# Semantic and Syntactic Interoperability in Transactional Systems

Chito Jovellanos
forward look, inc.
8 Faneuil Hall Marketplace, 3rd Flr
Boston, Massachusetts US 02109
+1 (617) 763-7011
chito@forwardlook.com

## ABSTRACT
This research note describes a middleware-based system that enables semantic and syntactic interoperability of transactional data exchanged in real-time between sending and receiving systems. Our system samples transaction streams, and statistically measures the variances in the semantics and syntax of the underlying data elements exchanged between sending and receiving systems. Automatic transformations of the data elements are applied when a context-specific Data Operability Threshold is crossed. The transaction is then transmitted onward by the sending system (or processed for consumption by the receiving system) in a format that ensures no manual reconciliation would be required to process the transaction. The System is being prototyped for cross-border securities trading and settlement.

## Keywords
Integration, Interoperability, Semantics, Syntax, Transformation.

## 1. INTRODUCTION
Enabling full data interoperability at the business-to-business level (defined as the ability of communicating business systems to generate and consume transactional data in a straight-through-processing (STP) manner) has never been satisfactorily addressed. In securities trade processing (as well as other financial industry verticals), the development of exhaustive message standards is rightfully envisioned as the primary enabler for STP. For example, the Society for Worldwide Interbank Financial Telecommunication (SWIFT) has established a number of standards (eg, ISO 15022), since its founding in the early 1970's, which stipulate how various securities and related payment transactions are to be communicated between investment institutions, brokerages and custodial banks. Despite committed and concerted efforts by industry participants to comply with the letter and spirit of the standards, a significant amount of manual reconciliation is still required to process these transactions. The latest industry research suggests that as much as 36% of securities transactions do not meet the goal of STP resulting in significant trade management and reconciliation costs [1].



The core reasons underlying sub-par STP rates are the semantic and syntactic differences with respect to how individual companies represent the specific data elements that populate the industry-compliant transaction formats they exchange in the normal course of business [2], [4]. Constrained systems implementations (eg, small mutual funds cannot often afford the data quality requirements for global securities cross-references), and proprietary business workflows (eg, an equity trading firm may not effectively support fixed income derivative business) often preclude full compliance with the rigorous application of the generic transaction standard. Four decades of experience with message standardization in the global securities industry shows that standards quickly evolve away from the idealized norm where the variances are conditioned by systems capabilities, internal workflows and shifting business imperatives.

## 2. SOLUTION FRAMEWORK
Our approach recognizes that small but significant differences in the application of messaging standards exist. The ability to achieve STP is tied intimately to compensating for the variances that continually arise in the semantics and syntax of data elements that comprise business transactions. We demonstrate the application of "integration on demand" [3], as a complement to "integration in advance" (ie, the naïve premise that data harmonization through message standards is a sufficient solution to the problem of business-to-business integration).

Our prototype implementation (System) is nested within existing middleware on the sender's computer, the receiver's computer, or both. Our System performs the following major operations :

- Analyze data streams of tagged transactions (NB, not necessarily XML-based) created by a computer system (eg, a company's internal order management systems) prior to outbound transmission to another computer system (eg, the counterparty's order management system). Similarly, complementary transactions from the counterparty's systems

(eg, trade confirmation systems) are also inspected. This universe of samples reflects a coherent collection of transactions that support an established business workflow, and provides a complete view of day-to-day "real-world" data usage.

- Derive transformation algorithms that enable automated restructuring of the data elements of the sender's transaction (ie, immediately prior to transmission by the sender's system) into a message format that a receiver's system can process without manual intervention. Conversely, transformation algorithms can also be derived (ie, for use by the receiver's system) which restructures the data elements of the counterparty's inbound transaction into a form that the receiving system can process without manual intervention. These transformation algorithms are enabled using XSLT.

The accuracy of our System improves when the sampled transaction is partitioned into logical semantic segments (Topic Domains) prior to analysis. Each Topic Domain comprises a discrete collection of data elements that delineate a generally understood and logically coherent subject matter area (eg, client information; regulatory compliance data).

The variances in the syntax and semantics of sampled transactions are tracked using the Data Interoperability Grid (DIG). The DIG is a set of three dimensional numeric arrays that captures the statistical similarity of data element usage within and across transactions.

The variance profile of every data element sampled from the transaction stream is accessible via a Universal Data Structure (UDS). One UDS is available for each field and counterparty combination (eg, "Net Amount" as used with "Goldman Sachs London"). The ensuing collection of UDS objects are inherently simple to use since each and every UDS is consistent in structure and uniformly addressable by an interrogating program.

Data transformations are triggered when the semantic and syntactic stability of a transaction stream is perturbed. The Data Operability Threshold (DOT) is a mathematical region (defined by the metrics captured in the DIG) that describes a stable (but non-stationary) optimum of transactional data usage. When the DOT region is pierced, the event triggers the corresponding data transformations (enabled via UDS calls) that dynamically restructure the data elements in the inbound or outbound transactions in order to accommodate the current (or a modified) stable optimum.

## 3. PRELIMINARY RESULTS

Preliminary results from our initial deployment of the System showed promising results (Table 1), and areas for improvement. We validated its capabilities in enabling data interoperability between counterparties' systems by processing corporate actions (eg, notifications for stock splits, mergers and other events that affect the capital structure of a company), and securities trade settlement instructions (eg, a purchase and sale effecting a change in ownership of stock on receipt of payment). Based on concurrent STP metrics obtained for the same set of transactions, usage of our System resulted in significant STP gains of approximately 33% for corporate actions and 12% for settlement instructions.

**Table 1. Improvement in STP Rates achieved using our System. STP Rate defined as percentage of transactions that did not require manual handling. Sample Size shows average daily number of transactions (daily peaks in parentheses). Results based on 30 contiguous business days. Data represents a global investment institution interacting with their custodian bank. The "training data" set consisted of the samples from the first five days, and were excluded from the "result set" (which comprised the remaining 25 days)**

| | STP Rate : Existing Automation | STP Rate : With Our System | Sample Size |
|---|---|---|---|
| **Corporate Actions** | 64% | 97% | 1000 (3000) |
| **Trade Settlement** | 86% | 98% | 40000 (180000) |

Corporate action transactions (which are semantically far richer than settlement instructions) each required 3.2 seconds on average to process, while each settlement instruction only took 0.8 seconds on average. Initializing the System for corporate actions required 33 hours of "learning time" based on the training set, while settlement instructions only required 14 hours in total. Follow-on optimizations to the underlying algorithms are now being undertaken.

## 4. REFERENCES


[1] Bryant, A. 2002. Presentation at SWIFT SIBOS Conference, Geneva, 30 September 2002.

[2] Rahm, E. and Bernstein, P.A. 2001. A survey of approaches to automatic schema matching. The VLDB Journal 10 : 334-350.

[3] Stonebraker, M. 2002. Too Much Middleware. SIGMOD Record 31:1 March 2002 (http://www.acm.org/sigmod/record/issues/0203/).

[4] Yan, L.L., Miller, R.J., Haas, L.M. and Fagin, R. 2001. Data-Driven Understanding and Refinement of Schema Mappings. Proceedings of the ACM SIGMOD 2001 Conf: 485-496.