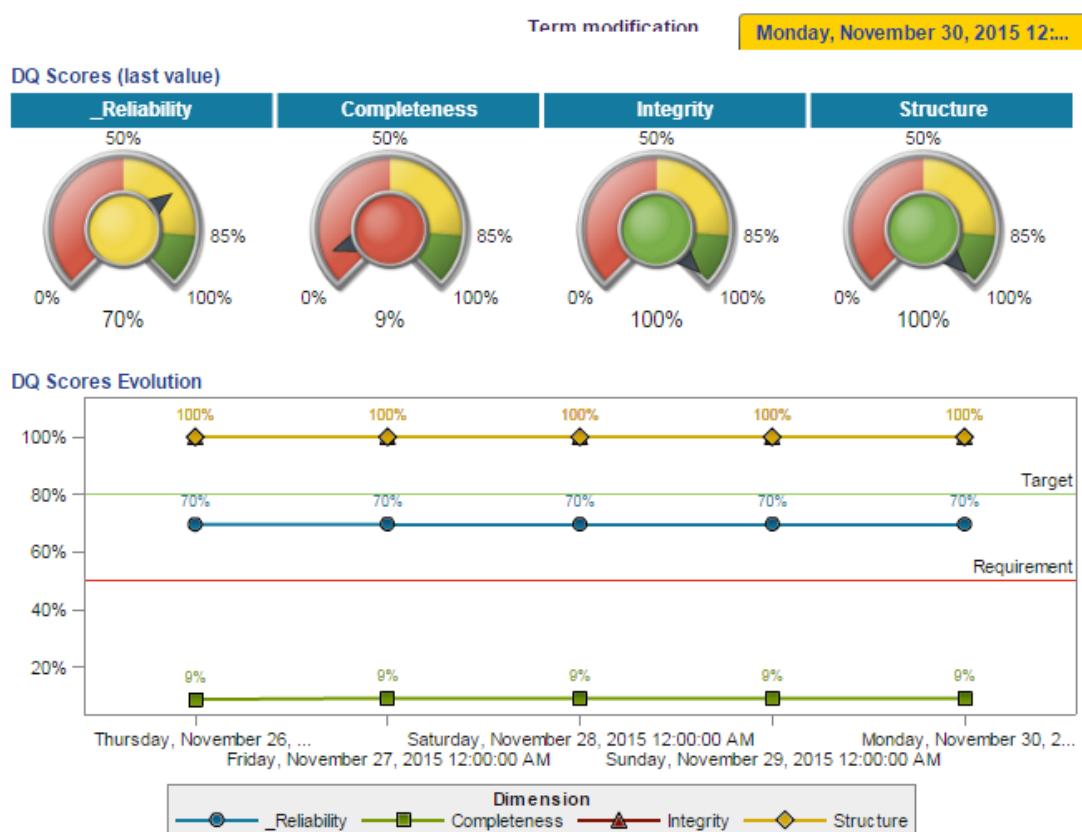


Display 19. DataFlux® Web Studio monitor report showing rules that were triggered, along with details on the records that caused the rule to trigger

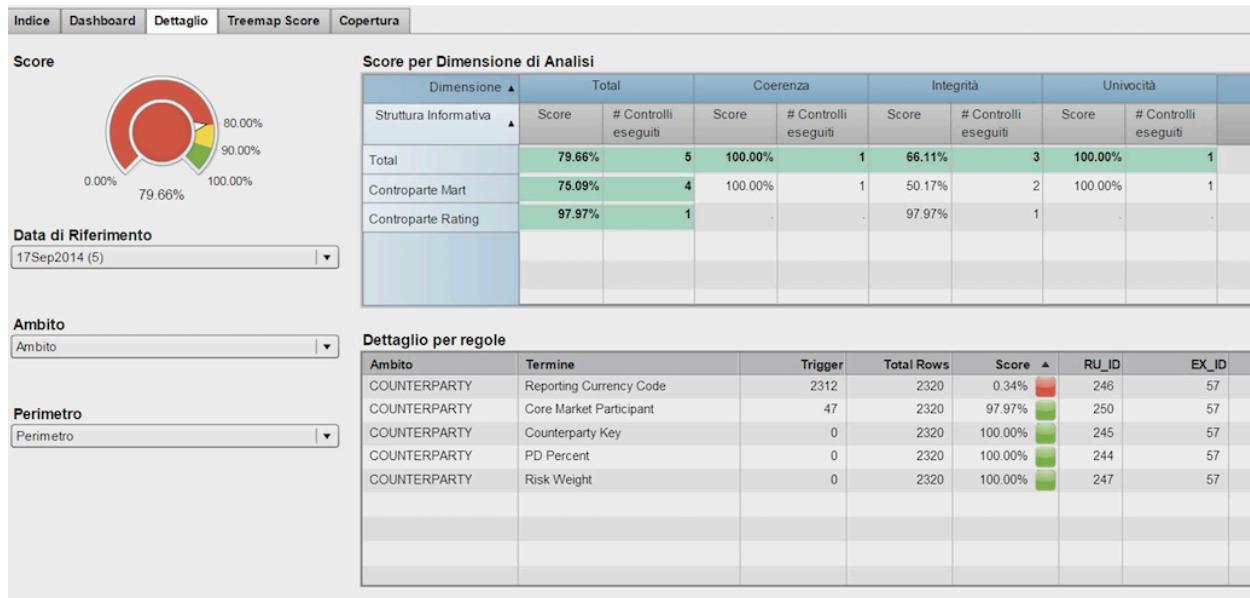


Display 20. DataFlux® Web Studio data remediation panel that allows you to document the reason for the trigger and assign action to take

Another popular method for reporting and analyzing data monitoring results is by using SAS® Visual Analytics. Below are two examples using SAS® Visual Analytics.



Display 21. A SAS® Visual Analytics dashboard showing scores and trends in data quality measures



Display 22. A SAS® Visual Analytics dashboard showing results of data monitoring rules that were triggered

In summary, you use SAS® Data Management Studio to create business rules to monitor your data, validate data, and enforce policy. The business rules are applied to the data using a data job. The data job is typically tested locally, then uploaded to SAS® Data Management Server. Usually the job runs on a regular basis using a schedule. The monitoring results are written to a repository. Then you use DataFlux® Web Studio or SAS® Visual Analytics to view the results.

CONCLUSION

Making business and technical metadata accessible is a sure-fire way to ensure your data is relevant, useful and trustworthy for everyone who needs it. SAS technologies provide the capabilities needed to meet the wide range of requirements including: business glossary, metadata collection and analysis, and data monitoring. By utilizing SAS, you can provide users in your organization with reliable information while meeting objectives around documenting key data flows.

CONTACT INFORMATION

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

JEFF STANDER
SAS INSTITUTE, INC
JEFF.STANDER@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.