XBRL in the Chinese Financial Ecosystem

Li Jimei, Hu Yuzhou, and Du Meijie, *Beijing Language and Culture University, China*

Business reports that use the eXtensible Business Reporting Language allow analysts to turn massive amount of financial data into well-structured information with software tools. The authors examine XBRL's implementation in China's financial ecosystem, highlighting the challenges and opportunities in analyzing XBRL-based big data.

The 1990s and 2000s witnessed several crises in international financial markets, partly caused by financial institutions' failure to adequately balance credit risks with the capital required to cover probable and (at least to some degree) unexpected losses.[1] Research into such failures uses business reports as a main source of information. Enterprises must periodically file reports with regulators, and such reports are increasingly appearing online as regulators and governments adopt the eXtensible Business Reporting Language (XBRL). In fact, more than 20 countries and regions have adopted XBRL, including the US, European Union, and China (www.xbrl.org/FRTAcknowledged).

Driven by the need to integrate data across countries and business sectors, business reporting strives to follow a set of common representation schemas and rules. The two major regulations are the International Financial Reporting Standards (IFRS) and XBRL. They're intended to standardize financial reporting to promote transparency and improve the comparability of business information. The IFRS Taxonomy (IFRS-T), provided by the IFRS Foundation, has become the accepted financial reporting standard worldwide (www.ifrs.org/XBRL/IFRS-Taxonomy/Pages/IFRS-Taxonomy.aspx).

Because businesses use XBRL to communicate digital business and financial data, it's important to understand the structure of that standard. XBRL can be divided into three levels: the specification level defines how XBRL works technically, the taxonomy level offers a

 Published by the IEEE Computer Society

computer-readable "dictionary" of business and financial data, and the instance level provides a business report that files the financial concepts defined in the taxonomy with actual data (see www.xbrl.org/how-xbrl-works-1). This helps users automate the handling of financial big data and cut out time-consuming and costly collating and rekeying of data. Furthermore, dedicated software can immediately validate, analyze, select, and process the XBRL-based data for further use.

In China, the Ministry of Finance (MoF) issued Chinese Accounting Standards (CAS), one of the generally accepted accounting principles, which is now converging toward IFRS. In October 2010, MoF also issued the CAS Taxonomy (CAS-T),[2] which is based on IFRS-T. The Shenzhen Stock Exchange (SZSE) and Shanghai Stock Exchange (SSE) are also playing vital roles in promoting XBRL in China. MoF, SZSE, and SSE are leading the XBRL implementation in China, yet they're using different taxonomies, disclosure requirements, and filing methods.

Here, we discuss the state-of-the-art of business analysis using XBRL in the Chinese financial ecosystem, highlighting the challenges faced and future potential.

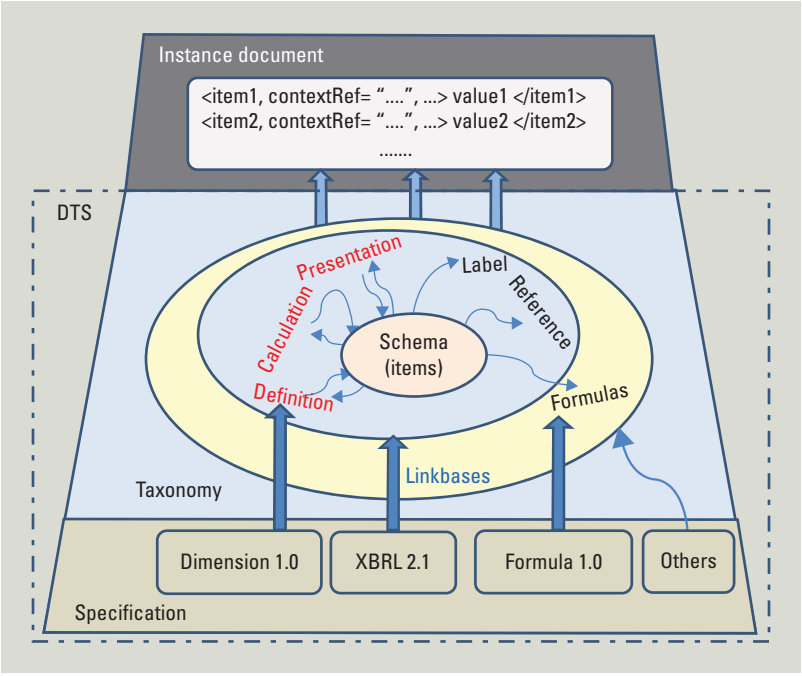

Figure 1. The three levels of the XBRL technology: instance, taxonomy (schema and linkbases), and specification. An XBRL instance contains the business facts to be reported, and the corresponding concepts of the facts are defined in the supporting Discoverable Taxonomy Set (DTS), which includes specifications and taxonomies.

A Brief Tour of XBRL

Charles Hoffman proposed the ideas behind the XBRL in 1998, and the language has drawn worldwide attention. The fundamental idea of XBRL, which is built on the top of XML, is to separate reporting fact values (data) from metadata, thus increasing flexibility and reusability (see www.ifrs.org/XBRL/Resources/Pages/Fundamentals.aspx for more information).

In XBRL, a fact value is tagged in a format such as <*ifrs:Assets*>100</*ifrs:Assets*> (XML syntax), such that it can be easily understood and processed by computers. Herein, <*ifrs:Assets*> represents a concept named *Assets* with the namespace *ifrs*. And the fact values of concepts are found in instance documents.

In this example, what computers understand is that something called *ifrs:Assets* has the content "100." Computers, however, have no built-in accounting knowledge, which is why we must teach them what a particular concept (such as *Assets*) means and explain its characteristics. We can do this with metadata using XML schema files. However, this isn't a trivial task. Indeed, *Assets* usually have a monetary value (*type* property) and a balance for this value (*balance* property). *Assets* can also be understood as an entity's resources or obligations at a particular point in time (*periodType* property, which, in this case, would have a value of "instant").

This description shows that at least three characteristics must be provided to a computer as information so that the computer can understand <*ifrs:Assets*> in an accounting manner. XBRL provides a way to describe and explain accounting concepts, and it can even store relationships between financial concepts in linkbases.

The XBRL Framework

Figure 1 illustrates the framework of XBRL technology, showing the relationships between the XBRL framework's three levels—the instance, taxonomy (schema and linkbases), and specification levels.

An XBRL instance contains the business facts to be reported, and these facts must be assigned to a context. The context defines the entity (for example, company or individual) to which the fact applies, defines the relevant time period (such as 2012.1.1–2012.12.31), and offers the other optional scenarios, such as segments of products. The corresponding concepts of the facts are defined in the supporting Discoverable Taxonomy Set (DTS), which includes specifications and taxonomies. An instance can be supported by more than one taxonomy.

A specification defines all of the XML elements and attributes that are necessary to create, exchange, and aggregate reports.[3] The current international base specification is XBRL 2.1. Dimension 1.0 is an optional, modular extension of XBRL 2.1 that lets XBRL taxonomy creators define and restrict dimensional information. Formula 1.0 is also a modular extension to XBRL 2.1. It defines a syntax for formulas that can be used to check the business facts and rules (for example, Assets = liabilities + equities) based on information obtained from XBRL reports.

A taxonomy provides definitions of and information about XBRL tags, and it organizes the tags in a meaningful structure. A taxonomy consists of a schema (or schemas) and linkbases. The schema contains definitions of elements (such as *ifrs:Assets*), whereas linkbases provide the relationships between elements, and the links from elements to specified external resources. There are six types of linkbases in CAS-T and IFRS-T.

The *label* linkbase contains human-readable strings of reporting elements in a given natural language. It can support multiple languages.

The *presentation* linkbase describes data structures such as the hypercube format and priorities for representing reporting elements. It can help the creator render or visualize the user interface.

The *definition* linkbase associates concepts with other concepts using a variety of arc roles expressing relations. Assuming there are two taxonomies, one follows XBRL 2.1, the other follows Dimension 1.0, and they both define different role types. If the taxonomy adheres to XBRL 2.1, then it can use four standard types of relationships: general-special, essence-alias, requires-element, and similar-tuples. If it follows Dimension 1.0, there are six standard types: all, not all, hypercube-dimension, dimension-domain, dimension-default, and domain-member.

The *calculation* linkbase aggregates the rules on reporting items that have common contexts.

The *reference* linkbase associates concepts with citations of authoritative literature, which are texts from legal documents or regulations applicable to a given reporting item.

Finally, the *formulas* linkbase contains the other business rules for reporting items, such as consistent rules and existence assertion, which can't be defined in calculation linkbase. Only those taxonomies that comply with Formula 1.0 have the formulas linkbase.

An essential attribute of XBRL is that it can be extended[4] to accommodate the complexity of different financial policies from different countries and the wide variety of business reporting needs for different industries and companies. For this reason, XBRL offers industrial- and enterprise-level extensions of general-purpose taxonomies, so entities can report information according to their particular needs.

In China, industry-level extensions include the Banking Taxonomy and the Oil and Gas Taxonomy. However, most of the filed financial reports using XBRL have been extended directly from CAS-T, because the companies filing the reports aren't in one of these two industries.

XBRL Business Reports and Big Data

There's a massive amount of online documents containing millions of rows of financial data. On 30 January 2009, the US Securities and Exchange Commission published a rule mandating the use of XBRL in reporting. Since 2000, all listed companies in China are required by the China Securities Regulatory Commission (CSRC) to post their annual reports on the Internet. SZSE and SSE started experimenting on listed companies' XBRL financial reports in 2004. Simply taking the electronic information disclosure of China's mutual fund industry as an example, we can see that between 20 July 2009—when the CSRC Fund Information Disclosure website (http://fund.csrc.gov.cn/web/login_do.login) went online—and 30 March 2012, the CSRC collected more than 510,000 instances.

XBRL is becoming increasingly popular among other Asian countries as well (such as India, Japan, Korea, Singapore, and Taiwan).

India created the XBRL Technical Advisory Committee to promote the development of the XBRL filing and dissemination platform. In Japan, all listed companies have been reporting full set of financial statements in the XBRL format since April 2008. The South Korea Financial Supervisory Commission's Data Analysis, Retrieval and Transfer (DART) reporting system (based on XBRL) officially went online in October 2007. In Singapore, the Accounting and Corporate Regulatory Authority (ACRA) is now leading an XBRL program. Since November 2007, ACRA has required all domestic companies, listed and unlisted, to file financial reports in the XBRL format. Taiwan has an XBRL demonstration platform (http://xbrldemo.twse.com.tw) with XBRL data for 37 listed companies.

All of these initiatives and XBRL-based platforms have generated a massive amount of financial data being stored on the Internet, contributing to the growth of big data.

XBRL in China

Here, we take a closer look at the role that the SSE, SZSE, and MoF are playing in the development of XBRL in China.

Shanghai Stock Exchange

The SSE first noticed the importance of XBRL when it started building its new information system in 2002. The following year, SSE chose 50 listed companies to be part of the XBRL annual report-filing pilot program. After a successful pilot project, the SSE decided to require all its listed companies to adopt XBRL in annual financial reports disclosures. The SSE also developed its own XBRL-report-filing software. The website, SSE XBRL Online (http://listxbrl.sse.com.cn/ssexbrl/companyInfoAction.do), has been providing the disclosure information of listed companies since 2008.

In 2005, the SSE issued the Chinese Listed Companies' Information Disclosure Taxonomy (CLCID-T), which has been approved by XBRL International. It defines 2,882 elements.

Shenzhen Stock Exchange

SZSE started formulating the Listed Companies' Electronic Information Specification (LCEIS) using XBRL in June 2003. After the Chinese Financial Standardization Technical Committee (CFSTC) approved LCEIS, SZSE immediately began fully implementing XBRL-based information disclosure. In 2005, the official website of SZSE opened a new section called XBRL Application Demonstration (XAD; http://xbrl.cninfo.com.cn/XBRL/index.jsp), letting the public download and browse instances and perform financial analysis online.

The CSRS has issued the LCEIS Taxonomy (LCEIS-T), which is mainly used by SZSE in its business reports filing and defines 2,679 elements.

Chinese Ministry of Finance

MoF is responsible for issuing accounting rules and making sure they're well applied in China. In 2010, MoF developed the CAS-T, a general-purpose taxonomy of CAS at the national level.

CAS-T is extended from IFRS-T using the *redefining method*.[5] In other words, CAS-T has only reused the core schema file of IFRS-T, and defined another 2,845 elements.

As noted, taxonomy extensions should be created for different industries. Furthermore, to make full use of XBRL decision-supporting benefits, extensions should be used not only in external reporting but also in internal reporting. Taking the banking industry as an example, China has created both the Banking Taxonomy and the Extension for Banking Supervision Statements. The latter has used the financial reporting (FINREP) and common reporting (COREP) taxonomies of the Committee of European Banking Supervisors for reference.

Challenges

In pushing for XBRL development, however, these institutions have faced both technology- and business-related challenges.

Taxonomies: Variety versus interoperability. Table 1 compares the three major taxonomies currently used in China.

Table 1 shows that the filers that adopt CLCID-T and LCEIS-T are only listed companies, whereas CAS-T is suitable for all organizations that must file financial reports. Regarding the specification, CLCID-T and LCEIS-T follow XBRL 2.1, whereas CAS-T additionally adheres to Dimension 1.0 and Formula 1.0. When looking into the taxonomies, we can see that the elements' naming

Table 1. Comparison between CLCID-T, LCEIS-T, and CAS-T.

	Chinese Listed Companies' Information Disclosure Taxonomy (CLCID-T)	Listed Companies' Electronic Information Specification Taxonomy (LCEIS-T)	Chinese Accounting Standards Taxonomy (CAS-T)
Issuing institution	SSE	CSRC	MoF
Implementation	SSE	SZSE	MoF
Filers	Listed companies in SSE	Listed companies in SZSE	Organizations that need financial reporting
Specification adherence	XBRL 2.1	XBRL 2.1	XBRL 2.1, Dimension 1.0, Formula 1.0, Versioning Base 1.0
Name rule of elements	Pinyin's C3 (Camel Case Concatenation) of the Chinese characters, such as ZiChan	Chinese characters, such as *资产*	C3 of the English characters, such as *FixedAssets*
Definition linkbase	Two pairs of essence-alias elements	No	3,833 definition arcs
Presentation linkbase	3,202 parent-child arcs	8,229 parent-child arcs	3,768 parent-child arcs
Calculation linkbase	753 summation-item arcs	1,433 summation-item arcs	1,168 summation-item arcs
Formula linkbase	No	No	Yes
Languages in label linkbase	English (2,863)—this means there are 2,863 English label values*	Chinese (2,687), English (1,038)**	Chinese (3,473), English (3,473)†
References in reference linkbase	CAS (prior version), guidelines (on content and format of information disclosure for listed companies)	CAS (new version), guidelines (on content and format of information disclosure for listed companies)	CAS (new version)

**There's no Chinese label value, only a Chinese annotation for every element in schema file of CLCID-T.*

***For the 2,687 label values, 1,038 elements have both English and Chinese label values. The other 1,649 elements have only defined Chinese label values.*

†There are 3,473 Chinese and English label values for 2,984 elements in CAS-T, and there are 119 Chinese label values and 118 English label values for 119 Extended Link Roles.

rules and labels' languages are different among these three taxonomies.[6] Consequently, even if there are several XBRL reports built directly based on CLCID-T, LCEIS-T, or CAS-T, it's still difficult to integrate the data of these reports.[7]

Due to the inability of XBRL to store the mapping rules among heterogeneous instances, there's a trend toward transforming XBRL into ontologies that can be easily linked to open data.[1,6,8]

Filing rules and comparisons among XBRL data. There are also differences between the filing rules used by the three actors. These differences, to some extent, weaken the comparability and interoperability between China's XBRL data.

When companies are preparing for the XBRL report filing, in some circumstances they must extend the general-purpose or industry-level taxonomy to report more effectively and precisely. However, both SSE's filing-software and SZSE's filing-website rarely let companies develop enterprise-level extensions. Therefore, the XBRL data in the SSE's or SZSE's information systems can be compared, because it has almost the same structure.

MoF encourages filers to build their own extensions using either the *reusing method* or *redefining method*.[9] The reusing method requires the filer to reuse not only the elements but also the linkbase definitions of CAS-T. The redefining method only requires the filers to reference the element definitions of general-purpose taxonomy. The reusing method is usually recommended, although most of the filers seem to prefer the redefining method.[8] Consequently, the extensions are heterogeneous, because they're developed by different accountants for different companies.

Different extensions sometimes define the same accounting terms with different elements when the terms aren't defined by CAS-T.

Business analysis using XBRL data. Business analysis identifies business needs and determines solutions to business problems by working on a trustworthy business dataset. XBRL is helpful to organize the massive business data and make it ready for analysis by tagging in advance.

High-quality data collection is the foundation of business analysis. Both SSE's software and SZSE's website provide users with a data validation function. Due to the real-time synchronization with SZSE's dedicated software, mistakes can be found immediately when data has been put into the online filing system. On the other hand, SSE's software offers only some built-in validation algorithms, so most of the mistakes will only be found by the more powerful validator on the server end or by a professional's manual work after data has been sent to the SSE's system. There are even fewer validating tools in MoF's XBRL-reporting system.[9]

To make the most out of the XBRL dataset, SZSE and SSE harness the power of the Internet to keep the data alive, thus providing the public with access to the details of companies' reports and letting people perform basic analysis tasks, such as comparisons. SZSE's XAD website (http://xbrl.cninfo.com.cn/XBRL/index.jsp) provides users with information services on listed companies' periodic reports and separate XBRL instance downloads, which is a great help in data distribution and reuse. SSE's XBRL Online also provides users with the latest XBRL reports and fundamental analysis. MoF has only published some financial reports of the year 2010 on its website.

Business Research Trends

XBRL makes it possible for analysts to turn massive amounts of financial and nonfinancial data into well-structured information using software tools. XBRL data is the source of new raw materials for business research, because it has the four "V" characteristics of big data: volume (tremendous data), variety (various data types), value (high value potential), and velocity (fast processing speed). Provided that improvements are made as far as standardization and interoperability, XBRL has potential in a several areas.

Integrated Data Analysis

The more comprehensive the data, the more it relates to and reports reality. Integrated analysis means that companies can potentially gain new insights from different perspectives, with business data integrated with processing details. However, the reality is that standardized and structured data accounts for less than 15 percent of the enterprise data. The rest is just a deep pool of unstructured and semistructured data. XBRL is currently one of the most able tools for integrating different sources of structured, semistructured, and unstructured data.

Full-Population Analysis

Traditionally, to make broader inferences at a low cost, organizations use a random sampling collection and analyze the sample. If and when the sample is representative of the population, the conclusion of the analysis will be more likely to be accurate. However, a bias in the sampling process will negatively affect the quality of the conclusions.

One way to bypass the challenge of sampling would be to dig out and capture all available data so that the sample will actually be the entire population of data. XBRL will standardize business reports so that population analysis can replace sample analysis of finances. XBRL can make the statement "sample = population" a reality.

Real-Time Analysis

Companies from various sectors rely on real-time information. In finance, constant access to data is vital. To analyze the tremendous amount of data synchronously, it's essential to achieve continuous monitoring, live reporting, and real-time auditing. Of course, this requires data interoperability across sectors. If XBRL taxonomies were to be further standardized or better mapped with each other, this would support real-time analysis. Users of XBRL would be in a position to analyze a large stream of events in real time, probing early into what happens and automatically responding to the changing world.

Forecast Analysis

Limited by data processing abilities, we tend to reason based on experience and thus understand the world based on the so-called "causality hypothesis," meaning that our conclusion is often

affected by our mode of thinking, inherent bias, or implicit knowledge in specific areas. However, the era of big data only requires us to know which of the data is correlated before making a forecast of the future. In fact, the data involved in enterprise business decisions is so complex that existing knowledge and experience aren't enough to fully understand the ins and outs. Therefore, exploring "what" are the relationships between XBRL data defined in the taxonomies—rather than "why" such relationships exist—can help us understand the future more thoroughly.

It's clear that big data, such as XBRL data, is becoming increasingly important in the business ecological chain. The value of XBRL data lies in combining the data with the major processes of a business, thereby helping decision makers achieve more thorough and timely analysis and solutions. In China, a significant effort is still required to standardize XBRL data, but the impact will be tremendous—integrated, full-population, real-time, and forecast analysis.

Acknowledgment

This article was supported by the National Science Fund of China (grant no.61272031) and the Humanities and Social Sciences Research Planning Foundation of Ministry of Education of China (grant no. 11YJA630012).

References

1. M. Spies, "An Ontology Modelling Perspective on Business Reporting," *Information Systems*, vol. 35, no. 4, 2010, pp. 404–416.
2. "Notification of the CAS-T" [in Chinese], Ministry of Finance, 2010; www.xbrl-cn.org/2010/1019/73285.shtml.
3. "Specification Recommendations," eXtensible Business Reporting Language International Community, 27 Feb. 2013; www.xbrl.org/SpecRecommendations.
4. H. Chang-Yin, "Research on Hierarchical Extension of XBR Financial Reporting Taxonomy" [in Chinese], doctoral thesis, Antai College of Economics & Management, Shanghai Jiao Tong University, 2012.
5. "Chinese Accounting Standards General Taxonomy Filling Rules" [in Chinese], Ministry of Finance, 2013; http://kjs.mof.gov.cn/zhengwuxinxi/zhengcefabu/201306/t20130603_902498.html.
6. L. Jimei, Z. Huizhou, and D. Meijie, "Detection and Resolution of Structural Conflictions in Heterogeneous XBRL Taxonomies," *Proc. 5th Int'l Conf. New Trends in Information Science and Service Science*, IEEE, 2011, pp. 312–317.
7. C. Jia-lan, "Using XBRL Taxonomies to Transfer Accounting Standards—Taking International Enterprise Consolidated Financial Statements as Example" [in Chinese], *Friends of Accounting*, no. 21, 2011, pp. 1–17.
8. L. Jimei, "Concept Similarity Assessment for XBRL Taxonomies Mapping," *Advances in Information Sciences and Service Sciences*, vol. 5, no. 10, 2013, pp. 339–348.
9. "Summary Report on the Implementation of CAS-T in the Local State-Owned Large and Medium-Sized Enterprises in 2012" [in Chinese], Ministry of Finance, 2012; http://kjs.mof.gov.cn/zhengwuxinxi/diaochayanjiu/201212/t20121228_723190.html.

Li Jimei *is an associate professor and department head in the Department of Information Management, College of Information Sciences, Beijing Language and Culture University, China. Her research interests include data mining, information systems, digital business reports. Jimei received her PhD in information systems from Beijing Forestry University. She is a member of the Association of Information Systems (AIS) and China AIS (CNAIS). Contact her at ljm7001@126.com.*

Hu Yuzhou *is a graduate student studying computer science at Beijing Language and Culture University, China. Her research interests include digital business reports, information retrieval, and semantic-based vertical search engines. Yuzhou received her bachelor's degree in information management and information systems from Beijing Language and Culture University. Contact her at huyuzhou0107@gmail.com.*

Du Meijie *is an associate professor and is the director of the accounting department in the International Business School at Beijing Language and Culture University, China. Her research interests include accounting, internal control, digital business reports, and auditing. Meijie received her PhD in accounting from Renming University of China. She is a member of the Accounting Society of China (ASC), China Audit Society (CAS), and China Data Management (DAMA China). Contact her at dumeijie@126.com.*