expressor ®

Data Integration Challenges in Onboarding Operational Data

Latest Update: September 21, 2011

Executive Summary

Organizations in a wide variety of industries must import, normalize, and integrate operational data from various internal and external business partners on an on-going basis. This process of *onboarding* operational data is often fraught with complexity due to variations in the way partners and systems represent data—and inconsistencies from one onboarding process to the next. These complexities often impact an organization's ability to scale its processes and support its business.

This white paper examines various challenges involved with onboarding operational data, and presents an effective solution based on a powerful enabling technology from expressor that simplifies and standardizes onboarding processes to enable organizations to overcome these business challenges.

Introduction

Onboarding operational data refers to the process of integrating data from various internal and external partners into an organization's information system infrastructure to support an organization's business processes. The ability to quickly and accurately onboard operational data is often a core, mission-critical requirement.

Many companies, in a wide range of industries, have business requirements for onboarding operational data from external sources—including from other departments within the company, as well as from third party business partners—into their structured data management environment. While some organizations onboard data that is delivered to them in a predictable format, many organizations must respond in real-time to new data formats provided by new partners—creating complexities, delays, and errors.

Examples of organizations that must quickly and accurately onboard operational data include the following:

- B2B manufacturers that must synchronize ERP and other data across a network of suppliers, subcontractors, and distributors.
- Pharmaceutical vendors that need to analyze the distribution of drugs by processing the daily transactions of the pharmacies that dispense its products.
- Marketing services providers that accept data from customers to create new campaigns, perform data enrichment and data cleansing, and other services.
- Discount retailers that depend on being able to rapidly integrate with new suppliers to quickly offer and sell their distressed inventory.
- Services providers such as health insurance vendors or 401-K administrators that must quickly support new enterprise customers.

While onboarding data is often a mission-critical business process, many organizations do not treat it as a mission-critical component of their enterprise data strategy. As a mission-critical business process, organizations can gain dramatic business benefits by improving accuracy and efficiency and reducing costs and complexities associated with developing, executing, reusing, and maintaining their onboarding processes.

Challenges

Onboarding operational data often requires handling data from a variety of sources and partners, represented in dissimilar formats. Coupled with a need to handle many different data sources and partners in a rapid and efficient manner, the process of onboarding operational data to support business partners can be challenging for organizations.

> **"Onboarding operational data is a source of pain for many organizations. We hear this again and again. Organizations need to normalize data from different clients or partners, and they need a solution that lets them adapt easily to each new onboarding requirement, which is where expressor can really help. expressor enables organizations to onboard new data much faster and more efficiently through a reusable data integration framework. With expressor, organizations will no longer need to write or modify one-off, custom programs and queries, which will simplify the onboarding process and reduce the maintenance associated with the traditional approach. Organizations will need fewer resources, while still reducing the time needed to add new customers or partners, enabling them to scale without adding significant costs to their business process."**
>
> *Wiqar Chaudry, expressor Product Manager*

The *data artifacts* that are exchanged between business partners may contain differences in the way that the data is represented, and there may be inconsistencies from one onboarding to the next, which make it difficult to develop standardized data integration applications to handle the process. For example, one partner may create an Excel spreadsheet with a data element called "Acct_Num," defined as a 16 character length alphanumeric field, which is the database representation of the "Account Number." Another partner may create a CSV file with "Account_Nbr" defined as an integer. While both data elements are functionally the same, they are represented in different formats that require different processing at the data integration level. These are the sort of inconsistencies that make it challenging to create efficient and reusable data processing applications that support multiple data sources and business partners.
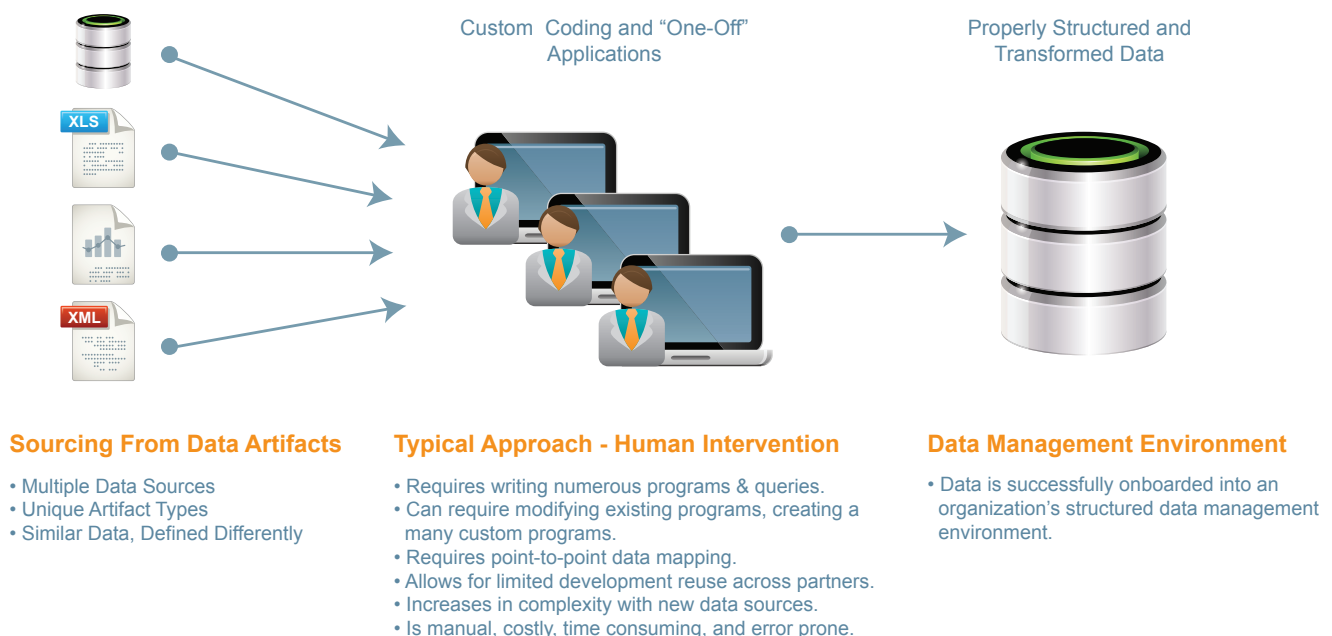
Custom Coding and "One-Off" Applications

Properly Structured and Transformed Data

Sourcing From Data Artifacts

• Multiple Data Sources
• Unique Artifact Types
• Similar Data, Defined Differently

Typical Approach - Human Intervention

• Requires writing numerous programs & queries.
• Can require modifying existing programs, creating a many custom programs.
• Requires point-to-point data mapping.
• Allows for limited development reuse across partners.
• Increases in complexity with new data sources.
• Is manual, costly, time consuming, and error prone.

Data Management Environment

• Data is successfully onboarded into an organization's structured data management environment.

Figure 1: Workflow and Challenges Onboarding Operational Data

In addition, apart from the challenges of integrating new data from new partners and sources, the metadata that defines the underlying data is often different as well.

As a result, it is difficult to standardize and reuse existing onboarding applications, whether developed using data integration tools or custom coding techniques, to accommodate new partners and data. Organizations are forced to repeatedly develop new applications to handle "one-off" data artifacts in customized ways for each new partner or data source. This growing library of one-off routines is cumbersome, expensive, and inefficient for organizations to manage and maintain. In addition, the subtle differences between the data elements require tedious attention to detail, often causing delays and errors. Traditional approaches to onboarding operational data are typically manually intensive, expensive, time consuming, and error prone.

Finally, as the volume of data, data elements, and frequency of onboarding events increase, the challenges and costs associated with processing these data artifacts are magnified—creating issues around an organization's ability to scale its process, and ultimately scale its support of the business.

expressor software: An Ideal Approach To Onboarding Operational Data

Data integration and ETL tools are universally recognized to be more efficient than custom coding and manual data manipulation techniques for the initial development and re-use of applications. However, the unique challenges presented by many onboarding tasks make application re-use difficult or impossible, even for most data integration tools.

expressor software provides a radically simplified approach to data integration that far surpasses the application re-use capabilities of every other data integration software (as well as those of custom coding), making it ideal for onboarding operational data. expressor has pioneered a unique *metadata abstraction framework* that enables unprecedented re-use of data, metadata, and data integration artifacts, to finally bring re-use and simplicity to the process of onboarding operational data from business partners. Perfectly suited to this use case, expressor is a fast, reliable, easy-to-use platform for accessing, transforming, and delivering data from virtually any business system in any format.

Benefits of Semantic Data Integration
• Creates a reusable onboarding infrastructure.
• Simplifies data mapping with reusable Semantic Types.
• Requires fewer technical resources.
• Creates a more manageable data integration environment.
• Supports a scalable process, which supports more files, more customers, and more data.

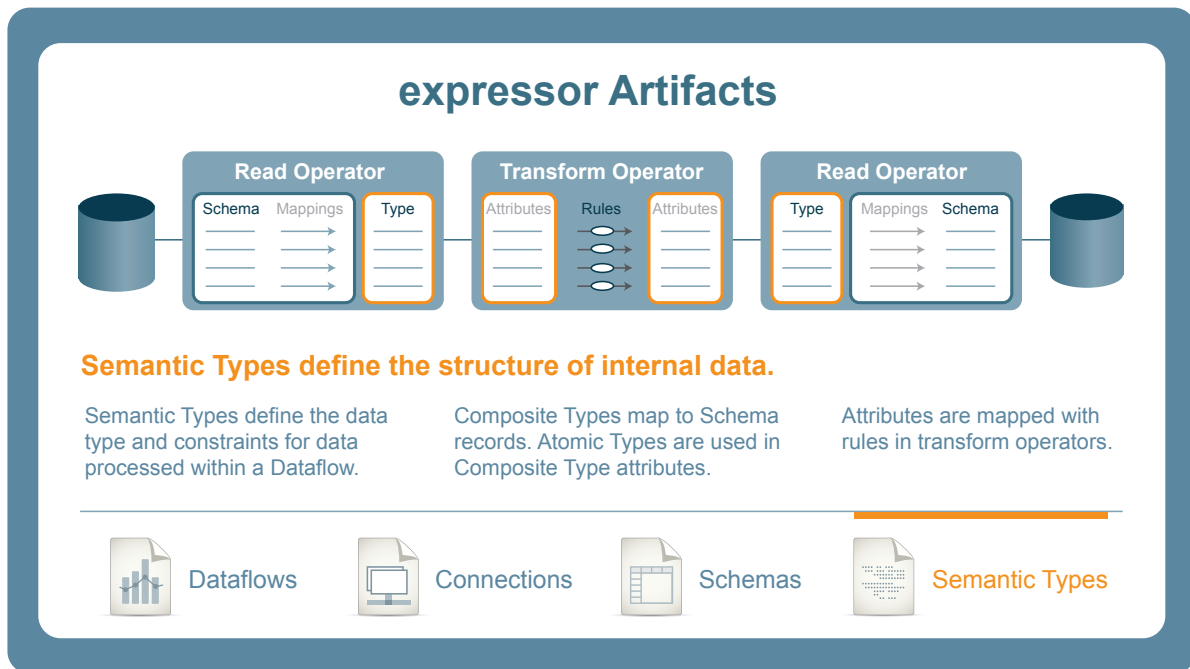*Figure 2: Semantic Data Integration simplifies Onboarding Operational Data*

Figure 3: Semantic Types in expressor

The foundation of the expressor platform is a powerful metadata layer called the *Semantic Framework*. The Semantic Framework simplifies and standardizes the complex and often messy process of data integration application design. It employs reusable data objects referred to as *Semantic Types* that enable organizations to define, in business terms, how the data should be represented. Semantic Type artifacts hold information such as field names and data types, which are automatically created and mapped to a schema artifact, which is a reusable piece of metadata that describes the external data layout when an external data source is read. The auto-generated Semantic Type inherits the characteristics of the schema by default, simplifying the task of working with disparate and dissimilar data sources. Semantic Types are reusable, abstracted, internal data structures that can be mapped to multiple data sources or targets. They enable external data types to automatically adapt to the internal data types defined in the Semantic Type, which enables the Semantic Framework to be easily modified to handle new data sources, and automatically insulates applications from schema changes that might occur in a data source or target.

Semantic Types Deconstructed

Semantic Types represent a reusable abstract internal data structure that can be user defined or automatically generated from the data's source schema. Users can further extend the capabilities of Semantic Types by adding constraints to each attribute of a Semantic Type. An application can utilize various Semantic Types:

Local Composite Type – A local type is an artifact that is auto-generated whenever a new data source is loaded. A local type represents an internal data structure that is mapped to a schema artifact, which defines the external data structure of the data source or target.

Shared Composite Type – A shared type can be a local type that has been promoted, or it can be a generic structure that can be defined within the Studio application. The benefit of this type of internal data structure is that it can be reused across many external data sources and targets without affecting any of the internal ETL logic written against it.

Shared Atomic Type – An atomic type is the fundamental building block of Semantic Types and represents a single attribute that can be utilized in one or more composite types.

Constraints – Constraints are defined on individual attributes of a composite type. Each attribute carries the following information as part of its definition:

- The name of the attribute, which should be descriptive for human consumption.
- A primitive data type which is assigned to the attribute (supported data types: string, integer, decimal, double, date-time, byte).
- Constraints, which are sets of rules that the data represented by the attribute must conform to based on data type (example constraints: minimum value, maximum value, minimum, regular expressions, allowed values).

Once a constraint is defined, the application can specify the corrective action that should be taken if the rule is violated. By default, all corrective actions are set to escalate errors to the dataflow, and the application can reject, skip, or redirect data and take appropriate actions. An application can be set to correct data within the constraint definition by providing default values or other available actions based on the attribute data type.
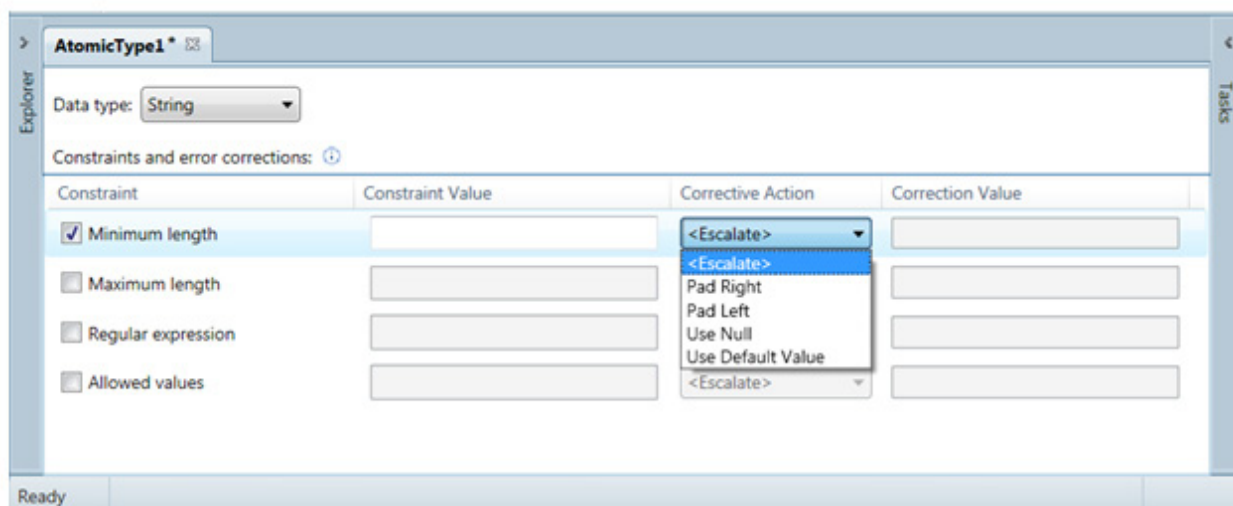

Figure 4: Constraints and Error Corrections Interface in expressor

Constraints are applied to attributes of composite types or atomic types and enable the constraint information to be carried forward wherever a given attribute is used. This streamlines the development process by decoupling data cleansing logic from any actual transformation that needs to occur to the data. All constraint-related logic is centralized and managed within the Semantic Type artifact enabling easier maintenance and artifact reusability, making it ideal for handling the wide variation in data artifacts during onboarding processing.

Simplified Data Mappings

The Semantic Framework simplifies the typically highly complex data mapping process by replacing the point-to-point mapping done in traditional ETL tools with a unique, streamlined mapping structure achieved through the use of Semantic Types. By mapping source and target datasets to business data objects, expressor automatically reconciles data type discrepancies.

expressor is purpose-built to allow organizations to reuse Semantic Type artifacts as many times as necessary, enabling organizations to build a single application that can be applied to the wide range of business partners and data formats dictated by its business requirements. Semantic Type definitions include mappings, data type conversions, constraint definitions and error handling actions. expressor's unique approach to data integration ensures that all changes are automatically and accurately propagated through all data integration applications.
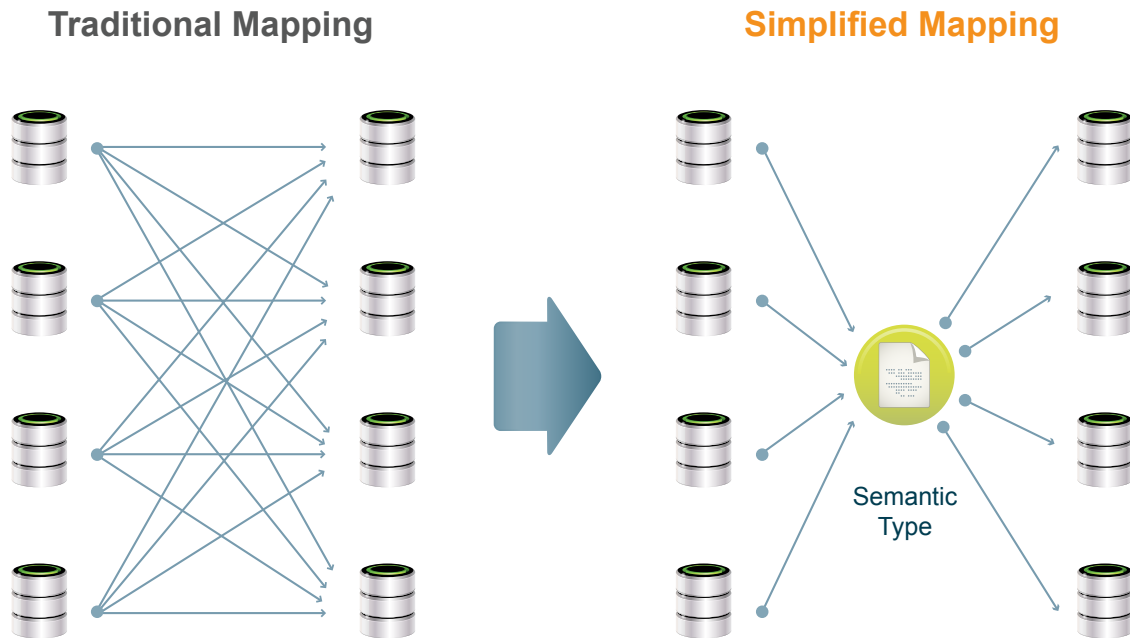
Traditional Mapping   **Simplified Mapping**

Figure 5: expressor's Simplified Mapping Process

Onboarding Operational Data Case Study: Managed Service Provider

A marketing services provider (MSP) serves a large number of financial institutions, providing various forms of email communications and campaigns to each financial institution's customers.

As part of each client engagement, the financial institutions submit their own customer files to the MSP, each in a unique format. The MSP must validate, normalize, and import the data in each customer file into its database before campaigns can be run. The company operates on tight timeframes, executing multiple campaigns for a range of clients across multiple time zones.

The company's initial, custom-coded solution for onboarding data from its clients was very rigid. Since new mappings and transformations were required for each new customer and campaign, "development was very slow, and mainte-nance was a nightmare," according to the company. Moreover, the company has seven employees that can function as analysts, but only one developer. The company needed a flexible onboarding solution that would allow the seven analysts to quickly customize the application for new customers and campaigns.

The company implemented expressor software to develop a flexible, generic application to read customer files, perform validation and error handling, map and transform the data in the incoming files to match the schema in the production tables, load new records to the production table, and write the prior records to a history table.

Whereas the MSPs prior onboarding application required 80 hours to develop, the expressor software application, built using expressor's graphical, drag-and-drop, color-coded interface was built four times faster.

With the previous application, each new customer or campaign required a new mapping, which took eight hours of the developer's time. With expressor and the use of Semantic Types, the new files are incorporated by re-using the existing application through expressor's graphical development environment. The new customer and campaign mappings are created eight times faster and are created by the company's analysts, freeing the developer to work on more critical tasks.

The company's seven analysts are now able to accommodate their new customers and campaigns using expressor's graphical environment, which requires no coding or compiling, allowing them to easily map new file formats to the common Semantic Type. And all of logic, including transformations, field validations, and error handling are automatically maintained—dramatically speeding and simplifying new client engagements and enforcing standardization.

Benefits

expressor software provides an ideal platform for companies that must perform onboarding of operational data to support multiple partners and critical business processes. Organizations using expressor software for onboarding operational data are experiencing measureable benefits:

Speed of development. With expressor, customers download, install, build, and run their first application in minutes. expressor is easy to download, install, learn, and use. It provides a familiar, purpose-built UI with an MS Office-like look and feel. It provides drag & drop configuration and uses simplified, graphical tools for defining reusable data mappings and business rules to speed and simplify the development of data integration applications.

Ease of re-use. With expressor, customers can reuse everything, making it ideal for handling the challenges and complexities of onboarding operational data from business partners. Organizations can define and reuse any project artifact from simple file connections to complex data flows and business rules. Users can design and reuse transformation operators, data flows, data source connections, and transformation logic. Moreover, all project artifacts can be shared and reused within a single workspace across multiple projects.

Accessible to a wider range of skill sets. Simplified re-use means that less technical users can adapt the applications to handle new data sources and business partners, freeing developers to work on more critical tasks.

Increased accuracy and standardization. Defining standard applications that enforce data mappings, constraints, and error handling actions ensures consistency with new data and partners, especially compared with custom coding and developing one-off applications.

Summary

Organizations across a wide range of industries that must effectively onboard operational data are experiencing dramatic business benefits by implementing expressor software's data integration platform, including improved accuracy, greater efficiencies, and reduced costs and complexity associated with developing, executing, reusing, and maintaining its onboarding processes.

About expressor

At expressor software, we're simplifying data integration by providing software that is easy to use, powerful, and afford-able—enabling organizations to accomplish more with fewer resources. We dramatically simplify the vast majority of basic data integration tasks—making it far easier to map and move data from one location to another. And, when data integration becomes more complicated—requiring various types of data transformations—we help make data integration and ETL projects more flexible and efficient by enabling the creation and subsequent reuse of powerful artifacts, con-figurations, and business rules. This means that when a complex business data challenge is solved once, the solution can be reused.

expressor has a rapidly growing community of over 12,500 unique expressor Studio download users today. Leading organizations such as SKECHERS USA, American Tower, Viverae, Western World, LABVANTAGE, and ClickRSVP have adopted expressor's affordable data integration platform for server deployment to help them integrate on-premises and cloud data sources with their business critical applications.

expressor Studio is the design component of the expressor Data Integration Platform, and expressor makes a free download of Studio available at www.expressorStudio.com.

expressor software is headquartered in Burlington, MA. For more information about how expressor is simplifying data integration, please visit our website at www.expressor-software.com or contact us directly at +1 (781) 505-4190 x222.

EXPRESSOR ® **Simplifying Data Integration**