SAS4223-2020

STEER YOUR HYBRID SAS® VIYA®/SAS®9 SHIP TOWARD THE "GOVERNED DATA" PORT

Bogdan Teleuca, SAS Institute Inc.

ABSTRACT

Imagine you are in charge of a ship marked SAS® Viya® on starboard and SAS®9 on port. The water is data. Navigation challenges: stay afloat, avoid the rocky shores of regulatory scrutiny, and circumvent the shallow waters of customer complaints. This session guides **you safely toward the "Governed Data" port. Learn when to steer to** SAS®9, and use SAS® Business Data Network as a chart and SAS® Lineage as your tracking system. Find out which SAS® Data Management REST APIs and command-line interface scripts can give you the extra steering. Turn to SAS Viya and accelerate data processing, preparation, and discovery. Improve decisions made with SAS® Intelligent Decisioning. Enter key data governance events in your ship's log, a SAS® Cloud Analytic Services (CAS) table. Watch key indicators on your SAS® Visual Analytics dashboards. Learn the most useful CAS actions and SAS Viya REST APIs. Finally, understand the protection and remediation kits available, in case of an emergency.

INTRODUCTION

Data governance can help you achieve your **company's business objective** in a faster, safer, and efficient way. Imagine Data governance as a journey on the SAS® Platform (SAS Viya and SAS® 9.4) with the following waypoints:

1. Start with the business objectives.

2. Access data and associate metadata.

3. Understand data.

4. Prepare data for business rules monitoring.

5. Use business rules to monitor and steer your business.

6. Trace data processing.

7. Report and visualize.

8. Remediation.

An example from banking is used to highlight the steps above.

The board of a bank has a goal to improve the loan portfolio's quality. The bank must reduce, by 5 percent, by the end of the year, the amount of loans lent to over-indebted clients.
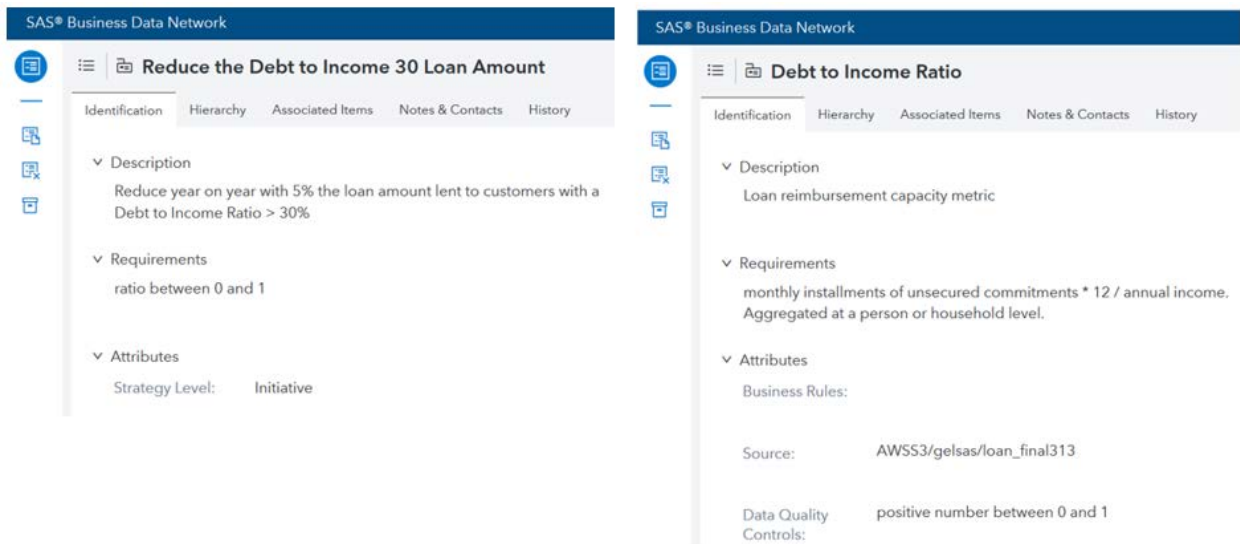
START WITH THE BUSINESS OBJECTIVES

Before you navigate, you should plan the journey: perform your boat pre-departure checks, assemble the crew, define the destination, and plot a course toward your destination.

On the SAS Platform, use the business glossary, SAS Business Data Network, to perform the following:

1. **Document the board's business objective as a business term.**

2. Translate the business objective into business terms.
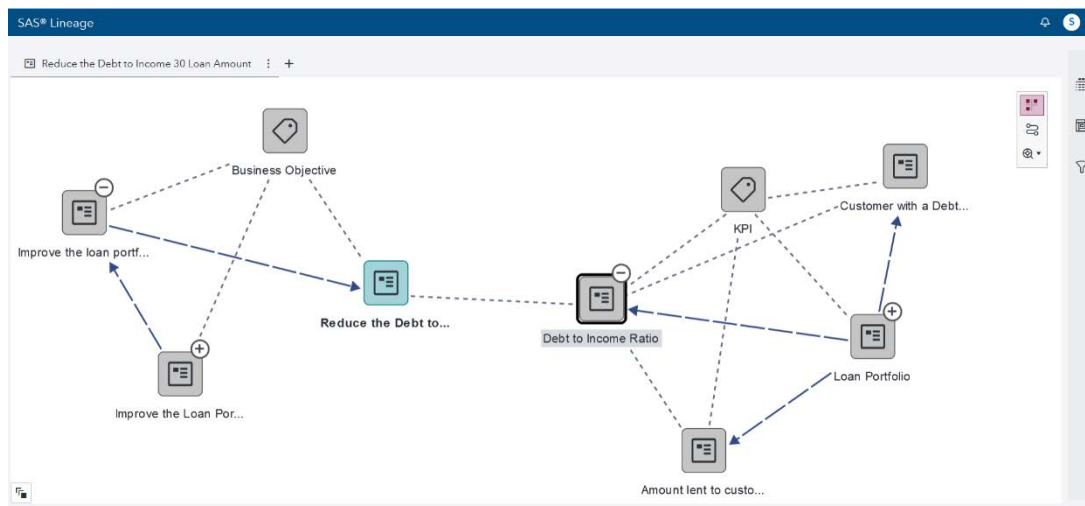
3. Link the business terms and the objectives.

In the banking example, the steps are the following:

- Document the business objective. Translate it into a key performance indicator (KPI) you can calculate and measure: 'Reduce the Debt to Income [Ratio] 30 Loan Amount'.

- Derive the term from the objective: **'Debt to Income Ratio'** is: *'an over-indebted retail customer is a person who has loan installments greater than thirty percent of his income'*. The ratio is calculated for the installment and income. The ratio serves the KPI above.

Display 1. Documented Objectives and Terms in SAS Business Data Network in SAS 9.4

- Link the business objective (left) with the business term (right).

Display 2. Linked Business Objectives and Terms Displayed in SAS 9.4 Lineage
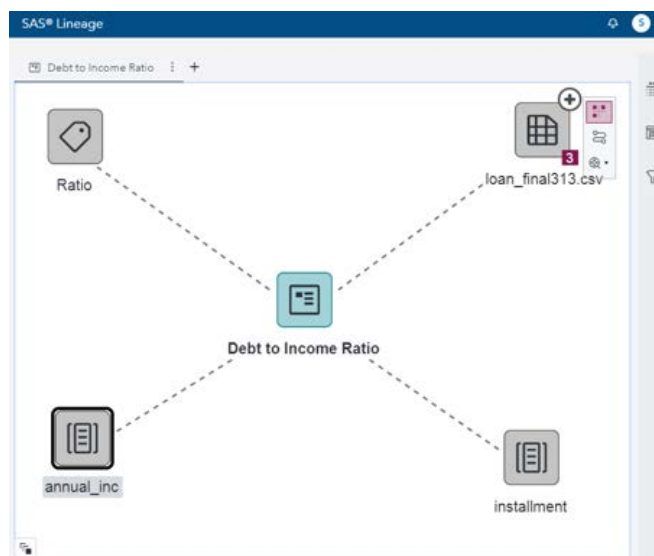
ACCESS DATA AND ASSOCIATE METADATA

When you start your navigation, you leave the safety of your port. You also plot a course toward the destination port. You might take bearings of familiar landmarks to help with orientation.

On the SAS Platform, you need to access a data source first. The **'familiar landmarks'** in SAS 9.4 are called associated items.

In the banking example, the business term 'Debt to Income Ratio' pinpoints to a file stored in Amazon Web Services Simple Storage Service (AWS S3). The file is loaded in CAS (SAS Viya) and simultaneously accessed in SAS 9.4 (proc s3).

After source access, you can associate in SAS Business Data Network the business term with metadata (the loan file and two columns).

Display 3. Business Term Associated with Source Metadata in SAS 9.4 Lineage
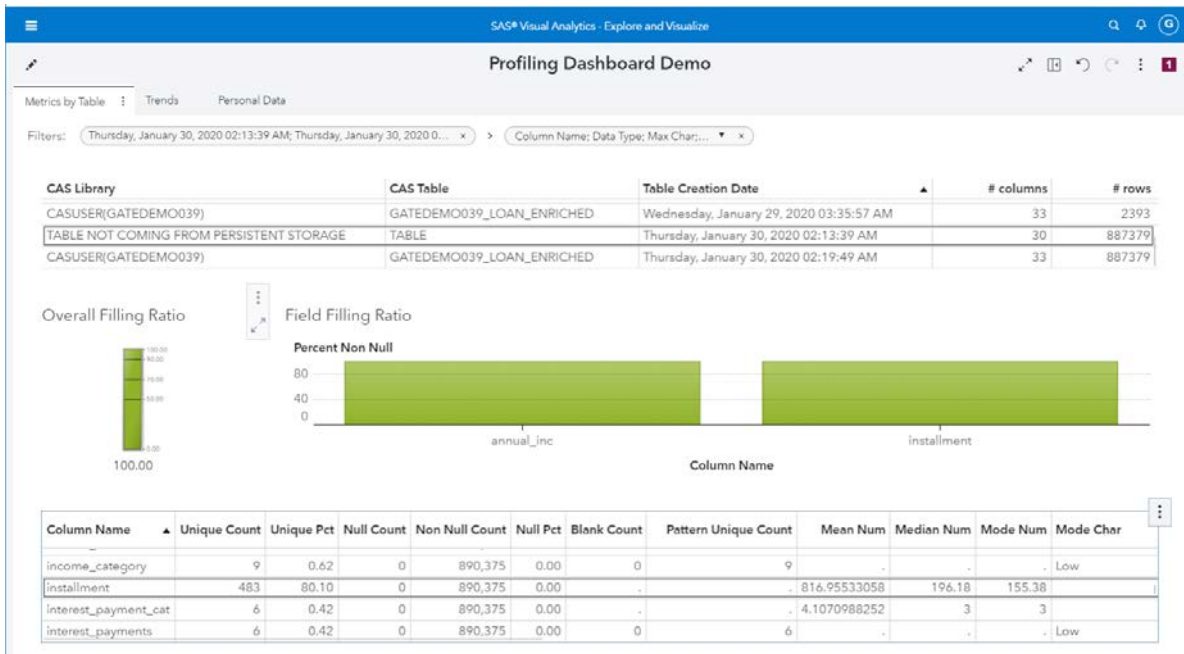
UNDERSTAND DATA

On a boat, you need to constantly discover and anticipate. To stay safe, you require discovery tools: a radio, compass, depth indicator, charts, GPS, lights, radar, night vision, and so on.

On the SAS Platform with SAS Viya you need to do the following:
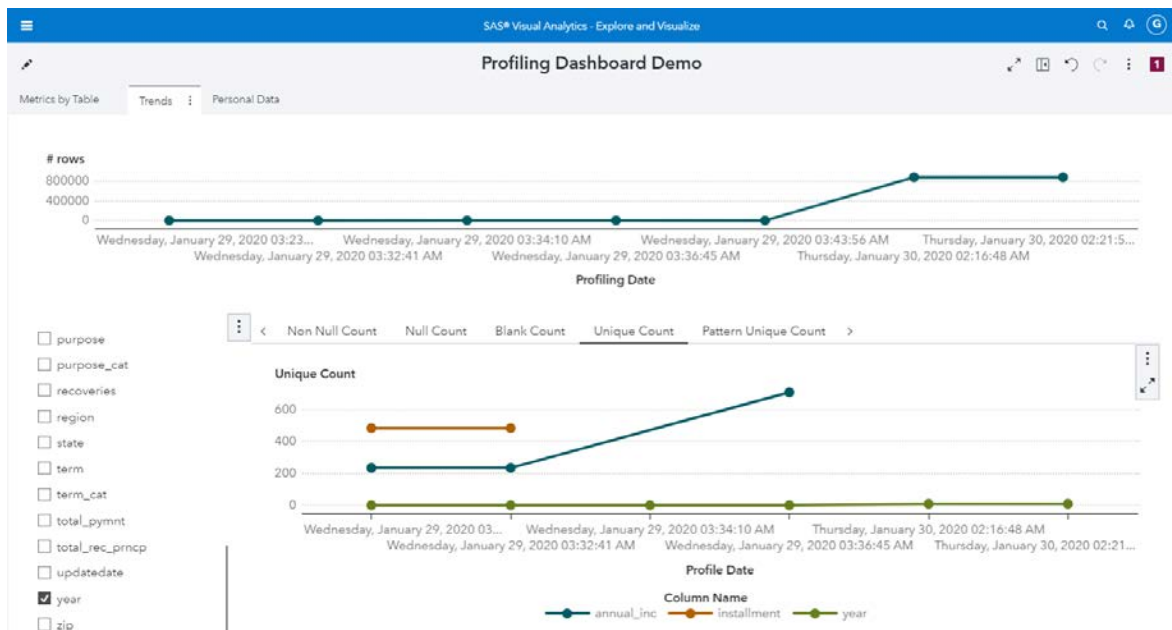
1. Find the data required to materialize the business term, if it is not known already. Use the profiling function in SAS® Data Explorer – Manage Data or the *discovery* CAS action set. Understand data. Examine key data metrics such as completeness, uniqueness, most frequent values, and so on.

2. Identify personal data in your source. Profiling can use the SAS® Quality Knowledge Base for Contact Information (QKB) to find personal information. Before you process any personal data, you must check you have a legal basis to do so.

3. Analyze the profiling results. Save the profiling results to a table. Using a script, append multiple profiling runs to a table. Visualize the data metrics, their evolution, and personal data identified with a SAS Visual Analytics profiling report.

In the banking example, the loan file is profiled. The granularity of the data is the loan contract ('id**'). The columns required for the 'Debt to Income Ratio': 'annual_inc', 'installment' are numeric, filled**, and appropriate for the ratio calculation.
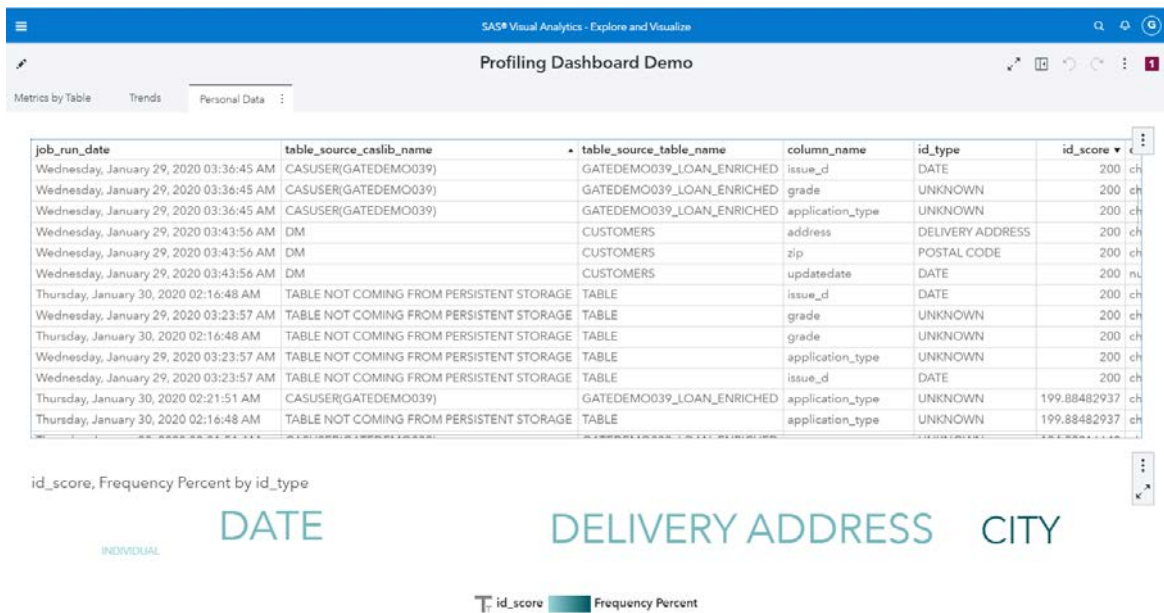
Display 4. Overview of the Profiled Source Tables and Columns in a SAS Visual Analytics Report

Their evolution is then assessed by year, as the business objective is assessed on a year-on-year basis.

Display 5. Source Column Metrics Evolution in a SAS Visual Analytics Report

The loan table loaded in CAS **is "scanned"** when profiling and assessed if it contains personal data identifying the customers. The '_LOAN_' table only contains dates and cities that do not identify customers.

Display 6. Personal Data Identified when Profiling Using a SAS Visual Analytics Report

PREPARE DATA FOR BUSINESS RULES MONITORING

Data preparation is like handling a boat on the water. You need to avoid obstacles and dangerous areas and adapt speed to match the conditions ahead.

Prepare and transform the data analyzed in the previous phase to facilitate the business rule monitoring.

In the banking example, a KPI and a ratio are required. Two different granularity levels justify the following two derived tables:

1. The first, prepared with a data plan, '_LOAN_ENRICHED', at a contract level, will contain several calculated items, including the 'Debt to Income Ratio'.

Display 7. Calculate the Ratio Defined in the Business Term Using SAS® Data Studio

2. The second, '_LOAN_COMPARE', at a year level, aggregates the previous table. Amounts are aggregated by year, and then compared with amounts of the

previous year. Computations can be performed in a data plan or when you write code (proc cas).

	year ↓	⚠ _NAM...	⊕ DIR30Amount	⊕ DIR30Amount_p	⊕ loan_amount	⊕ loan_amoun...	⊕ DIR30Customer
1	2015	_Sum_	562025	0	6417580175	3503830175	29
2	2014	_Sum_	0	0	3503830175	1981989225	0
3	2013	_Sum_	0	0	1981989225	718411025	0
4	2012	_Sum_	0	0	718411025	261683825	0
5	2011	_Sum_	0	0	261683825	131992550	0
6	2010	_Sum_	0	0	131992550	51928250	0
7	2009	_Sum_	0	16600	51928250	21119250	0
8	2008	_Sum_	16600	30750	21119250	4977475	1
9	2007	_Sum_	30750	30750	4977475	4977475	5

Display 8. Aggregated Loan Amounts, by year Using SAS® Studio V

USE BUSINESS RULES TO MONITOR AND STEER YOUR BUSINESS

On water, you must keep constant watch. Monitor your position, course, speed, and depth on your instrument console.
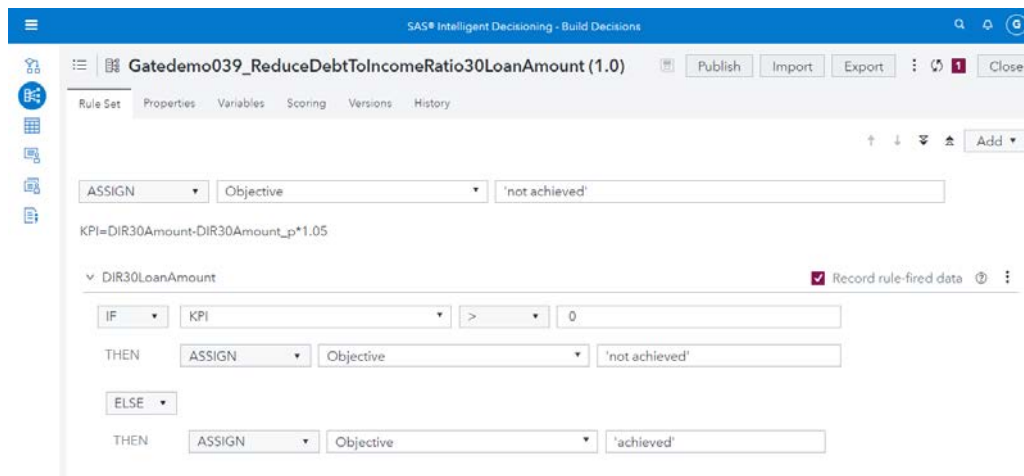
In SAS Viya, use SAS Intelligent Decisioning to monitor your business objectives with business rules.

You can adopt either of the following:

- a reactive approach to monitor the evolution, and then decide what the corrective actions are.
- a proactive approach. Embed the rule in an existing business process and steer it in the business objective direction.

REACTIVE: MONITOR YOUR BUSINESS OBJECTIVES

1. Define business rules. The bank needs to calculate the KPI described and assess whether the bank has achieved its goal.

Display 9. Calculate a KPI in a Rule Set Using SAS Intelligent Decisioning

2. Apply the business rules on data. Test the rule first. Secondly, publish the rule to a destination. The rule can run in SAS Viya : CAS or Micro Analytic Services (MAS), **or 'in-database' in** Teradata or Hadoop. Suppose you run the rule in batch, therefore you publish to CAS. Results will be logged in several CAS tables.

- Rule output: Example: the only year when the bank did not achieve the **board's objective was in 2015.**

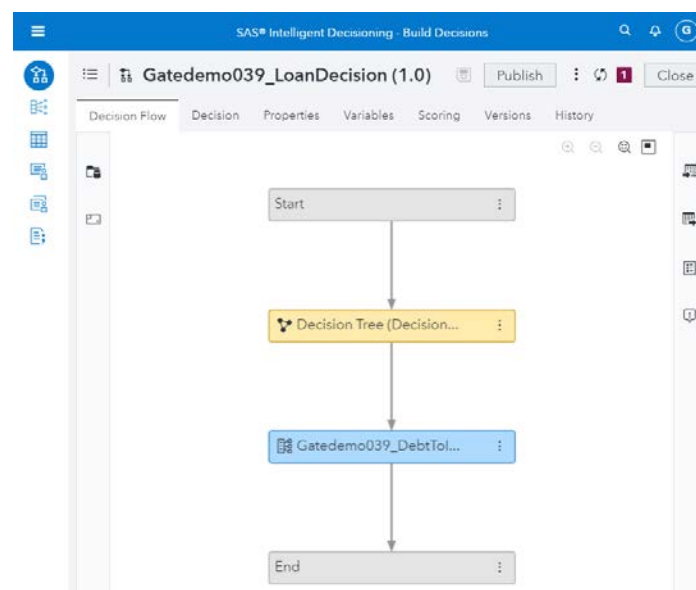| year ↓	Objective	Rules Fired Count	KPI	DIR30Amount	DIR30Amount_p	loan_amount	loan_amount_p
2015	not achieved	1	562025	562025	0	6417580175	3503830175
2014	achieved	1	0	0	0	3503830175	1981989225
2013	achieved	1	0	0	0	1981989225	718411025
2012	achieved	1	0	0	0	718411025	261683825
2011	achieved	1	0	0	0	261683825	131992550
2010	achieved	1	0	0	0	131992550	51928250
2009	achieved	1	-17430	0	16600	51928250	21119250
2008	achieved	1	-15687.5	16600	30750	21119250	4977475
2007	achieved	1	-1537.5	30750	30750	4977475	4977475

Display 10. Monitor the KPI – Rule Output in SAS Intelligent Decisioning

- Rule fired: rule metadata, records used by the rule, how many fired the rule, at which date, and so on.
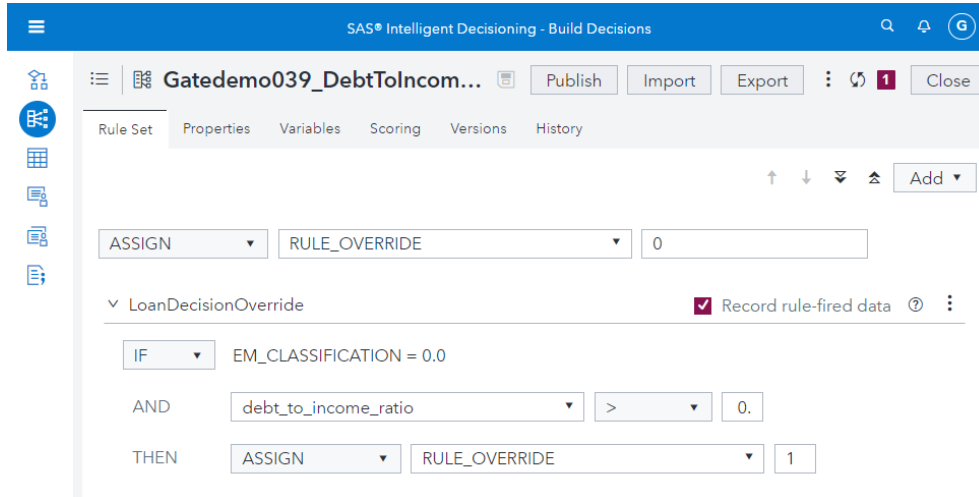
PROACTIVE: STEER YOUR BUSINESS

For example, the bank uses a model to grant loans to customers. The bank board wants to limit the loans granted to over-indebted customers. Suppose the process to review the model takes too long. What can you do?

1. Design a decision with the model and a business rule.

Display 11. Proactive Decision: Model and Rule Using SAS Intelligent Decisioning

2. Define a rule to override the model output.

Display 12. Rule that Overrides a Model in SAS Intelligent Decisioning

3. Interpret the output. The model assesses which customer should be granted the loan. The rule overrides the model when the customer has **a 'Debt to Income Ratio' higher than 30%**.

Display 13. Rule Output Overrides Model for Over-indebted Customers in SAS Intelligent Decisioning

In conclusion, you can use the business rule you developed to steer the business process toward the business objective. That is, do not grant any loans this year to over-indebted customers.

TRACE DATA PROCESSING

On a boat, several crew members keep constant watch: the first, on the starboard side, the second, on the port side. A third crew member oversees the stern. Each side has its own unique view. All the crew members use verbal communication when something needs attention. The helmsman looks ahead and integrates **the crew's observations** in a 360-degree simplified view.

SAS Lineage is useful for internal analysis and external compliance. With SAS Lineage you can track different technical processes, from source files or tables to models and reports, even across cloud, platforms, and technologies.

On the SAS Platform, SAS Lineage in SAS Viya acts as the crew member on the starboard side. SAS Lineage on SAS 9.4 is the crew member on the port side. SAS® Metadata Bridge [for a technology] is the crew member looking astern.

SAS Relationship REST API in SAS Viya and SAS 9.4 embodies the verbal communication. SAS 9.4 Lineage is the helmsman and can keep the 360-degree view.

In the banking example, the data source is in AWS S3. To integrate lineage from the cloud, SAS Viya, and SAS 9.4, the process can be broken-down into the following five steps:
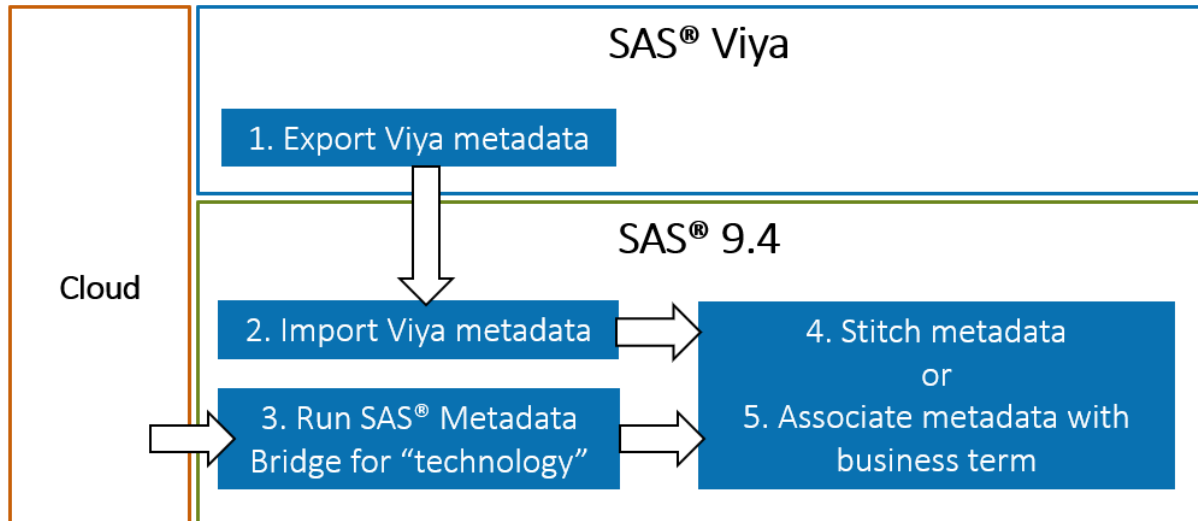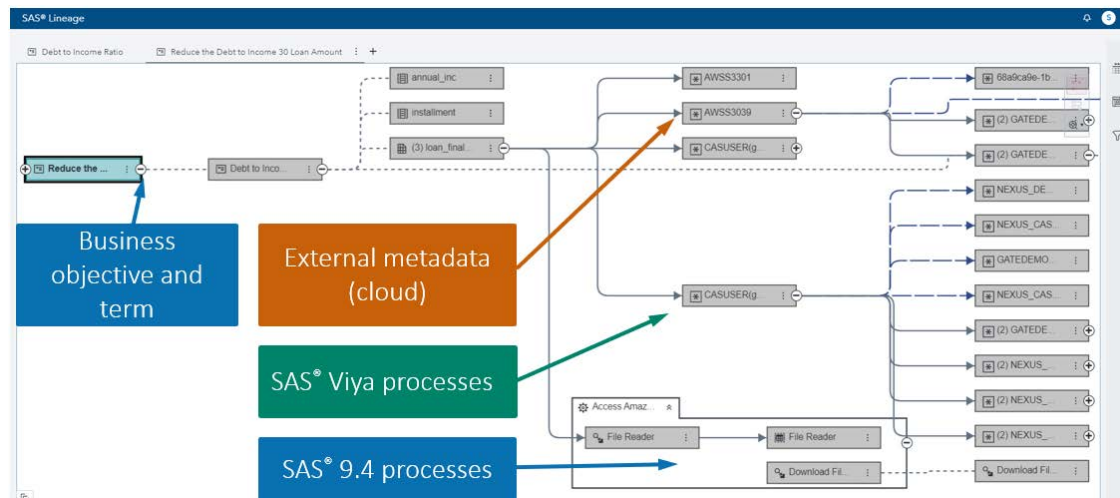

Figure 1. Integrate Metadata from Different Sources on the SAS Platform

1. Export SAS Viya metadata.
   - Retrieve objects from SAS Viya Lineage with the relationship REST API in SAS Viya.
   - Transform the export to an XML file accepted by SAS 9.4 Lineage.

2. Import SAS Viya metadata in SAS 9.4 Lineage.
   - Use the Relationship REST API in SAS 9.4.
   - POST the XML file and import the SAS Viya metadata into SAS 9.4.

3. Import external metadata. In the banking example, data is stored in AWS S3. Therefore, SAS® Metadata Bridge for AWS S3 should be used. (The bridge will be available in SAS 9.4 release M7, but the process described in step 2 is applicable to any other bridge).
   - Configure the metadata bridge.
   - Run the bridge and import external metadata in SAS 9.4 Lineage.

4. Integrate the views above in SAS 9.4. "Technical **Stitch" the metadata from** different sources.
   - Prepare an XML file containing the metadata IDs of the objects you want stitched.
   - Import with the SAS 9.4 Relationship REST API.

After you performed the steps above, SAS 9.4 Lineage could display the loan file linked to external metadata , SAS Viya metadata such as CAS tables, plans, rule sets, and SAS 9.4 metadata such as terms, or a SAS® Data Integration Studio job accessing the file.

Display 14. Integrated Metadata from Different Sources in SAS 9.4 Lineage

5. You could **"Business Stitch" the metadata as an alternative to step** four above. In SAS Business Data Network, simply associate the business term with the metadata generated by SAS 9, imported from SAS Viya, and imported metadata from your cloud or external source.

Your business objective now becomes traceable. You can centralize all the technical processes used to achieve that goal.

REPORT AND VISUALIZE

On a boat, thousands of data points are generated every minute. The instrument console keeps track of these data points and displays them so that you can take a decision.

On the SAS Platform, you can integrate, analyze, and visualize all the key elements from define until trace with a data governance dashboard. The data governance dashboard is not an out of the box component as it requires custom code to be written. The dashboard and the data model behind represent a recommended approach when addressing data governance.

With the dashboard you can close the loop, evaluate if the business objectives have been achieved. You can also communicate to different stakeholders:

- What prevents the achievement.
- What business rules have been violated and the detailed data behind.
- The evolution in time of the business objectives and rules.

If you need to build a dashboard you should take the following steps: extract, integrate, and visualize.

1. Extract:

- Business metadata: terms, contacts, and related terms. Use the SAS Business Data Network REST API.
- Technical metadata: links between business terms and business rules. Extract with the SAS 9.4 Relationship REST API.

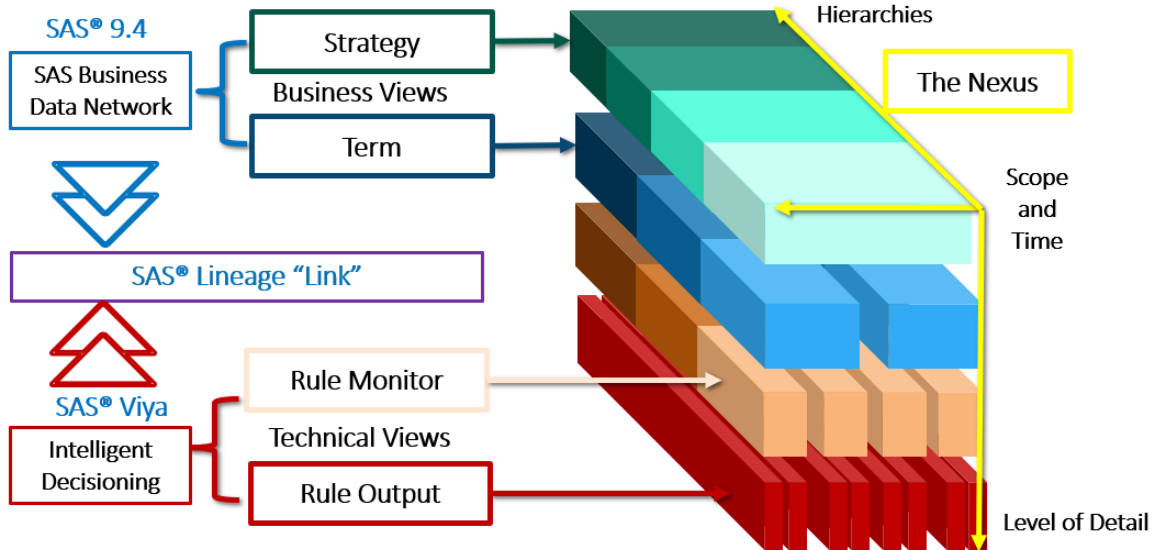- Rules output and rule fired data. Extract from SAS Intelligent Decisioning.

Figure 2. Extract Components Needed for Data Governance Reporting

2. Integrate the extracted components in a custom made data mart, **'The Nexus'**. It is a materialized as a CAS table and consists of the following:

- Facts: 'RULE_EXECUTION' logs the outcome of an executed business rule. The 'Score' is the ratio: ('Rows Processed' – 'Triggers') / 'Rows Processed'. In other words, the percentage of correct vs. total occurrences in the business rule.

- Dimensions: details of the rules applied, execution details or business terms, business objectives (strategy terms), associated items (metadata links between terms and business rules), and responsible persons (contacts and roles, such as Business Custodian, Data Steward, Technical Custodian, Data Owner, and so on).
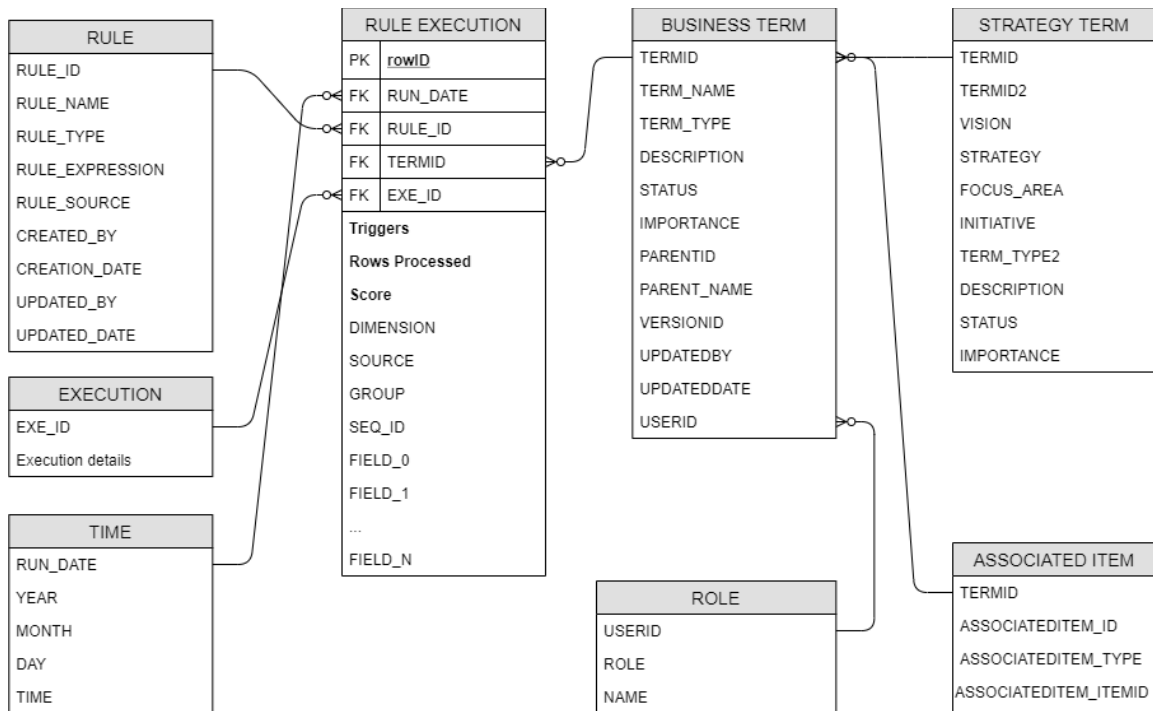
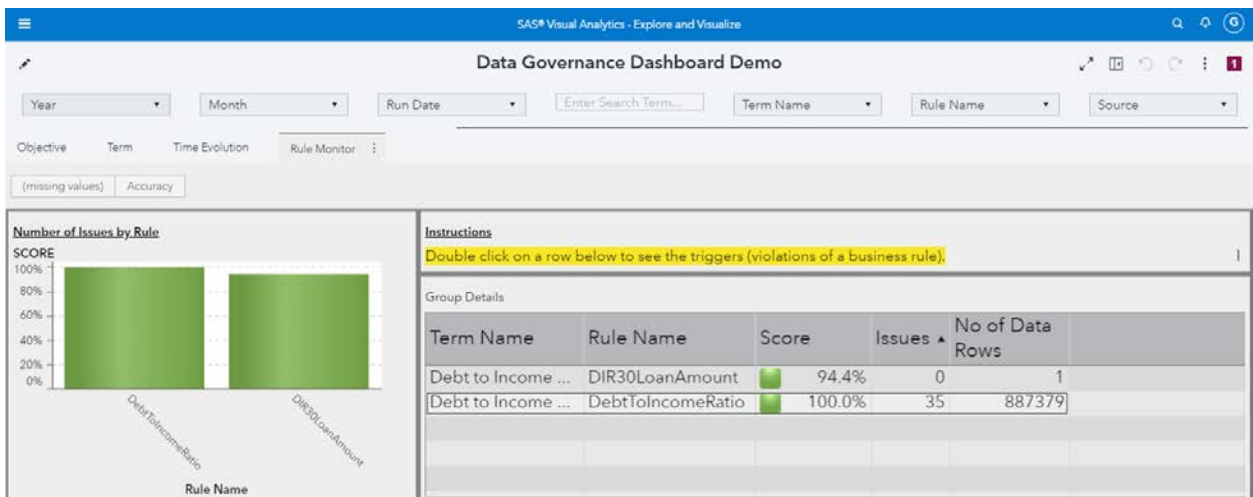

Figure 3. 'The Nexus' Data Model

3. Visualize, query, and interpret the facts and the dimensions in 'The Nexus' with a SAS Visual Analytics report**, the 'Data Governance Dashboard'.** You can follow the scores in different views.

- Rule Output: sample raw data from the rule output, at a moment in time. This is the lowest level of detail available.

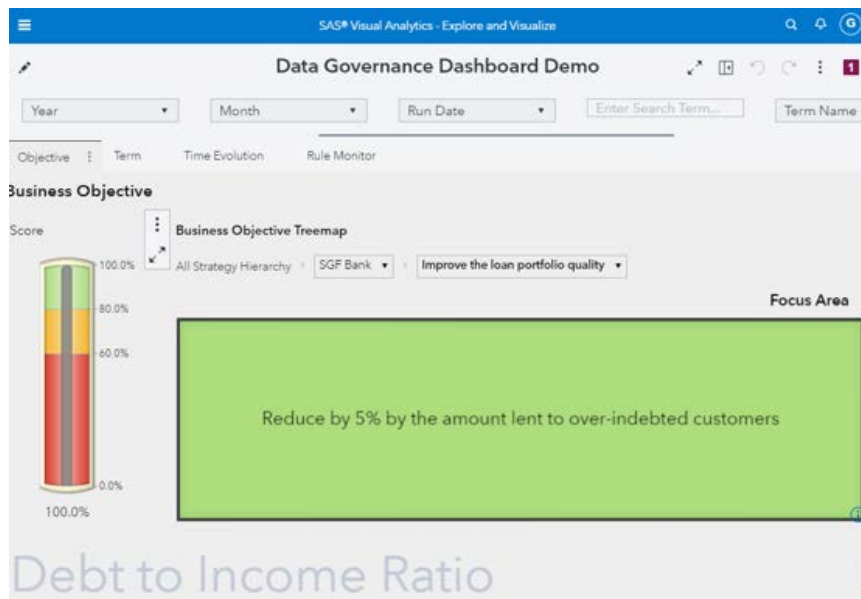**Display 15. Rule Output in a Data Governance Dashboard Using SAS Visual Analytics**

- Rule monitor: shows the score of each rule, the triggers, and the rows processed, at a moment in time.

Display 16. Rule Monitor in a Data Governance Dashboard Using SAS Visual Analytics

- Term: the business rule scores aggregated at a business term level. In typical implementations, you can use rule weights or lower and upper thresholds.

- Objective: the KPI derived from the business rule at a business objective level.

Display 17. Objective in a Data Governance Dashboard Using SAS Visual Analytics

- Time evolution: a rule can run multiple times and can have multiple run dates. The view tracks the evolution of the score across the dimensions in **'The Nexus'**.

By integrating different elements uncovered in the previous steps, you now have a solid base to analyze the issues, understand the cause, prioritize the issues, and plan the remediation process.

REMEDIATION

When you navigate, in case something unexpected happens, you need life-saving equipment: water pumps, fire-extinguishers, tools and kits you can

In SAS Viya, you can use the following remediation tools:

- SAS Data Studio is an interactive, self-service product that allows you to access, blend, shape, and cleanse data. SAS Data Studio makes now use of a Suggestion Engine, using machine learning to propose you the actions you should take. Once you accept the suggestions, Data Studio creates the steps in the data plan for you. You can be given the option to cleanse, standardize, de-duplicate, enrich. These are called data quality functions and make use of the SAS Quality Knowledge Base. Or you can perform one-hot encoding, center and scale, imputation, join, enrich, calculate, and so on.

- SAS Data Quality functions to cleanse your data: *dqstandardize, dqmatch, dqparse*. These can run in a data step, in CAS. More in the reference section: Nicolas Robert, SAS Institute Inc. *Data Quality Programming Techniques with SAS® Cloud Analytic Services.* 2019.

- CAS action to group records based on customized rules: *entityRes.match*.

- CAS action to update records in a CAS table: *table.update*.

- CAS action to delete records in a CAS table: *table.deleteRows*.

- CAS action to change a CAS table or CAS view: *table.alterTable*.

All the above, will write the results in a CAS table. If you need to correct the source, you need to save back into the data source associated with the caslib. You can use a PROC CASUTIL *SAVE* statement or a **table.save CAS action.**

In SAS 9.4 you can use SAS® **Data Remediation as your "data issue organizer". You can** assign issues to users and track them. In the interface users can visualize the issues, assign to other users, resolve or reject the issue, comment, etc. Data Remediation has a web-based interface for data administrators and REST API for system integration.

CONCLUSION

You should now have the confidence to embark on a journey **to the "Governed Data"** port, where business objectives drive data choices and data management activities, where data is trusted and supports business objectives.

To make the journey safer and more efficient, you need to start with the definitions, and then access, understand, and prepare the data. Create business rules and monitor your data. Steer your business with business rules. Trace every processing step from source to reporting. Quantify any issues and qualify and prioritize them with a data governance dashboard. Finally, plan your remediation strategy to improve data.

Your journey is supported by the SAS Platform. Some steps are easier and more efficient with SAS Viya, while others need SAS 9.4. Use the most appropriate to arrive at your destination.

CODE

Sample code for the steps described in this paper can be found on the SAS Global Forum 2020 GitHub page: https://github.com/sascommunities/sas-global-forum-2020/tree/master/papers/4223-2020-Teleuca .

REFERENCES

D'Alessio, G. and S. Iezzi. 2012. Household Over-Indebtedness. Definition and Measurement with Italian Data. *Proceedings of the Sixth IFC Conference on "Statistical Issues and Activities in a Changing Environment."* Basel, Switzerland: IFC. Available https://www.bis.org/ifc/events/6ifcconf/dalessioiezzi.pdf.

Dyché, Jill and Analise Polsky. SAS Institute white paper. "5 Models for Data **Stewardship."** Available https://www.sas.com/content/dam/SAS/en_au/doc/whitepaper1/5%20Models%20Data%20Stewardship%20White%20Paper.pdf.

McIntosh, Liz. 2016. New for SAS® 9.4: Integrating Third-Party Metadata for Use with Lineage and Impact Analysis. *Proceedings of the SAS Global Forum 2016 Conference.* Cary, NC: SAS Institute Inc. Available http://support.sas.com/resources/papers/proceedings16/SAS5562-2016.pdf.

Rausch, Nancy. **2020. What's new in SAS Data Management.** *Proceedings of SAS Global Forum 2020.* Cary, NC: SAS Institute Inc. Available https://www.sas.com/en_us/events/sas-global-forum/program/proceedings.html.

Rejany, Vincent. 2019. Data Governance: Harder, Better, Faster, Stronger. *Proceedings of the SAS Global Forum 2019 Conference.* Cary, NC: SAS Institute Inc. Available https://www.sas.com/content/dam/SAS/support/en/sas-global-forum-proceedings/2019/3247-2019.pdf.

Rejany, Vincent and Bogdan Teleuca. 2018. Enable Personal Data Governance for Sustainable Compliance. *Proceedings of the SAS Global Forum 2018 Conference.* Cary, NC: SAS Institute Inc. Available https://www.sas.com/content/dam/SAS/support/en/sas-global-forum-proceedings/2018/2052-2018.pdf.

Robert, Nicolas. 2019. Data Quality Programming Techniques with SAS® Cloud Analytic Services. *Proceedings of the SAS Global Forum 2019 Conference.* Cary, NC: SAS Institute Inc. Available https://www.sas.com/content/dam/SAS/support/en/sas-global-forum-proceedings/2019/3030-2019.pdf.

ACKNOWLEDGMENTS

CONTACT INFORMATION

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

Bogdan Teleuca
SAS Australia & New Zealand
bogdan.teleuca@sas.com