Industry models for enterprise data management in financial markets

J. G. Carney

The subprime credit crisis of 2007 and the ensuing liquidity crisis in global financial markets highlighted weaknesses in the risk management processes and the information technology infrastructure of many financial market organizations. There are various dimensions to this, but underpinning many of the issues that have arisen is a failure to manage critical data pertaining to financial instruments, counterparties, and liquidity in an integrated and readily accessible fashion. To address this, various industry bodies, vendors, and financial institutions have considered the adoption of industry data models and classification standards for financial instruments and other business entities. In this paper, we describe the industry models that IBM has developed for the financial market industry for this purpose. We illustrate how these models can be used to successfully underpin an enterprise data management (EDM) infrastructure in a financial market organization, thereby addressing some of the key issues arising from the crisis as they pertain to the integrity and management of critical enterprise data. We draw comparisons to related EDM technologies and to the data standardization efforts of the EDM Council, a leading industry body for the financial market industry.

Introduction

The subprime credit crisis of 2007 and the ensuing liquidity crisis in global financial markets have created a level of uncertainty and turbulence in the world economy unseen since the 1930s. The circumstances and events that led to this crisis are now generally well understood and documented in the literature. These include the search for yield enhancement in a low-interest-rate environment, poor investment management, lax underwriting standards, rating agency incentive problems, poor risk management by financial institutions, the lack of market transparency, the complexity of financial instruments, and the failure of regulators to understand the implications of the changing environment for the financial system [1–3]. In addition, in some cases, financial institutions paid insufficient attention to the inherent risks relating to balance sheet growth and were overconfident in their risk management and risk control mechanisms [4].

The financial institutions that survived the crisis, through either industry consolidation or government assistance, are now considering the steps they need to take in order to transform their risk management practices and supporting infrastructure. Although there is no consensus view in the industry on how to best approach this, there are some studies that offer useful recommendations relevant to many financial institutions (see, e.g., [1] and [4]). One recurring theme in many such studies is that financial institutions should adopt a more robust approach to enterprise risk management (ERM), which includes the sharing of quantitative and qualitative information in risk management committees that span all major lines of business. This may seem obvious, but it is in fact very difficult to implement in practice for many financial institutions. There are various reasons for this, but as we will explain in this paper, the absence of a robust enterprise data management (EDM) infrastructure that can guarantee the integrity of critical enterprise data and ensure its timely distribution to senior decision makers often plays a major role. Although there is nothing new in these observations (ERM and by extension EDM has been a goal or aspiration of the financial market industry since the early 1990s), the financial crisis has created a stronger impetus than ever to transform how data are managed across this industry. For example, there

Digital Object Identifier: 10.1147/JRD.2009.2038750

0018-8646/10/$5.00 © 2010 IBM

is early evidence that regulatory compliance requirements will change and become more rigorously enforced in some major jurisdictions, including the United States and the United Kingdom [5, 6]. Such developments will inevitably exert pressure on existing risk processes and supporting information technology (IT) infrastructure in many financial institutions, particularly those that have grown their various lines of business to date as vertical "silos of risk."

The main contribution of this paper is to highlight how industry models and related data standardization initiatives can play an important role in addressing some of the key challenges this industry is now facing. We start this paper by briefly reviewing some popular enterprise architectural frameworks that can be used to support EDM and summarize the evolution of the IBM information framework (IFW) and industry models. We then review some of the key information management challenges that the financial market industry has grappled with since the 1990s before summarizing the work of the EDM Council, a leading industry body engaged in data standardization initiatives for the financial market industry. The second half of this paper is focused on describing the details of how the industry models developed by IBM can be used to underpin the EDM infrastructure of a financial market organization. Here, we emphasize the importance of maintaining an enterprise business vocabulary and conceptual model, illustrating how the IBM models and supporting information management software can be used to support the generation, management, and deployment of such EDM assets across a financial market enterprise. The final sections draw comparisons between the industry models and related technologies, including the work of the EDM Council.

Frameworks for enterprise information management

In 1987, Zachman published a paper that defined "a framework for information systems architecture" [7]. This involved creating a descriptive framework from disciplines quite independent of information systems and then specifying an information system architecture based on this neutral framework. This became known as the "Zachman Framework" and has been extended [8] and studied in various contexts over the years (e.g., see [9] and [10]).

In modern-day terminology, the Zachman Framework is an "ontology," that is, a formal representation of a set of concepts within a domain that includes a description of the relationships between those concepts [10]. Zachman derived this ontology from analogous structures found in the disciplines of architecture for construction and engineering for manufacturing that classify and organize the design artifacts created in the process of producing complex physical products. It uses a 2-D classification model based on six basic interrogatives (what, how, where, who, when, and why) intersecting six distinct perspectives, which relate to stakeholder groups (planner, owner, designer, builder, implementer, and worker). The intersecting cells of the Zachman Framework correspond to models that, if documented, can provide a holistic view of the enterprise.

Since the early 1990s, a number of architectural frameworks and methodologies based on (or similar to) Zachman have been developed; some examples include the Open Group Architecture Framework [11], the Federal Enterprise Architecture [12], and the IBM IFW [13]. Although originally based on the Zachman Framework, IFW differs in several important respects. For example, at a fundamental level, IFW uses the analogy of a "city plan," whereas Zachman uses the analogies of classic building architecture and aircraft manufacturing to help define an information system. Over time, the experience of developing models and a methodology to support IFW in the field suggested that this alternative view provided a more effective way to gradually develop a complete "city of information," including information about individual applications and systems from other types of projects such as business process engineering [13]. Whereas the Zachman Framework is focused on information *system* architecture, IFW is focused on the management of *information* and, thus, is broader in scope.

Overall structure and characteristics of IFW

A detailed description of IFW and its benefits can be found in [13]. In this section, we summarize its overall structure and key characteristics in order to provide some background on the origins of IBM Industry Models.

IFW consists of a number of cells organized into a grid structure. As illustrated in **Figure 1**, these cells are organized into five rows and ten columns. The following summarizes the roles of the various components that underpin IFW:

- *Views*—The IFW columns are grouped into views that represent the different perspectives of the groups within an enterprise that use the framework, for example, strategists and managers (organization view), business analysts and designers (business view), or technical architects and builders (technical view).
- *Columns*—The columns of IFW represent the various ways to describe different types of information. This isolates one abstraction of the subject from other abstractions and, for example, allows information relating to critical business data to be analyzed separately from information relating to critical business processes.
- *Rows*—The rows of IFW show information at different levels of constraint that satisfy different purposes and objectives. In IFW, these include classification hierarchies, entity relationship models, organization diagrams, business strategy models, workflow models, and process dependency diagrams. Each row is labeled with a letter, where "A" represents the highest level of abstraction (e.g., a business concept of interest to the

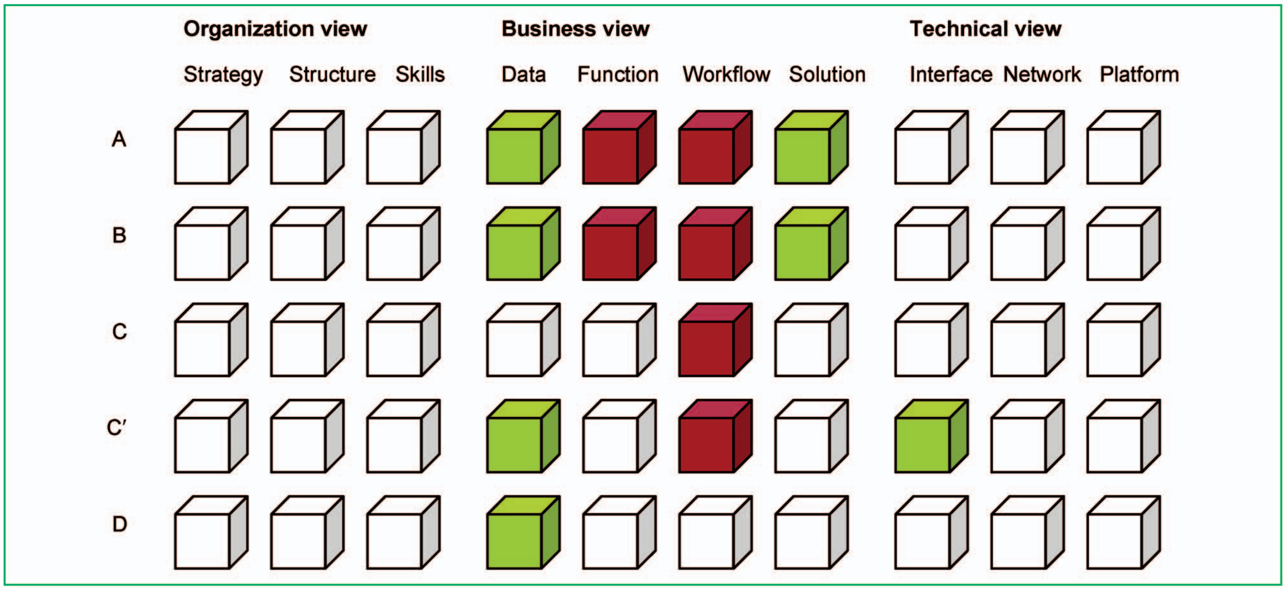

Figure 1

IFW cells populated with industry models for the financial market industry. Models that directly support EDM correspond to the green cells, and models relevant to business process management and service-oriented architecture correspond to the red cells. Note that the focus of this paper is on the models that support EDM.

enterprise) and "D" represents the lowest level of abstraction (e.g., a physical data model ready for deployment on a database management system).

- *Cells*—The cells of IFW represent a particular type of information defined at a specified level of constraint, e.g., information about data represented as an entity relationship model.
- *Dimensions*—The dimensions of IFW represent various architectural perspectives. The first three dimensions are the columns, rows, and cells of the framework. The fourth dimension represents the transition from one version of IFW to a subsequent version. The fifth dimension describes the "ownership" of content in a particular cell, and the sixth dimension describes project "route-map" views through IFW to assist in implementation.

The overall structure of IFW, its dimensions, and the ordering of the columns are designed to allow for *reuse of information* in any of the IFW cells. This is an important point: IFW can be used to analyze a piece of information (not necessarily data), break this information into its constituent parts, and then position these individual parts in the appropriate IFW cells. For example, information on a financial market transaction that involves a particular counterparty can be partitioned into information comprising the counterparty's identification information (*data* information) and information pertaining to the activities involved in executing the transaction (*workflow* information). Breaking information into these components makes it easier to define such information components once and *standardize* them across the enterprise.

Standardizing information and general business concepts in this way is an important first step in enabling the successful implementation of an enterprise architecture and operational EDM infrastructure.

From IFW to industry models

Possibly the most distinguishing feature of IFW versus alternative frameworks such as the Zachman Framework is that many of the IFW cells have been populated with business content representing business knowledge that is common to all organizations in a particular industry. This business content is described by a set of "industry models" represented in various forms depending on the IFW row to which it pertains. For example, the cell corresponding to the C′ row of the data column in the business view of IFW consists of a logical data model described by an entity relationship diagram.

In partnership with customers, IBM has developed industry models that populate the IFW cells for several industries, including banking, telecommunications, retail, and (most recently) financial markets. These are collectively branded as the "IBM Industry Models." The specific IFW cells that are populated by the various industry models vary for each industry. This is driven by a number of factors but primarily by the demand in that industry for a particular set of models. The color coding in Figure 1 summarizes which IFW cells have been populated for the financial market industry.

The overall goal of the IBM Industry Models is to capture the *information* requirements of an industry that are common to all participants in that industry. The extent to which there are common information requirements varies depending on the industry, but in financial services, regulatory compliance can be a key driver, particularly as it relates to data retention and reporting. Once these commonalities have been captured and modeled, they can then be used as an industry baseline to accelerate the development of systems or other projects that leverage the models. These may range from process transformation projects for process rationalization to data warehousing projects supporting regulatory compliance [14].

Evolution of the IBM Industry Models

Since their first incarnation in the early 1990s, the IBM models have matured significantly. With more than 600 customers across six industries, their impact in the field of enterprise architecture is notable. Experiences with customers over the years have also influenced the technical evolution of the models; the role of IFW, the terminology used to define the various model components, and the tools used to manipulate and customize the models have all changed.

Some of the more significant changes relate to tooling support and the role of IFW. Although IFW still exists in the background, and model components can be associated with IFW cells, these components are not usually described in terms of IFW rows or columns. For example, rather than describing a model as a C′-level model in the data column of the IFW business view, it is described as a "design model." Clearly, this alternative terminology is more user friendly and, in most cases, does not require prior understanding of IFW to understand its context.

At the time of writing this paper, the toolset used to manipulate the models is also in transition from legacy IBM software tools (called "*m1*" and "*mmm*," depending on the industry) to the IBM Rational standard toolset. With this, some substantive changes to the models and how they interrelate have also been introduced. In **Figure 2**, we highlight some of the changes in terminology and model components and relate these to IFW cells. The details of these model components and their context relative to each other are explained in the subsequent sections of this paper.

Before describing in detail the new industry models IBM has developed for the financial market industry, we first attempt to provide some context by considering some of the pertinent information management issues that dominate this industry today.

Information management in financial markets today

The management and exchange of critical business information is fundamental to the operation of modern financial markets. However, despite the investment the industry makes in IT, many firms are still challenged by the complexities of managing information in this industry. Indeed, the subprime credit crisis of 2007 and the ensuing liquidity crisis have put this issue into sharp focus for the whole industry.

There are many reasons why firms in the financial market industry have struggled with their EDM (and by extension ERM) infrastructure over the last number of years. However, there are three recurring themes in any discussion on this subject.

1. *Organizational misalignment*—Given the volatile nature and the rapid pace of modern financial markets, this industry has been characterized by a continuous run of mergers, acquisitions, and internal restructurings over the past 20 years. This inevitably places a strain on any enterprise infrastructure. To exacerbate this, there is also a culture of internal competition in the majority of firms that has led to an organizational structure that consists of various separate business units that operate as independent "silos of risk." The subprime crisis of 2007 and the subsequent worldwide financial crisis have highlighted some of the shortcomings of this approach, with many organizations suffering large exposures to the same counterparty across various business units without prior knowledge and that in aggregate exceeded their exposure limits [1]. There is no easy solution to this organizational misalignment, and as result, most firms have attempted to utilize existing organizational and IT funding structures. This has led to IT infrastructures that are genuinely complex, expensive to maintain, and in most cases inadequate as a tool to support enterprise decision making.
2. *Nonstandard data*—A significant proportion of critical business information in financial markets today is nonstandard and noncomparable. This is because the supply chain of information in financial markets is fragmented [15]; that is, information may start its journey in a precise and well-defined form, but as it is manipulated and relabeled by various suppliers and internal systems, it often becomes noncomparable, even among different lines of business in the same firm. Most firms and regulatory authorities in the industry recognize this as a significant problem and have recently made effort through various industry bodies such as the EDM Council to find appropriate industry-wide solutions. Most of this work has focused on developing precise industry standard identifiers for financial instruments, business entities, and data attributes and is still at an early stage of maturity. The goal is that these identifiers would go beyond existing coding schemes, such as the international security identifier code, International Securities Identification Number, and other International Organization for Standardization (ISO) efforts, to

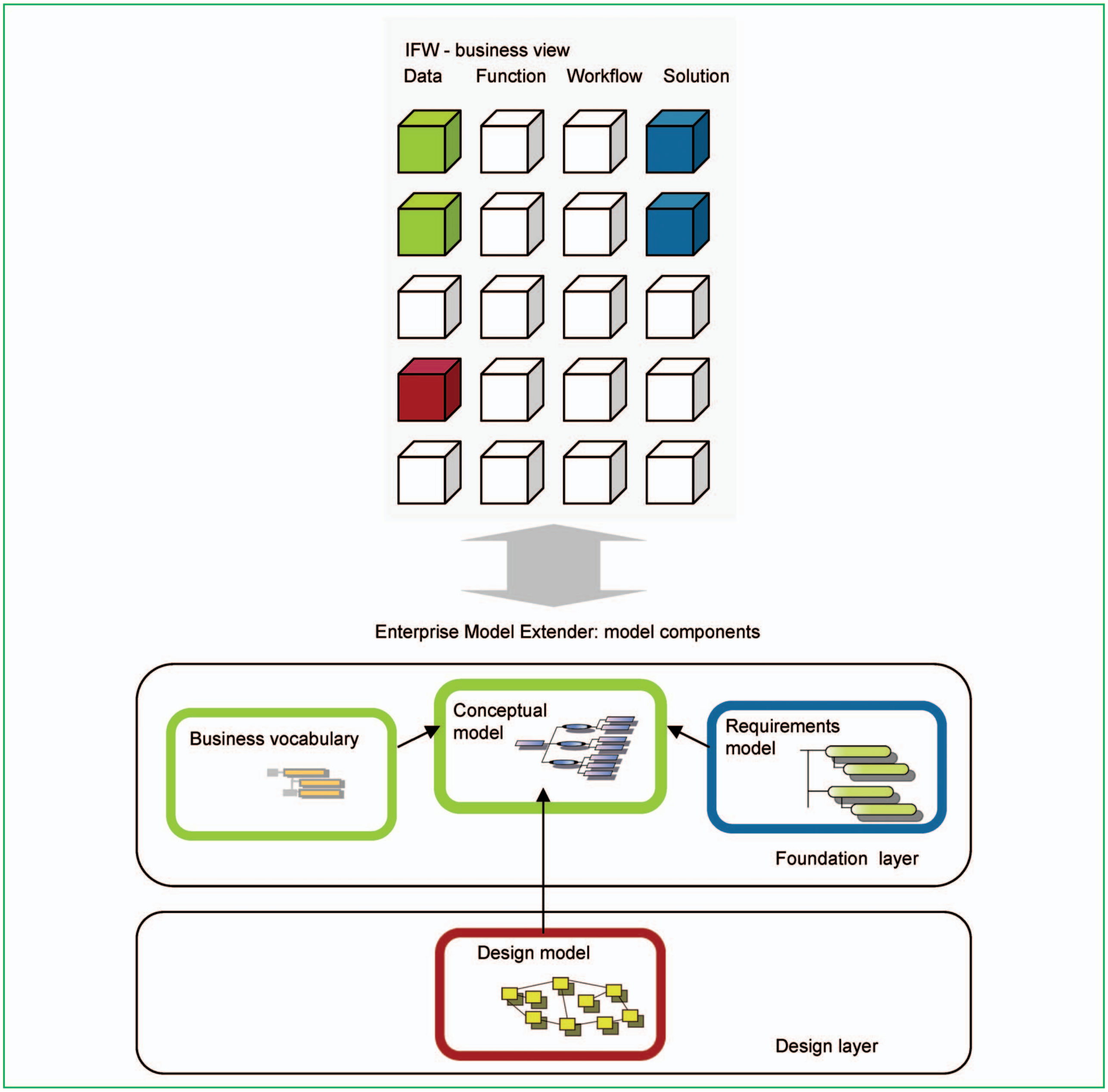

Figure 2

Industry models developed by IBM to support EDM in financial markets. These are available in the 2009 version of the IBM Infosphere Data Architect software tool. Color coding highlights the nearest equivalent IFW cells.

ultimately encompass *all* instruments and business or corporate entities involved in financial market transactions. We describe this work in subsequent sections of this paper and consider how the IBM Industry Models can play a role here.

3. *Market speed and a tactical mindset*—The Tabb Group estimates that some brokers during peak trading conditions can be bombarded with as many as 150 trades/second and more than ten times as many orders per second [16]. This imposes significant pressure on data management platforms that are used to disseminate and store these data and, in some cases, also represents a significant business risk that the broker may not be able to calculate trading limits as fast as its clients are placing orders. In this

environment, it is easy to see how many firms focus on tactical initiatives that discharge immediate business risks at the expense of strategic transformational initiatives such as enterprise-wide information management. This is reinforced by the short-term performance culture that pervades many firms in the industry that often discourages senior decision makers from championing infrastructure investments that may take several years to pay off.

There are other reasons why many financial markets firms have struggled to implement a robust EDM platform to support their business and ERM functions. However, most stem from, or relate in some way to, the three reasons summarized above.

An important point to note here is that although the vast majority of firms recognize the importance of EDM and have made an effort to build an EDM infrastructure [15], there is little evidence to suggest this extends to an enterprise *information* infrastructure (as defined by IFW). In this scenario, all critical business information is managed on an enterprise-wide basis, not just *data*. This would include critical business process flows (e.g., for trade settlement) with embedded data flows that leverage standardized enterprise data and business terms. If optimized appropriately and deployed globally, then such processes could make an important contribution to a firm's effort in moving toward the "mature stage" of ERM adoption [17]. As we now describe, a good first step in achieving this goal is to develop what the EDM Council calls a "Semantic Repository."

Semantic Repository for financial markets

In May 2008, the EDM Council authorized and funded a project to standardize the terms and definitions of reference data attributes for financial markets. This is being managed as an open resource for the industry, and a Web-based mechanism has been developed to support participation by industry practitioners.

The goals of the Semantic Repository (as defined by the EDM Council itself [18]) are threefold:

1. To serve as a *common data dictionary* for internal mapping within financial institutions.
2. To serve as a *common language for data vendors* by reducing the amount of mapping required for both inbound and outbound data transfer.
3. To provide a *common set of tags for source markup*. The long-term goal here is to tag data attributes at the point of origination and reduce the need for multiple transformations at every stage of the data manufacturing process.

The repository is structured to consist of two key components: 1) a conceptual model and 2) a set of terms and definitions. The conceptual model serves as a template for the repository and is intended to facilitate communication for data analysis and establish a consistent naming scheme. It is divided into a number of domains, each of which has a diagram that describes the model in Web ontology language. These domains and diagrams can be viewed and navigated via a Web interface. Spreadsheet and Unified Modeling Language representations of the model can also be downloaded for manipulation by tools that support these formats.

The primary focus of the Semantic Repository to date has been to standardize data pertaining to financial instruments. As such, the repository has been designed to follow the ISO 10962 Classification of Financial Instruments (CFI) standard [19], and many of the terms and definitions for financial instruments were directly taken from various ISO, Depository Trust Company, and other financial institution data dictionaries. These terms and definitions can also be viewed and navigated via a Web interface and downloaded as spreadsheets.

Full details on the content and operation of this repository are provided by the EDM Council [18]. The primary relevance of the repository to this paper is how its conceptual model and terms and definitions relate to similar model constructs in the industry models developed by IBM. The overall goals and purpose of the repository versus the IBM models are also relevant to this paper.

Industry models for financial markets

In this section, we describe some of the pertinent details of the industry models IBM has developed for the financial market industry. We start by describing the various architectural levels of abstraction of the models as they are implemented in the IBM data-modeling toolset and then describe the key characteristics of each model component. To illustrate how the various model components implement business content and interact with each other, we describe key features of financial market content developed to support the ISO CFI standard [19] and Basel II regulatory compliance disclosure requirements [20].

Note that in the interest of controlling the scope of this paper, the focus of this section is on data models only. A detailed examination of the IBM process models developed for financial markets merits a separate paper.

Levels of abstraction

In 2008, an extension to the IBM Infosphere Data Architect software application (part of the IBM Rational toolset) was launched, which is called Enterprise Model Extender. This software implements a meta-model that describes the overall architecture of the IBM models and mapping rules that enforce how each model artifact interacts with other model artifacts. The primary purpose of this software is to enable end-users to extend or customize the IBM data models according to the specific requirements of their enterprise in an environment that is both user friendly and architecturally consistent.

Since the launch of the Enterprise Model Extender, the IBM models have been described in terms of three levels, or "layers," of abstraction: 1) the foundation layer; 2) the analysis layer; and 3) the design layer. These layers are analogous to the rows of IFW. At the time of this writing, IBM had developed data models for financial markets that populate two of these layers: 1) the foundation layer and 2) the design layer (see Figure 2 for an illustration). The purpose and characteristics of each layer are as follows:

- *Foundation layer*—This layer represents a high-level repository of knowledge about an industry (or enterprise) and operates at the same level as the EDM Council's Semantic Repository. Model artifacts that are supported in this layer for financial markets include business concepts, business concept relationships, business concept classifiers, and business concept descriptors.
- *Analysis layer*—This layer represents an abstract specification of how conceptual elements in the foundation layer are supported, independent of any specific design. Model artifacts that normally reside in this layer include entities, attributes, classes, use cases, relationships, generalizations, methods, and service definitions. To use the example of a data model, in the analysis layer, this might be described by an entity relationship diagram that contains no attributes to indicate the ultimate physical manifestation of the model, e.g., the model might ultimately be used to build a data warehouse for the storage of historical data or an operational data store for the storage of real-time operational data. The primary application of models that reside in this layer is to serve as a logical integration point between models in the design layer that have the detailed attribute and other information required for the development of specific solution designs.
- *Design layer*—This layer represents the various aspects of a specific solution design, defining the manner in which the analysis elements are supported. Model artifacts that are supported in this layer for financial markets include entities, attributes, entity relationships, and entity generalizations. To use the example of a data model, in the design layer, this may be described by an entity relationship diagram that contains detailed attribute information to support the ultimate physical implementation of the model, for example, attributes required to support the storage and summary of large amounts of historical data.

It is important to note that these layers do not operate in isolation. There are interfaces defined that integrate the layers.

- *From analysis to foundation*—In implementation scenarios where an analysis model has been developed, typed mappings are provided that specify the conceptual model element that is being supported by an analysis model element. This reflects the idea that an analysis model entity models a data structure that carries data-related instances of the concept.
- *From design to foundation*—In the absence of an analysis model, typed mappings are provided that specify the conceptual model element that is being supported by a design model element. This reflects the idea that a design model entity models a data structure that carries a data-related instance of the concept but in a particular implementation context, e.g., data warehousing.
- *From design to analysis*—Where there is an analysis model, typed mappings are provided from design model elements to analysis model elements in order to support data traceability and model consistency.

As previously noted, the models that are supported vary for each industry. This is primarily driven by the demand for particular models in that industry. As illustrated in Figure 2, a conceptual model, a business vocabulary, a requirements model, and a design model have been developed for the financial market industry. The following three sections describe the key characteristics and usage of these models.

Conceptual model and business vocabulary

In this section, we describe the key characteristics and goals of the IBM conceptual model and business vocabulary. To provide a practical example of how they operate, we use the example of new model content developed to support the ISO CFI standard (10962). This standard is also implemented in the EDM Council's Semantic Repository.

The conceptual model is an *ontology* that describes business concepts and business concept relationships of interest to an enterprise. Although the financial market content in this model is relatively new (first released in 2008), the content in the model relating to other more generic business entities e.g., accounting, arrangements, involved parties, etc.) is very mature and the product of many person–years of IFW development [13]. The 2009 version of the conceptual model supports more than 4,100 business concepts and 900 classifiers spanning banking and financial market concepts.

The model artifacts supported in the conceptual model are the following:

- *Business concepts*—These are business terms that describe an area of knowledge for the business. Business concepts are organized into hierarchies and based on different criteria through a classification mechanism. At the top of a hierarchy, a *core concept* resides that does not belong itself to the classification of any other (higher-level) business concept. As illustrated in **Figure 3**, there are nine core concepts in the model spanning business areas of fundamental importance to banking and financial markets: 1) Arrangement;

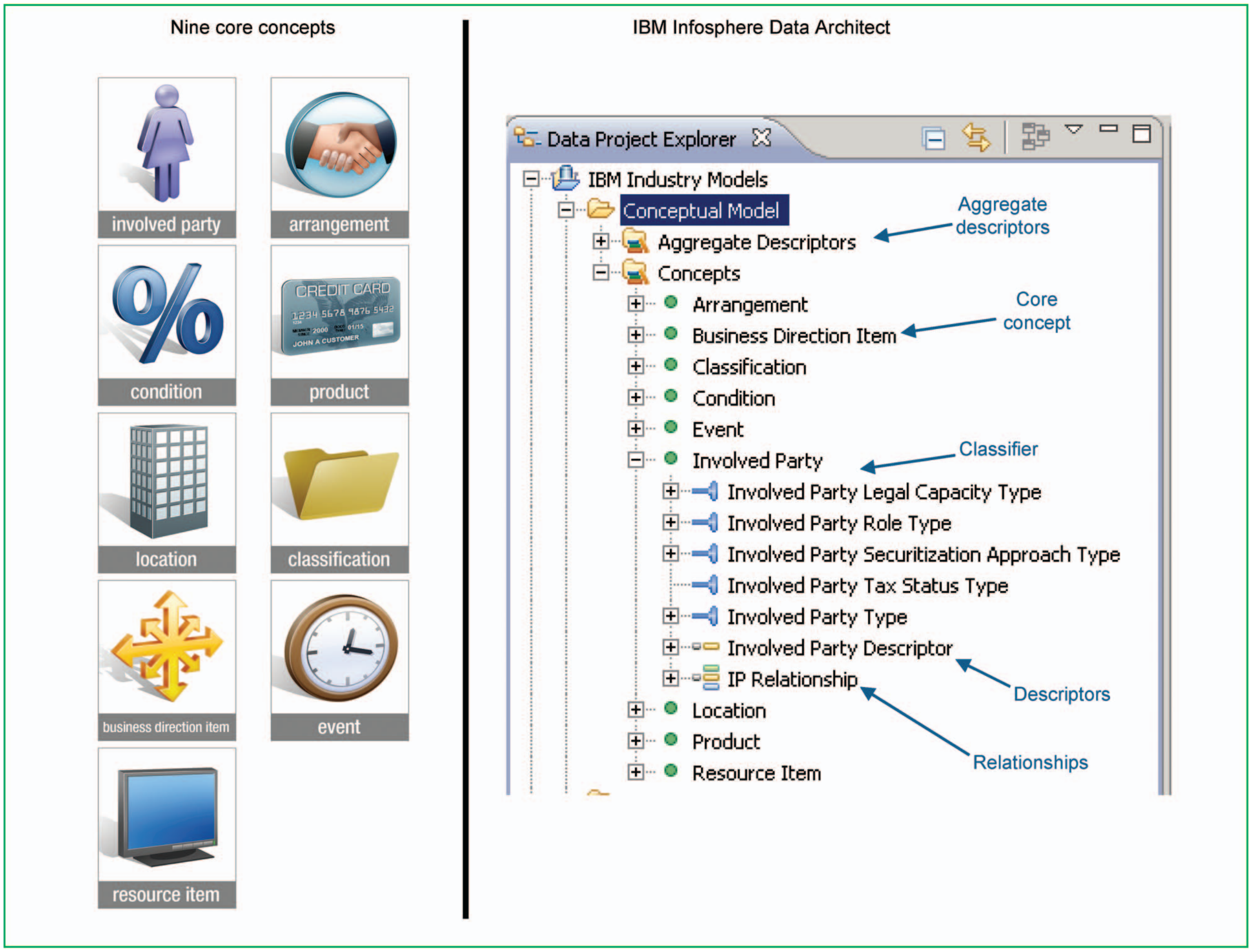

Figure 3

Nine core concepts in the conceptual model and a screenshot of their implementation in the IBM Infosphere Data Architect software tool. Note that all conceptual model artifacts, including core concepts, expand into fully populated hierarchies. The content developed to support the ISO CFI standard represents a subset of this conceptual model.

2) Business Direction Item; 3) Classification; 4) Condition; 5) Event; 6) Involved Party; 7) Location; 8) Product; and 9) Resource Item. The origin of these can be traced to an early version of IFW [13].

- *Classifiers*—Classifiers define the way a particular set of business concepts can be divided into logically related subsets of concepts using a given criteria. For example, the set of Persons can be broken into subsets of Female and Male according to the criteria of gender, so gender is one of the many possible classifiers of Person.
- *Relationships*—These define relationships between two business concepts. A relationship is itself a business concept and can therefore also be classified.
- *Descriptors*—These are business concepts that describe other business concepts on which they fully depend. For example, an involved party may be described by a name and an address. Descriptors can represent numeric or nonnumeric data.
- *Aggregate descriptors*—These are numeric quantities defined using an expression on one or more measurable properties or other aggregate descriptors. The expression used to define an aggregate descriptor has to be an aggregation expression, for example, "sum" or "average."

Figure 4 summarizes new conceptual model content developed to support the ISO CFI code and illustrates how business concepts and classifiers can be used to capture business knowledge. This code represents the detailed characteristics of a financial instrument as a six-letter code [19]. In the example provided here, an ordinary share with enhanced voting rights and other CFI attributes is represented by the code "ESEUFR." Each letter of this code is

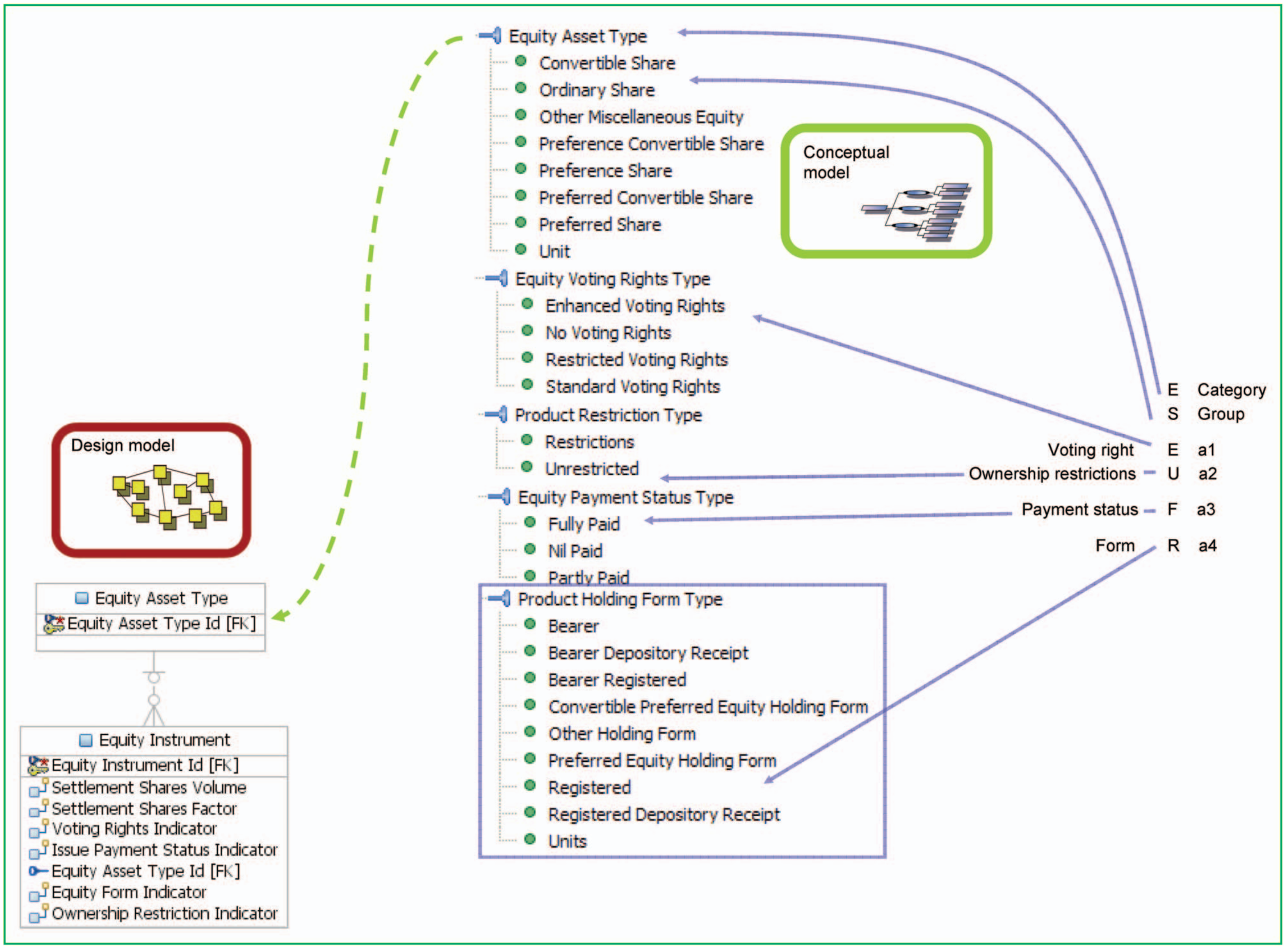

Figure 4

This diagram should be read from right to left. Here, we illustrate how the six-letter CFI code is modeled by the conceptual model as classifiers and business concepts. Classifiers are used to model the instrument category, and business concepts are used to model the instrument group and attributes. We also illustrate how the conceptual model hierarchies are supported by entities and attributes in the design model. Here, a classification entity is used to model the instrument category and instrument group, and indicator attributes implemented in a fundamental entity with foreign key connections to the classification entity are used to model the CFI attributes.

implemented in the conceptual model as a business concept (or a classifier if it is the first letter). The significance of implementing the CFI code in the conceptual model in this way is that it enables the conceptual model to serve as the enterprise-wide business definition of the CFI code for *all* downstream data repositories (and process models). For example, if the enterprise is building a data warehouse for regulatory compliance and an operational security master for a single view of securities traded, then design models can be defined that inherit the same hierarchical structure and definitions as defined in the conceptual model, thus standardizing how financial instruments are represented in both repositories. The IBM Infosphere Data Architect software tool has specific support for this functionality, that is, typed mappings can be defined that specify the conceptual model element that is being supported by a design model element.

Relationships in the conceptual model can also play an important role in supporting the standardization of data across an enterprise. For example, the different ways involved parties are associated with products are defined as relationships in the conceptual model and can be applied across an enterprise. Associations such as "Involved Party is Market Maker for Product" and "Involved Party Trades Product" have been implemented for financial markets. Relationships that describe how the involved parties relate to each other are also important and are defined in the conceptual model. Associations such as "Involved Party is Customer of Involved Party" and "Involved Party is Broker for Involved Party" have also been implemented for financial markets. As above for the CFI

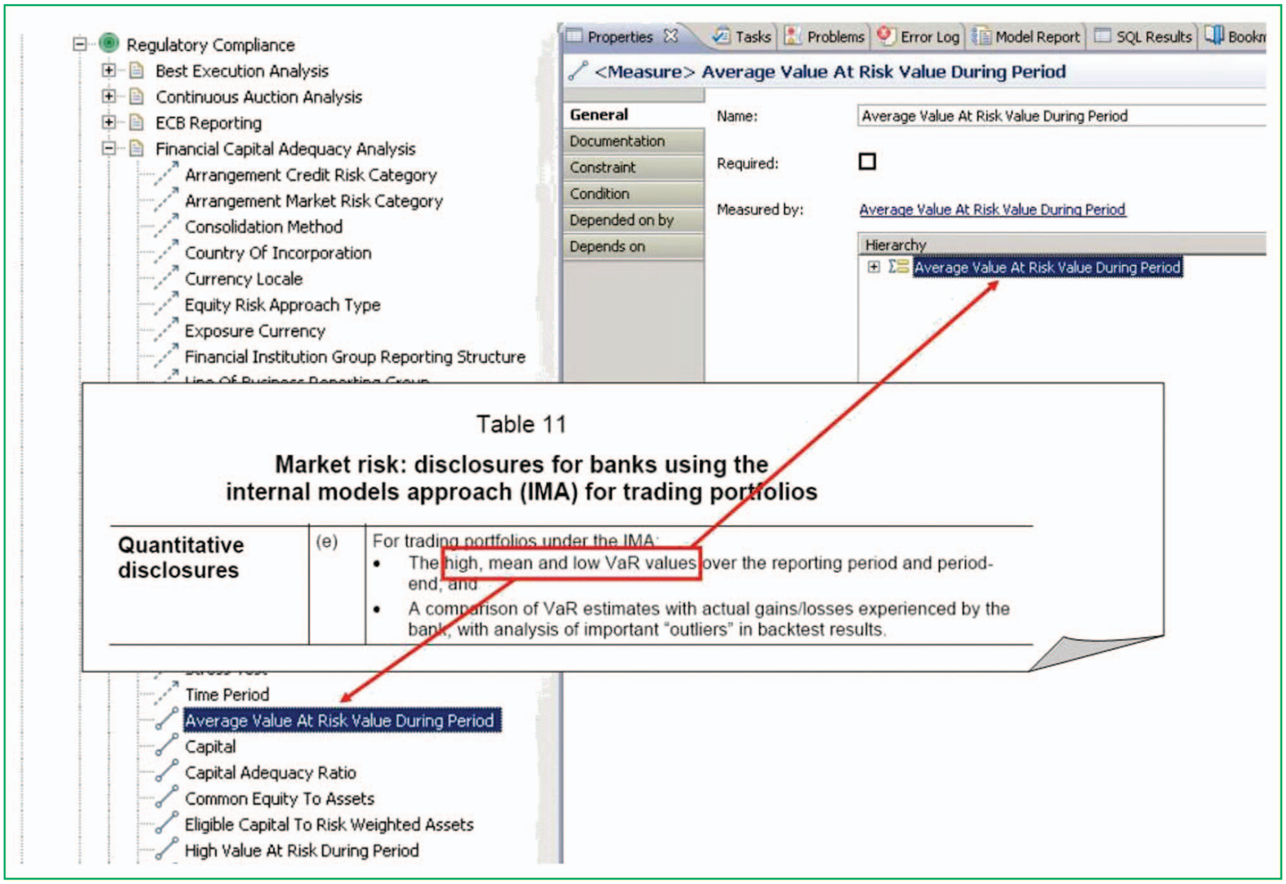

Figure 5

Example of how the requirements model is used to define the market risk regulatory reporting requirements of Basel II in the IBM Infosphere Data Architect software tool. The highlighted measure "Average Value At Risk Value During Period" is measured by (i.e., maps to) an aggregate descriptor in the Conceptual Model with the same name, where the full details of the aggregation of value at risk values for a portfolio over a reporting period is defined. Note that this represents a description of the implementation of only one measure in one analytical requirement template. In total, more than 1,200 measures across 70 templates have been developed relevant to financial markets.

content, the significance of this is that the conceptual model definitions for business concept relationships can serve as the enterprise wide business definition for such relationships across *all* downstream data repositories and process models.

Descriptors are used to define the characteristics of business concepts and can be in numeric or nonnumeric form. For example, an involved party "name" would be an example of a nonnumeric descriptive property, whereas an involved party "current assets" would be a numeric measurable property. Measurable properties can be aggregated into "aggregate descriptors" to define enterprise wide aggregations of numeric data that are important to the enterprise, for example, key performance indicators or risk measures aggregated over time.

The business terms defined across the conceptual model (and requirements model below) can be separated from the remainder of the conceptual model and deployed as a separate business vocabulary. This has specific support in IBM Infosphere Data Architect. The business vocabulary is represented as a separate hierarchy and can also be exported to the IBM Infosphere business glossary software application for further manipulation and deployment in a production environment.

Requirements model and design model

The requirement model supports the management of analytical requirements expressed in the form of groupings of measures and dimensions. These requirements may be for business intelligence or regulatory reporting purposes, for example. The purpose of the requirement model is to serve as an enterprise-wide repository of critical measures and dimensions, where each measure, dimension, aggregation, and business term defined in the requirement model complies with the semantic and business vocabulary standards defined in the conceptual model and business vocabulary described above.

Figure 5 illustrates how the requirement model works with a practical example in the area of regulatory compliance reporting. Here, we show how the quantitative disclosure requirements of Basel II for market risk as defined in Table 11 of the Basel II Comprehensive Document [20] are represented as measures and dimensions in the "Financial Capital Adequacy Analysis" analytical requirement template. Each measurable quantity, in this case the mean (average) value at risk for a portfolio over a measurement period, is captured by a measure. The requirement to report these measures over different reporting periods is also captured by a dimension. It is important to note here that the requirement model does not merely provide a mechanism to collect requirements in a convenient and standardized fashion; rather, all measures and dimensions have typed mappings to appropriate aggregators, descriptors (measures), and classifiers (dimensions) in the conceptual model. If measures are aggregates of atomic level data, e.g., sum, average, etc., then they map to aggregate descriptors that define the aggregation in precise terms. Likewise, all design model elements (entities and attributes primarily) have typed mappings to the conceptual model; this process "closes the loop" in terms of following an audit trail of data from reporting requirements on the front-end to low-level atomic data stored in a data warehouse or operational data store.

As previously described, the design model represents a specific solution design, i.e., a model optimized for operational data storage (e.g., a securities master) or historical data storage (e.g., a data warehouse). The 2009 version of the design model developed for financial markets consists of a full set of logical and physical structures for warehouse design, consisting of almost 1,000 entities and more than 5,500 attributes. The structure of this model is subject oriented and fully aligned in terms of hierarchy, classifiers, and business terms with the conceptual model, requirement model, and business vocabulary.

Comparisons with the EDM Council's Semantic Repository

As previously described, in 2008, the EDM Council authorized and funded a project to develop a Semantic Repository of standardized terms and definitions of reference data attributes for financial markets. This is being managed as an open resource for the industry, and a web-based mechanism has been developed to support participation by industry practitioners. The following are some of the key differences between the industry models IBM has developed for the financial market industry and the EDM Council's Semantic Repository:

- *Development approach*—As described above, a key goal of the EDM Council's Semantic Repository is to establish a common set of standards across the financial market industry that can be leveraged by all market participants, including financial institutions, vendors, and market infrastructure providers. The route to achieving this involves all major industry participants engaging in a process facilitated by the EDM Council to mutually define and agree these standards. The IBM approach is different in that an IBM development team collects information that is common across an industry (e.g., reporting requirements for regulatory compliance and ISO standards), and then a process of iterative development continually refines and expands this content in partnership with customers.
- *Corporate versus industry standards*—Although the IBM models could become the de facto standard for certain applications in an industry if a critical mass of organizations use them, their purpose is not to standardize at the level of an entire industry but rather to provide the basis for defining corporate standards [10]. This is in contrast with the EDM Council's Semantic Repository, which aims to provide the basis for industry-wide standards.
- *Scope of models*—As detailed above, the scope of the IBM models extends beyond a conceptual model and business vocabulary. The 2009 financial market release of the models (as deployed in the IBM Infosphere Data Architect software) includes a conceptual model, a requirement model, a business vocabulary, and a design model. The architectural basis for more models is already established and bounded only by the "layers of abstraction" (foundation, analysis, and design) described previously.
- *Supporting tools and deployment software*—The IBM models can be manipulated and customized using standard IBM tooling. From here, the design model can be transformed from logical to physical and deployed on a standard database management system. The business vocabulary can also be exported to meta-data management platforms such as the IBM Information Server and leveraged by run-time applications across the enterprise. Although it is conceivable that model content from the EDM Council's Semantic Repository could be deployed by enterprises in a similar fashion, the exact mechanism by which this would be enabled is not yet clear.

The differences and similarities summarized here raise the question of whether or how the EDM Semantic Repository could work in tandem with the IBM models and supporting IBM tools and software. One obvious option worth considering here is that, in a scenario where industry-wide standards have been firmly agreed via the EDM Council process, organizations might consider harvesting the EDM Council content and integrating it with the IBM industry model content not supported by the EDM Council (e.g., the design model and requirements model) using IBM supporting tools and software. This approach would enable enterprises to leverage industry-wide standards agreed through the EDM Council process in the same development and runtime deployment environment as the *corporate*-wide standards developed through customizations of the IBM models. Clearly, there are many other scenarios worth exploring here and issues to consider; these are the subject of future work.

Comment on related EDM technologies

There are many software vendors and consulting firms operating in the financial market industry that offer solutions that either directly or indirectly address EDM in some context. Vendors that are most aligned to this market can be split into two groups: 1) EDM software providers and 2) ERM software providers.

A feature of the majority of solutions offered by these vendors is that they normally leverage embedded "enterprise data models" built to support a specific business application and software implementation. For example, since the 1990s, off-the-shelf systems have been available that can be used to maintain an enterprise view of (operational) security data. These "securities masters" have enjoyed widespread adoption in the industry. However, although they certainly address important aspects of EDM, their scope is limited to a particular application, which in this case is the enterprise management of operational security data. The problem with building solutions to address one aspect of EDM this way is that without an enterprise business vocabulary, conceptual model, model governance process, and overall architecture, there is a greater risk that multiple "enterprise" repositories and applications proliferate throughout an organization. For example, the front and middle office may build a security master to support trading activity using a particular representation of security data, whereas the back office may build a data warehouse to support regulatory compliance using a different representation of security data. Clearly, this can compromise data integrity, create significant integration challenges, and increase implementation costs.

In contrast to this, the industry models IBM has developed for this industry (and, to a degree, the EDM Council's Semantic Repository) offer enterprise models at various levels of abstraction, plus an architectural framework and toolset that enables the integration of each model component across the various levels of abstraction in a technically consistent fashion. This way, the models define the architectural blueprint for an enterprise in a manner that is both comprehensive in scope and independent of any particular software implementation. These models can then be customized and extended as required by clients and leveraged by various applications and repositories in a manner that maintains the overall architectural integrity of the EDM infrastructure.

Concluding remarks

The financial market industry has experienced a great deal of uncertainty and turbulence since the subprime crisis of 2007. This has highlighted major weaknesses in the risk management infrastructure of many firms, including the failure to manage critical data pertaining to financial instruments, counterparties, and liquidity in a robust and integrated fashion. In this paper, we have explored important aspects of this, focusing on the efforts of IBM to develop industry models that can support the establishment of corporate-wide standards for financial market organizations. We compared this approach with the efforts of the EDM Council to develop a Semantic Repository for the financial market industry and also commented on related technologies offered by software vendors in the EDM space. The main contribution of this paper is to highlight the role that such industry models and data standardization initiatives can play in addressing some of the key challenges this industry is facing.

For financial institutions that have grown their various lines of business to date as vertical "silos of risk" (and suffered the consequences of this through the financial crisis), the IBM industry models can provide a roadmap to assist in the transformation of their data management infrastructure to better support ERM. Most firms will start this process by using the IBM industry models and supporting software tools to create an enterprise business vocabulary that provides a common definition of all critical business concepts and terms. Once agreed by all relevant stake holders and deployed across the various lines of business, the next step for most firms will be to use the IBM industry models together with the deployed business vocabulary to create an enterprise conceptual model that specifies how the business concepts defined in the vocabulary are classified or segmented and relate to each other. This conceptual model can also be used to define commonly used aggregations of data for business intelligence or other reporting purposes. If deployed and governed appropriately, then the conceptual model can serve as the basis to define all downstream data repositories developed across the various lines of business. This way, the conceptual model can play a leading role in enhancing the integrity of all data used for enterprise risk reporting and other critical business decision making.

In conclusion, there is little doubt that the financial market industry still has many hurdles to overcome in addressing the EDM challenges it is facing. Some of these will require industry-wide solutions, such as the data standardization efforts of the EDM Council, and others will require enterprise-wide solutions, such as the establishment of corporate standards and the deployment of enterprise data as facilitated by the IBM industry models and supporting software. Other hurdles lie beyond the scope of technology and will require regulatory intervention to address—not least among these are the organizational alignment issues that dominate most financial institutions in this industry and can lead to the creation of "vertical silos of risk"—and the failure of many ERM initiatives.

Acknowledgments

The IBM Industry Models represent the fruits of a significant collective effort, and this paper would not have been possible without the hard work of the development team and support of the management team. The authors would like to thank V. Lister, E. Hanley, and N. Ryan for their considerable technical expertise and also B. O'Hara, B. Hallahan, P. O'Sullivan, and R. Goenka for their leadership and continued support on the financial market project.

References

1. M. G. Crouhy, R. A. Jarrow, and S. M. Turnbull, "The subprime credit crisis of 2007," *J. Deriv.*, vol. 16, no. 1, pp. 81–110, 2008.
2. V. Acharya and M. Richardson, Eds., *Restoring Financial Stability: How to Repair a Failed System*. Hoboken, NJ: Wiley, 2009.
3. D. W. Diamond and R. G. Rajan, "The credit crisis: Conjectures about causes and remedies," *Amer. Econ. Rev.*, vol. 99, no. 2, pp. 606–610, May 2009.
4. Swiss Federal Banking Commission, *Subprime Crisis: SFBC Investigation Into the Causes of the Write-Downs of UBS AG,* 2008.
5. B. S. Bernanke, "Reflections on a year of crisis," in *Proc. Federal Reserve Bank Kansas City's Annu. Econ. Symp.*, Jackson Hole, WY, Aug. 2009.
6. U.K. Financial Services Authority, *The Turner Review: A Regulatory Response to the Global Banking Crisis,* 2009.
7. J. A. Zachman, "A framework for information systems architecture," *IBM Syst. J.*, vol. 26, no. 3, pp. 276–292, 1987.
8. J. F. Sowa and J. A. Zachman, "Extending and formalizing the framework for information systems architecture," *IBM Syst. J.*, vol. 31, no. 3, pp. 590–616, 1992.
9. W. H. Inmon, J. A. Zachman, and J. G. Geiger, *Data Stores, Data Warehousing and the Zachman Framework: Managing Enterprise Knowledge*. New York: McGraw-Hill, 1997.
10. J. A. Zachman, *Zachman International Enterprise Architecture*. [Online]. Available: http://www.zachmaninternational.com
11. The Open Group, *TOGAF Version 9*. Zaltbommel, The Netherlands: Van Haren Publishing, 2009.
12. U.S. Office of Management and Budget, *FEA Consolidated Reference Model Document Version 2.3*. [Online]. Available: http://www.whitehouse.gov/omb/assets/fea_docs/FEA_CRM_v23_Final_Oct_2007_Revised.pdf
13. R. Evernden, "The information framework," *IBM Syst. J.*, vol. 35, no. 1, pp. 37–68, 1996.
14. M. Delbaere and R. Ferreira, "Addressing the data aspects of compliance with industry models," *IBM Syst. J.*, vol. 46, no. 2, pp. 319–334, Apr. 2007.
15. M. Atkin, *EDM: The Road Ahead*. [Online]. Available: http://www.edmcouncil.org/PDFs/20090401_EDM_The_Road_Map_Ahead.pdf
16. A. Sussman, "Managing risk in real-time markets," in *Tabb Group Report*, Feb. 2005.
17. C. Abrams, J. von Kanel, S. Muller, B. Pfitzmann, and S. Ruschka-Taylor, "Optimized enterprise risk management," *IBM Syst. J.*, vol. 46, no. 2, pp. 219–234, Apr. 2007.
18. EDM Council, *Semantic Repository Overview*. [Online]. Available: http://www.edmcouncil.org/default.aspx
19. International Standards Organization, *Securities and Related Financial Instruments—Classification of Financial Instruments (CFI code)*. [Online]. Available: http://www.iso.org/iso/home.htm
20. Bank of International Settlements, *Basel II: International Convergence of Capital Measurement and Capital Standards: A Revised Framework—Comprehensive Version*. [Online]. Available: http://www.bis.org/publ/bcbsca.htm

Received July 17, 2009; accepted for publication August 18, 2009

John G. Carney *IBM Software Group, Information Management Division, Ireland Software Labs, Building 6, IBM Technology Campus, Damastown Industrial Estate, 15 Dublin, Ireland (jcarney@ie.ibm.com)*. Dr. Carney received the B.A. and Ph.D. degrees in computer science from Trinity College, University of Dublin, Dublin, Ireland, in 1996 and 2000, respectively. He subsequently cofounded a software company specializing in risk management analytics for financial markets, where he served as the company's chief technology officer for more than four years. He is the author of three patents and seven peer-reviewed technical papers and regularly speaks at industry conferences and workshops on the topics of enterprise data management, risk management, and regulatory compliance in financial markets. He is a member of the IBM Industry Models team, where he is the Product Manager responsible for financial services models and tooling software. He previously held the position of Enterprise Architect and led the development of the IBM Industry Models for Financial Markets.