Modernizing Financial Data Collection with XBRL

Linda F. Powell and Andrew Boettcher[1]

Abstract
In 2003, three U.S. banking regulatory agencies combined resources to revolutionize the collection, editing, storage, and dissemination of Commercial Bank Reports of Income and Condition. The regulatory agencies relied heavily on web-based technology and the Extensible Business Reporting Language (XBRL) transmission protocol. This paper will review the creation of an interagency data collection and dissemination facility. It will focus on the business problem that needed to be solved, the evolution of the technology that enabled the project, what XBRL is, and why was it selected as the transmission protocol. The paper will also review the challenges and benefits associated with using a standard transmission protocol versus creating a customized XML transmission facility.

Introduction
In 2003, U. S. bank regulators, the Federal Deposit Insurance Corporation (FDIC), Board of Governors of the Federal Reserve System (FRS), and Office of the Comptroller of the Currency (OCC) combined resources to revolutionize the collection, editing, storage, and dissemination of the Commercial Bank Report of Income and Condition (Call Report). The system development process took approximately two years and was completed in March 2005. The new process utilizes internet technology and XBRL protocols. The benefits of the new process include easier distribution of reporting requirements, fewer revisions to reported data, and easier dissemination of data to the market.

The Call Report is a regulator-specified report which collects primarily financial statement data. This means that the data collected are standardized across the industry and predefined by the banking regulators. The financial data conform to U.S. generally accepted accounting principles. However, the data are not synonymous with data that are collected from other market regulators such as the Securities and Exchange Commission (SEC). Many of the data variables collected are comparable to those collected by other regulators or published in the financial statements, but additional variables are uniquely defined for and used specifically by the banking regulators.

The Call Report is filed by all U.S. insured commercial banks and state chartered savings banks, which currently includes over 7,700 institutions. Each bank is required to file a quarterly report, which contains over 2,000 variables, to its regulator. Although all banks must file a Call Report not all banks are required to file the exact same set of data. For example, banks with foreign operations are required to provide details about their foreign activities that are not applicable to banks that operate solely in the domestic realm. The criteria for which variables a bank must report are detailed in the instructions.

Historical Data Collection Model
The collection of Call Report data has evolved over the years from a manual collection process to the current internet submission process. Figure 1 shows the regulators' data collection, editing, and distribution model in use before the 2003 Call Report Modernization project. The historical model required all banks to file their Call Report electronically but involved multiple regulators maintaining multiple copies of the Call Report data and, individually contacting banks regarding any questions or modifications to that data. The historical model included the following data flows:

1. Reporting form changes are communicated to banks and software vendors via PDF and Word files.

2. Each regulator and bank modifies its internal applications to receive or send data.

3. Banks compile the required data and send a standardized formatted file to a central collection vendor.

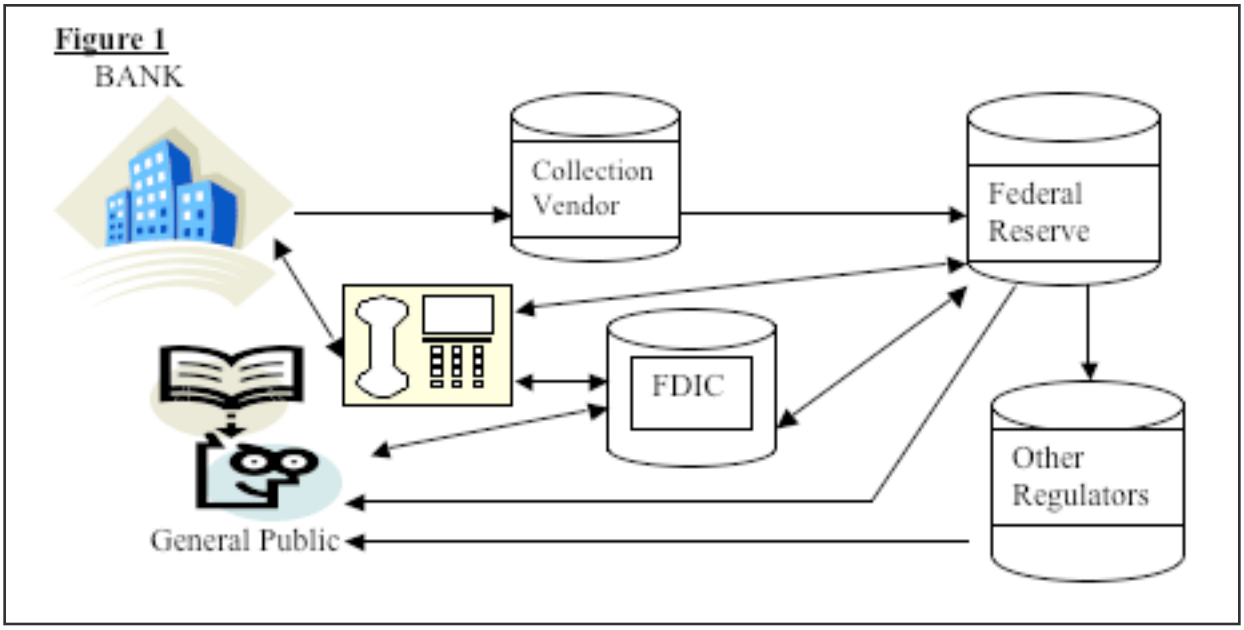
4. The collection vendor compiles the banks' filings and transmits the data to the Federal Reserve System (FRS).

5. The Federal Reserve parses the data and transmits the data to the FDIC.

6. Each banking regulator (FRS, FDIC, OCC) maintains its own copy of the most current data.

7. The FRS and FDIC edit the data for their banks and call the banks to resolve data anomalies and process revisions. As revisions are processed, the regulators transmit revisions amongst one another.

8. Non-confidential data are released to the public in various forms from the regulators



Business Problems with the Historical Collection Model

The historical business reporting model, outlined in Figure 1, presents some obvious and not-so-obvious risks and challenges. The challenges that needed to be addressed fell into three categories: 1) multiple collection and storage sites, 2) difficulties for the industry in implementing changes to the data collection requirements, and 3) improvements to data quality.

Multiple data collection and storage sites coupled with the ability of the banks to revise their data introduces the risk of inconsistencies in the data maintained and published by the various regulators at any one time. The inconsistencies were the result of timing differences, human error, or communication breakdowns between organizations. Additionally, redundancies in overhead existed across the regulatory agencies. Separate databases and editing applications were maintained by each regulator. Although the regulators collaborated and combined efforts where feasible, such as in communicating with the banks, activities such as database administration, edit identification, and metadata maintenance were performed separately at each agency.

A larger but less obvious challenge was the need for multiple organizations (regulators, banks, and software vendors) to maintain metadata to support electronic filing. Early in the Call Report Modernization process, industry participants (banks and software vendors) identified the need for quality metadata that could be electronically communicated and consumed. Necessary metadata include the variables collected, variable definitions, and data edits (business rules that verify the accuracy of the data).

Revisions to the reporting structure, including adding and removing variables as well as definitional changes and clarifications, can occur multiple times a year. Under the historical model, the regulators, banks, and software vendors had to each make revisions to their applications based on changes that were communicated in hard copy, text, or PDF documents. Electronic distribution of metadata was appealing because it can be easily altered to reflect reporting requirements.

Additional business needs included making the editing process more efficient and improving the data quality for the industry. As noted above in the historical model, all data editing was done after the data were received by the regulator. After editing the data the regulator would call the bank to determine if data revisions were necessary or to gather explanations for data anomalies. It was much more difficult to resolve data anomalies and edit failures after the data were submitted to the regulators than at the time the data were compiled by the banks. For example, if a total variable does not equal the sum of the component variables, it is much easier for the bank staff to determine the reason for the difference and fix the data or note the cause of the difference when compiling the data than several weeks later after the data have been submitted to the regulator.

Evolution of Technology Enabling XBRL

Collecting, distributing, and sharing data have always been very labor and time consuming and are often further complicated by incompatible software and the need to provide volumes of text that explain the data and define the data file layout. Although it is currently hard to imagine professional communication before the development of the internet, Call Report data were collected and disseminated long before common acceptance and use of the internet, which occurred around 1995.

In the late 1990s the innovation of XML (eXtensible Markup Language) enabled web developers to move from static HTML documents to interactive web pages. In other words, beyond simply displaying preset reports, XML supports data exchange and machine-actionable processing over the internet. Unlike HTML, XML also enables descriptions of the data to be associated with the data and embedded in the web applications. These descriptive data are commonly referred to as metadata. Shortly after XML was introduced, business users began to harness the power of XML and apply standardized data descriptions and naming conventions. The compilation of naming conventions and metadata is referred to as taxonomy. Since the creation of XML, several transmission protocols were created and have become industry standards. XBRL (for financial data), SDMX (for time series statistical data), and DDI (for social science data) are three internationally recognized data exchange formats capitalizing on the XML format.

Specifically, XBRL marries new internet-based technology across dissimilar computer systems, with business rules and practices. XBRL can be thought of as a set of accounting standards coupled with information technology (IT) standards that simplifies the exchange of data. This marriage of accounting and IT is accomplished through the taxonomy, which is an integral part of the XBRL standards.

XBRL taxonomies provide definitions for tags that are identified in the XML schema. Within the taxonomy, there are a number of required data elements that provide information about the data characteristics. XBRL examples include the currency of monetary variables, if the variable is from the balance sheet or income statement, and if the variable is a debit or credit. Using the XBRL taxonomy facilitates solving the business problems of distributing data collection requirements.

The New Collection Model

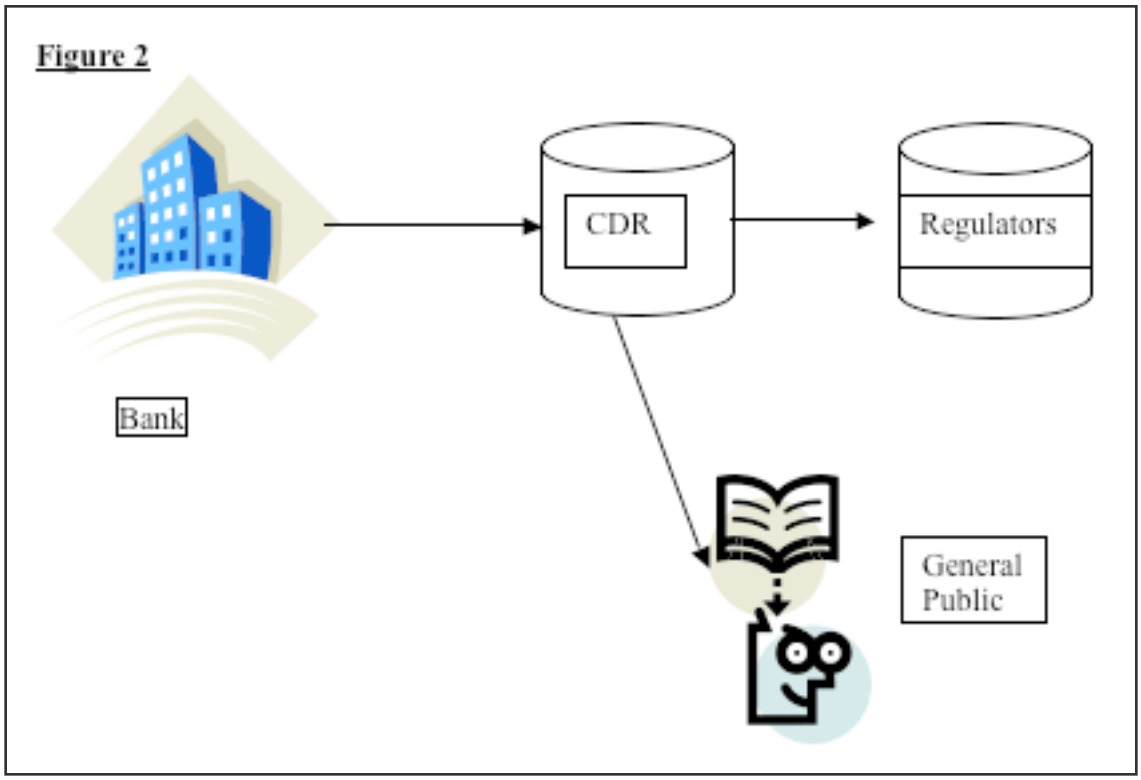
The collection model that resulted from the Call Modernization project is outlined in Figure 2. In the new model, XBRL supports the collection, editing, and distribution of Call Report data. Each quarter, the regulatory agencies create one taxonomy that includes all of the variables to be collected, metadata, and edit calculations. The taxonomy is downloaded by banks or bank vendors and is used to programmatically update their internal reporting systems. The data are then compiled by the banks and submitted to a Central Data Repository (CDR).

As noted above, before the implementation of XBRL, every time the Call Report was revised each agency manually updated its collection, editing, and data storage systems. In addition, each bank or bank vendor manually updated its individual regulatory reporting system. With the use of XBRL, the taxonomy can be programmatically imported into vendor software, while all three agencies use one system to collect and edit Call data. XBRL also makes it possible for edits to be distributed to the banks because the taxonomy contains data validation routines. Thus, each bank edits its data before submitting to the CDR, which then reedits the data upon receipt. If any edits fail or anomalies are not adequately explained, the report submission is automatically rejected and the bank is notified immediately. This pre-editing routine reduces the number of follow-up phone calls the agencies must make to analyze questionable data, thereby reducing the burden on both the banks and the regulators.

An added benefit of the XBRL taxonomy is the ability to use the metadata to create data presentations. The presentation metadata arrange the Call data into a human readable format. For the Call Reports, it organizes the data into balance sheet and income statement formats. Once the presentations are defined, web reports can be created dynamically from the metadata, enabling the CDR to easily publish Call data to the general public. Although the agencies have made the data available to the public on the internet for years, the process to maintain the web reports was time consuming and the reports were only published in bulk when all of the banks' data were finalized. Under the new XBRL model, each bank's data are published three business days after being submitted to the CDR. The CDR also has a web service that allows the public to retrieve all Call Report data. This service is an efficient method to disseminate more timely banking industry data to the public and allows public users to repackage and easily disseminate data[2]

The creation of an internet-based data collection and dissemination facility is a tremendous undertaking. The U.S. banking agencies had two advantages in place that helped facilitate the new XBRL collection model: well structured data and variable level metadata. The Call Report data are well structured because comparable information is collected from all entities with well defined rules regarding which banks must report what data. Although some of the rules are quite involved, they are structured enough to define programmatically.

The variable-level metadata predated the CDR and are housed in a metadata warehouse know as the Micro Data



Reference Manual (MDRM) at the Federal Reserve Board. The MDRM was originally built as a tool for end users who access the data. However, its well structured naming convention and nomenclature are well suited for the data collection process. Although the MDRM did not contain all of the XBRL metadata variables (which have since been added manually), it did provide variable naming conventions, variable history, and detailed descriptions of every variable.

Pros & Cons of Using an Industry-Wide Protocol VS. a Custom Built Web Interface

All data exchange formats provide structure for sending and receiving data. The XBRL protocol includes metadata for each data variable, such as if the variable is from the balance sheet or income statement, what currency it is denominated in, and if the variable is an integer, monetary value, ratio, or text item. The detailed metadata remove any doubt about the meaning of data variables. For example, within XBRL, the variable "total deposits" can be defined in the metadata so there is no confusion whether the deposit is a liability (such as for a bank) or an asset (such as for an individual with a savings account).

Clearly, an internet based collection system could have been built without using a data exchange protocol. There is even an argument that not using an industry-wide protocol might have been easier to build since having standards requires adherence to its ensuing rules and guidelines. For example, one hurdle encountered by using XBRL is that Call Report data are reported in thousands of dollars while the XBRL standard requires that data be transmitted in single dollars. To meet the XBRL standard, additional steps are taken to convert thousands of dollars to dollars and then back to thousands of dollars for the reporters and end users who expect the data to be in thousands of dollars.

In spite of the added hurdles, the benefits of using the industry-wide protocol outweigh the challenges of meeting the requirements of the protocol. Having the standard ensures that any user of the data can access and understand the information as long as he or she has basic knowledge about the standard's data elements and definitions. One of the earliest examples of the benefits of standardization is the standard rail gauge. The standard rail gauge defined the distance between the rails on railroad tracks. Before the standard rail gauge was adopted in the 1870s trains could not travel across the country and cargo had to be unloaded and reloaded to a new train each time the rail gages changed. Adopting the standard rail gauge allowed trains to travel on any track providing tremendous economic benefits[3]. When data are transmitted under different formats or standards the data need to be transformed each time the data are transmitted. Just as the rail gauge facilitated the shipping of freight over long distances, transmission protocols allow for an easier flow of data along the information highway.

Where XBRL is Going

There is a growing body of research on the benefits of XBRL. The published research falls into three main perspectives: financial statement consumers, financial statement creators, and financial statement auditors. Financial theory suggests that more frequent, reliable, and readable financial statement reports will result in a healthier marketplace. The first published studies focused on market participants' decision making. Recently, research is moving toward XBRL's impact on reporting costs and regulatory enforcement.

Financial theory suggests that a market with more informed participants would price risky assets closer to the assets' true economic value. Market participants use reported financial statements to assess a company's financial strength and adjust their investment in the company accordingly. However, there is a prohibitive search and learning cost to understanding financial statements. As a result, risky assets are priced by fewer investors or worse, misguided investors. While all statements retain core structure (i.e. Equity = Assets - Liability), line items and footnote disclosures can differ significantly from industry to industry and even company to company. Therefore, early research focused on the financial statement consumers and whether tagged financial data improves transparency and results in a healthier marketplace.

One of the first XBRL studies (Hodge, Kennedy, Maines 2002)[4] tested investors' judgments regarding the reporting method of stock option compensation data. Stock option compensation is allowed to be reported either on the face of the financial statement or disclosed in the footnotes. Investors were shown tagged financial statements to see if XBRL reduces the difference between the financial statement recognition and the footnote disclosure reporting methods. The authors found that without tagged data, the influence of footnote disclosures diminished, resulting in investors making less informed decisions. When using tagged data, investment decisions did not change between reporting methods, suggesting that the XBRL format equalizes recognition of the footnote and the financial statement disclosures. Furthermore, investors reported trusting tagged statements more than the untagged data. While the authors found XBRL tagged data eliminated the financial sophistication required to understanding financial statements, XBRL only had added benefits for investors who used the available search technology. The authors found that almost 50% of the survey participants did not use the search technology to the extent that XBRL enabled.

New research is examining XBRL's impact on financial statement creators. The current statement filing process is expensive and time consuming, requiring audited statements to be filed annually and unaudited statements to be filed quarterly. An XBRL-based reporting system offers the potential of more frequent, potentially continuous, financial reporting. A paper by Hunton, Wright, and Wright (2003) argues that there is an optimal reporting frequency somewhere between quarterly and continuous. The authors claim that, on appearance, continuous reporting would align fair market value with asset prices resulting in reduced price volatility but that psychological factors could undo and further exacerbate price volatility. While further research is needed to determine the ideal frequency for financial statement reports, XBRL can report any time frequency desired.

Successful XBRL implementations, growing international use, and positive academic research influenced the Security and Exchange Commission (SEC) and the Chartered Financial Analyst Institute to publicly support adoption of the XBRL standard in the United States. Both the SEC Commissioner and CFA president are promoting XBRL reporting standards and its importance to the industry.

With these inherent benefits of XBRL, government regulators worldwide are either already using or pursuing the use of XBRL. Other significant adoptions of XBRL standards include, but are not limited to: 1) The Australian Prudential Regulation Authority (APRA) to collect data from Australia's super funds, insurers, and banks, 2) The U.K Inland Revenue Services, 3) The Bank of Japan to gather data from financial institutions in February of 2006. 4) Japan's Financial Services Agency's launch of the Electronic Disclosure for Investors' NETwork (EDINET) system in March of 2008, 5) The Netherlands' data collection process for corporate tax, business financial, and business statistical data in 2007. 6) United States Generally Accepted Accounting Practices (GAAP) and The International Accounting Standards Committee (IASC) taxonomies.

There is a sustained need for faster, better, and easier data retrieval in the financial industry. New technology has opened doors for financial reporting that were unimaginable as recently as the early 1990s. Although the need and desirability of financial data reporting on a continuous basis is debatable, the current technological environment of XBRL facilitates that option. Regardless of the optimal frequency, governments and regulators have an ongoing desire to improve the efficiency of data collection and consumption. In addition to helping the government regulators, improved data flows may also help to improve market efficiencies, as they rely on current and accurate data. Using the evolving technology while enforcing compliance with transmission protocol standards may take our information highway to destinations that we are currently unable to envision.

Bibliography

Australian Prudential Regulation Authority. "Statistics Project – Data Submission – XBRL," APRA, http://www.apra.gov.au/Statistics/XBRL.cfm (accessed April 2008).

BNET Business Network. “Business Services Industry,” BNET, http://xbrl.squarespace.com/japan-edinet-reporting/ (accessed April 2008).

EDGAROnline. “Home page and various content pages,” EDGAROnline, http://www.edgar-online.com/ (accessed April 2008).

Hodge, Frank, Kennedy, Jane Jollineau, and Maines, Laureen (2002). “Recognition versus Disclosure in Financial Statements: Does Search-facilitating Technology Improve Transparency?,” November, 2002.

Hunton, James, Wright, Arnie, Wright, Sally (2003). “The Supply and Demand For Continuous Reporting,” Publications of the Univesiteit Van Amsterdam (Netherlands). PriceWaterhouseCoopers LLC.

Roohani, Saeed J, eds. (2003) Trust and Data Assurances in Capital Markets: The Role of Technology Solutions. PriceWaterhouseCoopers LLC.

Securities and Exchange Commission. “Recent Publications – various publications,” SEC. http://www.sec.gov/ (accessed April 2008).

XBRL International Home Page. “Extensible Business Reporting Language,” XBRL, www.xbrl.org/Home/ (accessed May 2008).

Xtech Conference (2007). “Large scale XBRL introduction for Dutch eGovernment Presentation”, Xtech, http://2007.xtech.org/public/schedule/detail/202 (accessed May 2008).

* Linda F. Powell and Andrew Boettcher, Board of Governors of the Federal Reserve System, Washington, D.C. This was presented at the IASSIST conference in Stanford 2008. Contact: Linda.Powell@frb.gov.

Fotenotes

1 The opinions are of the authors and not the Federal Reserve Board.

2 The taxonomy and data are publicly available at cdr.ffiec.gov/public/default.aspx

3. Encarta 97 Encyclopedia, “Railroads”. CD.

4. ”Recognition versus Disclosure in Financial Statements: D3oes Search-facilitating Technology Improve Transparency?” Hodge, Frank; Kennedy, Jane Jollineau; Maines, Laureen A November, 2002