

# Modernizing Your Data Strategy:

## Understanding SAS<sup>®</sup> Solutions for Data Integration, Data Quality, Data Governance and Master Data Management

Gregory S. Nelson  
ThotWave Technologies, Chapel Hill, NC

Lisa Dodson  
SAS Institute, Cary, NC

### Abstract

For over three decades, SAS has provided capabilities for beating your data into submission. In June of 2000, SAS acquired a company called DataFlux to add data quality capabilities to its portfolio. Recently, SAS folded Data Flux into the mother ship and with SAS 9.4, the SAS Enterprise Data Integration (and baby brother Data Integration) solutions were upgraded into a series of new bundles that still include the former DataFlux products, but those products have grown. These new bundles include data management, data governance, data quality and master data management and come in advanced and standard packaging.

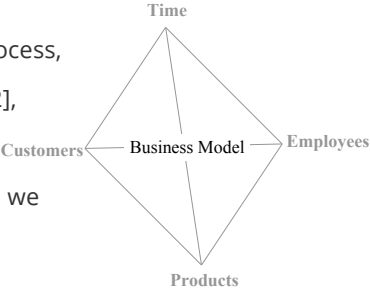
This paper will explore these offerings and help you understand what this means to both new and existing customers of the Data Integration and DataFlux products. We will break down the marketing jargon and give you real world scenarios of what customers are using today (pre-SAS 9.4) and walk you through what that might look like in the SAS 9.4 world. Each scenario will include what software is required, what each of the components do (features and functions) as well as the likely architectures that you may want to consider. Finally, for existing Data Integration customers, we will discuss implications for migrating to the new version and detail some of the functionality that may be new to your organization.

<b>INTRODUCTION .....</b>	<b>2</b>
DATA INTEGRATION .....	3
DATA QUALITY .....	5
MASTER DATA MANAGEMENT .....	5
DATA GOVERNANCE .....	6
<b>SAS TECHNOLOGY LANDSCAPE .....</b>	<b>7</b>
<b>REFERENCE ARCHITECTURES.....</b>	<b>9</b>
TYPICAL ARCHITECTURES.....	9
MODERNIZATION STRATEGIES.....	10
<b>SUMMARY .....</b>	<b>15</b>
<b>REFERENCES .....</b>	<b>16</b>
BIOGRAPHY .....	17
CONTACT INFORMATION .....	17

# Introduction

For decades, the traditional SAS aficionado has used the SAS language to beat data into submission. Using the fundamentals of the DATA STEP and PROC SQL, programmers have coaxed data out of systems and quietly transformed them into beautifully orchestrated visualizations. Sometime in the late 1980's, people talked of data warehousing and we upgraded our vocabularies to maintain pace with the state of the art.

But more than just words, data warehousing introduced us to a discipline – improving process, quality, collaboration and reuse. Whether you are a proponent of Inmon [1] or Kimball [2], we learned how to describe our craft in a modern light. Using phrases like “time variant”, “facts and dimensions”, “change data capture” and “extract, transform and load”, we began to improve our resumes and how data flowed through our organizations.



Fast forward two decades and we realize that fundamentals of data have not changed. We still struggle with making the right data available to the right people at the right time in the right form. This is in part due to our appetite for knowledge, which is only further whetted by the volume, velocity and variety of data in and outside of our organizations. In fact, if we look Gartner's *The Top 10 Strategic Technology Trends for 2013* [3], we can clearly see the threads of data throughout each of the major trends – each generating or describing data and the need for analytics to understand, improve and optimize:

1. Mobile device battles
2. Mobile applications and HTML5
3. The personal cloud
4. The Internet of Things
5. Hybrid IT and cloud computing
6. Strategic big data
7. Actionable analytics
8. Mainstream in-memory computing
9. Integrated ecosystems
10. Enterprise app stores

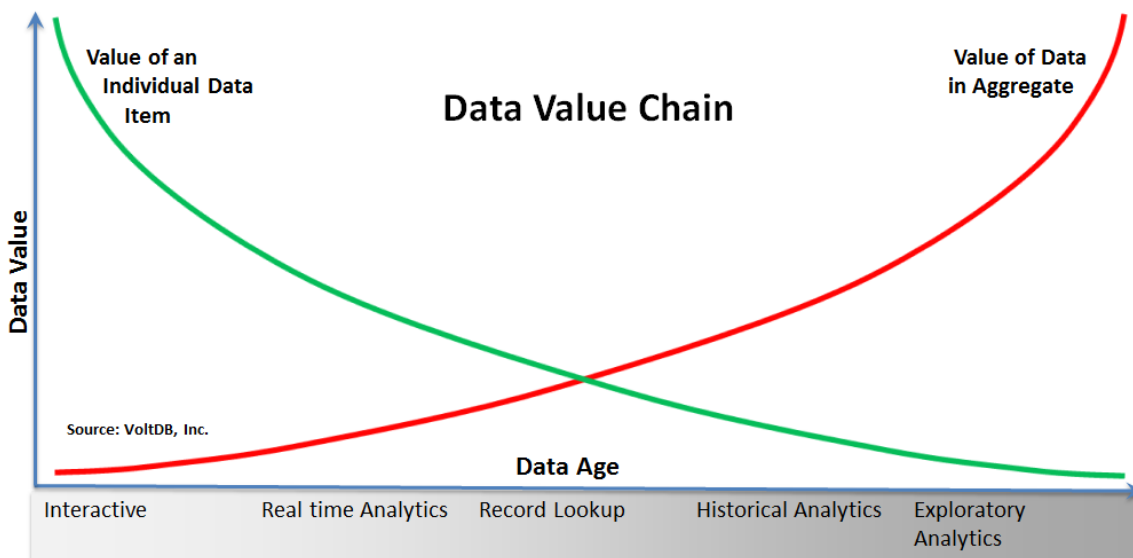
Except for the sleepless or voracious, keeping up with the methods and technologies for managing data can be daunting. In this paper, we wanted to take you through an exercise where we first explain some of the concepts in modern data systems and then relate them to the technologies found in the SAS portfolio.

We recognize organizations, and the industries in which each operate, are unique in how they think about and use the terms data integration, data management, data quality, master data management and data governance. This is further exacerbated by the fact that the vendor community often disagrees on their relative importance, how the tools function and their implementation (for example, in memory versus in database.) So let's start by characterizing these concepts in terms of their function in a typical organization, regardless of technology.

## Data Integration

To address these technologies, it is prudent that we start with the terms “data management” and “data integration”. While we often hear people speak about these in the same context, for our purposes they are two different concepts.

Data integration means bringing data from two or more sources together into a single view for analysis and reporting. Common examples from industry might include bringing together customer sales data with warranty claims or integrating patient results stored in an electronic health record with fitness activity tracking data from something like Fitbit. Data that is integrated is often far more valuable than data that stands alone – especially as data tends to decay over time. See the diagram below for a visual depiction of this relationship.



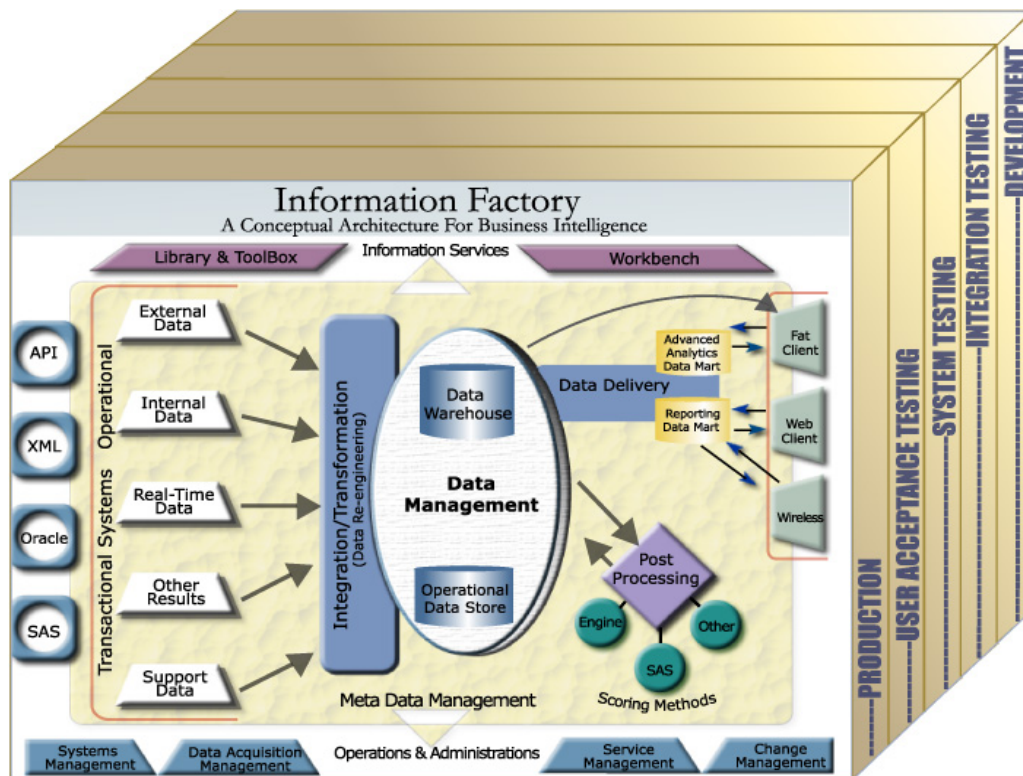
There are a number of vendors and technologies in the data integration space that support the construction and implementation of data access and delivery. Common applications include data acquisition for data warehousing, business intelligence and analytics; integration of master data in support of master data management (MDM); data migration or conversion (common when integrating systems, companies or legacy system retirement); data sharing where information is exchanged beyond the corporate firewalls with partners/ suppliers, customers or regulatory agencies; or in the delivery of data throughout an organization (enterprise application or a service-oriented architecture (SOA)).

While data integration is seen as more of a tactical component of an overall data architecture in an organization, data management can be thought of as the global set of practices that govern how data strategies are designed, executed and governed within an organization. Think of data management as the guiding principles, architectures, policies, practices and procedures for managing data within

enterprise and data integration as the set of tools that support the extraction, transformation and loading processes.

In SAS architectures, data integration can be accomplished with SAS/Access Engines, SAS Foundation components (e.g., DATA STEP, PROC SQL) or through the use of SAS solutions like SAS Data Integration and SAS Enterprise Data Integration ([4] for a comparison). In a previous paper, one of the current authors collaborated on a paper [5] describing how BASE SAS and SAS Data Integration Studio compared relative to Ralph Kimball's benchmark standards for the functions that an ETL (or data integration) tool should provide. Generally, the functions of a data integration solution include the following features/ functions:

- Design and development environment
- Metadata management and data modeling
- Source data extraction (connectivity/adaptor, change data capture)
- Transformation and loading (transformation and data delivery)
- Interoperation with data governance via data quality and profiling
- Deployment (multiplatform, cloud, in-memory, in-database, virtualization)
- Operations and administration (deployment, flow / process control, auditing and exception handling/management, traceability/ lineage)



Later in this paper, we will map how these functions are implemented in the modern SAS Solutions.

## Data Quality

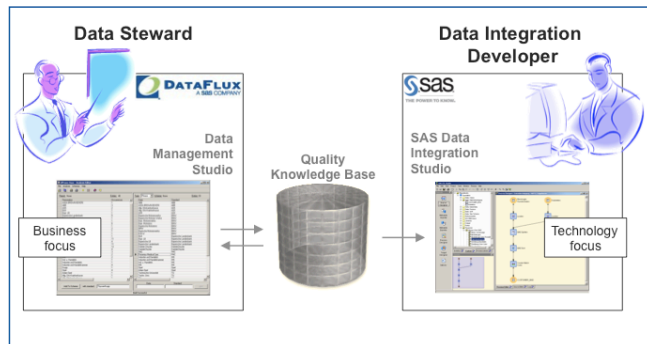
"Quality information is not the result of doing work.", as Larry English states in his book *Information Quality Applied*, but rather "[comes] as a result of designing quality (error-proofing) into processes that create, maintain, and present information...". [6]

While many of us have little control over the quality of data before it gets to us, we can certainly apply sound techniques to evaluating, remediating and communicating the results of our data quality investigations. Data quality assurance and data quality control can both be supported by technologies and is a discipline founded on the proposition that data should be fit for use. Typically, organizations evaluate the quality of their data in terms of various attributes (such as quality, consistency, completeness, retention) in an attempt to reach a "single version of the truth." But, as Redman ([7]) suggests, "getting everyone to work from a single version of the truth may be a noble goal, but it is better to call this the 'one lie strategy' than anything resembling truth." Instead, what we often hope for is at least a consistent version of the truth ([8]).

As an overall framework, data quality comprises much more than software — it also includes people (roles, responsibilities, organizational structures), processes (monitoring, measuring, reporting and remediating data quality issues) and, of course, technology.

There are a number of features/ functions that characterize data quality solutions, these include:

- Data profiling
- Data quality measurement
- Parsing and standardization
- General "cleansing" routines
- Matching
- Monitoring
- Data enrichment



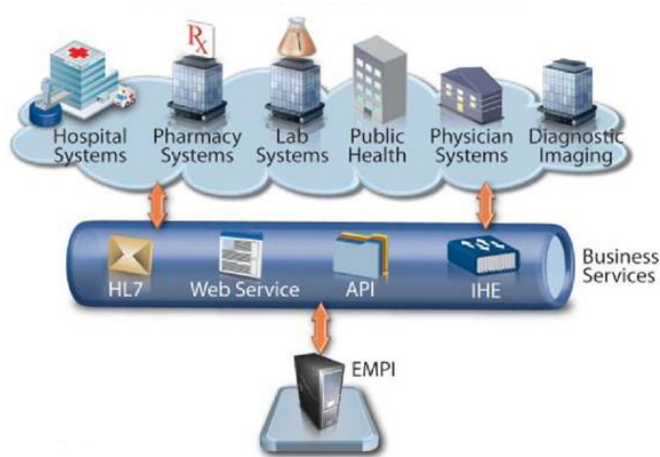
When we talk about the differences between data integration and data quality, it is often not clear which should be done first – integration or data quality (see for example, [9]). Suffice it to say that we view this as an iterative process so often during the construction of a data system you will go back and forth between data profiling, data design, matching, standardization, transformation, enrichment, metadata management and monitoring. You can clearly see the value of technologies that have strong interoperation between the data integration and data quality.

## Master Data Management

One of the approaches to proactive combat data quality issues is through master data management (MDM). Gartner defines MDM is a discipline in which an organization comes together to ensure the

“uniformity, accuracy, stewardship, semantic consistency and accountability of an enterprise's official, shared master data assets”. [10]

Think of MDM as the strategy that an organization uses to maintain consistent versions of data that are important across the enterprise. For example, a hospital system will want a single place where a patient is defined so that her information remains throughout the enterprise – from her primary care provider recording in the electronic health record to the pharmacy to the laboratory where specimens are evaluated to billing and insurance reimbursements.



There are a variety of approaches for managing this rationalized, integrated view of what the single truth is for a given domain. Across industries and organizations, specific implementations often vary in how they connect master data to their use but often rely on one or more of the following strategies: a master index or registry, warehoused or consolidated” view of data, or centralized application and associated workflows that govern the master.

### **Data Governance**

Earlier we defined data management as “the guiding principles, architectures, policies, practices and procedures for managing data within enterprise”. One of the most critical aspects of this is creating organization processes around how data is defined in the context of the business – it’s use, interpretation, value and quality – are managed. Data governance is all about creating an organizational mandate to ensure confidence (assurance) around data. As a discipline, we often see the specific focus by the organization on data quality, data management, data policies, business process management, and risk management. [11]

In the 1980’s we saw a dramatic rise in the role of quality in improving processes in manufacturing, finance and healthcare. Similarly, data quality is being elevated to the boardroom and executive leadership is becoming personally involved in making data quality as important as the products and services it deliver. This is especially important given the role of information as a tangible asset in the modern organization. As we saw with the quality movements of Total Quality Management (TQM), Six-Sigma, Plan-Do-Check-Act (PDCA), the concept of quality has evolved to mean far more than the integrity of a manufactured product. Quality now represents a management philosophy, a system of policies, methodologies and practices and an ongoing commitment to excellence.

So you might be asking yourself - *if data governance is all about the people, processes, and accountabilities* - *what does technology have to do with this?* Just as the six-sigma black belt supports his research with analytics tools, checklists, monitoring systems and communications, data governance can be supported through enabling technologies that help enable the collaboration among the players that participate in the data governance life cycle.

### **Data Governance Life Cycle**

As data moves through the system, we see the various actors as they interact with the data quality issues. The data governance platform can help bring the data and collaboration together.

As Carol Newcomb describes in her blog [12] "The process of data governance is fundamentally very simple." These steps are those that are supported in the governance platform.



- Identify the data quality issues to address
- Prioritize the portfolio of issues to isolate/tackle the most important
- Perform root cause analysis to determine the true source of the data issue
- Design the corrective action
- Formalize the correction through consideration and approval by the data governance organization
- Implement the fix
- Monitor the results

In the next section, we will take some of the functional components of the data integration, data quality, master data management and data governance and outline how these are implemented with SAS solutions.

## **SAS Technology Landscape**

SAS has breadth of technologies that address an organizations needs around 'data'. The following capabilities are addressed with that set of technologies and are collectively referred to as 'SAS Data Management':

- **Data access** refers to your ability to get to and retrieve information wherever it is stored. Certain technologies can make this step as easy and efficient as possible so you can spend more time using the data – not just trying to find it.
- **Data quality** is the practice of making sure data is accurate and usable for its intended purpose. This starts from the moment data is accessed and continues through various integration points with other data – and even includes the point before it is published or reported.
- **Data integration** defines the steps for combining different types of data (ETL). Data integration tools help you design and automate the steps that do this work.
- **Data governance** is an ongoing set of rules and decisions for managing your organization's data to ensure that your data strategy is aligned with your business strategy.
- **Master data management (MDM)** defines, unifies and manages all of the data that is common and essential to all areas of an organization. This master data is typically managed from a single location or hub.
- **Data streaming** involves analyzing data as it moves by applying logic to the data, recognizing patterns in the data and filtering it for multiple uses as it flows into your organization.
- **Data federation** is a special kind of virtual data integration that allows you to look at combined data from multiple sources without the need to move and store the combined view in a new location.

SAS sells these technologies in 'bundles' that addresses a range of needs from very specific to a much broader set of needs. The intent behind the bundles is to simplify and enable a step-wise approach to an enterprise level of data management. The bundles are grouped as follows:

- Data Governance
  - SAS Data Governance
    - Business Data Network
    - Reference Data Manager
    - Web Monitor
    - Dashboards
  - Data Quality
  - SAS Data Quality Desktop – no server component
  - SAS Data Quality Standard/Advanced

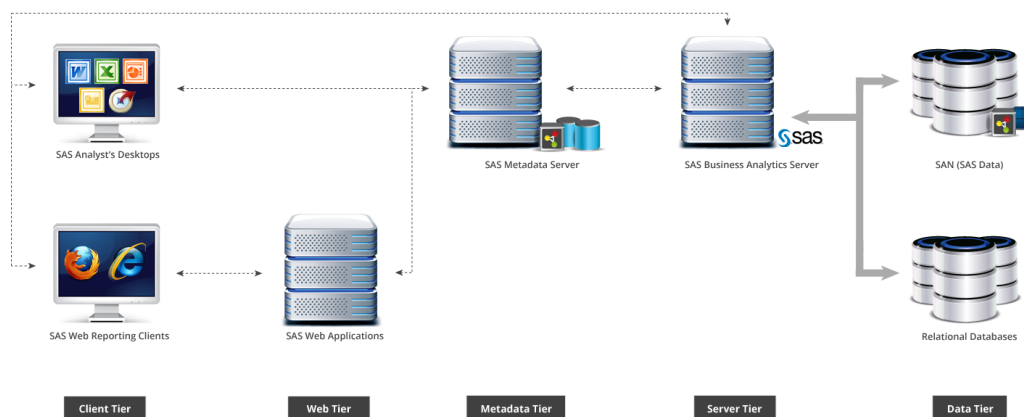


- Replaces most traditional former DataFlux a la carte offerings
    - Advanced includes all Standard features plus Data Governance
  - SAS Data Quality Accelerator for Teradata – requires BASE, SAS/Access to Teradata and the Data Quality Standard bundle.
- Data Integration/Management
  - SAS Data Integration Server -Standard ETL package
  - SAS Data Management Standard/Advanced
    - Replacement for SAS EDI Server
    - Combines former DataFlux data quality capabilities with DI Studio data integration
    - Advanced includes all Standard features plus Data Governance & new DM Console \*\*\* [LISA – What do these asterisks mean?]
- Master Data Management
  - Master Data Management Standard/Advanced
    - Complete offerings for Master Data Management
    - Advanced includes all Standard features plus Data Governance & Business Rules Manager

## Reference Architectures

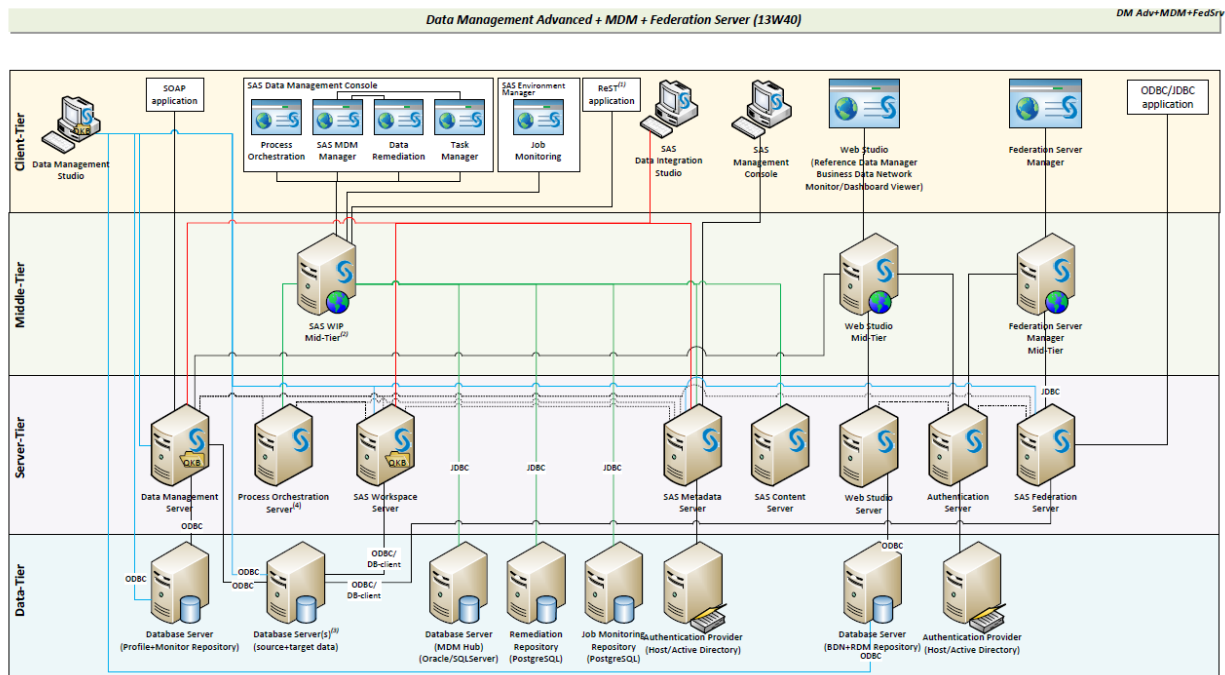
### Typical Architectures

While there is no “one” SAS architecture that can be used across all industries, companies or even business units, we do often start with a common set of components as seen below.



What most people think of as “foundation SAS” sits on the SAS Business Analytics Server (see the SAS Logo above). For SAS Data Integration Server or SAS Enterprise Data Integration Server, the “server” parts of these solutions would sit there and the clients, of course, would reside on the desktop machines.

The number and types of servers can grow to accommodate the usage patterns and the specific products that are in use. In the diagram below, we show what a full-blown SAS Data Management Advanced, Master Data Management and Federation Server solution architecture would look like (logically.)



### Modernization strategies

Many organizations are suffering from data overload or the impact of ‘big data’. One of the best definitions of ‘big data’ is when your traditional data management technologies and processes can no longer meet the needs of the data consumers. Modernizing those ‘traditional’ technologies and processes will better enable organizations to manage their growing data (along with the variety and velocity!) Organizations that truly treat data as a corporate asset will, as stated previously, have mandates to ensure confidence in corporate data. Outlined below are a few modernization strategies that, in a step-wise approach, build data confidence

### MOVING FROM PC SAS OR SAS ENTERPRISE GUIDE TO SAS DATA INTEGRATION

Many SAS customers relying on SAS programs written in display manager or created by Enterprise Guide to handle the data preparation and ‘management’ of the data within their environment. While these programs and processes are doing the job today, there is no easy way to understand what’s happening inside those programs, especially for non-SAS programmers. Lack of confidence and mistrust can result from not knowing what data is being used as input, how/if that data is being changed, and what data are produced. This is where the value of metadata comes into play. Metadata

offers greater visibility through the ability to searching and to analyze the lineage of data as it was transformed. Automatically created metadata helps document the data sources and elements that are used by various jobs that ultimately drive reports that guide fact-based decision-making.

Before SAS Data Integration, programs that have to be manually screened and documented, managing 100-1000s of SAS programs in this fashion was challenging and presents risk to the organization:

```
libname ditest 'c:\DISdata';

data temp.burgers;
input where $ 1-18 food $ 19-34 calories fat $ sodium $ id $;
cards;
Burger King      cheeseburger    380  19g 780mg 1
Hardees          cheeseburger    390  20g 990mg 10
Jack In The Box  cheeseburger    320  15g 670mg 0
McDonalds       cheeseburger    320  14g 750mg 35
Wendys           cheeseburger    320  13g 770mg 20
;
run;

data temp.lesscalories;
```

Using SAS Data Integration Server, the code can be imported and metadata automatically created. With the creation of metadata and importing of code into SAS Data Integration Studio, I now have a process flow diagram with associated metadata objects (jobs, tables, columns, libraries). I can see impact analysis as well as drillable reports about the metadata objects. Because the metadata objects get stored, they are accessible to others for better collaboration.



**Burgers (Generated)**

**Location:** /AmericasDemoPackages/Code Import  
**Repository:** [Foundation](#)  
**Created:** February 6, 2014 3:43:29 PM EST  
**Updated:** February 6, 2014 3:43:29 PM EST

**Description:**  
 Generated by SAS Data Integration Studio

**Source Tables**

Name	Type	Location
<a href="#">BURGERS</a>	DATA	/AmericasDemoPackages/Code Import

**Target Tables**

Name	Type	Location
<a href="#">BURGERS</a>	DATA	/AmericasDemoPackages/Code Import
<a href="#">LESSCALORIES</a>	DATA	/AmericasDemoPackages/Code Import

**Responsible Parties:**

Name	Role
SAS Power User	Created By
SAS Power User	Modified By

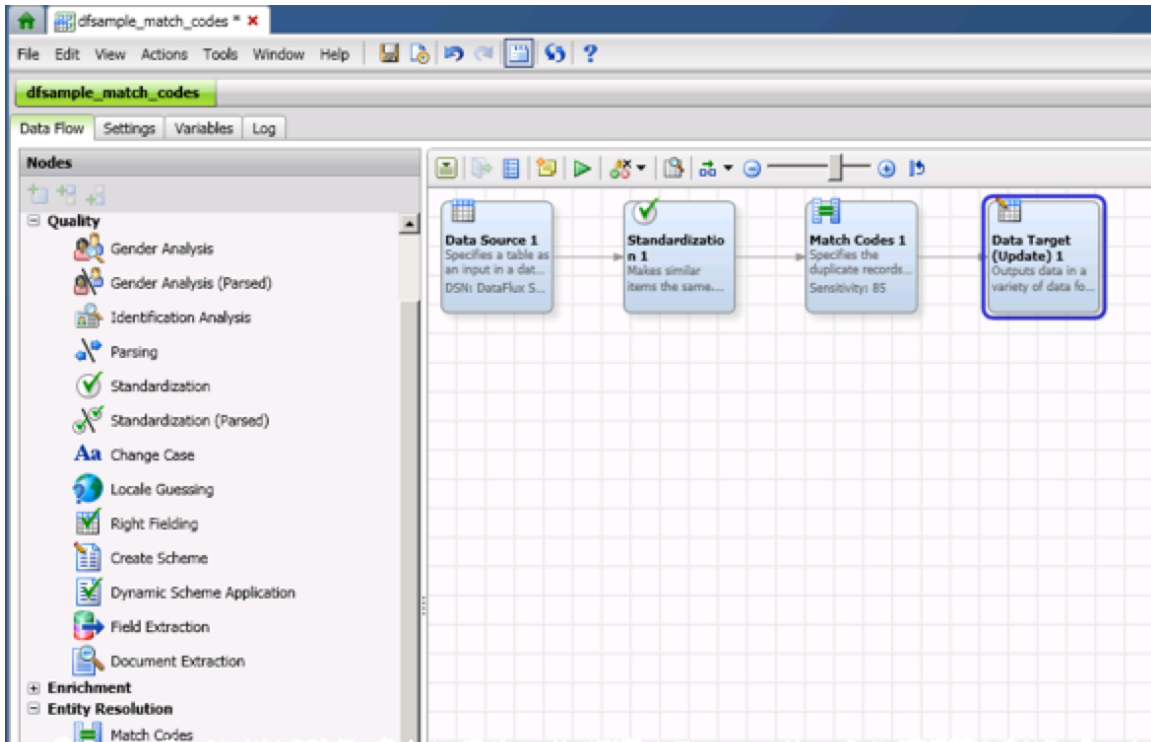
**DATA INTEGRATION PLUS DATA QUALITY = DATA MANAGEMENT STANDARD**

Thanks to the previous modernization efforts of importing my SAS programs and creating metadata objects, the organization now has better insight into what data is being utilized by jobs to create particular reports or analysis. The next step in ensuring confidence in the data and treating the data as a corporate asset is to evaluate the quality of the data we are using. SAS Data Quality provides the ability to evaluate the current quality of the data with 'out of the box' quality checks and the ability to create custom quality checks that may be specific to your organization. Data stewards can profile operational data and monitor ongoing data activities with an interactive GUI designed specifically for their needs. Once the quality of the data is determined and the quality issues are understood, SAS Data Quality provides the ability to fix the data issues.

Field Name	Count	Null Count	Blank Count	Unique Count	Pattern Count	Data Type	Actual Type	Data Length	Minimum Length	Maximum Length	Minimum
ADDRESS	3276	124	0	2389	1048	VARCHAR	string	100 chars	3	30	12121 Central
CITY	3276	0	0	2804	289	VARCHAR	string	30 chars	1	28	30 Rockefeller ?
COMPANY	3276	0	0	1407	1004	VARCHAR	string	50 chars	3	40	-Ford
CONTACT	3276	0	0	3243	176	VARCHAR	string	30 chars	5	23	Brob Beckett
DATABASE	3276	107	0	95	76	VARCHAR	string	25 chars	3	25	MS sql Server
DATE	3276	0 (not applicable)	0	323 (not applicable)	0	DATETIME	datetime	19 chars	(not applicable)	(not applicable)	5/23/1997 12:00
DELETE_FLG	3276	3276	0	0	0	VARCHAR	(no data)	50 chars	(no data)	(no data)	
ID	3276	0	0	3276	5	VARCHAR	integer	9 chars	1	5	1
MATCH_CD	3276	3276	0	0	0	VARCHAR	(no data)	50 chars	(no data)	(no data)	
OS	3276	3	0	42	32	VARCHAR	string	20 chars	2	16	95
PHONE	3276	11	0	3220	2	VARCHAR	string	15 chars	12	14	(287) 473-4617
STATE	3276	37	0	61	21	VARCHAR	string	15 chars	2	14	CA

SAS Data Quality delivers the ability to perform data cleansing and matching in native languages for more than 38 regions worldwide. Out-of-the-box standardization rules conform data to corporate standards, or you can build customized rules for special situations. The cleansing and matching processes

can be embedded as batch, near-time and real-time processes as the functions are available in both operational and reporting environments.

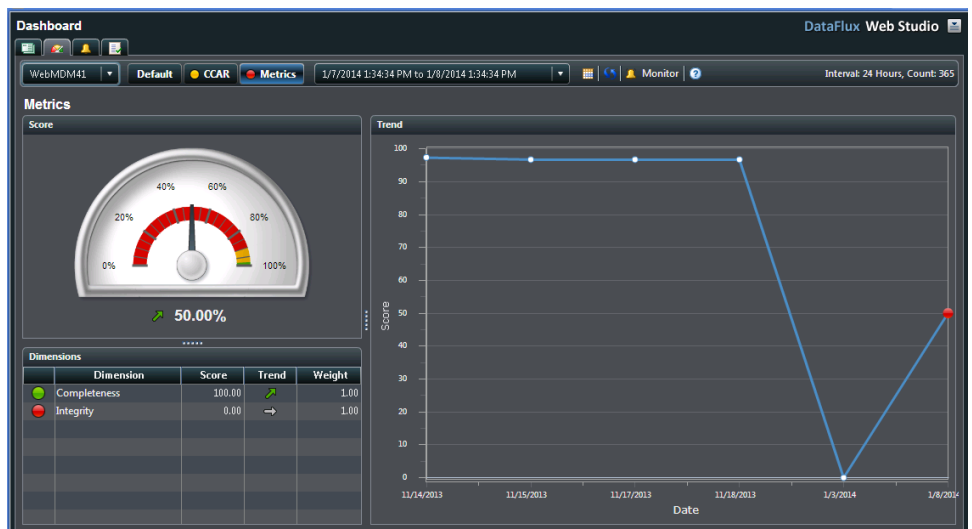


**DATA MANAGEMENT STANDARD + DATA GOVERNANCE = DATA MANAGEMENT ADVANCED**

Technology alone won't solve an organization's data governance challenges, as the discipline requires the heavy involvement of people to establish policies and processes around data. SAS Data Governance provides the ability to link the people, policies and processes to the actual data lifecycle so that when it comes to knowing what's happening with your data, you can know. With built-in reporting, monitoring and validation, you can see when you're succeeding or whether you need to make changes. There is an intuitive dashboard to monitor trends, and you can know when policies are being followed - and trace the ones that aren't.

**SAS Data Governance**

provides a business data glossary to facilitate the creation and management of business terms. In the following example we see a business term called 'Acceptable risk', it's description and definition,



as well as its requirements as defined by a third party regulation/compliance committee.

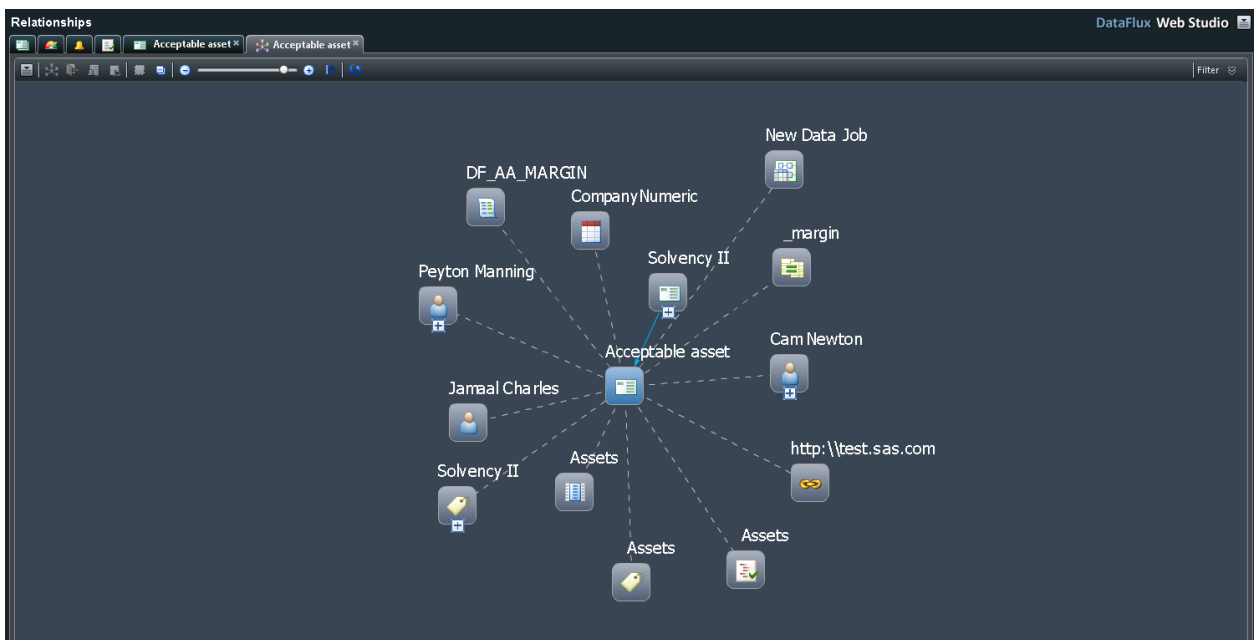
The screenshot shows the 'Term Business Data Network' interface. The main content area is divided into several sections:

- Description:** Capital (either on or off-balance sheet) which, under regulatory rules, may be taken into account (fully or partially) to cover insurance obligations.
- Requirements:** Must be the sum of available margin and the different between eligible capital and acceptable risk.
- Attributes:**
  - Related applications: Senri details spreadsheet Weekly statements
  - Locations: France
  - Locations: Great Britain
- Details:**
  - Status: Under Review
  - Importance: Critical
- Links:** Solvency II Terms
- Tags:** Assets, Solvency II
- Related Terms:**

The term is 'linked' to associated items like business data rules, data elements, other business terms, and data jobs.

Name	Location	Description	Note	Type
_margin	dfr://BDN/Shared Data/_margin(Rule)		margin value rule	DataFlux Business Rule
Assets	dfr://DataFlux Sample.ODBC///CompanyNumeric/...		Actual persisted values	Column
Assets		This is the set of assets that are allowed per the solve...		Reference Data Hierarc
CompanyNumeric	dfr://DataFlux Sample.ODBC///CompanyNumeric/		Asset values of the company	Table
DF_AA_MARGIN	dfr://BDN/Shared Data/DF_AA_MARGIN(Task)		Margin calculation	Monitor Task
New Data Job	dfr://BDN/batch_jobs/New Data Job(DataJob)		Job that calculates and alerts margin rule errors	Data Job

The relationships of the associated items, and more, can be visualized so anyone can understand how people, data, terms, policies and jobs are related when it comes to 'Acceptable Asset'.



SAS Data Governance provides data remediation through workflows for governing stewardship and other business processes. Allowing for run-time drill-through of tasks associated with active workflows if issues are flagged during monitoring processes and enables data stewards to review records and resolve issues once problems are identified during a load process.

**TASK MANAGER**

NAME	ACTIVITY	START DATE
INVALID OR MISSING DATA ...	Edit	01/03/14 03:27:37 PM
INVALID OR MISSING DATA ...	Edit	01/03/14 03:27:36 PM
INVALID OR MISSING DATA ...	Edit	01/03/14 03:27:35 PM
Bad Data - SHARON KLEN...	Edit	12/05/13 07:31:01 AM

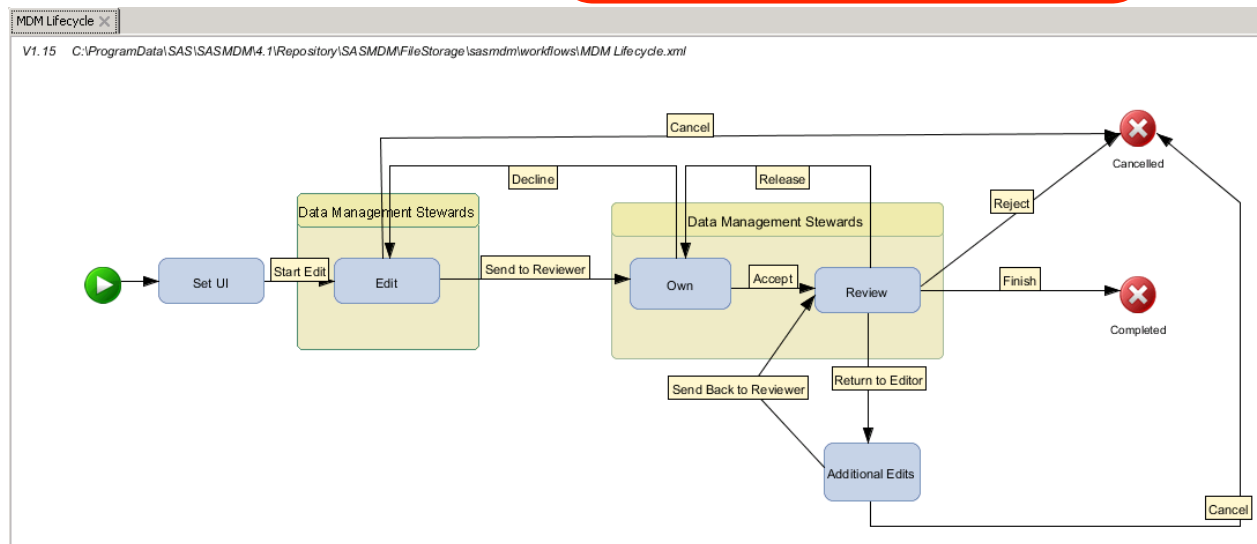
**MONITORING CENTER**

0 total 24 hrs

- dashboard\_stats.djf
- add\_update\_individual.djf
- add\_update\_part.djf
- add\_update\_company.djf
- Update MDM Dashboard Stats
- MDM Data Load

**DATA REMEDIATION**

ISSUE	ITEM	APPLICATION	AGE (DAYS)
INVALID OR MISSING DATA	SAS ADMIN	SAS MDM	2 days
INVALID OR MISSING DATA	SAS ADMIN	SAS MDM	2 days
INVALID OR MISSING DATA	SAS ADMIN	SAS MDM	2 days
INVALID OR MISSING DATA	SAS ADMIN	SAS MDM	2 days
INVALID OR MISSING DATA	SAS ADMIN	SAS MDM	2 days



## Summary

While sometimes confusing the portfolio of SAS products continue to evolve to meet the needs of the SAS ecosystem. Given that DataFlux and the SAS Data Integration family of solutions are being managed as a single portfolio of products, the ability for customers to evolve their use of these capabilities has been enhanced.

As we have outlined, there are a number of terms that get used across the industry including data integration, data quality, and master data management and data governance. Having a suite of tools that allows you to mature your organization's ability to deal with the increasing complexity and volume of data is paramount. On the maturity curve, we typically see traditional uses of SAS aided by programmer tools such as Enterprise Guide move to more standard metadata-based solutions like Data Integration Studio. From there, organizations can continue to improve their processes by implementing either data governance strategies or master data management solutions (or both.)

## References

1. Inmon, W.H., *Building the data warehouse*. 1992, Boston: QED Technical Pub. Group. xi, 272 p.
2. Kimball, R., *The data warehouse toolkit : practical techniques for building dimensional data warehouses*. 1996, New York: John Wiley & Sons. xxviii, 388 p.
3. Cearley, D.W. and C. Claunch, *The Top 10 Strategic Technology Trends for 2013*, in *The Top 10 Strategic Technology Trends*, G. Group, Editor 2013, The Gartner Group. p. 12.
4. Nelson, G., *Best Practices for Managing and Monitoring SAS® Data Management Solutions*, in *SAS Global Forum*, S.G.U. Group, Editor 20112, SAS Institute: Orlando, FL.
5. Grasse, D. and G. Nelson, *Base SAS® vs. SAS® Data Integration Studio: Understanding ETL and the SAS Tools Used to Support It*, in *SAS Users Group International 2006*, SAS Institute: San Francisco, California
6. English, L., *Information quality applied : best practices for improving business information, processes, and systems*. 1st ed. 2009, Indianapolis, IN: Wiley Pub., inc.
7. Redman, T.C., *Data driven : profiting from your most important business asset*. 2008, Boston, Mass.: Harvard Business Press. xiii, 257 p.
8. Dyché, J. and E. Levy, *Customer data integration : reaching a single version of the truth*. 2006, Hoboken, N.J.: John Wiley & Sons. xxvi, 294 p.
9. Power, E. and G. Nelson, *ETL and Data Quality: Which Comes First?*, in *SAS Global Forum*, S.G.U. Group, Editor 2008, SAS Institute: San Antonio, TX.
10. White, A. *Defining MDM – again*. Gartner Blog Network 2009; Available from: <http://blogs.gartner.com/andrew.white/2009/07/01/defining-mdm-again/>.
11. Gidley, S. and N. Rausch, *Best Practices in Enterprise Data Governance*, in *SAS Global Users Group*, S. Institute, Editor 2013, SAS Global Users Group: San Francisco, CA.
12. Newcomb, C., *A data governance primer, part 1: finding the root cause*, in *The Data Roundtable*, SAS, Editor 2013, SAS.



## ***Biography***

### **Greg Nelson, President and CEO, Thotwave Technologies, LLC.**

Greg is a global healthcare and Business Intelligence (B.I.) executive with over two decades of experience and leadership in the field. Greg is a prolific writer and speaker interested in healthcare analytics and the strategic use of information technology.

He received his BA in Psychology from the University of California at Santa Cruz and advanced his studies toward a PhD in Social Psychology and Quantitative Methods at the University of Georgia. Recently, Greg completed his Masters degree from Duke University in Clinical Informatics from the Fuqua School of Business. His academic and professional interests include helping organizations mature their analytic capabilities. Founder, President, and CEO of ThotWave Technologies, a niche consultancy specializing in healthcare analytics, Greg is particularly interested in how lessons from across other industries can be applied to help solve the challenges in healthcare.

With certifications in Healthcare IT, Project Management, Six Sigma and Balanced Scorecard, Greg is also a prolific writer and has presented over 200 professional and academic papers in the United States and Europe. He won the Best Paper Award in 2013 at the Pharmaceutical SAS Users Group Conference and sits on the board of the SAS Global Users Group. In 2011, Greg was selected by SAS into their loyalty partner group. "This program acknowledges and supports individuals who are recognized experts in their fields and have a long-term relationship with SAS."

Married to wife Susan and living on a small "farmlet" in rural North Carolina, Greg is an avid woodworker, enjoys photography, rides a Harley-Davidson Motorcycle, and strives to be a lifelong learner.

### **Lisa Dodson, Manager, Data Management – Americas Technology Practice (SAS)**

Lisa has been with SAS for 14 years and is a recognized expert in the information management, data governance and data management space within the organization. She holds a Master's Degree in Information Quality, and has affiliations with many data management/governance organizations including as a former board member and President for the International Association for Information and Data Quality and organizing committee member for MITIQ's Industry Symposium. Through job roles including, account executive, systems engineer, product manager, technical trainer and solutions architect she's developed a deep understanding of the SAS software architecture. In her current role she leads the Americas Data Management Practice.

## ***Contact information***

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

Greg Nelson     [greg@thotwave.com](mailto:greg@thotwave.com)

ThotWave Technologies, LLC

1289 Fordham Boulevard #241

Chapel Hill, NC 27514 (800) 584 2819

<http://www.thotwave.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.

*thinking data*® is registered trademark of ThotWave Technologies, LLC.

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