PERVASIVE

INTEGRATION

A Metadata Management Strategy Checklist: **What Metadata Really Matters?**

A Pervasive Software White Paper

What Metadata Really Matters?

The short answer is that all metadata matters. Not just the metadata about database schemas, but also metadata about information processing flows, integration interfaces, data quality and on and on. Metadata is the key to providing answers to a great many questions, both for executive decision-makers, and for IT professionals in the trenches who will be asking questions like:

- "If I buy this great new CRM application, how much is it going to cost my business to get our processes switched over?"
- "What's a realistic time estimate for getting customer, prospect, invoice and billing data stored in four different applications synced up to eliminate duplicate data entry?"
- "Will the labor and error-reduction savings justify the expenditure for the project?"

Metadata is the essential blueprint for what data a company has, what it means and where they can get it. Without clear accessible metadata, an enterprise could have all the data they need to make sound decisions sitting on their servers and hard drives, and still end up making a best guess because they can't find that data or making decisions that they later regret because they only found a fraction of the data they really needed, or couldn't make sense of what they had.

The trouble is, there's been a lot of talk for quite a while now about the need for and value of creating a warehouse for metadata that makes this critical information easily accessible, but businesses have spent too much money and seen far too little results

Traditional Approaches Add to Your Challenge

Traditional approaches to metadata management have been ad hoc at best. Simple oral knowledge transfer was not uncommon. In the past, database administrators or IT professionals would abstract metadata from data sources by hand, using Excel spreadsheets or even doc files to document data properties and relationships. This type of documentation wasn't exactly designed for answering queries, and locating these documents could be a challenge in itself since their location was generally undocumented. But, the biggest shortcoming of this pile-of-documents approach was that it had no strategy for dealing with the natural drift and change of data structures, value sets and relationships. Even assuming that the Excel spreadsheets or docs were accurate in the beginning, they would become increasingly inaccurate over time. It would be rare for overworked database administrators to consider it their job to keep all of those scattered spreadsheets up to date.

The other option was to pay a million dollars or so to a consulting company or tool vendor for an ambitious multiyear project. For large sums of money, a vendor would cheerfully build a slick metadata repository with a great interface that provided a sharp, clear window into some fraction of the essential metadata that you had two years ago. Not quite what you need to make dynamic, adaptive decisions for this year, or this quarter.

More Recent Approaches Have Led to Silos and Inconsistency

Now that we have begun to move out of the dark ages of metadata management, various tool vendors across the enterprise are attempting to solve the metadata puzzle by storing the metadata that is relevant to the type of problem that tool is designed to solve. This is definitely a step in the right direction, but it's far from ideal. Multiple isolated silos of metadata create redundancy, inconsistency, and incompleteness of metadata across the enterprise. The matter is complicated by the sheer volume of data regularly processed in businesses today and the exponentially increasing number of new data and metadata sources.

In one place, data profiling and quality tools store statistics about the anomalies in the data, value ranges and other valuable information. In another, integration tools store design-time information about source and target data structures, transformation lineage, business rules and definitions and process flow as well as run time information about production lineage, which versions were executed, and success, failure and error logs. In still other locations, database management systems—either stand-alone or embedded in various applications—store field rules, names, types, and table structures, relationships and interdependencies. Each tool stores metadata in its own proprietary format. And of course, there are still many applications that don't store their metadata as an entity at all, and require metadata to be extrapolated either by hand or by custom coding.

If an important big picture inquiry or an in-depth technical question requires more than one of these types of metadata, there is no common place to go for the information. The information is scattered across multiple metadata repositories in multiple locations—and multiple storage methods—rendering the accessibility of cross-functional metadata an unavailable option. On top of that, these tools and metadata storage mediums are part of the constantly changing business landscape. In today's world of continual mergers and acquisitions, changing business initiatives, and constantly increasing variety of applications, sources of both data and metadata are unstable, moving targets.

The one lesson that both the hand documentation and the overpriced, insufficient metadata solutions of the past teach us is that capturing metadata—while it may seem like a daunting task—is relatively easy. Modern tools have made that aspect of the job even easier. Achieving a unified view of all metadata across the enterprise without bankrupting the company, and keeping that metadata synchronized with the actual properties of the data sources, are the real challenges.

Metadata Management Strategy Checklist

The secret to meeting those challenges is to build the metadata solution on top of a versatile integration platform that exposes its own metadata for consolidation, automatically extrapolates metadata from sources that don't provide it, and provides the connectivity capability to extract metadata from a wide variety of data sources. Modern integration toolsets frequently include metadata management capabilities as part of the package. Since a good metadata strategy requires bringing together a huge variety of disparate data sources and since a fair number of the benefits of metadata management are directly related to data and application integration projects, it makes sense to not just build the metadata warehouse with integration processes, but to build the metadata management into the warehouse and processes.

1. **Consolidate Metadata** - The first step in a good overall metadata strategy is to extract the metadata out of its isolated silos and bring it all together. This will reduce data redundancy, duplication and inconsistencies. This is, in its essence, an exact-transformation-load (ETL) problem; using a good ETL tool to get the metadata out of various repositories, cleanse and aggregate it and load it into a metadata warehouse makes perfect sense.
2. **Automate Synchronization** – A scheduled or change-event driven automated integration process can make certain that the metadata warehouse is regularly updated and will remain synchronized over time with the changing sources, without adding to anyone's ongoing workload. In addition, when future data sources are added, with an integration toolset that has a decent user interface, it's a simple matter to extend already established integration processes to include new data sources.

3. **Make ALL Enterprise Metadata Easily Searchable** - A comprehensive metadata warehouse with a well-documented star schema or other query optimized structure, can make a world of difference in terms of the speed and quality of answer. The data warehouse should be designed for optimal use with your existing or desired business intelligence (BI) reporting tools, but not be limited to only BI-related metadata. It should enable you to retrieve all the metadata that matters to everyone across the enterprise, whether business or technical, front office or back office, into a single location, and then provide multiple windows into that data. This will provide a clear global view of the enterprise and enable knowledge transfer and information sharing. It also means that the metadata that matters to any particular user will be easy to find when needed. You will be able to find answers to technical, business, or cross-boundary issues that haven't even been conceived of yet in a good metadata warehouse.
4. **Get your ROI perspective in shape** - So, you've got a good idea of how to build an enterprise-wide robust metadata management strategy. The question now is, "Why bother?" It's tough to show the ROI for metadata initiatives in hard numbers up front, especially since a lot of the benefits show up over a fairly significant length of time. It's essential to have a low-cost option that makes ROI a no-brainer. A low cost point can make the difference between months of justification, and a low effort, back-of-an-envelope kind of calculation. Start tactically. Build a small, viable data warehouse that includes just the information for a particular business segment, such as a division or department. This allows the demonstration of clear ROI with low risk. It also gives you a chance to build model integration and synchronization processes that can then be re-used or extended to other departments, at an even lower cost.

A solid metadata strategy can show a clear increase in ROI and decrease in cost point for a variety of other important initiatives as well. For instance, it can help to show adherence to data governance and compliance requirements by providing a clear audit trail.

Benefits of Robust Metadata Management

Metadata management is critical for organizations that rely on multiple data sources for business intelligence, especially if some of those data sources are older, or less easily accessible. Including even hard-to-get metadata improves the knowledge base available for BI queries, giving more relevant, accurate and useful analyses and reporting. The ability to view and analyze both technical and business metadata also provides a mechanism that ensures that the value being extracted from the data continues to meet business objectives. This can improve IT decision-making by validating technical processes with business goals.

For integration projects, metadata provides data lineage throughout various stages of transformation for future error checking, compliance auditing, and data quality improvements. Good metadata can help with any business policy or infrastructure change by providing information that helps companies gauge the complexity of changes and plan the best use of resources. This can significantly reduce development and maintenance costs. Management of integration design metadata can also provide the basis for module re-use. Enabling collaboration and component re-use significantly accelerates project timelines. The creative use of metadata can provide innovative approaches to long-standing integration problems that provide completely unforeseen benefits. A good example is the Vision Award-winning process of LifeMasters Supported SelfCare, Inc. Using metadata on top of a versatile data integration platform, LifeMasters was able to reduce patient data processing time from five days to two hours, and new patient on-ramping from 7-10 days down to 1-2 days. (http://www.prweb.com/releases/LifeMasters_Vision_Award/Business_Impact/prweb1531744.htm)

Overall, the goal of a metadata management strategy is to reduce IT costs and increase corporate productivity and agility, and that always translates one way or another to increased ROI. With a solid, broad-spectrum, easily searchable metadata warehouse and automated updating processes in place to keep it current, an enterprise will undoubtedly see benefits, including those the organization may not have thought of. Executives, data analysts and developers will also genuinely understand their data descriptions, definitions, lineage and relationships. The metadata that really matters will be at their fingertips.

... What our customers are saying

"We had many data feeds and data types that needed to be quickly implemented in order to be able to assess our overall portfolio's adherence to risk exposure guidelines. Custom code couldn't meet our need to quickly get policy data into our data warehouse and analysts' hands. Pervasive integration gives us cost-effective, wide-scale connectivity that complements our move to support greater automation and analysis."
Andrew Wild
Head of IT
Scottish Re Holding Limited

"Pervasive's software has allowed Eloqua to focus on our key value proposition without having to develop custom code, so that we can bring up new customers with complete CRM integration in a fraction of the time formerly required. The integration agent is a very elegant solution for reaching legacy applications behind the firewall."
Steven Woods
Chief Technology Officer and Co-founder
Eloqua

"Pervasive has helped us move from manual data conversion to a quick-hitting embedded software solution, and I next see Askesis using Pervasive's integration engines in helping us automate formatting and compliance processes."
Devendra Rao
Vice President, Development
Askesis

"Pervasive Data Integrator is now fully employed, really narrowing the time window I would have found in a back-out custom-code rebuild. Cost-wise, Pervasive has saved us significant outlays. We saved in what we paid an offshore vendor to write and maintain code, the costs associated with correcting data that was mangled by bad code, and the costs associated with delaying data transfers by using Pervasive over custom code."
Bill Seay
Director of IT
Revenue Recovery Corporation

"With the Process Designer embedded within Pervasive Data Integrator, we have reduced the processing steps from 75 to only 6. Previously with this customer, if we wanted to change a file format, we had to manually change it in 75 different places. The Process Designer enables us to make changes in only one place. Using Pervasive Data Integrator, we have reduced our maintenance times by more than half and the customer receives more usable data much faster and with far less risk of error."
Jay Graves
President
SmartDM

"Pervasive has enabled us to far surpass our existing data gathering capabilities. Pervasive Data Integrator offers out-of-the-box connectivity and instantly improves productivity, which directly benefits hospitals and their patients."
Kevin Conway
Vice President of Health Information
Nebraska Hospital Association

"Pervasive software eliminates the need for our developers to spend time manually coding and recoding each data transformation and writing applications to move data. Instead, developers can create and reuse a single map that can be minimally modified for separate transformations. Once the original map is developed, the majority of the data transfers will be automated, saving us hundreds of hours of coding and writing."

David McMath
Database Administrator and Software Engineer
Payment Processing, Inc.

"I can say that Pervasive allowed us to achieve results in considerably less time than it would have taken to build our own integration engine. The cost of Pervasive's solution is far less than what it would have cost us to build and maintain our own."

Claire Annechini
Chief Information Officer
MEDecision, Inc.

"We considered several tools for data transformation and some were asking $70,000 to $100,000 to do a simple conversion. We looked at the functionality Pervasive could provide for a lower price and the decision was easy. Pervasive supports many different formats and with the merging of different systems, that versatility has been very valuable for us,"

Ted Sabino
Senior Network Consultant
Northrop Grumman

"I looked for an integration solution that would give the company exactly what Pervasive provided: direct connectivity to numerous legacy systems, extensive mapping capability and fast processing. In terms of value, there was no comparison."

Peter Murdock
Senior Database Designer
AEGON Equity Group

Contact Information

Pervasive Software Inc.
12365 Riata Trace Parkway, Building II
Austin, Texas 78727
http://www.pervasivedataintegration.com

United States
888.296.5969
512.231.6000
Fax: 512.231.6010
info@pervasivedataintegration.com

Europe
Toll Free: 00800.1212.3434
Phone: 0032.2290.5391
Fax: 0032.2.709.7883
integration@pervasive.com