

# SAS Data Management Update / Information Catalog

Günter Greulich, Sr Solutions Architect

17.September 2020

Plattform Netzwerktreffen | Online Live Session (via WebEx)

# Better Together

Data Management

Analytics

*"I didn't win without  
SAS Data Management,  
and that's why I consider  
him my best teammate  
of all time.*

*He helped me so much in  
the way I approached  
the game, in the way I  
played the game.*

*Whenever they speak  
SAS Analytics, they  
should speak  
SAS Data Management."*



# SAS Data Management on SAS 9.4

## Product Themes

1

### Cloud Connectivity Enabled

Access data in any format and any location

2

### Updated & Secured

SAS 9.4 products are Flash free and stable

3

### Streamlined

Data Management portfolio is rationalized and simplified

# Now

## As part of SAS 9.4M7 – 20W34

- SAS/ACCESS Interface to DB2 – Support for IBM Integrated Analytics Server (IIAS)
- SAS/ACCESS Interface to BigQuery – Bulk Unload
- SAS/ACCESS Interface MongoDB – GA
- SAS/ACCESS Interface to MSSQL – Support for Microsoft Azure Synapse
- SAS/ACCESS Interface to PostgreSQL – Support for EnterpriseDB
- SAS/ACCESS Interface to SalesForce – GA + Support for Proxy
- SAS/ACCESS Interface to Spark – GA & LAX only
- SAS/ACCESS Interface to Yellowbrick – GA
- SAS In-Database Technologies for Aster, DB2, Greenplum, Netezza, Oracle, SAP Hana, Teradata - Updates
- SAS In-Database Technologies for Spark – GA (CDH 6.x, HDP 3.x, MapR 6.x, AWS EMR 5.x)
- SAS In-Database Technologies for Hadoop – Updates

## As part of SAS 9.4M7 – 20W47

- SAS/ACCESS Interface to BigQuery – BigQuery Storage API (Hotfix)
- SAS/ACCESS Interface Hadoop – Support for CDP 7.1 (in-cloud and on-premises)
- SAS/ACCESS Interface to Netezza – Add Support for IBM Performance Server for PostgreSQL
- SAS/ACCESS Interface to Snowflake – Support Bulk Load on Microsoft Azure (ADLS)
- SAS In-Database Technologies for Spark – Support for CDP 7.1 (in-cloud and on-premises)

# Future

## 21WXX or HotFixes

- SAS/ACCESS Interface to DB2 – Support for external table loading technique
- SAS/ACCESS Interface to PostgreSQL – Support for CockroachDB
- SAS/ACCESS Interface to SalesForce – Support for Updates, Inserts
- SAS/ACCESS Interface to Snowflake – Support for bulk load/unload on GCP Storage
- SAS/ACCESS Interface to Spark – Support for Google DataProc
- SAS In-Database Technologies for Aster, DB2, Netezza, Oracle - Updates
- SAS In-Database Technologies for Spark – CDH 6.x, HDP 3.x, MapR 6.x, AWS EMR 5.x, CDP 7.X
- SAS In-Database Technologies for Hadoop – Updates
- SAS QKB Definition Editor

# SAS Data Management on SAS Viya

## Product Themes



1

### Together with Analytics

Streamline data preparation & integration,  
and analytics transparency

2

### Together with the Cloud

Access data in any format and any location

3

### Together with Automation

Scale and Democratize Data Management  
with AI

# Together with the Cloud

**teradata.**

Amazon Web Services  
Microsoft Azure  
Teradata IntelliCloud  
VMWare



Amazon EMR  
Azure HDInsight



Amazon Aurora  
Amazon RDS MySQL  
MariaDB, MemSQL  
**GCP MySQL**  
**Azure MySQL**



**Microsoft SQL Server**

Azure SQL Database  
**Azure Synapse**  
**Azure Big Data Cluster**  
Amazon RDS MS SQL

**ORACLE**

Oracle Cloud Platform  
Amazon RDS Oracle

**PostgreSQL**

Amazon Aurora  
Amazon Redshift  
Amazon RDS  
PostgreSQL  
**GCP PostgreSQL**  
**Azure PostgreSQL**



# Released

## SAS/ACCESS Interfaces to

- BigQuery
  - MongoDB
  - SalesForce
  - Snowflake
- SAS Viya 3.5*
- SAS Data Preparation
  - SAS Data Quality
- SAS Viya 3.5*

# Now

## SAS/ACCESS Interfaces to

- DB2
  - BigQuery
  - MongoDB
  - MSSQL
  - PostgreSQL
  - Spark
- 2020.0.7*
- BigQuery
  - Hadoop
  - Netezza
  - Snowflake
  - Yellowbrick
- 2020.1.1*

## SAS In-Database Technologies for

- Hadoop
  - Spark
  - Teradata
- 2020.0.7 - 2020.1.1*

- SAS Studio “Flow”
  - SAS Data Preparation
  - SAS Data Quality
  - SAS QKB CI 32
- 2020.0.7*

# Next

## SAS/ACCESS Interfaces to

- MongoDB
  - BigQuery
  - PostgreSQL
  - SalesForce
  - Snowflake
  - Spark
- 2020.1.1*

- SAS Studio “Flow”
  - SAS Information Catalog
- 2020.1.1*

## SAS In-Database Technologies for

- Spark
- 2020.1.4*
- SAS Data Enrichment
  - SAS Data Quality
- 2021.1.2*

# Future

## SAS/ACCESS Interfaces to

- DB2
  - MongoDB (CosmoDB)
  - REST
  - Snowflake
  - Spark
- 2020.1.4*

- SAS Studio “Flow”
  - SAS Information Catalog
- 2020.1.4*

- SAS Studio “Flow”
  - SAS Information Catalog
  - SAS Data Quality
  - SAS In-Database Technologies for Spark
  - SAS In-Database Technologies for MSSQL
- 2021.1.2*

- “Business Glossaries” on SAS Viya
- 2021.2.2*



SAS Viya

# Together with Analytics

## SAS Studio - Unified Flow Designer on SAS Viya 4

SAS® Studio - Develop SAS Code

Start Page \* Flow.flw x +

Run | Add | View | Apr 20, 2020 05:18 PM

Flow Generated Code Submission

Tasks

Filter tasks

SAS Tasks My Tasks

Flow Tasks

- File
- Import
- Query
- SAS Program
- Table

Prepare Data

SAS Viya Cloud Analytic S...

SAS Viya Econometrics

SAS Viya Evaluate and Im...

SAS Viya Forecasting

SAS Viya Machine Learning

SAS Viya Optimization an...

SAS Viya Prepare and Expl...

SAS Viya Statistics

SAS Viya Text Analytics

beers

Import

breweries

Import

Query

Table

SAS Program

...

SAS Program

Code Node

```
1 ⊖ proc freq data=&input1;
2      tables beer_name;
3      where state = "ND";
```

The screenshot shows the SAS Studio interface with a flow diagram and generated SAS code. The flow diagram consists of nodes: 'beers' and 'breweries' (both represented by purple icons) connected to 'Import' nodes (blue square with a white icon), which then connect to a 'Query' node (blue square with a white circle). The 'Query' node has two outputs: one to a 'Table' node (black circle with a white grid) and one to a 'SAS Program' node (yellow square with a white icon). Below the flow diagram, the 'Generated Code' tab is selected, showing the following SAS code:

```
1 ⊖ proc freq data=&input1;
2      tables beer_name;
3      where state = "ND";
```

# SAS Studio - Prioritized Themes

- 1
- 2
- 3
- 4
- 5

Standard usability features (Status indicators, Run/Stop, Step by Step execution)

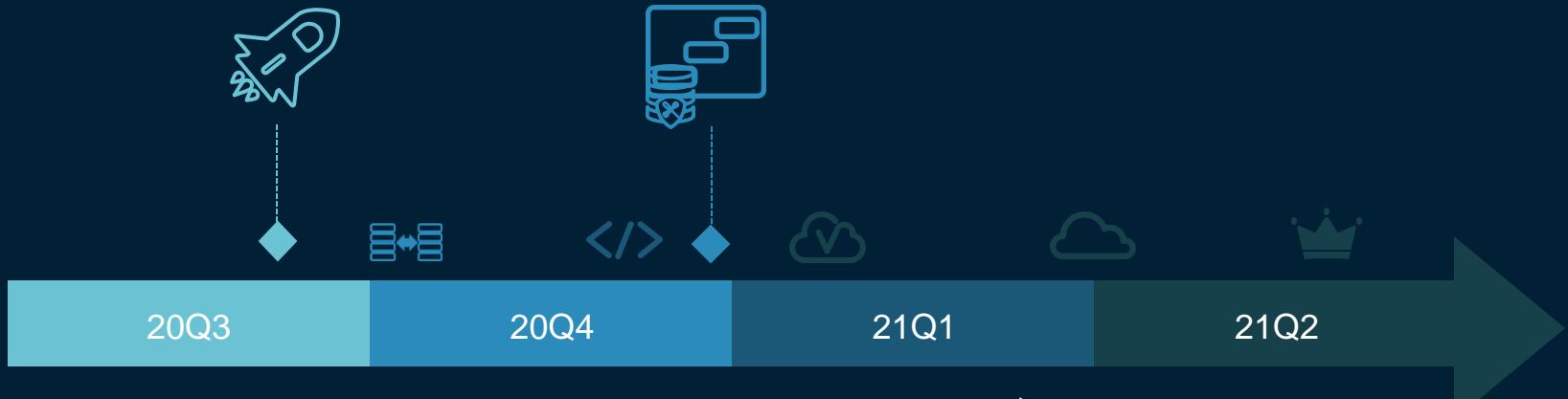
Custom Transformations

Data Movements (Insert/Delete/Update, Union, Distinct)

Enterprise Guide Projects Migration

Read/Write files from/to Azure ADLS

# High Level Plan (Proposed)



*Information is subject to change*



SAS Information Catalog

An intuitive self-service  
Data Catalog portal to  
Analytics and Data  
Management.

# Together with Automation

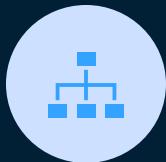
## A new perspective on your metadata

The screenshot displays the SAS® Catalog - Find and Collect Assets interface. At the top, it shows the catalog path: Catalog Home > Search Results > Cars. Key metadata for the asset 'Cars' is listed: Columns 9, Rows 15, Size 495, Value score: 100, Completeness: 82%, Risk Score: Low, Status: Approved. The asset is categorized under Vehicles. Below this, there are tabs for Overview, Column Analysis, Sample Data, and Collaboration. A message box indicates that a steward can post a message or notifications to a specific asset or group of assets. The Properties panel on the right lists ownership information (Andrew Bird - Curator, Beth Thompson - Technical Steward), provider (CAS In-memory), source (Cars\_Merged.sashdat), location (Cary, NC, USA), encoding (UTF-8), version (3.2), and modification details (Last modified: April 30, 2019 03:54 PM, Modified by: Andrew Bird). The Usage section features a donut chart showing 387 total usage across categories: Reports (green), Projects (blue), Decisions (yellow), Models (purple), Jobs (orange), Plans (cyan), and Other (red). A line chart tracks the Value Score over time from November to June. The Top Users section lists Gopichand Sana, Shadrius Pearson, Ingo Schimpff, Freddy Kauschke, Aneta Škodová, Otmar Doležal, Leelah Leatherbarrow, Teng Hu, and Dan Sheldon, each with their profile picture and title. A 'Notify Users' button and a link to see all 36 users are also present. The bottom right corner features the EGERIA CONFORMANT logo.

# SAS Information Catalog Project Goals



Deliver foundational  
metadata services for  
SAS Viya



Support SAS Customers  
needs in terms of Data  
Cataloging & Metadata  
Management



Provide populators to  
crawl, index, classify,  
and profile content



Provide intelligent, fast,  
ranked, natural  
language search

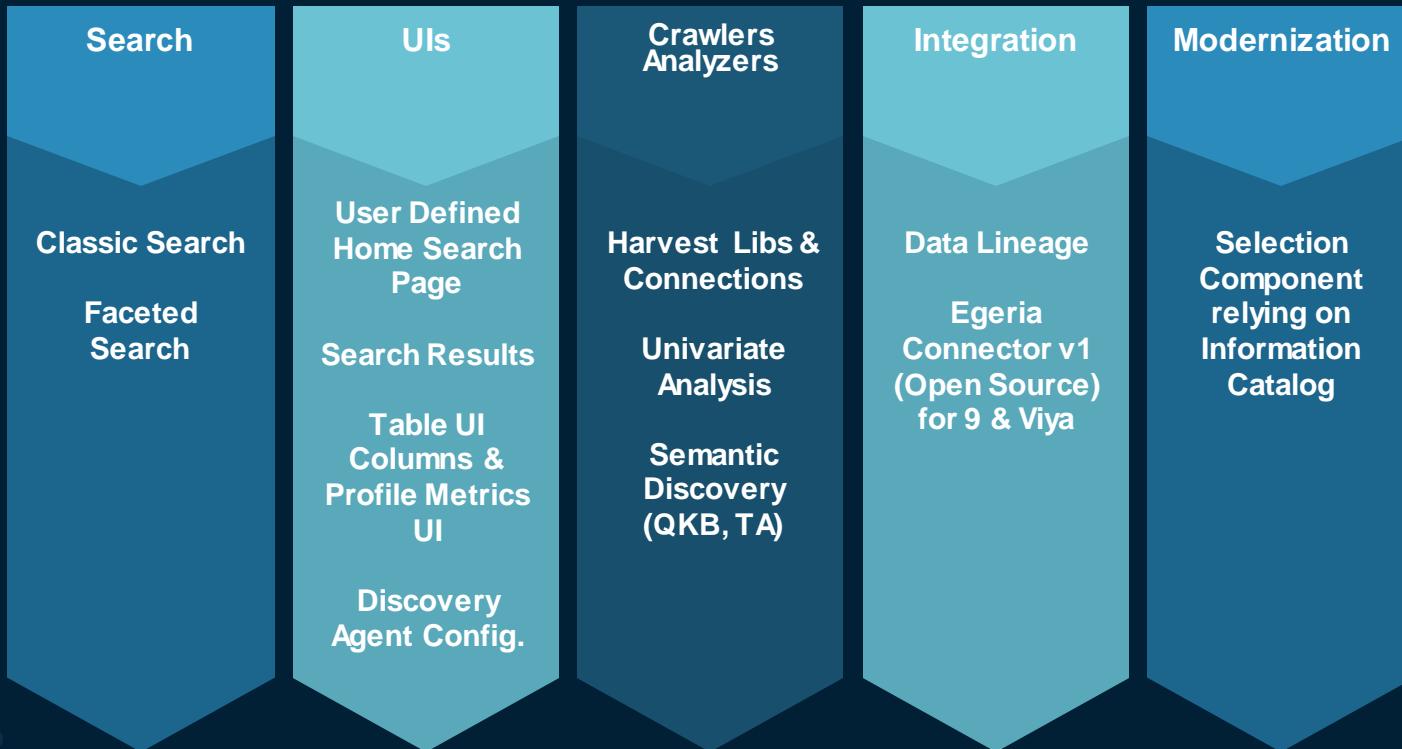


Support federation of  
metadata with external  
systems



Support an upgrade  
path for metadata from  
SAS 9 to Viya

# SAS Information Catalog - Scope



## Looking Ahead

# ODPi Announces Egeria for Open Sharing, Exchange and Governance of Metadata

Industry's First Open Metadata Standard Helps Organizations Better Understand, Manage and Gain Value from Data



Automation

Availability

Open & Accessible

Discovery



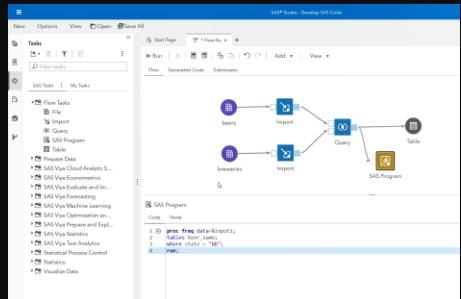
"Egeria and its open approach to metadata management and integration only underscores further the need for metadata standards to promote responsible data exchange across varied technology environments."

# Coming 2020-2021

## Together with Analytics

### SAS Studio

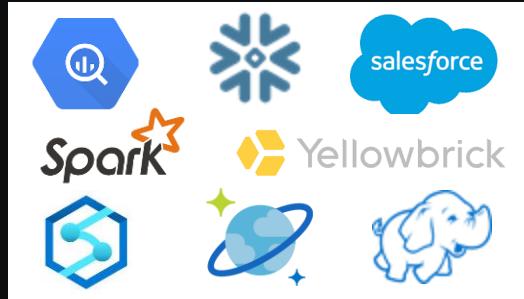
- New unified flow designer on SAS Viya
- Role based driven user interface and transformations
- SAS Data Integration functions embedded in new SAS web-based data flow designer
- Multi-language code generation angle
- AI Driven Optimization Code/Execution location features



## Together with the Cloud

### SAS Connectivity

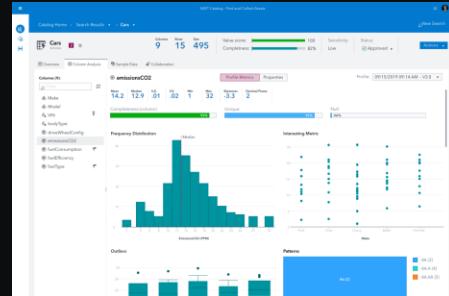
- Machine Learning Driven Data Quality functions
- Faster implementation of identification definitions
- Provides a solution for customers who have proprietary data types
- Faster introduction of support for new domains



## Together with Automation

### SAS Information Catalog

- Augmented repository including information on entities, attributes, relationships, and domains.
- Discover, classify and document how data fits together as well as provide a common vocabulary around data.
- Integrated with Data Profiling, personal data identification
- Open Metadata Integration through ODPI Egeria





# Fragen ??

sas.com



Vielen Dank !

QUARANTINE  
& CHILL

