

1st International Conference on Data Science, ICDS 2014

Framework Formation of Financial Data Classification Standard in the era of the Big Data

Shuang Yang^a, Kun Guo^b, Jianping Li^{c,*}, Yingjiong Zhong^d, Rui Liu^d, Zili Feng^b^a College of Engineering and Information Technology, University of Chinese Academy of Sciences, Beijing 100049, China^b College of Management, University of Chinese Academy of Sciences, Beijing 100049, China^c Institution of Policy and Management, Chinese Academy of Sciences, Beijing 100190, China^d Economic Information System Department, Communicating Technology Bureau, Xinhua News Agency, Beijing 100803, China

Abstract

Deeper excavation of relevance of data and a top-down thinking to take apart financial data into blocks for more efficient analysis are essential for the big data, as well as to eliminate data noise and to remove data redundancy in the process¹. The financial data classification standard, which always performs excellently in these aspects, is an essential premise for data mining and analysis in the big data². To find a method to form the classification standard framework that can meet diverse purposes is very important. This research proposes a way to form the framework of financial data classification standard based on uniform classification standard and relative books, improved by comparing with classification standard of existed financial database and verified by practice with financial data sources. This framework can adapt to trends in the era of the big data and improve data storage mode.

© 2014 Published by Elsevier B.V. Open access under [CC BY-NC-ND license](http://creativecommons.org/licenses/by-nc-nd/4.0/).

Selection and peer-review under responsibility of the Organizing Committee of ICDS 2014.

Keywords: big data; financial data standard; data source; classification standard

1. Introduction

In the era of the big data, we need deeper excavation of relevance and a top-down thinking to take apart the financial data into blocks for more efficient analysis. Meanwhile, to eliminate data noise and to remove data redundancy in the process become increasingly important³. The financial data classification standard, which always performs excellently in these aspects, is an essential premise for data mining and analysis in the big data. The most typical case is the development and implementation of index pointers. Nowadays, various indices of our life fill in our vision, such as the consumer price

* Corresponding author. Tel.:86-10-59358805
E-mail address: ljp@casipm.ac.cn

index (CPI) and the stock index. Another representative is financial statistics system of the International Monetary Foundation (IMF) with Monetary and Financial Statistics Manual (MFS)⁴. It covers a series of financial data standards to guide the harmonization of national data and statistics, publications and amendments of each country.

How to efficiently play the role of premise of data mining and analysis is one of the important goals of creating financial data standards. To find a method to create the classification standard that can meet diverse purposes is very important. However, there have not been any researches to explore this problem. Existed financial data standard researches mainly focus on their disadvantages and comparisons between regional classification standards and international standards or practice aspects, especially in China. Like Yang C⁵, this research is one of the earliest to discuss the problem of statistic standards in China. Recently, many researchers have carried on some analogous and further work, like Yang FJ⁶, Huang XY⁷, et al. This research not only concentrate our attention on differences and combination between data standards, but also explore a way to integrate them with practices, then to create a classification framework that is more suitable for data mining and analysis in the big data.

As the preparation for the way of creation, we should deeply learn mainstream international uniform classification standards and their connections as well as the existed classification standards of typical financial databases. Because uniform standards can paved the way for the train of thought and give the theoretical basis. Meanwhile, existed standard of typical financial databases can perform as references to improve the standard framework. But the theory is often an idealized pattern. Even if it has specificity and irreplaceability, it also needs to adapt to practice. Then the classification standard will be verified by data sources, and a corresponding table should be created.

2. Unified international financial data classification standards

Mainstream financial classification reference of the world can be roughly divided into two categories. One major is in accordance with the product or industry categories, such as the Central Product Classification (CPC)⁸, the International Standard Industrial Classification of All Economic Activities (ISIC)⁹ and the Classification by Broad Economic Categories (BEC)¹⁰. This type is simple and clear, based mainly on the main participants in the process of economic to carry on statistics, division and finishing. The other classification system is strict statistical classification standard like the System of National Accounts (SNA)¹¹ and the Monetary and Financial Statistics (MFS), these strict statistical classification principles are closely related to the process of economic and make the rules in strict accordance with relevant statistical codes and tables

2.1. International financial statistics classification standard and contrast

2.1.1. System of National Accounts(SNA)

System of National Accounts (SNA), also known as the System of National Economic Accounts, is based on economic theories in the West accounted for the Gross National Product (GDP) as the core indicators of national economic activity. The SNA system is the most important classification criteria of United Nations Statistic Division (UNSD) statistical classifications and also the core reference of established account classifications of the major statistical agencies. From the world view, SNA, on the consistency of the statistical basis for economic activities, plays a vital role and is convenient on comparing to each other's national accounts for all countries. SNA system is designed for all states as to adapt to the needs of every country in different stages of economic development. It provides an overall framework, which can combine with other areas of statistical standards, and promote the integration of the statistical systems to achieve the consistency with all the national accounts.

2.1.2. The Monetary and Financial Statistics Manual (MFS)

The Monetary and Financial Statistics Manual, prepared by The International Monetary Fund (IMF), is the basic methods and principles of international monetary and financial statistics and the main base of financial statistics by central banks. It is the first time to establish for the conceptual and methodological guidelines in monetary and financial statistics, becoming principles and basic methods of international monetary and financial statistics. From the perspective of process and management of financial operation, MFS absorbed the balance sheet and funds flow table of expressed way in SNA and designed various financial total indicators to guide statistics, such as statistics on currency, credit and the other financial flow of space distribution, classification statistics on assets number changes by causes, statistics in the form of a memo contingent assets caused by guarantees and commitment and statistics reflect on all of stock in the type of balance sheet. In order to effectively measure and analyze these elements, MFS is in preparation of the relevant tables, forming the basic framework for monetary and financial statistics.

2.1.3. The contrast between SNA and MFS on Financial Statistics

On the whole, MFS's statistical standards is based on the SNA system, and SNA statistics standard is the full framework of organization, measurement and processing economies in the non-financial and financial activities. Therefore, basic principles of the SNA determine the statistical principles of MFS. However, due to both the object and scope of different statistics, SNA's monetary and financial statistics and MFS have major differences both on architecture and in the details. The contrast between them is reflected in Table 1.

Table 1. The contrast between SNA and MFS on Financial Statistics

Contrast	SNA	MFS
Statistic purpose	The overall operation of the business accounting of national economy	The expression of national monetary and financial statistics
Theoretical basis	The social reproduction theory	Financial theory
Statistical content		
1.	Total index (stock and flow)	All kinds of total financial statistics
2.	Assets change accounts	The financial flow changes
3.	Revaluation account	Contingent assets memos
4.	Foreign direct investment memos	
5.	Financial output and application	
The basic framework		
1.		Broad overview of the financial department
2.	(Included in the indexes of SNA)	SNA balance sheets and flow meter
3.		(Divide into financial assets and other financial instruments. The classification of financial assets is different from SNA.)
Classification of financial instruments	(Without clear classification, financial instruments can be divided into seven different categories in according to the stock and flow.)	
Department classification		
1.	Non-financial sectors	Be basically the same with the SNA.
2.	Financial firms	Financial firms are further subdivided.
3.	General government department	
4.	Nonprofit organizations providing service	

2.2. International financial classification standards by products or Industries and their associated

2.2.1. Statistical Data and Metadata Exchange (SDMX)

Statistical Data and Metadata Exchange (SDMX) ¹² was initiated by seven organizations of authoritative

metadata standard, and seven organizations are the Bank for International Settlements (BIS), the European Central Bank (ECB), Eurostat, the International Monetary Fund, the Organization for economic cooperation and development (OECD), the United Nations (UN), the World Bank. It has been certificated by ISO to become a common data standard and has been adopted by many states and combined with its integrated application status to ensure their proper functioning of data collection, analysis and dissemination.

2.2.2. International Standard Industrial Classification of all Economic Activities (ISIC)

Currently, international common industry classification standard is the International Standard Industrial Classification of all economic activities (amended in 2006, referred to as ISIC Rev.4), which is prepared by and referred by the Statistics Division of the United Nations (UNSD). ISIC classification is not intended to replace national industry standards, but provides a reference framework for international comparison of statistics. The United Nations recommends every state to develop national industrial classification through making full use of the same principles and definitions to make their standards be coordinated with ISIC as much as possible.

2.2.3. Central Product Classification (CPC)

Central Product Classification (CPC), covering the complete product classification of goods and services, is intended to act as an international standard through pooling and tabular lists various requirements give a breakdown of the data, including industrial production, national accounts, services, domestic and foreign trade in goods, international trade in services, balance of payments, consumption, and price statistics. It also intends to provide a framework for international comparison, promoting harmonization of statistics on a variety of goods and services. Its fundamental purpose is to provide a tool for coordinating economic statistics on goods, services and assets for international comparisons, and provide useful guidelines for the country to develop products for the initial classification or amend the existing approaches, making their classification in accordance with international standards.

2.2.4. Classification by Broad Economic Categories(BEC)

Classification by Broad Economic Categories (BEC) is a commodity classification system for international trade commodity statistics, which is a very important part of UNSD. BEC is compiled by rearranging the standard international trade classification (SITC)¹³ project number, under economic categories of merchandise integrated summary data in the development of international trade, in accordance with the main end-use of international trade commodities. Through BEC classification, can make prepared trade data by international trade standard classification under SITC converse into three basic goods categories in national economy accounting system under (SNA) framework: capital products, and middle products and consumer. This operation is convenient for putting trade statistics and national economy accounting and the industrial statistics, other basic economic statistics to combining up for analysis of country economic, and regional economic or world economic.

2.2.5. Associations between these standards

Both CPC and ISIC are the general classifications, but ISIC is on behalf of these two interrelated categories of activities of the United Nations. Each subprogram of the CPC is by the vast majority of a specific level or several levels of ISIC in the production of goods or services. In order to take care of users who wish to clear relations between CPC and ISIC, each secondary reference to ISIC industries generally produce most of the relevant goods or services. Next to the secondary about CPC listed in a table, list the corresponding 4-digit ISIC code to show the leading level of the international standard industrial classification.

The relationship between CPC and BEC is similar to the relationship with SDMX, because BEC also uses classification of SDMX as basic components to build more suitable for economic analysis of trade grouping of commodities. About BEC and CPC correspondence in transportable goods, under normal circumstances, BEC's 5-digit project is all contained within the CPC within a single secondary sector from 0 to 4. Thus, CPC secondary is consisting of one or more items from BEC. Since BEC deals only with transport goods like SDMX, in terms of the financial sector of CPC categories, there is no correspondence between CPC and the BEC as well as that between CPC and ESMX.

3. Major financial database classification standards and their data sources

Covering the world's major financial databases, in micro-financial data, classification standards are mainly in accordance with product classifications- CPC and BEC to gain the specific distinctions; in terms of macroeconomic data, they are basically based on SNA and MFS, whose sector classifications for the collection and collation of data are essential references. Their data show obviously and clearly, reflecting layers of depth, we call it "deep-in-generally". Characterization of it is by clicking on the associated markup into the next interface for the need step by step until reaching the desired interface. Advantages of this approach are clear interface, cause affiliation clear and easy search for redirect lookups. Disadvantage is that the complex progressive relationship, data within the database easily clutter redundantly and associated data is not accurate.

Principle data of databases and their corresponding sources are listed in Table 2. The main data sources of the mainstream database are shown in Table 3.

Table 2. The principal data of the main financial database and the corresponding source

Principal data	Data content	The main source of data
Real-time financial information services	Source of information around the world will bring together a variety of real-time news politics, finance, commodities, etc. to edit center	Central Bank Government Exchanges
Stock Market	Real-time and historical data of the world's major markets	Exchanges Trading Centers of Foreign Exchange Market between banks BIS
Macroeconomic and financial information	Understanding the macro statistics source release cycle of the world, timed release information	Bureau of Statistics IMF BIS World Bank
International Trade Information	Specific data on each country's trade	IMF OECD
Trend Analysis	Financial information operators will employ many senior economists, bankers, financial experts and analysts responsible for writing daily financial market commentary and trend analysis, and they will adopt relevant financial institutions analytical data	Own analysis IMF World Bank Major financial institutions
Technical chart analysis	Financial information providers provide users with diagramming tools. Users can use the diagramming tools to draw various technical charts.	Own technical support

Table 3. The main data sources of mainstream databases

Source	Reuters	Bloomberg	Wind	CEIC
I.	Financial information from 242 exchanges and direct trading markets	The world's major large trading venues	Chinese three major stock exchanges	121 National Bureau of Statistics
II.	Statistics information of major countries and regions	Bureau of the major developed countries and regions	Major foreign exchanges	IMF - World Economic Outlook
III.	The database information of major organization in the world	Database of major organizations in the world	The world's major NBS	United Nations Economic and Social Commission for Asia-Pacific
IV.	Its own powerful news gathering and analysis capabilities	Its own powerful news gathering and analysis capabilities		

4. Financial data standard establishment under the perspective of the era of big data

4.1. The principle of establishing the financial data standard in the big data

In the era of the big data, a relevance relationship attracts increasingly more attention than a causal relationship. In a traditional statistical analysis, the key reference is the reliability of the causal relationship. Google is walking in the forefront of this road. The Google Trends developed by Google is used by researchers to predict the stock market's gyrations. Through tracing 98 search keywords, they found the waist keywords about the stock market up or down¹⁴. Establishing new financial data standard is conducive to help us alter our traditional modes of causation, but the premise is that the standard is still built on the basis of extensive knowledge of experience accumulated and data correlation by the deeper mining. Not only can we adopt to knowledge of precipitation and lessons of essence that present financial data classification criteria creating experiences gain, but also analyze and mine existing data with more scientific instruments. Thereby, establish a different standard for data classification framework, which will be different from previous graphical tree structure, or may be a complicated network forms of organization.

4.2. Lay a foundation for the financial data standard establishment

According to the second part of the introduction to international uniform data classification standards, there are correlations between each other. The harmonization of concepts and statistical classifications is existed between MFS and SNA. Meanwhile, the financial classification criteria among CPC, ISIC and BEC can be correspond to and combined with each other. Likewise, the harmonization of classification standards can apply to the finance area. Therefore, correlations between reference standards have reached a level to lay the foundation for the financial data standard establishment.

4.3. The formation steps of the financial data classification framework

- Determine the first and the secondary title and their serial number, then correspond the detailed specific categories to the referenced international uniform data classification standards¹⁵. The correspondent is shown in Table 4.

Table 4. Financial data classification and its corresponding referenced standard

No.	Main category	Sub-categories	Classification and description of the corresponding reference
A	Economy	A1 Globaleconomy	Classification by Broad Economic Categories(BEC)
		A2 Macroeconomic	SNA's eight categories and corresponding indicators
		A3 Sectors of the economy	International Standard Industrial Classification of all Economic Activities (ISIC)
		A4 Microeconomic	Company Fundamentals
B	Financial	B1 Currency	Financial classification have two dimensions: 1.MFS Product categories; 2.MFS sector categories. The two dimensions form stereo graph core index, the core part of stereo graph is connected, the other parts are associated with this component a large three-dimensional metadata interface. But MFS also has some characteristics of account data in classification, so still need to learn the product classification of CPC to adjust and feedback, and also to update with the external force.
		B2 Bank	
		B3 Stock	
		B4 Derivatives	
		B5 Fixed Income	
		B6 Fund	
		B7 Exchange	
		B8 Others	
C	Trading	C1 International trade	SITC(Standard of International Trade Classification)
D	Government	D1 Government Finance	GFS(Government Financial Statistics)
E	Analysis	E1 Volatility	Correspond to the appropriate analytical tools, including the introduction of existing and own.
		E2 Trend	
		E3 Icon	
F	Information	F1Economic information	Mainly base on the classification of Xinhua news agency itself, then to do a little adjustment.
		F2Financial Information	
		F3Government financial information	

- Referenced on the corresponding uniform financial classification standard and relative books, like Finance¹⁶ and Investments and Portfolio Management¹⁷, gain more detailed categories.
- Compared with the main stream financial database about data storage and data sources among all the elements shown in Table 2 and Table 3, to make the data standard framework be more comprehensive and more convenient for further completing standard.

4.4. Realizing the correspondence between data storage and data sources

It is needed to implement the correspondence between data storage and data sources and complete the verification. Through learning experience of deconstructions of the mainstream financial databases, the components concluded for the financial database are shown in Table 5 that gives Chinese and American financial data source as well as overall world data source as examples for verification.

Table 5. The correspondence between data storage and data sources

Main category	Sub-categories		China	United States	Overall World
economy	World economy			University of Maryland University of Michigan PACAP	World Bank Database Yale UN database United Nations
	Macroeconomic		Bureau of Statistics China Economic Information Network	University of Michigan U.S. Census Bureau Fojiliya University SDC Department of Commerce	World Table OECD
	Sectors of the economy		Bureau of Statistics	U.S. Department of Commerce Haver	OECD
	Microeconomic		CSMAR Database	S&P Compustat、IBES	
Finance	Currency		People Bank of China	Fed	IMF
	Bank		People Bank of China	Fed	IMF、BIS
	Stock	High frequency	Exchange	NYSE TAQ	IMF、BIS
		History	SFC	S&P Compustat、CRSP	IMF、BIS
	Futures		Exchange, Commission	CME	IMF、BIS
	Fixed Income		Debt board	GovPX	IMF、BIS
	Fund		China Fund Network	S&P Compustat	IMF、BIS
	Exchange		China Foreign Exchange Trading Center	SDC The Fed.	IMF BIS
Trading	Others		Bureau of statistics (NBS)	The Fed. the U.S. Commission	IMF、BIS
	International trade		Bureau of statistics (NBS)	Department of Commerce American University	World Table OECD
Standard	Data Standards		SAI GLOBLE (SAI World standard library)		
Government	Government Finance Data		IMF's Government Finance Statistics (GFS)		
			NetEYEInc (a global government databases created by StreetEYE)		

The financial databases in the figure above are the database defragmented only one or two times. To some degree, they mostly didn't experience data standardization and processing of major database, thus, the data retain original features and structure. It makes it easy for data standardization from the basic level as well as the collation and analysis of data from the view of our own financial data classification standard.

5. Discussions and conclusions

This research makes financial data classification standard framework actually involve in mutual relationships with databases in the content and adapt to trends in the Big Data for its abilities to deeply excavate relevance of data and take apart a great number of data into blocks, which is easy and convenient for data mining and further analysis. At the same time, it proposes a new way for financial data storage, eliminating data noise and removing data redundancy in the process.

Even though, in this research, uniform standards, experience of existed mainstream databases and data sources are main factors¹⁸ to influence the data classification standard's utilization efficiency, there are still various other factors that should be involved in the establishment of financial data classification standard in the era of the big data. More attention ought to be concentrated on the other factors in the completeness of the total standard.

Acknowledgements

This research is supported by Xinhua News Agency and China Finance Corporation.

References

1. Mayer-Schönberger V, Cukier K. *Big data: A revolution that will transform how we live, work, and think*. Boston:Houghton Mifflin Harcourt;2013.
2. Shiller R J. *The new financial order:Risks in the 21st century*. Princeton, NJ: Princeton University Press; 2013.
3. Haag S, Cumming M. *Management information system for the information age*. 9th ed.New York: McGraw Hill Higher Education ; 2012.
4. International Monetary Fund. *Monetary and Financial Statistics(MFS)*. Washington: International Monetary Fund Publication Services; 2008.
5. Yang C. A comparative study of SNA and British or American accounts system. *Economic Issues in China* 1996;(5):57-63.(in Chinese)
6. Yang FJ. The research for differences between MFS2006 and MFS2000. *Statistics and Decision* 2010 ;(3):152-153.(in Chinese)
7. Huang XY. The differences between principles of MFS and Chinese Practice.*Economic Research Journal* 2011,31-32.(in Chinese)
8. United Nations Statistic Division. *Central product classification (CPC)*.Rev.2.New York: United Nations Publication; 2007.
9. United Nations Statistic Division. *International standard industrial classification of all economic activities(ISIC)*.Rev.4. New York: United Nations Publication; 2008.
10. United Nations Statistic Division. *Classification by broad economic categories(BEC)*.New York: United Nations Publication; 2008.
11. United Nations Statistic Division. *System of National Accounts2008(SNA)*;2008.
12. BIS, ECB, Eurostat,UNSD,IMF,OECD,World Bank. *ISO/TS 17369:2005 Statistic Data and Metadata eXchange*; 2005.
13. United Nations Statistic Division. *Standard international trade classification (SITC)*. Rev.4. New York:United Nations Publication; 2008.
14. Tobias P, Helen SH, Eugene S. Quantifying trading behaviour in financial markets using Google trends. *Scientific Reports* 2013; 1684:1-6.
15. CFSTC. *GB/T 21076-2007 Securities and related financial instruments—International securities identification numbering system*; 2007.
16. Huang D. *Finance*.9th ed. Beijing: Renmin University of China Press; 2008.
17. Zvi B, Alex K, Alan JM. *Investments and portfolio management*. 9th ed. New York: McGraw Hill Higher Education;2011.
18. Xu DL, Li ZL. Research on monetary and financial statistics in SNA.*The Theory and Practices of Finance and Economics* 2002 ;(5):75-79.(in Chinese)