Traditional Chinese Medicine Information Digitalization Discussion

Qi Xie, MSc; Meng Cui, MSc; Zhen-Dou Wu, MSc; and Hong Zhao, MA

Abstract

**Background:** With the rapid development of information science, the ancient Traditional Chinese Medicine is combining with it rapidly, and forming a new discipline: Traditional Chinese Medicine (TCM) Informatics. TCM information digitalization is the process of digital processing, which uses modern information technology to obtain, process, store, and analyze TCM-related data, information, and knowledge. It gathers research, application development, and service in an integrated whole.

**Methods:** This article systematically analyzes the key research issues of TCM informatics (e.g., on data resources, data standard, data system construction). Also, the methodology and technology of TCM information digitalization research are thoroughly discussed.

**Conclusions:** The starting point of the research on traditional Chinese medical information digitalization was in question. The research from the current study research was drawn from collected information that was stored, transferred, and utilized. This process helped to place an emphasis on the topic, as well as extending its research areas. In addition, an innovative TCM information virtual study center was set up to support a great deal of fundamental work.

Introduction

The worlds of science and technology have developed dramatically in the 21st century. The integration of life sciences, information science, complexity science, and systems sciences has created many new branches such as Traditional Chinese Medicine (TCM) Informatics. TCM information digitalization involves collecting, processing, analyzing, and storing TCM-related data digitally. It combines research, application development, and service. This is a complicated and open project, which involve many aspects of TCM. It is also a long-term task that needs several generations’ efforts. In recent years, by the joint efforts of state government and national TCM information researchers, the research on TCM information digitalization has made substantial progress.

TCM information digitalization plays an important role in gathering TCM information services. It has become the indispensable part of TCM science and technology innovation, and is greatly supported by state government. The field of research includes four major areas: data resources, data standard, data system construction, and data utilization.

Research on Data Resources

Research on data resources is the premise of TCM information digitalization. Its emphasis is on science data collection and study of literature resources. Science data collection centers on research of resources used during the innovation of science and technology. Over the last 50 years, all levels of government have greatly supported massive TCM science and technology projects. Those data results are very valuable. To organize and utilize these data resources adequately could avoid duplication, advance research levels, and speed up science and technology development. Currently, the study of science data collection has been piloted in many Chinese colleges (Fujian College of Chinese Medicine, Shanghai University of Chinese Medicine, and China Academy of Chinese Medicine Science). It involves three basic principles of policy and technology, mainly focusing on the collection policy, platform, and data.

TCM literature is the essence of science and technology innovation achievements. It has been reported that 90% of all TCM information comes from literature. To classify this literature better, building a united catalog became the foundation...
of the overall planning. They can be categorized by type (periodicals, books, and other TCM resources), time (modern literature resources 1949 to today, the Republic of China literature resources 1911–1949, and the ancient literature resources before 1911), and geography (domestic and overseas).2

Besides the classification of TCM information, extensive developmental research has been done. Recently, the fastest progress has been made with regard to TCM ancient books. Since 1950s, there have been three instances of nationwide investigation on TCM literature resources. During the 2005 investigation, we obtained the bibliographic data from 151 different kinds of TCM literature, which were published before 1949. From those, we acquired the collection and distribution of many TCM ancient books (published in the Complete Bibliography of Ancient Chinese Literature of TCM).

The other achievement is to actively promote the “Memory of the World” program, thus driving more resources for investigation. “Memory of the World” was established by UNESCO in 1992, aiming at preserving valuable assets and library collections worldwide. The Chinese government established a special agency to carry out this work. Since 2005, the State Administration of Traditional Chinese Medicine has led an initiative to further survey resources in collaboration with Chinese libraries related to TCM literature.

Data Standard Research

Data standardization is involved in the collection, processing, and treatment of data. It not only ensures that the right data are inputted into the system, but also sets the foundation for utilizing and sharing. At present, research work has been focusing on the following items.

“TCM Subject Index” divides TCM data into 15 large classes and 68 different subclasses. It includes 8307 standard terms and 5598 searchable words, similar to what will be used by the Medical Subject Headings (developed by National Library of Medicine in the United States).

“TCM Language System” aims to convert text from the ordinary language of TCM to computer languages so it can be searched within a computer database. This system’s development is based on the framework of the Unified Medical Language System (UMLS; developed by the National Library of Medicine in the United States), with full consideration of TCM’s unusual language. At present, the system has already collected 320,000 vocabularies, 130,000 definitions, and 1,270,000 word relations. It can already support literature retrieval and database creation. However, those vocabularies still cannot cover all of the database literatures. The research for full coverage is still in progress.

“TCM Clinical Terminology” digitalizes clinical medical records to meet clinical research needs. Its framework is influenced by the internationally recognized Systematized Nomenclature of Human and Veterinary Medicine Reference Terminology (SNOMED). However, because of the difference between TCM clinical practice and Western medicine, this resource’s “principal axis” is quite different. Lately, it has collected 130,000 words, and continues to make progress.

“TCM Metadata Standard” includes many important TCM metadata, in additional to the current standards of “Metadata in Science Data Sharing” and “the Metadata Standard in Medical and Health Science Data Sharing.” The standard is still being enhanced.

“TCM Data Resource Classification Standard” organizes the TCM data resources into separate categories. It avoids data overlapping, and covers all TCM data resources. The first draft plan includes 5 primary classes/39 secondary classes/124 three-level classes. This will require up to 329 databases to complete. The primary classes include the TCM program’s resources, data resources of TCM, data resources of Chinese Material Materia (CMM), data resources of acupuncture, and data resources of TCM ancient books. The classification covers all of TCM data resources and will be a prioritized area for the next 2 decades.

Research on Data System Construction

Data system construction includes single-table database, structure database, and database interface. There have been many attempts that were beneficial in the overall database design. With Chinese medicines as the focal point, a massive database was created to connect the Chinese medicine database group and TCM database group. This project started in 1999, and is continuing to be improved.

Single-table database construction has made substantial progress. The periodical database has included some TCM literatures from 1911 to 1949, all from 1949 to present, and all abstracts from 1984 until now. In total they account for more than 800,000 records, and are categorized into 15 database with special topics. There are 15 special-subject databases provided by the Chinese TCM periodical literature database, including a CMM documents database, TCM geriatric disease documents database, CMM chemical documents database, TCM famous doctors experience database, CMM pharmacologic documents database, TCM clinical diagnosis and treatment documents database, CMM adverse reaction and toxicological documents database, TCM clinical trials documents database, acupuncture documents database, TCM history documents database, tumor documents database, TCM research subjects database, sexually transmitted disease (STD) TCM-intervention documents database, TCM digests database, and Human Immune Deficiency Virus/Acquired Immune Deficiency Syndrome TCM-intervention database.

Furthermore, more than 30 different kinds of TCM factual databases have been built, covering almost every aspect of TCM (including medicine database for TCM, China Tibetan, China Yao, and CMM database, Chinese formula database, modern clinical application of TCM formula database, chemical components of TCM medicine database, medicine reaction database, disease diagnosis and treatment database, etc.). The “Fujian University of Traditional Chinese Medicine database” includes 19 TCM-related database mainly focused on the Taiwan area (such as Taiwan wild edible plant database, Taiwan medical plant resource information database, Taiwan TCM scientific research database, Taiwan TCM database, Taiwan TCM database, and Taiwan TCM dissertations).

The Structure database consists of the CMM foundation database, Chinese medicine pharmacology tests database, CMM chemical experiment database, and TCM clinical database group (contains 17 disease databases right now). All the collected data have been queued for simple statistical analysis, which will help new medicine development and clinical decision-making.
The current database interface allows the users to obtain the necessary information from different interlinked databases. It includes the decision supporting interface of public health emergency response plan, “toxic” CMM data interface, TCM ancient books data interface, TCM prevention and intervention of tumors data interface.

**Research on Data Utilization**

The purpose of TCM information digitalization is to enhance the overall utilization of the TCM data. Improvements include data sharing, analyzing, and customer’s special needs.

There are three existing research works to stimulate the sharing of TCM scientific data:

- **TCM scientific data grid service applied research**: Perfecting the data grid project/its service tools, knowledge management analysis system, objective analysis, and defining system.
- **Study of TCM clinic optimization and efficiency evaluation interface based on clinical literature**: Treatment research plan optimization of 22 diseases, clinical efficiency evaluation research of 22 diseases.
- **Study of independent evaluation on TCM clinical efficiency based on clinical literature**: Pilot research on evaluation of periodicals, literatures, data, plans, efficiency, and safety.

TCM information digitalization research provides powerful information support for scientific study. According to statistics, from 2002 to 2006, 80 GB of scientific data services were offered to the public. So far up to 3,000,000 Internet data search requests, and up to 1200 GB of data have been downloaded by more than 10,000 personnel with around 1000 organizations. Analysis shows that among them, educational institutions make up 34% of all of these organizations, scientific institutions 28%, enterprises 23%, social groups 5%, foreign organizations 5%, and others 5%.

**Conclusions**

To ensure the research quality and its continuous development, TCM information digitalization has created a national working team, which is called “TCM scientific data center,” built with the support of “scientific data sharing project of national science and technology infrastructure platform—national medical scientific data sharing network.” An experienced team has been established and trained for TCM digitalization. More than 300 TCM professionals from 35 TCM colleges and academies all over China work together to form a TCM information virtual study center. Participants worked closely together to set up a virtual workspace. They have standardized data flow and strengthened procedure management. Considering the weak capacity of the TCM information sector that hinders it from undertaking large projects, this is a major breakthrough, which will greatly encourage the development of TCM information work.

**Acknowledgments**

Funding for this study was provided by Science Data Management and Sharing Service Center of Chinese Material Materia (Grant 2004Dka20250 from the Ministry of State Science and Technology of the P.R.C. Platform item of science basic condition), A Study of Traditional Medicine Meta Conceptual Model with Domain Ontology Method (Grant 30973716 from National Nature Science Fund).

**Disclosure Statement**

No competing financial interests exist.

**References**


Address correspondence to:

Meng Cui, MSc
Institute of Information on Traditional Chinese Medicine
China Academy of Chinese Medical Sciences
No. 13, Baijiazhuang dong, Chaoyang district
Beijing 100700
People’s Republic of China
E-mail: cm@mail.cintcm.ac.cn