stone bond
technologies

WHITE PAPER

Data Virtualization and Federation
Metadata Comes of Age with Enterprise Enabler®

Pamela Szabó
Stone Bond Technologies
February 2011

Background

The time has come for data virtualization and federation to step up to the plate. There have been waves of hype around the subject for many years, and although most approaches have had solid theoretical foundations, they have simply been too complex to implement and maintain in an ever-changing environment. Today's demand and expectation by business end users for information has come in part from the overwhelming surge of social media and information access. SaaS applications, portals, and ubiquitous tools like SharePoint can be configured by non-programmers who need to access information and data to address their own needs. ISVs are producing best-of-breed line of business (LOB) applications at an alarming speed, as business software players strive to mimic consumer technology trends. Let's call all of these uses collectively, "bizapps."

Imagine if each of these bizapps were dependent on integration directly to a database or application, and if a separate database were built and populated to have information from more than one source, just for that use. As we all know, over time either requirements change or the sources change in some way that impairs the quality of the application unless the dependencies are repaired. Businesses and government spend billions of dollars annually to create and maintain integrations across bizapps. How many companies are still suffering severe consequences right now of fragile integration that cannot adjust to change?

Suppose one of these bizapps needs customer information and gets it directly whenever it needs it from Dynamics CRM. Most typically, a database is designed and integration is written to populate it from CRM. After a few months, it turns out that they really should be using data from SAP. An integration that is based on data virtualization would have an abstracted layer of metadata that could be easily redirected to SAP instead of CRM, as opposed to rewriting the entire integration. Data federation adds to that the ability to merge data from multiple sources, so while treating it as a single set of data, the bizapp could bring part of the customer data from SAP and part from CRM. Thus empowered, the bizapps can eliminate all data staging, pulling directly from the sources. In fact, since the data connectivity is live and direct, the bizapp does not even need to save its own copy of the data, but can present it to end user who can interact with the virtual presentation of the data; changes would be written back directly to the sources, with proper security and transaction assurance.

Data virtualization with federation is an essential aspect of Agile Integration Software (AIS), one that is a natural result of the end-to-end metadata found in Stone Bond Technologies' Enterprise Enabler®.

The Importance of Data Virtualization and Federation

How long does it take for a business user to get a dashboard with the bizapp data he needs? It's not that the dashboard itself is hard to build; in fact there are plenty of tools that he can use himself to build it. Or he can use a spreadsheet and make nice graphs to display the data. Usually the IT department must get involved in providing the data for him, but how do they do that? Very often they analyze where it should come from, design a data model, set up the database, and then write code to get the data into the database and keep it up to date. This all constitutes a project big enough to need to put in a queue, which sometimes means an eternity. In some cases a Data Warehouse has been put in place to serve such needs. If the business user is lucky, his data is already being captured in the data warehouse and he's even luckier if there are tools he can use himself to get the data out easily.
 Over time, these staging databases and data warehouses begin to degrade the overall data management abilities of the IT department.

Mergers and acquisitions often happen faster than the IT department can really keep up with as far as offering flexibility to identify and act on business opportunities or even deliver a consolidated view of the company's financials.

The table below identifies some of the impediments to smooth flow of data in almost every company. The root causes are handily addressed with the implementation of data federation and virtualization. There are other things that support a company truly becoming "agile," but without this, it is virtually impossible.

Why do we need federation and virtualization?

Symptom	Issue	Root Cause
Multiple copies of data.	• Don't know which is correct. • May not be synchronized • Data security can easily be broached	• Integration tool can only handle cross-application relationships by staging in a database. • Integration or display tool can only access a single database or a specific type of database • Over time, IT (and "shadow IT") projects have built ad hoc data marts for specific point solutions. • The primary purpose of most data warehouses is to bring the data from disparate sources together, meaningfully aligned for use by multiple applications.
Expensive to maintain Integration across the enterprise.	• As applications and databases change, the integration requirements change • Time, cost, and complexity of creating and maintaining integrations become impediments, especially to maintenance	• Most integration requires hard-coded programming. Finding, fixing, compiling, testing, and deploying for every change is expensive
Security breaches	• End users can access data that they do not/should not have permission to use.	• Data access mechanisms do not honor end user security. Data is coming from lots of different places, even data stores built just for the specific end user application. The applications and portals are built without any way to control indirect access.
Proliferation of databases, data marts, spreadsheets that have key information	• Many are not kept current • Often rely on manual data entry or re-entry • Not reliable • Important decisions often made from information in these data stores	• Integration tools cannot deliver data aligned and merged live from multiple sources. • For applications that require ad hoc queries, most have no way to do federated queries. • The only and easiest way to get the data is to build a staging database and populate it.
Dashboards are read-only	• These decision support tools would be a natural mechanism for correcting or adding data and for executives to make and convey decisions, if they were not read-only. • Commercially available dashboards integrate live to few bizapps, so most dashboards report off staged data that may not be current.	• Data is usually staged in databases or data warehouses, making write-back impossible. • Managing transaction rollbacks and security is impossible.
Huge effort to implement data federation/virtualization solutions even with single vendor suite	• Many steps, separate and different tools, custom programming to get them to work together. Hard to debug and maintain.	• "Single-vendor" designation by analysts does not mean single product, but rather multiple disparate products that collectively qualify for the scope.

What is Data Federation? Data Virtualization?

Data Federation

The idea of Data Federation has been around for a very long time, with various interpretations of meaning, approach, and solution. The common thread is that it is a method to make data that is spread across multiple sources meaningful together and usable as a single data entity. The enticement of data federation is powerful. It can eliminate a large percentage of fragile data integrations as well as a huge number of staging databases whose sole purpose is to bring data together so it can pretend to be from a single source.

Most recently, it seems that data federation is used to mean federated data queries, which is the ability to make a database query as if it were querying a single database, but behind the scenes the data is actually gathered from multiple databases.

Historically, the ideas around data federation have always assumed that the federation is applied against relational databases. Stone Bond's Enterprise Enabler, on the other hand, is agnostic to the format of the sources, and can combine information from heterogeneous formats via federated data queries.

Data Virtualization

Today the distinction of where data federation ends and data virtualization begins is quite indeterminate. The separation of the two is only useful now for narrow slivers of the full range, such as federated data queries, none of which can alone make even a ripple of improvement in the general state of enterprise data management.

The general idea of virtualization is to be able to use data without necessarily knowing where the data resides or even what it looks like, by using a layer of abstraction called metadata (data about data). Though never called "virtualization du jour," over the years there have been many trendy upheavals offering great expectations of getting a handle on corporate data. The brave but naïve fall for the consultants (once again) to undertake a major MDM (master data management) project. Even with the best of the MDM software, implementation is expensive and complex, and the jury is still out as to the success of these projects. Is MDM any more successful than the "Corporate Dictionary" projects of the '90s, and all the interim manifestations?

Federation and Virtualization Together

Both data federation and data virtualization deliver significant steps forward for data accessibility in a meaningful way, but the two working closely together are the epitome of synergy. While the grandfather integration providers all at least nominally handle both, it is important to factor in that those vendors have achieved a "checkmark" in many categories by simply acquiring a company with a qualifying product. Since we are talking about consolidating, at design time and run time, multiple capabilities, the only way to provide federation and virtualization is to have shared user interfaces and shared metadata across the products with a single execution engine. It is only with a true Agile Integration Software like Enterprise Enabler that the full synergy can be realized.

For Enterprise Enabler,
- Federation is not just across multiple databases, but also any applications, SaaS, data sources, etc.
- Federated data queries include relational and non-relational data as well as applications.
- All activities are completely metadata driven.
- Data is moved live without staging. Federation significantly reduces the overall need for staging databases.
- Federation is bi-directional; that is both aggregating data from multiple sources and writing back to them.
- Write-back logic is auto-generated.
- The same virtualized federation can be used behind the scenes for orchestrated data or for presentation and interaction with end users.
- End user security is honored (read and write) in cases where federation is presented to end user.
- Federation is not a replacement for integration, but rather a feature of comprehensive integration.
- Includes complex transformation capabilities across multiple sources.
- Supports, but is not limited to web services or XML formats.

If any of the above features are missing, the solution will not deliver the full promise of virtualized federation.

Increasing the Mandate for Security

Great strides are being made in Federated Identity, which becomes a critical aspect of implementing data federation and virtualization. For the class of usage in which the data is federated and moved among applications and data stores using some kind of orchestration behind the scenes (on premise), generally it can be assumed that the participating applications are maintaining the security with respect to end users. Moving data among existing databases and applications that end users log into does not change the concept that the application itself manages end user permissions.

With the surge of portals and tools like SharePoint that make it possible for business users to build Line of Business (LOB) applications, it becomes increasingly important for the integration layer to acknowledge and enforce end user permissions. Of course, that means that the internal endpoint applications and databases must support security models and work in tandem with the middleware to establish such secure hand-offs via APIs. An end user of a SaaS application may be accessing data that is actually derived from an on-premise backend system. One of the concerns about security in data integration stems from a potential and common reality of a break in the security chain when interim data stores are and populated with generic administrative credentials. This is all too common in custom integration.

SaaS has been a strong force for increasing awareness of the need for federated security with WebSSO establishing a growing profile. In parallel, we are seeing a burgeoning need within companies to handle end user security in a federated manner. Enterprise Enabler is able to pass end user authentication credentials to backend systems at each read and write.

A separate layer of security that Enterprise Enabler brings inherently to the mix is the locked down environment for building and modifying integration components, which means that a rogue programmer will not be able to modify or divert data flows without permissions to do so. With a single interface for end-to-end integration development, a complete audit trail is maintained of "who did what when."

From Federation and Virtualization to MDM

Over the years, there have been various approaches, mostly separate paths, for solving data federation and virtualization issues. Early federation was handled by trying to define a single data model that included all corporate data, with all levels of granularity and interpretation. The idea was that all the corporate applications would be re-written to use the central data base. Obviously, every year they spend designing the database, the more applications were developed, and it became impossible, even if the data model could have been defined. Later we saw virtualization with multi-year "corporate Dictionary" projects, which were attempting to identify all important corporate data and define a standard view of that data. Those projects, too, were destined to failure for the same reasons. Corporate metadata management and Master Data Management (MDM) approaches are evolving in a way that is retro-fitting organizations to existing metadata and/or defining another layer of abstraction. Either way, they appear to also be hugely complex and expensive, and likely to quickly become, themselves, unwieldy to support.

The disciplines of managing metadata for correct sourcing and optimum reusability are essential going forward. It turns out that data virtualization and federation are based on the concept of creating metadata and layers of abstraction. Metadata management can mean creating even more additional layers along with perhaps unnecessary complexity. The metada behind Enterprise Enabler offers a clean basis from which to begin managing metadata.

The paths of federation and virtualization began as separate disciplines, which are being conceptually merged together with technologies that were designed to solve different problems. Both must have the same architectural roots, as is the case with Enterprise Enabler, in order to maintain flexibility. The approach of AIS allows metadata to grow organically and in parallel to establish universal corporate metadata abstraction without requiring changes to the consuming bizapps.

Enterprise Enabler (EE) as a Solution

While various of the range of buzzwords and acronyms are associated and dependent on specific cross-sections of technology, Stone Bond Technologies' Enterprise Enabler decouples itself from such dependencies. It is not limited to handling, for example, only relational databases, or only XML, or dealing only with web services. Instead it supports and enables these subsets of the whole picture without imposing requirements on the participating elements.

Data virtualization and federation are inherent benefits of the overall architecture of Enterprise Enabler as opposed to a new set of features added because of the current hype around them.

Enterprise Enabler distinguishes itself in multiple ways from other integration environments. It is a single secure "end-to-end" product with a single User Interface and a single consolidated metadata stack. All instructions are executed by a single run-time engine. It is extensible in that any user can add reusable functionality to the base product. The data mapper and transformation engine handle complex data alignment and manipulations accessing data directly in native format from the sources, and with the added bonus of embedded code editors and compilers, virtually anything can be handled in-line without ever leaving the product. All of these features taken together significantly reduce integration implementation costs and eliminate the requirement for staging databases.

The same product is being used as an embedded and often private labeled, integration capability for Independent Software Vendors (ISVs) in various forms such as ADO.Net drivers. In most cases, the ISVs are eliminating the requirement for a staging database to feed their application. Instead, they are using virtualized federated data and becoming considerably more competitive in doing so.
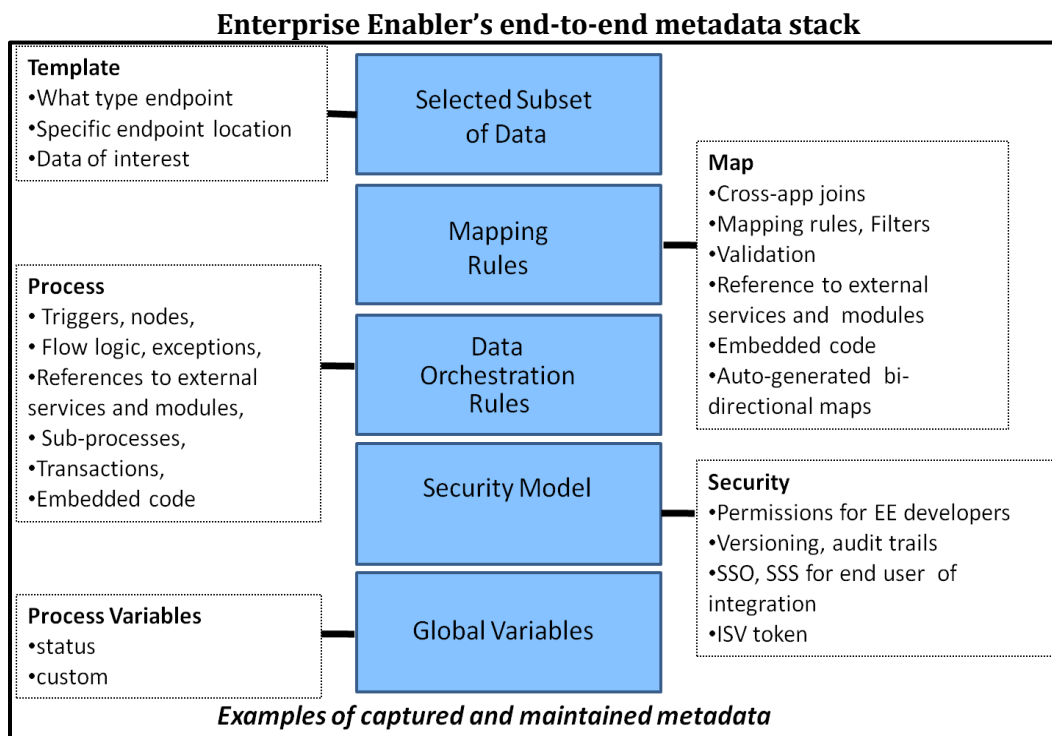
Enterprise Enabler offers a single environment for end-to-end development:
- Design
- Discover
- Develop
- Test
- Deploy
- Monitor
- Maintain

And a single environment for end-to-end integration:
- EAI
- ETL
- EII
- Data Orchestration
- Data Federation
- Data Virtualization
- Federated Queries
- Web Services
- SaaS Integration
- Integration Integrity monitor-impact analysis-resolution
- Source of Record Management

A powerful feature of Enterprise Enabler is its end-to-end internal metadata stack. Having a consolidated metadata stack means that any of the metadata values from the stack are available at run time for use in data flow logic or in transformations, adding to the fluidity and maintainability of the integration, on cloud or on premise.

Enterprise Enabler's end-to-end metadata stack

Template •What type endpoint •Specific endpoint location •Data of interest	Selected Subset of Data	
	Mapping Rules	**Map** •Cross-app joins •Mapping rules, Filters •Validation •Reference to external services and modules •Embedded code •Auto-generated bi-directional maps
Process • Triggers, nodes, • Flow logic, exceptions, •References to external services and modules, • Sub-processes, •Transactions, •Embedded code	Data Orchestration Rules	
	Security Model	**Security** •Permissions for EE developers •Versioning, audit trails •SSO, SSS for end user of integration •ISV token
Process Variables •status •custom	Global Variables	

Examples of captured and maintained metadata

Conclusion

Data Federation and Virtualization are not just the next trendy hype cycle. Together they address the root causes of a huge cross-section of the difficulty and complexity of moving data smoothly among line of business applications, both on-premise and in the cloud, as well as to end users in dashboards and information exchange with business partners. Of course, unless the technology used to deliver this capability is a single product, such as Enterprise Enabler, with all contributing features operating off a single internal metadata stack, the benefits will be dramatically reduced because of the complexity to implement.

It is not necessary to undertake implementation as a huge, corporate-wide (and therefore high-risk) project, but rather to apply the model to new projects and high-maintenance or "broken" but high impact situations first. This new agility can then be applied across the organization. Unlike big projects that require a huge effort to implement all kinds of new databases and other heavy complexity, think about the end result as eliminating databases and complexity.

Customers, end users, executives, programmers, the budget, and Shareholders alike will benefit.