THE DATA CHALLENGE OF BASEL II

David M. Rowe & Dean Jovic

THE BURDEN OF HISTORY

Data fragmentation is an uncomfortable fact of life in every organization. This reflects the wide adoption over the past 25 years of decentralized decision making regarding information systems. Before the advent of the personal computer, data processing was a heavily centralized corporate function. It was characterized by "the guys with the white coats in the glass air conditioned room with the raised floors." Business users were at the mercy of an elite core of knowledgeable technologists. The result was a high degree consistency in technology at the enterprise level but great difficulty in keeping up with the changing needs of individual business units.

This situation began to change with the introduction by Digital Equipment Corp. of the PDP-10 mini-computer in the mid-1970s. For the first time, acquisition of serious data processing equipment need not be a multi-million dollar proposition. Mini-computers brought the price of significant computing power within the budget authority of department managers, many of whom were only too ready to escape the clutches of the "high priesthood" at the corporate data center. The options for localized computing expanded further with the advent of the personal computer. The trend accelerated when IBM entered the field, giving PCs a mainstream acceptability that had previously been lacking. It is no exaggeration to say that these developments were the beginning of a revolution. The ability to implement new analytical capabilities quickly allowed for better local decision making and was even the basis for whole new products and markets such as derivatives.

While localized computing brought greater nimbleness to those who adopted the new technology, it also had its dark side. Quality control suffered as applications were deployed without the strict quality assurance that characterized centralized computing. Backup and archiving of data were inconsistent or non-existent. Perhaps most serious, however, even fully accurate and reliable data became scattered across multiple platforms. These data were on different machines (sometimes with different operating systems) in inconsistent formats, with limited documentation and usually with no means of external access.

It is an old insight that a firm's technology reflects the organization's priorities. Throughout much of the past two decades, decentralization and local empowerment have been popular management trends. The needs of individual business units to meet dynamic changes in their markets was considered to be of primary importance. Enterprise-wide information was mainly for the limited purpose of financial reporting and long-term planning. These management trends interacted with the availability of increasingly powerful localized technology to accelerate the fragmentation of information processing and data storage.

THE RISE OF ENTERPRISE-WIDE RISK MANAGEMENT

More recently risk management has moved from a localized to an enterprise-wide function. This reflects the broadening realization that risk is inherently a portfolio concept. To measure risk accurately at the enterprise level requires analysis not just of individual local risks but of how these interact with each other. To accomplish this does not require simultaneous access to every scrap of data throughout the firm. It does, however, require the ability to analyze a minimum core of information on a unified basis. In addition, for credit risk the problem is even more fundamental. Even localized analysis for individual obligors often requires consolidating data across regions, products and organizational divisions. The following graphics illustrates the point.

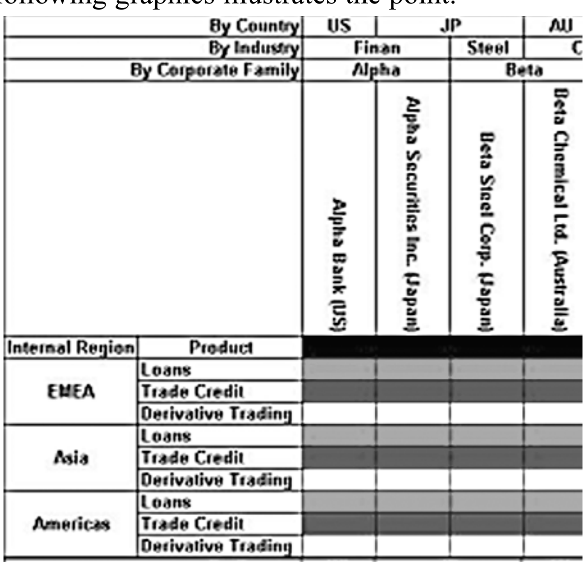


Figure 1

As shown in figure 1, data consistency tends to run parallel to organizational segments and these generally reflect regional and product

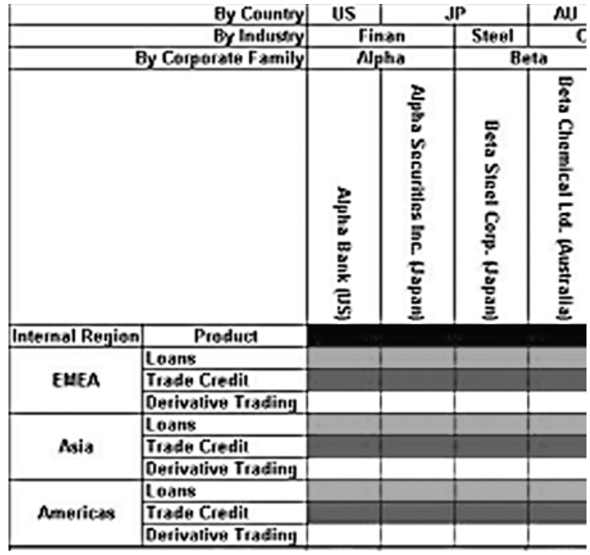


Figure 1b

categories. Such data include credit exposures to all customers dealing with that product and region. To be sure, data fragmentation is not exhaustive. In the above illustration there is one integrated global system for loans and a common trade credit system for two of the three regions. Nevertheless, it is a rare organization in today's environment that can claim full data integration on a single platform for all products across all regions.

To analyze the credit risk for a single obligor, on the other hand, requires a "single customer view." As shown on figure 1b this demands consolidation of data that cut directly across all the individual systems that are, at least partially, fragmented by region and product. Thus data consolidation is necessary just to achieve an integrated customer view before even considering the analysis of portfolio dynamics in estimating enterprise-wide risk.

BASEL II AND DATA MANAGEMENT

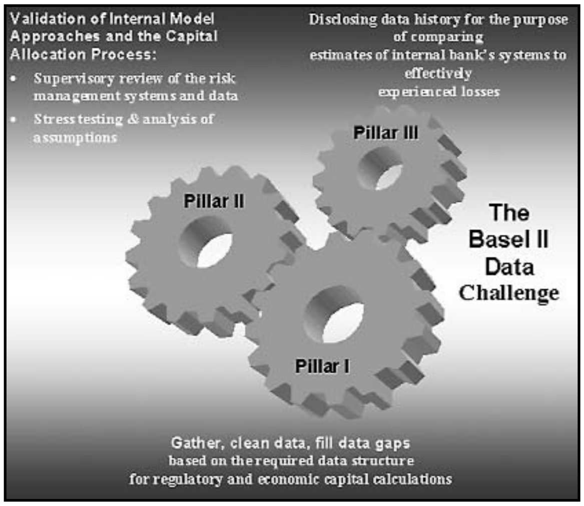
The New Basel Capital Accord (commonly referred to as "Basel II") was proposed in 1999 by the Basel Committee on Banking Supervision. This revision to the 1988 Accord ("Basel I") was prompted primarily by serious inconsistency between required capital amounts under the current accord and the true economic risk of the associated assets. In April 2003 the third Consultative Paper (CP-3) was published. The final implementation of the Accord is currently scheduled for the beginning of 2007. By then, larger banks are expected to adopt the more sophisticated alternatives for determining regulatory capital for their credit and operational risks based on Pillar I of the Accord. Apart from that, Pillars II and III imply further requirements regarding the supervisory review process and disclosure of information on risk measures and capital adequacy. Essentially, Basel II poses some very challenging requirements for data integrity (accuracy and completeness), data aggregation and consolidation as well as the analysis and archiving of these data. Ultimately, for capital calculation purposes in a large financial institution, the relevant data must be gathered consistently from many different source systems and "data silos" in a global environment.

The demands for data integrity and data integration are among the most fundamental challenges in meeting the Basel II requirements. Consistent collection of data often takes years and is certainly one of the main reasons for the complexity of such projects. Data requirements can be reflected under the three Pillar Framework of Basel II as follows:

▲ For all the proposed capital calculation approaches, Pillar I requires input data to be accurate and complete. In order to receive meaningful risk figures and to use them as the basis for the internal capital allocation process, banks need to check consistency, timeliness and reliability of the data sources used. Also, regular and independent validation should be part of the banks' internal audit process (see requirements for the standard approach to credit risk in CP-3, § 136. Data requirements for the Internal Ratings-Based (IRB) Approach are set out in § 233, § 425 and § 428.) Specific data administration issues (§ 391) refer to what data need to be stored to reflect relevant borrower characteristics and businesses in rating systems. A further dimension is that data need to be sufficiently detailed and granular (e.g. to allow for ex post re-allocation of borrowers to rating grades). While the application of the Foundation IRB Approach requires banks to build up a data history in order to support probability of default (PD) estimates, banks implementing the Advanced IRB Approach will have to structure their data archives more comprehensively to be able to provide historical data for the validation of Loss Given Default (LGD) and Exposure At Default (EAD) estimates (see CP-3, § 393). Regarding calculation of regulatory capital for operational risks, Pillar I describes the Advanced Measurement Approaches (AMA, § 622) as the most sophisticated methods. The CP-3 paper states clear guidelines for the collection of meaningful op risk-related loss data and requires data flows to be transparent and accessible (§ 626). Some further detailed requirements refer to the collection and assessment procedures for internal loss data (§ 630-633) and external loss data (§ 634-635).

▲ Pillar II of the Accord states that banks should validate their risk management systems on a continuous basis (CP-3, § 703). While accuracy and completeness of data should be ensured, analysis and stress testing of the underlying assumption must also be conducted on a regular basis. This relates to the bank's risk measurement and management system as well as the enterprise-wide capital allocation process.

▲ Pillar III: Banks applying the IRB Approach for credit risk need to archive data over a long time horizon in order to be able to disclose a comparison between estimates of the rating system for PD, LGD and EAD and effectively experienced losses. This should allow for a meaningful assessment of the performance of the bank's internal rating process. An example would be if a financial institution disclosed average of default rates over a 10-year-period for each of its rating classes (see quantitative disclosures in the Pillar III chapter of CP-3, section (g), table (b)).

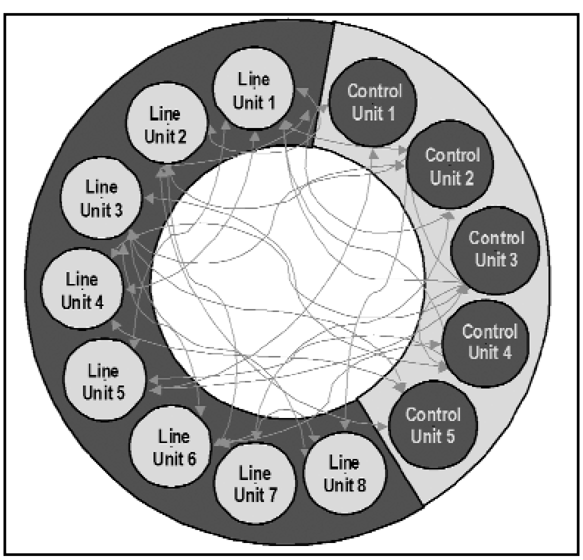


Data Integration Strategies

This leads naturally to the question of the best strategy to achieve the required degree of data integration both to achieve best practice risk management and to meet the requirements of Basel II.

Default Solution

The first approach could be called the default solution. It is what generally occurs in the absence of a corporate strategy for dealing with this issue. It is comprised of a large number of point-to-point data feeds developed on a bespoke basis as the need arose. The following graphic illustrates this approach.

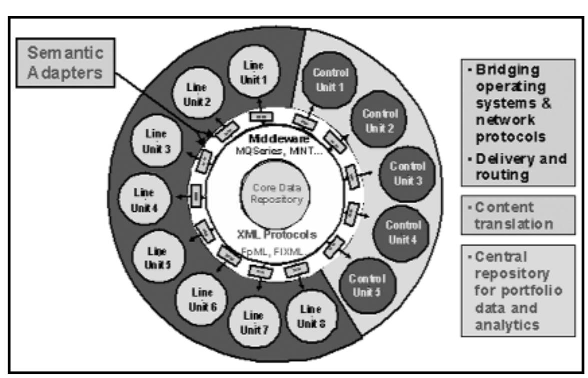


Such point-to-point data links are often the fastest and easiest way to solve an immediate problem. The downside is that there is no consistency in the format and the logic to create recurring data files is usually unique to each instance. In the end, maintaining such data feeds becomes a major hidden (or sometimes not so hidden) cost of keeping the whole process running. As new features are added to existing products, new products are introduced and new business units are added via merger and acquisition, the structure of these feeds needs to be revised accordingly. It has been estimated that just maintaining such links consumes well over half the data processing maintenance budget of many large organizations. In addition, this is inherently a batch update strategy. Having initially solved the data transfer problem via point-to-point file transmissions, it is very hard to move from a batch orientation to a real-time event driven architecture.

An "Ideal" Strategy

The vision of an ideal strategy centers around self-describing messages built on the foundation of the eXtensible Mark-up Language (XML). Properly speaking, XML is not a mark-up language at all, but rather a meta-language, i.e. a syntax within which a true mark-up language can be developed. Such mark-up languages include the Financial products Mark-up Language (FpML) and the Financial Information eXchange Mark-up Language (FIXML). These define specific semantic content for describing products, transactions and events for a specific business domain.

This approach is illustrated in the following graphic.



The disorganized snarl of point-to-point connections is replaced with a corporate information backbone. There are three broad functions that need to be implemented for this approach to be successful. The first is bridging disparate operating systems and network protocols as well as assuring that information that information being transmitted is properly routed and received by the appropriate recipients. This is the task of standard middleware products such as MQSeries from IBM and MINT from SunGard. The second, and more troublesome, requirement is content translation. First this demands a well-defined mark-up language such as FpML or FIXML. Second, it demands a series of adapters to translate content between the individual local systems and the standard XML-based mark-up language. Finally, to achieve effective portfolio modeling requires that core data be consolidated in a central repository for ease of analysis and assured archiving in a consistent format.

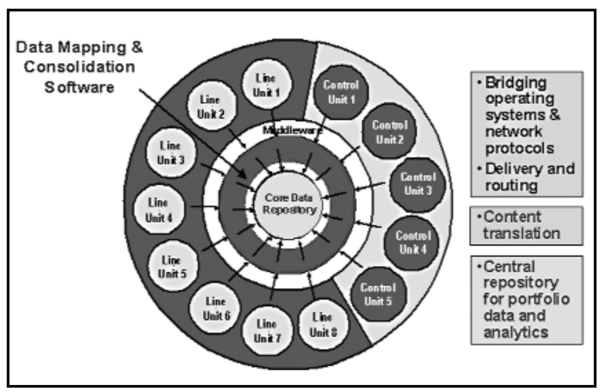
The advantages of this approach are that it is extensible. New features can be added to the mark-up language without disturbing the existing messages that do not use these new features. In addition, once the adapters are written a single piece of information can be transmitted to multiple destinations by placing it on the ring in standard form. Furthermore, this approach is inherently modular and lends itself naturally to an event driven environment.

There are, however, some serious disadvantages to this approach. First, the industry standard mark-up languages have been understandably slow to develop. This has left many institutions reluctant to forge ahead on their own knowing they will have to make major revisions when an industry standard is established. Second, the process of developing the adapters is a significant investment and these would have to be modified if the structure of the core mark-up language is revised. Furthermore, the payoff from these investments tends to be broadly distributed across the organization rather than accruing mainly to the business units that develops them. A less serious drawback is that transmissions based on this approach are much more "verbose" than with traditional approaches such as fixed format files. The global overhang of unused communications capacity, however, is likely to minimize this problem.

The drawbacks noted here have resulted in only a slow acceptance and implementation of this ideal approach. While we suspect that this will be the standard approach to this problem in ten to fifteen years, it is unlikely to represent a viable alternative in the time-frame of Basel II compliance.

A Practical Alternative

Fortunately, there is a practical middle ground between bespoke point-to-point file transfers and an ideal self-describing messaging environment. This involves inserting a layer of content translation software between the local systems and the central data repository. The following graphic illustrates this approach.



This software layer plays a dual role. First it provides a visual data mapping environment that is useable by a business analyst who does not have to be a programmer. It allows the user to define the appropriate correspondence between fields in the remote database (or flat file output from the remote system) and the central data repository. Having defined these correspondences, the software creates a standard translation file to preserve this correspondence. The second role of the software is to perform periodic transformations and transfers of actual data from the local systems to the central database. For this task it uses the meta data in the translation file created and maintained by the business analyst.

For a variety of reasons this is an attractive middle ground between the other two approaches. Compared to bespoke point-to-point links it introduces much greater discipline and consistency in the process of creating correspondences between data in the remote systems and their counterparts in the central database. Moreover, changes in the local data formats can be spotted relatively easily and the resulting problems corrected in a timely manner. This ability can be strengthened by defining "sanity checks" on the values of inputs as part of the metadata in the correspondence table. This often can allow issues to be trapped in the translation process even if the local change has not been communicated to those in charge of the central data consolidation.

The biggest advantage of translation software relative to immediate adoption of the ideal approach is that it does not require a comprehensive markup language to be workable. In addition, the translation tables can be created easier and faster than the semantic adapters. Finally, this strategy can support an event driven messaging approach to data transfer. This is done by allowing the translation software to read and write information in self-describing messages, typically formulated in XML-based markup languages. Hence it can exist comfortably with, indeed can support, a gradual evolution toward the ideal event-driven approach.

Obviously an approach to data consolidation based on translation software does not create virtual many-to-many interoperability, which is the ultimate goal of the ideal system. Nevertheless, it greatly streamlines the process of creating and maintaining a central repository of the data needed to perform meaningful enterprise-wide risk analysis. As such, it should be seriously considered as an option for any organization struggling with the data consolidation and analysis requirements of Basel II.

David M. Rowe, Ph.D. is Group Executive Vice President for Risk Management, SunGard Trading and Risk Systems

Dean Jovic, Ph.D. is the Managing Director, Basel II, SunGard Trading and Risk Systems

1 - Reportedly a saying among corporate technology staff in the 1980s was, "The users are revolting, in both senses of the word."