Financial Information Mediation: A Case Study of Standards Integration for Electronic Bill Presentment and Payment Using the COIN Mediation Technology

Sajindra Kolitha Bandara Jayasena, Stéphane Bressan, Stuart Madnick
Singapore-MIT Alliance, National University of Singapore, Massachusetts Institute of Technology

***Abstract* - Each player in the financial industry, each bank, stock exchange, government agency, or insurance company operates its own financial information system or systems.**[1]

By its very nature, financial information, like the money that it represents, changes hands. Therefore the interoperation of financial information systems is the cornerstone of the financial services they support. E-services frameworks such as web services are an unprecedented opportunity for the flexible interoperation of financial systems. Naturally the critical economic role and the complexity of financial information led to the development of various standards. Yet standards alone are not the panacea: different groups of players use different standards or different interpretations of the same standard.

We believe that the solution lies in the convergence of flexible E-services such as web-services and semantically rich meta-data as promised by the semantic Web; then a mediation architecture can be used for the documentation, identification, and resolution of semantic conflicts arising from the interoperation of heterogeneous financial services.

In this paper we illustrate the nature of the problem in the *Electronic Bill Presentment and Payment* (EBPP) industry and the viability of the solution we propose. We describe and analyze the integration of services using four different formats: the IFX, OFX and SWIFT standards, and an example proprietary format. To accomplish this integration we use the COntext INterchange (COIN) framework. The COIN architecture leverages a model of sources and receivers' contexts in reference to a rich domain model or ontology for the description and resolution of semantic heterogeneity.

***Index Terms* – Financial systems, interoperation, mediation, standards.**

[1] S. K. B. Jayasena was in the Computer Science program of the Singapore-MIT Alliance (SMA). (e-mail: sajindra@mit.edu).

S. Bressan. is with the School of Computing, National University of Singapore, 3, Science Drive 2, Singapore, 117543 (phone: (65) 68 74 35 43; fax: (65) 67 79 45 80 e-mail: steph@nus.edu.sg).

S. E. Madnick is with the Engineering Systems Division, School of Engineering and Information Technologies Group, Sloan School of Management, Massachusetts Institute of Technology, Room E53-321, Cambridge, MA 02142 USA (phone: 617-253-6671; fax: 617-253-3321; e-mail: smadnick@mit.edu).

1. INTRODUCTION

Effective and transparent interoperability is vital for the profitability and sustainability of the financial Industry. Adhering to a standard is not feasible because different institutions often utilize different standards. Even when within one standard or when standards seem to agree, one often finds different possible interpretations originating in the particular practices and cultural background of the various players.

Typically, a Financial Institution (FI) that is involved in *Electronic Bill Presentment and Payment* (EBPP) Industry, for instance operating in a European Union country, is faced with a multitude of standards such as IFX (Interactive Financial Exchange protocol)[10], OFX (Open Financial Exchange Protocol)[9] and the world wide inter-bank messaging protocol, SWIFT [11]. Making matters worse, the FI may have its own semantics for its internal information systems that represent the same business domain in a different context. In the rest of this paper we would be referring to the set of assumptions about the representation, syntax and interpretation of data according to IFX, OFX, and SWIFT as IFX, OFX and SWIFT contexts and the assumptions of the internal financial system of a Financial Institution as internal context.

The *Price* and *Invoice* concepts may be represented in different ways, e.g., excluding tax, with tax and fees, and even with inter-bank charges, resulting in definitional conflict [1]. Interoperability of such definitional conflicts is vital in distinguishing intra-bank and inter-bank payment across borders. Further, different contextual heterogeneities exist on the *currency*, where in certain contexts like IFX and OFX it is implicitly based on where the funds are directed. As a result of different *Account types* and BANK/*BRANCH code,* financial institution would need to maintain complex mappings between different contexts. In addition, there can be data level heterogeneities like date formats and representations. Examples of possible conflicts are summarized in table 1. The columns for OFX, IFX, and SWIFT represent actual real-life conflicts and similarities that exist between those standards, while the conflicts addressed under the internal schema column refer to a hypothetical, but realistic, financial system that would interact between OFX, IFX and SWIFT standards.

Table 1: Some conflicts in different standards

Property	Internal Schema	OFX	IFX	SWIFT 103/103+
Price	1000 FFR (French Franc)	1000 USD + 1000 5%	1000 USD + 1000 * %	1000 USD + 1000* 5% + 0 USD (inter-bank charge if utside EU)
Currency	FFR	Currency of country f incorporation of ayee bank i.e. USD	Currency of country of ncorporation of payee ank i.e. USD	Specified in message – can e the payee or payer's urrency
Account types	CHK,SVG, INYMRT	CHECKING, AVINGS	DDA,SDA	N/A
Bank and branch ode	Internal ID	Dependent on the ountry i.e. clearing ,sort #	Dependent on the ountry i.e. clearing ,sort #	BIC / BEI (branch ID + bank Id)
Invoice	Net	Net + fees + tax	Net + fees + tax	Included in *Amount* – N/A
Due date	*23022002*	*20020223*	*2002-03-23*	*030223*

The objective of this research is to analyze how COIN mediation technology [2, 3, and 8] could be applied to provide a declarative, transparent yet effective mediation solution to the sources of heterogeneity and conflicts that exist within and among the existing financial standards. Further we discuss the extension of our work in mediating the conflict that cannot be addressed in the current COIN implementation.

The organization of the following sections is as follows. First we look at the plethora of financial messaging standards that infest the financial world followed by a review of related work in mediation technologies and specific work related to interoperability in the financial industry. Then we look at the intricate details of the COIN mediation framework. Next, the bulk of the paper focuses on how COIN can be applied in one of the critical industries in the financial world – The *Electronic Bill Presentment and Payment* (EBBP) industry. In the final section, we summarize and briefly discuss future research.

2. BACKGROUND AND RELATED WORK

2.1 Financial Standards

The standards addressed herein are involved in business banking, Electronic Bill Presentment and Payment, Securities and Derivatives, Investments, Economic and Financial indicators, straight through processing and other *over the counter* derivatives. As a whole, the financial industry is cluttered with numerous protocols and standards that are utilized in different segments in the financial industry. Prominent ones are Financial Information Exchange protocol (FIX), S.W.I.F.T., Interactive Financial Exchange (IFX) and Open Financial Exchange (OFX). SWIFT is the leader in inter bank transactions, and also has gained a significant market holding on Securities and derivatives, payments as well as investments and treasury after introducing a set of messages for securities and derivatives industry. OFX is the leader in Intra-bank transaction systems followed by its successor, IFX. IFX is opting to replace OFX, through its rich and extended messaging standards. Both of these standards are widely used in business banking, Electronic Bill Presentment and Payment, ATM/POS Industry. FIX is the leader in securities and derivatives market, used by major stock markets around the world. Most of these protocols use XML as the medium of messaging. Non-XML based standards like FIX and S.W.I.F.T have come up with XML versions, namely *FIXML* and '*SWIFTStandards XML*'. In addition to these major players, some of the other protocols are *RIXML* – Research Information exchange and *IRML* – Investment research markup , focusing on fixed income securities and Derivatives market, *MDDL* - Market Data Definition and *REUTERS* in economic and industrial indicators, *STPML* – Straight through processing markup language - a superset protocol to replace *FIX,SWIFT ISITC* and *DTC ID*, *FinXML* – Financial XML which focuses on Capital market instruments and straight through processing (STP) and finally *FpML* - Financial products markup language focusing on interest rate swaps, forward rate agreements, Foreign Exchange and other *over the counter* derivatives.

2.2. Different mediation strategies

When institutions exchanging financial information subscribe to different standards, a mediator can be used to translate from one encoding scheme to another. The problems that the mediator needs to solve are similar to those in data integration of heterogeneous sources, where the potential variety of encoding schemes can be arbitrarily large in the latter case. The approaches addressing the issue of interoperability of disparate information sources have been categorized in the literature as static vs. dynamic [14], global vs. local schema [15], and tightly vs. loosely coupled [16, 17] approaches. These groupings can roughly be thought of referring to the same distinction characterized in [16] by:

· Who is responsible for identifying what conflicts exist and how they can be circumvented; and

· When the conflicts are resolved.

We briefly look at these approaches under the categories of tightly and loosely coupled approaches.

In tightly coupled approaches, the objective is to insulate the users from data heterogeneity by providing a unified view of the data sources, and letting them formulate their queries using that global view. In *bottom up* approaches the global schema is constructed out of heterogeneous local schemas by going through the tedious process of schema integration [18]. In *top-down* approaches global schema is constructed primarily by considering the requirements of a domain, before corresponding sources are sought. In tightly coupled approaches, data heterogeneities between sources are resolved by mapping conflicting data items to a common view. Early prototypes which have been constructed using the tight-coupling approach include Multibase [19], ADDS [20], and Mermaid [21]. More recently, the same strategy has been employed for systems adopting object-oriented data models (e.g. Pegasus [22] based on the IRIS data model), frame-based knowledge representation languages (e.g. SIMS [17] using LOOM), as well as logic-based languages (e.g. Carnot [23] using CycL, an extension of first-order predicate calculus).

Loosely coupled approaches object to the feasibility of creating unified views on the grounds that building and maintaining a huge global schema would be too costly and too complex. Instead they aim to provide users with tools and extended query languages to resolve conflicts themselves. Hence, instead of resolving all conflicts *a priori*, conflict detection and resolution are undertaken by receivers themselves, who need only interact with a limited subset of the sources at any one time. MRDSM [15] is probably the best-known example of a loosely-coupled system, in which queries are formulated using the multidatabase language MDSL. Kuhn et al [24] have implemented similar functionalities in VIP-MDBS, for which queries and data transformations are written in Prolog. They showed that the adoption of a declarative specification does in fact increase the expressiveness of the language. Litwin et al [25] has defined another query language called O*SQL which is largely an object-oriented extension to MDSL.

In the past two decades, various mediation strategies have been developed attempting to tackle these semantic heterogeneity problems. We will not give a detailed review of these approaches; interested readers are referred to [ZD04, W*01, L02] for recent surveys. For example, the authors of [28] use a domain model and source modeling to realize and optimize queries to distributed and heterogeneous sources. Generally, under these strategies, the mediator needs to be rebuilt when the underlying sources or user requirements change, which hinders the extensibility of the approach. We will discuss a middle ground approach that overcomes these drawbacks in Section 3.

2.3. Current integration and mediation strategies in Financial Standards

Due to intricacies and inefficiencies in using and integrating multiple standards and additional overhead, financial institutions as well as government organizations have put effort in merging different standards or coming up with new super set standards to replace the existing diverse standards.

One example is the effort by FIX, SWIFT, OPEN APPLICATIONS GROUP and THE TREASURY WORKSTATION INTEGRATION STANDARDS TEAM (TWIST) to outline a framework of cooperation and coordination in the area of the content and use of a core payment kernel XML transaction.

Also the Organization for the Advancement of Structured Information Standards (OASIS) is carrying out research on one XML based super set protocol that would cover all business areas. But all these effort are focused on futuristic direction rather than the problem at hand. The effort of migrating the diverse world-wide standard to a common standard would be an enormous task. Current business integration efforts like the Microsoft™ BizTalk Server support diverse messaging standards integration through its rich messaging and mapping framework, but lack the sophistication in mediating complex ontological and contextual heterogeneities.

3. The COntext INterchange (COIN) approach

The COntext INterchange (COIN) framework is neither a tightly coupled nor a loosely coupled system; rather, it is a hybrid system. It uses a mediator-based approach for achieving semantic interoperability among heterogeneous information sources. The approach has been detailed in [8]. The overall COIN approach includes not only the mediation infrastructure and services, but also wrapping technology and middleware services for accessing source information and facilitating the integration of the mediated results into end-users' applications. The set of context mediation services comprises a context mediator, a query optimizer, and a query executioner.

The context mediator is in charge of the identification and resolution of potential semantic conflicts that exist in a query. This automatic detection and reconciliation of conflicts present in different information sources is made possible by general knowledge of the underlying application domain, as well as the informational content and implicit assumptions associated with the receivers and sources.

The declarative knowledge is represented in the form of a domain model, source descriptions, a set of elevation axioms, a set of context definitions, and a conversion library. The result of the mediation is a mediated query. To retrieve the data from the disparate information sources, the mediated query is transformed into a query execution plan, which is optimized, taking into account the topology of the network of sources and their capabilities. The plan is then executed to retrieve the data from the various sources; results are composed as a message, and sent to the receiver.

Domain model: The domain model defines the different elements that are needed to implement the strategy in a given application: The domain model a collection of rich types (*semantic types, attributes, etc.*) and relationships (*is-a relationship*) defining the domain of discourse for the integration strategy. This declarative knowledge about the domain ontology is represented independent of the various information sources and represents the generic concepts associated with the domain under consideration. Semantic types resemble the different entities in the underlying domain. For example *Account, Person* can be entities in a financial domain. The attributes represents the generic features those semantic types can have. i.e. *bankBalance, creationDate* attributes of *Account* semantic type. Further, attributes can be used to infer relationships between different entities. For example the *holder* attribute of an *Account* could refer to a *person* semantic type. In some instance the attribute can constitute a basic type; either a string or a numeric value represented by the super semantic type, *basic*.

Context: Context axioms are used to capture different semantic, contextual, and ontological representations that the underlying data sources contain. The context definitions define the different interpretations of the semantic objects from the different sources' or receiver's point of view. We use a special type of attributes, *modifiers*, to define the context of a data type. For example *currency* modifier may define the context of objects of semantic type *moneyAmount,* when they are instantiated in a specific context (i.e., currency is USD in that specific context).

Sources: All sources are represented using the *Source* concept where the type of the sources could be any data source ranging from a relational table, an XML stream, to a web page. Different wrapper implementations, including the web data extraction engine *Cameleon* [12], provide different interfacing mechanisms to diverse sources.

Elevation Axioms: The sources and the domain model needs to be linked in order to facilitate mediation. This is achieved through the definition of *Elevation axioms*. Its usage is two fold. First, each source is given a Context definition. Second, each attribute of the source is elevated to a particular semantic object (instances of semantic types) that is represented in the Domain Model. This facilitates in bridging the link between the context-independent, '*generic'* domain model and the context dependent sources.

Conversion library: Finally, there is a conversion library, which provides conversion functions for each *modifier* that specifies the resolution of potential conflicts. The relevant conversion functions are gathered and composed during mediation to resolve the conflicts. No global or exhaustive pair-wise definition of the conflict resolution procedures is needed. The mediation is performed by a procedure, which infers from the query and the knowledge base a reformulation of the initial query in the terms of the component sources. The procedure itself is inspired by the abductive logic programming framework [27]. One of the main advantages of the abductive logic-programming framework is the simplicity in which it can be used to formally combine and to implement features of query processing, semantic query optimization, and constraint programming.

In the next section we would show how these concepts are applied in our case study.

4. CASE STUDY

4.1. Electronic Bill Presentment and Payment domain

In order to demonstrate the usage of the COIN framework, a subset of the standards, namely the '*Electronic Bill Presentment any Payment – (EBPP)*' domain is selected. The EBPP domain is a rich subset of the financial services messaging frameworks that have considerable amount of heterogeneities. The main standards are OFX, IFX for intra-bank payment schemes and SWIFT for inter-bank payment and funds transfer.

The overall functionality can be visualized from Figure 1 in using those standards. The focus is on the analysis of various heterogeneities that lie among these standards and financial systems as well as how they are handled using COIN. The key intermediaries in an EBPP scheme are as follows:

- Biller Payment Provider (BPP) is an agent (usually a financial institution) of the Biller that originates and accepts payments on behalf of the Biller.
- Biller Service Provider (BSP) is an agent of the Biller that provides an electronic bill presentment and payment service for the Biller
- Customer Payment Provider (CPP) is an agent (usually a financial institution) of the Customer that originates payments on behalf of the Customer.
- Customer Service Provider (CSP) is an agent of the Customer that provides an interface directly to customers, businesses, or others for bill presentment. A CSP enrolls customers, enables presentment, and provides customer care, among other functions.
- Financial Institution (FI) is an organization that provides branded financial services to customers. Financial Institutions develop and market financial services to individual and small business customers. They may serve as the processor for their own services or may choose to outsource processing.

Both IFX and OFX provide XML based messaging framework for individuals as well as businesses in bill payment and presentment electronically. But the most acclaimed inter-bank fund transfer framework, SWIFT uses a non XML messaging protocol and recently went through a major restructuring in phasing out one of the most utilized messaged for inter-bank customer fund transfer, the M100, and introduced modified versions of MT103 and MT103+.

In order to depict the usage of COIN in EBPP mediation in a practical scenario, we have broken down the analysis to three main areas that spans from a customer initiating a Bill payment to its subsequent verification by the Biller. In addition to analyzing these standards separately we address how they are utilized in practical scenarios. All three standards inevitably require interfacing with the internal accounting and financial system of a financial system to make a successful payment from the customer to the Biller as in figure 1.

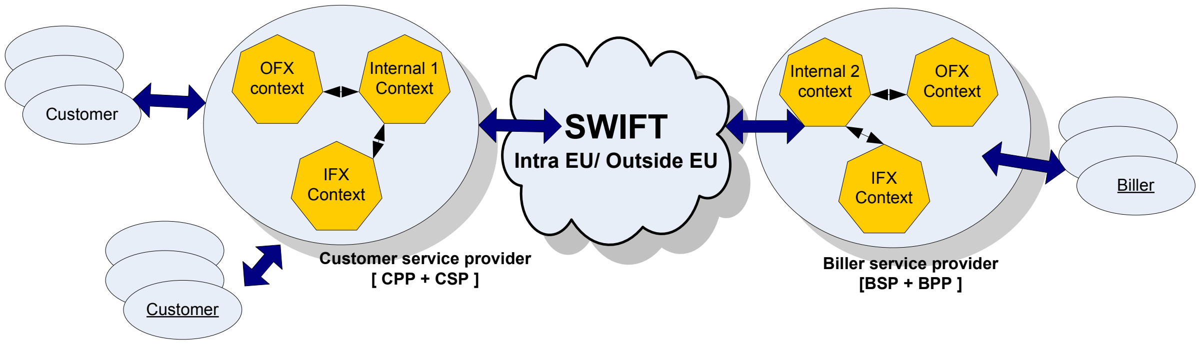

Fig. 1. Interfaces in EBPP

For example a bill payment from a customer might interact with an FI's IFX based system which in turn has to interface with customer's bank's internal accounting system. Then to facilitate the inter-bank funds transfer to the Biller's bank, a separate interfacing is required with a global inter bank messaging framework like SWIFT. Finally at the Biller's bank it needs to be represented and stored in its proprietary financial system. Finally the biller should be able to view the payments through its bank's bill presentment system which possibly utilizes OFX standard, where the internal representation needs to be transformed according to OFX's syntax and semantics. Therefore we have introduced a hypothetical internal system that represents the true nature of a realistic situation to bridge the gap between the financial standards as modeled in a real-life situation. The conflicts analysis and mediation with the diverse financial standards have been analyzed with respect to the hypothetical internal system of a Financial Institution which could be an in-house developed system or third-party (off the shelf) system. This internal system is represented by the term 'internal context'. Following are the three main areas analyzed in the case study.

- Mediation between an internal context and OFX context.
- Mediation between an internal context and IFX context.
- Mediation between an internal context and SWIFT context.

The IFX, OFX and SWIFT contexts represent the semantics and definitions adopted by IFX, OFX and SWIFT messaging frameworks respectively. SWIFT distinguishes intra European Union (EU) fund transfer and outside EU fund transfers for accounting for inter-bank charges.

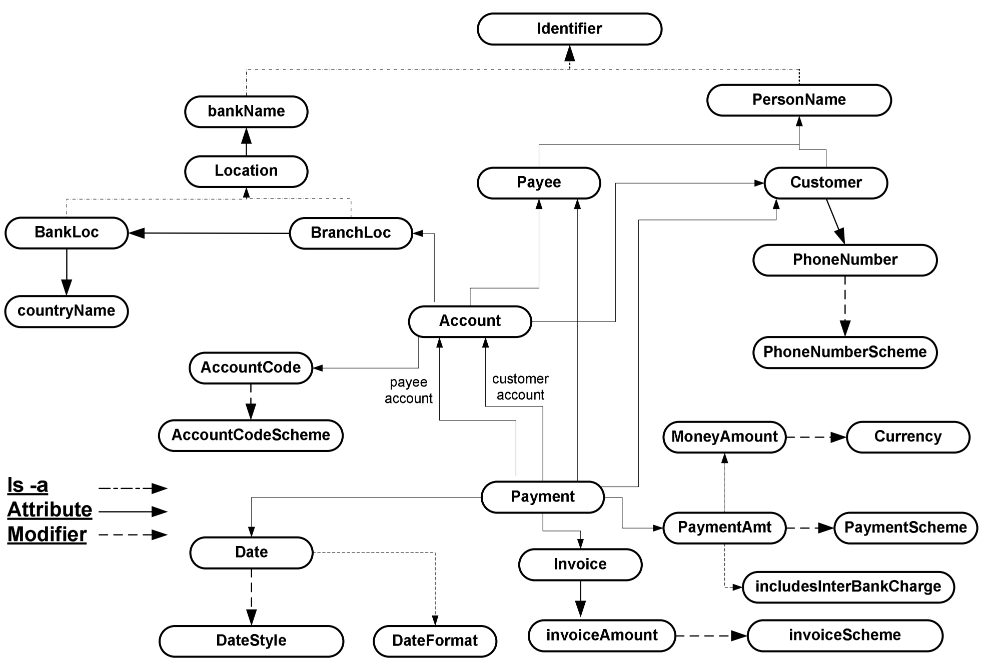

Fig. 2. Domain Ontology for EBPP

Figure 2 represents the context independent, COIN domain ontology for the EBPP domain denoting some of the key the concepts used by IFX, OFX, SWIFT and financial institution's own internal schema. This was constructed by exploring the business domain in EBPP and the relevant message handling semantics used in these diverse standards. The semantic types (entities) represents the business entities that encompass the main functionalities in EBPP Industry The sources and their conflict are mapped to these semantic types (entities). The semantic types denote the entities and their relationships in the EBPP domain like *Payment*, *Account* etc. *is-a* relation denotes an inheritance relationship between semantic types. A semantic type may have certain *Attributes* (e.g., payment has payee, payer accounts, amount etc). The entities that constitute conflicts in these contexts are modeled through *modifiers*. As an example, the *paymentAmount* can include/exclude various taxes in different contextual representations and in SWIFT it would incur an additional inter-bank service charge. These are represented by COIN modifiers *paymentScheme*, *includesInterBankCharge* respectively. Further, monetary amounts could be expressed in different currencies. This is modeled using the *currency* modifier for the super-semantic type *moneyAmount*. This represents how COIN models inheritance of contextual knowledge for different entities.

In an actual scenario, the heterogeneities of the standards and the mappings needed for mediation would be analyzed and formulated by a business analyst or a person working for the respective Financial Institution. The following sections addresses each of these three cases separately.

4.2. Internal Schema vs. OFX

First we will look at the mediations between OFX and an internal schema of a financial institution. Table 2 summarizes the heterogeneities identified in the two schemas. As denoted in COIN's mediation strategy, the modifiers and relevant conversion functions are the main ingredients in facilitating the mediation for a particular heterogeneity exiting between two different contexts. As shown in the table, there are different types of heterogeneities between the two contexts. The significant conflicts are payment amount, currency type and Account code reference identifiers. They are discussed below.

Table 2: conflicts in Internal and OFX contexts

Conflict	Internal Context	OFX Context	Mapped modifier (refer Fig 2)
Payment amount	Net amount without tax	Net + tax amount	*PaymentScheme*
Account Location dentifier – BANK reference	Bank identifier represented the internal scheme	Bank Identifier depends on e Bank's country of ncorporation.	*BankLoc*
Account Location dentifier – BANK BRANCH eference	Branch identifier of the ccount	Branch Identifier dependent n the bank's country of ncorporation.	*BranchLoc*
Payment due date ormat	European format	US format	*DateFormat*
Payment due date Style	dd/mm/yyyy 03/03/2003	Yyyymmdd 20030303	*DateStyle*
Account type code	CHECKING,SAVINGS etc	CHK,SVG etc	*AccountCodeScheme*
Currency type Exchange rate)	"EUR"	Currency of country of ncorporation of payee bank	*currency*
Phone number format	415.445.4345	1-415-445-4345	*PhoneNumberScheme*

Payment amount - The mediation strategy for payment amount is as follows. The mediator needs to apply two conversion functions in order to obtain the mediated payment amount, namely the currency conversion inherited from the *moneyAmount* super semantic type, and the tax adjustment for the payment. For simplicity let's assume that in both contexts the currency is denoted in three letter ISO 4217 format (i.e. USD, GBR, and EUR etc).

Assume that the query '*select AMOUNT FROM PAYMENT*' is called in *OFX* context;

First, payment amount is adjusted for the tax inclusion. Let us assume that the applicable tax is 'GST'. Then;

$$\text{Payment}_{\text{OFX}} = (\text{payment}_{\text{INTERNAL}} + \sum \text{GST amount for payment}_{\text{OFX}} * \text{payment}_{\text{INTERNAL}}) * \text{Exchange Rate ("EUR", OFF\_CUR,DATE\_OF\_TRANSACTION)} \quad (1)$$

In the COIN framework, the mediation formulas are translated into logical expressions of the COIN theoretical model [1]. Later these expressions are implemented in prolog and evaluated by an abduction engine implemented in the same language [13]. The following describes the logical representation of the formula (1) for this example.

The formula below describes a non-commutative mediation of *paymentType* object depending on its modifier *paymentScheme*, which in this case holds the values "*noTax*" or "*withTax*". The *Ctxt* defines the destination context. The conversion in simple terms would be to retrieve the Rate for the tax "GST" from the elevated relation '*OFX_TAX_TYPES_p'* which is an elevation mapped to relation '*OFX_TAX_TYPES'* under *OFX* Context (The destination context in this case) and utilizes in the tax calculation. The *value* predicate in the formula defines a value of a particular semantic object under a certain context.

```
cvt(noncommutative,paymentAmt,_O,paymentScheme,Ctxt,"notax",Vs,"withtax",Vt) ⇐
value(TaxName,Ctxt,"GST"),'OFX_TAX_TYPES_p'(TaxName,_,Rate),
value(Rate,Ctxt,RR),
(Vtemp is RR * Vs),
(Vt is Vs + Vtemp).
```

Further, this resembles an *Equational ontological heterogeneity* addressed in [5], which is a clear example of differences in the contexts of OFX and internal contexts. But the ontological conflict has been transformed into a contextual heterogeneity by way of matching the definitional equations as in [5].

Then, this tax adjusted payment needs to be mediated to the currency of *OFX* context. This requires a *dynamic modifier* to extract the currency value depending on the official currency in the incorporated country of the payee's bank as given below.

$$OFF\_CUR_{OFX} = Currency_{OFX}(payment) \Leftarrow AID = Payee\ Account\ of\ Payment_{INTERNAL} \quad (2)$$

$$BRANCH_{OFX} \Leftarrow Branch\ of\ Account\ AID_{OFX}$$

$$BANK_{OFX} \Leftarrow Bank\ of\ BRANCH_{OFX}$$

$$COUNTRY_{OFX} \Leftarrow country\ of\ Incorporation\ of\ BANK_{OFX}$$

$$OFF\_CUR_{OFX} \Leftarrow official\ currency\ of\ COUNTRY_{OFX}$$

The following logical representation describes how the value of modifier *currency* for *paymentAmount* is obtained for *OFX* context dynamically through the relationships between semantic objects.

```
  modifier(paymentAmt,_O,currency,ofx,M) ⇐    ( attr(_O,paymentRef,Payment),
attr(Payment,payeeAct,Account),
                                attr(Account,location,Location),
                                attr(Location,bank,Bank),
                                attr(Bank,countryIncorporated,Country),
                                attr(Country,officialCurrency,M))).
```

For example the predicate *attr (Payment,payeeAct,Account)* defines the attribute relationship '*payeeAct*' between the *Payment* and *Account* semantic objects. This relation can be mapped to underlying relationships in different contexts as shown in the following logical representation.

```
attr(Payment,payeeAct,PayeeAcct) ⇐
                     ('INTERNAL_PAYMENT_p'(Payment,_,_,_,_,_,PayeeAcct,_).
attr(Payment,payeeAct,PayeeAcct) ⇐
                     ('OFX_PAYMENT_p'(Payment,_,_,_,_,_,PayeeAcct,_).
```

The two statements correspond to how the attribute relation *payeeAcct* has been elevated to two elevation relations with their attributes, mapped in *INTERNAL* and *OFX* contexts.

Account type code - This is represented as heterogeneity in enumerated data types in defining the account type codes in the three contexts. The following summarizes the enumerated data mapping in the three contexts. Since there can be more than two types of financial standards, rather than having mappings between each standard , we adopt a 'Indirect conversion with ontology inference' strategy [13] where we represent the different account types in the ontology itself and providing mapping between the context independent ontology's enumerated type and the context sensitive type codes. The context model would then map each security type context construct into its corresponding security type ontology construct.

Therefore the conversion from *INTERNAL* to *OFX* would be,

$$Account\_type_{OFX}(\ Account\_type_{INTERNAL}(\text{'CHK'})) \Leftarrow ONTOLOGY\_TYPE_{INTERNAL} = \text{'CHKA' [table INTERNAL]} \quad (3)$$

$$ONTOLOGY\_TYPE_{NONE} = \text{'CHKA' [table Ontology]}$$

$$ONTOLOGY\_TYPE_{OFX} = \text{'CHKA' [table OFX]}$$

$$OWN\_TYPE\ (\text{'CHK'})_{OFX} = \text{'CHECKING' [table OFX]}$$

Ontology : Account types Table Ontology	
ONTOLOGY_TYPE	**Description**
CHKA	Checking account
SVGA	Savings account
MNYMRTA	Money Market Account
CRLINEA	Credit Line Account

Mapping between Internal and Ontology - Table INTERNAL	
ONTOLOGY_TYPE	**OWN_TYPE**
CHKA	CHK
SVGA	SVG
MNYMRTA	MNYMRT
CRLINEA	CRLINE

Mapping between OFX and Ontology - Table OFX	
ONTOLOGY_TYPE	**OWN_TYPE**
CHKA	CHECKING
SVGA	SAVINGS
MNYMRTA	MONEYMRKT
CRLINEA	CREDITLINE

Mapping between IFX and Ontology - Table IFX	
ONTOLOGY_TYPE	**OWN_TYPE**
CHKA	DDA
SVGA	SDA
MNYMRTA	MMA
CRLINEA	CDA

4.3. Internal Schema vs. IFX

After looking at some of the interoperability issues between internal context and OFX, now we would delve into the newer standard, IFX, which has more features and detailed representations. Table 3 shows the different types of heterogeneities. The conflicts of *account type*, *date format*, *phone number format* and *currency types* are similar to the OFX scenarios. The new conflicts are the extended conflicts identified in *payment amount* and introduction of *invoice* related conflicts.

Both IFX and OFX handle complex business payment transactions for business customers. This requires incorporating multiple invoice details attached to the payment aggregates when both the biller and customer are business entities. The older OFX provides a basic mechanism of incorporating invoice details like invoice discounts, line items in invoices etc. But the newer IFX extends this by providing more elaborate aggregates constituting different tax schemes as well as fees (late fees, FoRex fees, etc.) that are applicable to invoice.

Mediating Invoice Amount

Each payment can have at least one invoice aggregate that represent the different invoices paid through a particular invoice. In an internal schema the invoice amount might be represented as the net amount, where the taxes and fees would be aggregated when the bill is presented or invoiced. But the IFX context, the Invoice amount consists of the various taxes and fees that could be added to the net amount.

Table 3: Conflict between Internal and IFX contexts

Conflict	Internal Context	IFX Context	Mapped modifier Refer Fig 2)
Payment amount	Net amount	Net + $\sum$ tax amount + $\sum$ Fees	*PaymentScheme*
Payment due date format	European format	US format	*DateFormat*
Payment due date Style	dd/mm/yyyy 03/03/2003	Yyyy-mm-dd 2003-03-03	*DateStyle*
Account type code	SVG,MNYMRT,CRLINE, CHK etc	SDA,MMA,CCA,DDA etc	*AccountCodeScheme*
Invoice Amount	Net amount	Net + $\sum$ tax amount + $\sum$ Fees	InvoicePayment-Scheme
Currency type (Exchange rate)	"GBP"	Currency of country of incorporation of payee bank	*Currency*
Phone number format	415.445.4345	1-415-4454345	*PhoneNumberS-cheme*

The mediation between the two invoice amounts represents an equational ontological conflict (EOC) [5] that would be resolved through introduction of a set of modifiers that would match the two different definitional equations. Each invoice would have multiple fees .i.e. an invoice would have FoRex, late payment fees, import fees as well as multiple taxes like GST, withholding taxes etc

Therefore the relationship between the two definitional equations for invoice amount is:

$$InvoiceAmount_{IFX} = InvoiceAmount_{internal} + \sum (InvoiceAmount_{internal} * FeeRate_{IFX}) + \sum (FixedFee_{IFX}) + \sum (InvoiceAmount_{internal} * TaxRate_{IFX}) + \sum (FixedTax_{IFX}) + \quad (4)$$

Let us say we executed the query '`select INVOICE_AMOUNT from INTERNAL_INVOICE`' in *IFX* context where the relation `INTERNAL_INVOICE`' is defined for *internal* context.

The following shows the mediated SQL query automatically generated by the COIN mediation framework considering all the conflicts associated between *internal* and *IFX* contexts:

```
  select
(internal_invoice.INVOICE_AMOUNT+(((internal_invoice.INVOICE_AMOUNT*ifx_tax_types.AMOUNT)+(internal_invoice.INVOICE_AMOUNT*ifx_tax_types2.AMOUNT))+(ifx_fees_types.AMOUNT+(internal_invoice.INVOICE_AMOUNT*ifx_fees_types2.AMOUNT))))
  from   (select 'GST', TYPE, AMOUNT from   ifx_tax_types
         where  TAX_NAME='GST') ifx_tax_types,
```

```
        (select 'IMPORT', TYPE, AMOUNT from   ifx_tax_types
         where  TAX_NAME='IMPORT') ifx_tax_types2,
        (select 'LATE', TYPE, AMOUNT from   ifx_fees_types
         where  FEES_NAME='LATE') ifx_fees_types,
        (select 'DELIVERY', INVOICE_NO from   ifx_invoice_fees
         where  FEE_NAME='DELIVERY') ifx_invoice_fees,
        (select 'LATE', INVOICE_NO from   ifx_invoice_fees
         where  FEE_NAME='LATE') ifx_invoice_fees2,
        (select 'IMPORT', INVOICE_NO from   ifx_invoice_taxes
         where  TAX_NAME='IMPORT') ifx_invoice_taxes,
        (select 'GST', INVOICE_NO from   ifx_invoice_taxes
         where  TAX_NAME='GST') ifx_invoice_taxes2,
        (select INVOICE_NO, PAYMENT_ID, INVOICE_AMOUNT, DESCR, INVOICE_DATE,
         DISCOUNT_RATE,DISCOUNT_DESC from   internal_invoice) internal_invoice,
        (select 'DELIVERY', TYPE, AMOUNT rom   ifx_fees_types
         where  FEES_NAME='DELIVERY') ifx_fees_types2
where  ifx_invoice_fees.INVOICE_NO = ifx_invoice_fees2.INVOICE_NO
and    ifx_invoice_fees2.INVOICE_NO = ifx_invoice_taxes.INVOICE_NO
and    ifx_invoice_taxes.INVOICE_NO = ifx_invoice_taxes2.INVOICE_NO
and    ifx_invoice_taxes2.INVOICE_NO = internal_invoice.INVOICE_NO
```

Some readers may have so far considered that identifying and resolving semantic heterogeneity is a small matter of handling date formats, currency exchange, and other accounting conventions. We observe now that the net effect and accumulation of such small matters makes the programmer's task impossible. A programmer not equipped with the COIN mediation system must devise and create the above query. A programmer using the COIN mediation system can type the original query: '*select INVOICE_AMOUNT from INTERNAL_INVOICE*' in *IFX* context and rely on COIN to automatically mediate the query. The application gains in clarity of design and code, as well as in scalability. The sharing of domain knowledge, context descriptions, and conversion functions improve the knowledge independence of the programs and their maintainability.

4.4. Some insight to conflicts analysis between internal and SWIFT contexts

The SWIFT protocol is mainly involved in inter-bank cross border transactions. It uses globally unique identifiers for bank code like BIC, BEI. For e.g. the BCI code comprise of concatenation of bank code, country code and location code (defined by ISO 9362), compared to just a bank code representation used in internal schema. This peculiar heterogeneity requires a non-commutative building up of a composite bank identifier when mediating from *internal* to *SWIFT* context. The following represents a logical formula for the mediation for the concatenation. The predicate notations were discussed in a previous example.

```
cvt(noncommutative,bankLoc,O,idType,Ctxt,"single",Vs,"composite",Vt) ⇐
              ('SWIFT_BANK_BCI_p'(BANK, LOC, COUNTRY),       value(BANK,Ctxt,Vs),
               value(LOC,Ctxt,Locc),   value(COUNTRY,Ctxt,Countryc),
(Vtemp is Vs + Locc), (Vt is Vtemp + Countryc))).
```

Usage of sub contexts

Under the *SWIFT* context, depending on whether the transaction is between financial institutions inside the EU or outside, a bank handling fee is credited to the payment amount. This can be modeled using the *sub context* concept of COIN. A sub context derives all the super context based modifier values while having specialized modifier values for extended features. The following logical formulas denote how this can be modeled in COIN

```
is_a(swift_intraEU,swift)
is_a(swift_outsideEU,swift)
```

Then a query like '*select amount from payment'* in *outsideEU* context, called on a relation defined for internal context, is resolved by adding the handling charges on top of the local applicable tax (inherited from *SWIFT* context) as denoted in the following mediated datalog.

```
answer('V15'):-
              'INTERNAL_PAYMENT'('V14', 'V13', 'V12', 'V11', 'V10', 'V9', 'V8', 'V7'),
              'TAX_TYPES'("GST", 'V6', 'V5'),        'V4' is 'V5' * 'V12',
              'V3' is 'V12' + 'V4',   'SWIFT_CHARGE_TYPES'("outsideEU", 'V2', 'V1'),
              'V15' is 'V1' + 'V3'.
```

Note that although datalog and prolog representations are used internally within COIN and shown in this paper, the actual COIN system provides a graphical and user-friendly interface so that data administrators setting up the knowledge representations (e.g., domain models, context) need not know anything about these internal representations.

5. CONCLUSION AND FUTURE WORK

We identified different semantic, ontological heterogeneities that exist in different financial messaging standards. It showed that indeed mediation between these is not a trivial task, yet is critical and important to the globalization of the financial industry. Further we show that an effective answer is to have a mediation service that provides automatic and a transparent mediation without requiring engineering new standards.

We have shown that the COIN approach is capable of mediating the different heterogeneities that exist in different financial standards and internal contexts of Financial Institutions. Our approach in modeling a business domain and mapping different contextual representations and values through a declarative manner demonstrates the extensibility, flexibility and user-friendliness of the COIN framework.

One aspect that is lacking in COIN and that we are currently investigating is the modeling temporal heterogeneities like the examples denoted in table 4. We are currently studying different aspects of temporal heterogeneities which are sources of conflicts among financial standards.

Table 4: Temporal heterogeneities

Property	Internal Schema	OFX	IFX	SWIFT 103/103+
Price	Net	Net + tax of 5% on and efore 2000 , Net + tax of 2% after 000	Net + tax of 5% on and efore 2000 , Net + tax of 2% after 000	(Net + tax of 5% on and efore 2000 , Net + tax of 2% after 000) + inter-bank charges.
Currency	FFR on and before 000, EUR after 2000.	Currency of country of ncorporation of payee ank	Currency of country of ncorporation of payee ank	Explicitly mentioned- ISO 217

ACKNOWLEDGEMENTS

The authors wish to acknowledge the extensive help of Aykut Firat, Hongwei Zhu, Philip Lee and Allen Moulton of MIT. The research reported herein has been supported, in part, by the Singapore-MIT Alliance.

REFERENCES:

[1] A.Firat ."Information Integration using Contextual Knowledge and Ontology Merging", PhD Thesis, MIT,2003

[2] C.H. Goh, S.Bressan.S.Madnick,M.Siegel, "Context Interchange :New Features and Formalisms for the Intelligent Integration of Information", ACM TOIS, vol. 17,pp 270-293,1999.

[3] A.Bressan , C.H. Goh, "Answering Queries In Context", Proceedings of "Flexible Query Answering Systems". Third International Conference, FQAS, 1998, Roskild,Denmark.

[4] S.Madnick,A.Moulton,M.Siegel, "Semantic Interoperability in the Fixed Income Securities Industry: A Knowledge Representation Architecture for dynamic integration of Web-based information", HICSS,Hawai,2003

[5] S.Madnick, A.Firat, B.Grosof, "Knowledge Integration to overcome Ontological Heterogeneity: Challenges from Financial Information Systems",pp. 183-194,ICIS,Barcelona,Spain, 2002

[6] S.Madnick,A.Moulton,M.Siegel, "Context Interchange Mediation for Semantic Interoperability and Dynamic Integration of Autonomous Information Sources in the Fixed Income Securities Industry", (WITS), Barcelona, Spain, December 14-15, 2002, pp.61-66

[7] S.Madnick,S. Bressan, C.H. Goh, T. Lee, and M. Siegel "A Procedure for Mediation of Queries to Sources in Disparate Context", Proceedings of the International Logic Programming Symposium, October 1997

[8] S.Madnick, S. Bressan, C. Goh, N. Levina, A. Shah, M. Siegel ,"Context Knowledge Representation and Reasoning in the Context Interchange System" , *Applied Intelligence: The International Journal of Artificial Intelligence, Neutral Networks, and Complex Problem-Solving Technologies,* Vol 12, Number 2, September 2000, pp. 165-179

[9] Open Financial Exchange Specification OFX 2.0.2, Open Financial Exchange, http://www.ofx.net/ofx/de_spec.asp

[10]Interactive Financial Exchange –IFX version 1.5, IFX Forum, Inc, http://www.ifxforum.org/ifxforum.org/standards/standard.cfm

[11] Society for Worldwide Interbank Financial Telecommunication (S.W.I.F.T), Standard Release 2003, http://www.swift.com/index.cfm?item_id=5029

[12] S.Madnick, A. Firat and M. Siegel, "The Caméléon Web Wrapper Engine", *Proceedings of the VLDB2000 Workshop on Technologies for E-Services,* September 14-15, 2000

[13] S.Madnick, A. Moulton and M. Siegel "Semantic Interoperability in the Securities Industry: Context Interchange Mediation of Semantic Differences in Enumerated Data Types", *Proceedings of the Second International Workshop on Electronic Business Hubs: XML, Metadata, Ontologies, and Business Knowledge on the Web* (WEBH2002), Aix En Provence, France, September 6, 2002

[14] Kuhn, E., Puntigam, F., Elmagarmid A. (1991). Multidatabase Transaction and Query Processing in Logic, Database Transaction Models for Advanced Applications, Morgan Kaufmann Publishers.

[15] Litwin, W., Abdellatif, A. (1987), "An overview of the multi-database manipulation language MDSL". Proceedings of the IEEE, 75(5):621-632.

[16] Goh, C. H. (1997), "Representing and Reasoning about Semantic Conflicts in Heterogeneous Information Systems, MIT Ph.D. Thesis.

[17] Arens, Y., Knoblock, C., Shen, W. (1996). Query Reformulation for Dynamic Information Integration. Journal of Intelligent Information Systems 6(2/3): 99-130.

[18] Batini, C., Lenzerini, M., Navathe, S. B. (1986) "A Comparative Analysis of Methodologies for Database Schema Integration", ACM Computing Surveys 18(4): 323-364.

[19] Landers, T., Rosenberg, R (1982) "An Overview of MULTIBASE", International Symposium on Distributed Data Bases", 153-184

[20] Breitbart, Y., Tieman.L. (1984), "ADDS - Heterogeneous Distributed Database System", Proceedings of the Third International Seminar on Distributed Data Sharing Systems, 7- 24.

[21] Scheuermann, P., Elmagarmid, A. K., Garcia-Molina, H., Manola, F., McLeod, D.,Rosenthal, A., Templeton, M. (1990), "Report on the Workshop on Heterogeneous Database Systems" held at Northwestern University, Evanston, Illinois, December 11-13,

[22] Ahmed, R., De Smedt, P., Du, W., Kent, W., Ketabchi, M., Litwin, W.,, Rafii,A.,,Shan, M. (1991)." The Pegasus Heterogeneous Multidatabase System". IEEE Computer 24(12): 19-27.

[23] Collet, C., Huhns, M. N., Shen, W. (1991), "Resource Integration using a large knowledge base in Carnot", IEE Computer, 24(12):55-63.

[24] Kuhn, E., Ludwig, T. (1988), "VIP-MDBS: a logic multidatabase system", Proceedings of the first international symposium on Databases in parallel and distributed systems, p.190-201, December 05-07, Austin, Texas, USA.

[25] Litwin, W. (1992). "O*SQL: A language for object oriented multidatabase interoperability". In Proceedings of the Conference on IFIP WG2.6 Database Semantics and Interoperable Database Systems

(DE-5) (Lorne, Victoria, Australia), D. K. Hsiao, E. J. Neuhold, and R. Sacks-Davis, Eds. North-Holland Publishing Co., Amsterdam, The Netherlands, 119-138.

[26] Baral, C., Gelfond, M. (1994). "Logic Programming and Knowledge Representation", Journal of Logic Programming, 19,20:73-148.

[27] Kakas, A. C., Michael, A. (1995). “Integrating abductive and constraint logic programming”, To appear in Proc. International Logic Programming Conference.

[28] S. Patel and A. Sheth, "Planning and Optimizing Semantic Information Requests using Domain Modeling and Resource Characteristics," Proceedings of the 6th Intl Conf on Cooperative Information Systems (CoopIS), Trento, Italy, September 5-7, 2001, pp. 135-149.

[ZD04] P. Ziegler and K R. Dittrich. (2004)., “User-Specific Semantic Integration of Heterogeneous Data: The SIRUP Approach”. International Conference on Semantics of a Networked World (IC-SNW 2004), Paris, France, June 17-19, 2004

[W*01] Wache, H., Vogele, T., Visser, U., Stuckenschmidt, H., Schuster, G., Neumann, H., Hubner, S., (2001), “Ontology-Based Integration of Information – A Survey of Existing Approaches”, Proceedings of the IJCAI-01 Workshop on Ontologies and Information Sharing, Seattle, USA, 4 –5 August, 2001.

[L02] M. Lenzerini. (2002). “Data integration: a theoretical perspective”, Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems, pp233 - 246