# Statistical Metadata Research

Name: Lu Chunlian
*Address: No.75 Yuetannanjie Sanlihe Beijing P.R.C.*
*Email: lchl@stats.gov.cn*

Abstract
This article expounds the production, function and evolution of the metadata. It expatiates the conception, contents and progress of the semantic metadata, business rules metadata and operation rules metadata of the statistical metadata. It illustrates a method to build a statistical metadata bank and shows its role for statistical data processing, retrieving and disseminating.

## 1. Introduction

Since 60d the metadata are mentioned as the data that describing the digital data. The statistical metadata are the semantic definitions of the statistical data. In fact the metadata display the multidimensional meaning of the statistical data. The metadata research is developing as the data processing and exchange technique, database technique, data warehouse technique, oriented objects technique, and so on. It has been expanded what metadata are today. The metadata are classified differently in different application area. I think the statistical metadata should be classified into three kinds in official statistical system.

## 2. Statistical semantic metadata

The semantic metadata are the basic metadata of the statistical data of cause. The essential meaning for any statistical data is the basic semantic metadata. We usually called it as the name of the index. The name can be the very detail term or the only master term. By my experience the index name should be the master term. Other describing parts should be decompounded into essential components. Every component should be the steady meaning and be used for many different master indexes. According to the stability of the statistical semantic metadata when they are used in statistical data processing and retrieving I classified the statistical metadata into seven kinds: the master index, the time character, the spatial character, the statistical classifying character, the statistical grouping character, the used measurement, the scale of measurement. The character from one to five above called as the category character and the character from six to seven above called as summery character. And the statistical classifying character of them is not only one kind of classification. One master index may have several kinds of classifications. The same situation is for the statistical grouping character. Therefore it does the reason that the number of dimension of the practical statistical index is not changeless.

## 3. Statistical business rules metadata

As you know, the statistical data is the measurement of the activities of the economic, technique and society. There are many kinds of statistical activities and measurement methods. For example, the value of one index is obtained from the statistical activity of government statistics or the survey activity of industry statistics. Or the two values of one index are obtained from different statistical deducing methods. The people want to know more character from the data produced rules then the semantic character of the statistical data. The analyzers of statistical data want know who define the meaning of the indexes, how the data obtained, what arithmetic method is used to derive the data, and so on. We called those characters of statistical data as business rules metadata. The management and use of those metadata are progressing as the database technique and data warehouse technique developing.

## 4. Statistical operation rules metadata

When you do data processing you need to know the edition rules and auditing rules for data pre-processing and to know the formula of summery and condition rules of accumulation for data machining. When you retrieve the database or data-warehouse you need to know the theme keys and the query paths and you want produce the own favorite report tables. The convenient and intuitionistic operation to processing or retrieving statistical data by users own opinion is welcomed more and more. So, the program developers want to storage, manage

and use those operation rules neatly.

**5. Statistical metadata databank**

The statistical data processing must include the operation of metadata. The statistical data would be nothing of meaning without relative semantic metadata. The metadata are stored as a data file when the data processing adopts the file system fashions. The same metadata are exhibited in many different data files for many different applications accordingly. The metadata are stored as several relational tables when the data processing and retrieving based on the relational database management system. The metadata in relational tables could be used in many different applications that are constructed upon the same relational database. The same metadata could be stored as different relational tables in different relational database systems constructed by different developers. This situation is not a good status for the statistical information exchange and disseminating conveniently and cost-effectively. It is obvious that the independent statistical metadata databank should be constructed. It should provide the uniform statistical metadata structure, uniform definitions of metadata, and uniform maintaining methods and rules for those metadata. In fact, it will become one of the most important soft establishment facilities for the whole statistical data exchange, data machining, and data disseminating, and so on.

The structure of metadata database is growing rapidly. Many data-warehouse commercial products the semantic metadata database is a multi-dimension database generally. In 1991-1993 year my developing team and I developed such as a metadata database parts in the project of Comprehensive Statistical Database Application system. It supports the willful accessing to the database and making a freedom format-reporting table. From 1999 year we construct an independent statistical metadata databank through revising the original semantic metadata structure, replenishing the statistical tables describing operation rules, supplement the edition and auditing operation rules and. summery or machining rules. Last year we began to add the business rules into the statistical metadata bank. The inspecting should be obtained for this work from the ISO 11179 standards and revising some part of the metadata bank.

**6. Conclusion**

The statistical metadata databank is the most important system for the statistical information processing and application system. The metadata databank will spur the reform of statistical survey plans and activities. It will be the basic level to support the structure the on-line statistical business processing system and the statistical data warehouse. Finally it will improve on the dissemination of statistical data to the all level of government and the public taxpayers.

References:

1. Designing Research of Statisticaldatabase

Bulletin of the 50th International Statistical Institute Beijing 1995

2. Research to Basic Structure of State' s Statistical Information System Proceedings of 1998 2nd International Conference on Information Infrastructure ICII' 98-Beijing

Publishing House of Electronics Industry

3.Evolving data collection techniques and technologies

Seminar on Application of Information Technology in National Statistical Offices 1998

Sommaire

Cet article traite sur les méta-données statistiques. L' article couvre la conception, le contenu et le progrès des méta-données sémantiques, les règles d' affaires des méta-données et les règles opérationnelles des méta-données statistiques. Cet article illustre une méthode pour construire une banque de méta-données statistiques et démontre son rôle pour le traitement, l' extraction, et la diffusion de données.