Studies of Chinese Digital Health Resources

Metadata (CDHR-m)

Zhu Dong-ping

Medical information Department, Fuzhou Dong-fang Hospital, Fuzhou, China, 350025

Abstract

Background: Using metadata to solve the discrepancy between the exploding medical information and the low efficiency of organization. Retrieving such information is an interesting topic.

Purpose: To Establish a set of metadata application schema by using DC metadata in organizing and dealing with Chinese digital medical information metadata (named CDMI-m), in accordance with the characteristics of contemporary medical information and can be applied to network and digital medical information resources.

Method: (1) This thesis gives a general survey of the development of medical information organization and data description. (2) Four major medical metadata projects in the world are being introduced and their similarities and differences are explained. (3) Put forward the framework of CDMI-m based on the theories and practice stated above. (4) To construct the CDMI-m, elements of medical information are embedded in DC metadata semantic framework. Detailed description and analysis are given to all core elements by using DC qualifier. By way of reusing, extending and self-defining, definite metadata elements and self-defined elements are created, and the outline of the application of Chinese medical information metadata is thus established. (5) Through analyzing the features of the main elements in CDMI-m, this thesis for the first time proposes that EBM and HL7, both of which are closely related to clinical medicine, should be consulted and combined together.

Conclusion: The design of CDMI-m aims at seeking for new methods and ways to organize and acquire all kinds of Chinese network and digital medical information.

Background: Metadata, a new type of information-structured data was created by high speed development of network resources and its low efficiency usage. “Data about data” is a common definition for metadata [1]. Dublin Core Metadata Element Set (DC) was initiated in March 1995 during the first workshop session on metadata organized by the OCLC (USA Online
Library Center) and NCSA (the United States Center for Supercomputing Applications) in the United States. The aim of the creation of DC [2] is to be able to search and access to the needed information rapidly through description, collection and organization of information resources by using simple identifications. In a short span of 10 years, DC metadata has been developed rapidly and has been applied in many fields of many resources organizations [3]. The standard metadata scheme for health resources was first created in 1988, such as MCM[4] [5] 、CISMeF[6] [7] 、EBM metadata[8], and NLM metadata[9]. However, up till now there is no standard metadata scheme profile for Chinese health resources being reported yet [10]. With the rapid development of Internet and the computer technology, the Chinese Biomedical literature has become digitized more and more quickly. The electronic versions of medical resources take up a larger proportion of all health resources. The dynamic and disorganized nature of those digital health resources has made it difficult for search and utilization of Chinese health resources. This paper intends to study and design a set of Chinese digital medical information metadata (CDMI-m) resources based on the DC metadata, suitable for the characteristics of Chinese health resources in the hope that an adapted new way of organizations, retrieval and utilization to Chinese digital medical information resources will be explored.

Methods:
1. Principles in selecting a standard metadata scheme for Chinese health resources

It is the best way to establish a standard metadata for Chinese digital health resource based on the DC-metadata. There are three reasons to support this proposal: One of them is that DC-m has been repeatedly revised, and improved to become an appropriate and mature international metadata standard and it has been widely used in many fields for the past tens years. It has been in line with the regularization, standardization, internationalization and the principle of interoperability [11]. The second reason: there are a number of foreign-based medical meta-data DC programs [12] that are sophisticated enough to be used as a model for Chinese digital health resources in the aspects of consultation, reference, reusing and nesting model when creating CDHR-m. The third reason: The trial of “Chinese Journals metadata to describe the rules of the recorded” has entered the national phase [13], which is also
based on the DC metadata standard. Therefore, there is a ready-made sample DC-based local metadata to provide CDHS-m of the reference re-usage.

2. **Establishment of the contents of CDHS-m coding schemes**

   In recent years with the continuous improvement and promotion of the use of information resources and a huge demand for network, DC has become international standards of network resources developed by OCLC. Given that it is an international standard, the advantage it has is that it is a highly condensed summary. But the disadvantage of it is that it is not "exhaustive." Therefore, only using DC metadata 15 core elements to interpret and recapitulate highly specialized resources is obviously inadequate for description in the medical literature. For example: DC's "Subject" element subject to describe the concept of subject in the face of medical resources, can not meet the requirement of the depth and width demanded by medical information resources. So in accordance with the rules of the DC elements Content encoded [14], and in the framework of DC provided by the standard coding system which direct toward the characterized of the medical profession, the CDMR-m uses ways of those foreign several medical metadata program(MCM, CISMeF, EBM metadata and NLMmetadata), and quote the medical MeSH Vocabulary as encoding scheme in the "Subject" element of the subject coding system, so as the "EBMC" and "EBMS" (meta-data are EBM metadata used to describe the evidence-based programs medical coding system), CLC ("Chinese Library Classification 4th edition") and HL7 (Health Information transmission standard).

3. **Selection and determination of the description object of CDHR-m programm**

   For organization and description digital medical literature and health resources by use of DC-m programm, it is quite important description to select and define the medical resources as the analyzed metadata resources, which possess representative and mainstream. Even though various forms of medical literature have been created, it's an undeniable fact that the medical journals that carry the results of modern medical science and technology core thesis are still taking the main role in the medical information resource because they contain sophisticated and reliable source of medical information [15].

   As the process of the digital health information resources has been speeding up, many electronic versions of medical journals, books and full text (or abstracts) databases are
booming and emerging on in the network. In addition, because of medical journals also have such advantages as the normative body in the literature version, better standard in the written format and same or similar expression characteristics in almost all major terms. As a result, medical journals are regarded to be available representation in all types of medical resources. CDHR-m Metadata select issue medical journals as major description object, and at the same time the other types of medical information carriers are being taken into account, such as: Medical multimedia, network resources, electronic medical records, medical conferences, various types of agreements, tele-medicine, and medical imaging, etc. The medical literature of all kinds of resources relevant to the scope of the object is being strived to the fullest possible consideration.

4. Factors Analysis of CDHR-M metadata contents description Objects

Factors Analysis of CDHR-M metadata contents description objects should fit in with the major characteristics of the medical information literature and make track for the newest point of view on development of medical science. The following four factors must be taken into account:

The first factor: Evidence-based medicine (EBM) is the integration of the best research evidence with clinical expertise and patient values. An important process of EBM practices is to find strong evidence supported by clinical studies (both primary studies and secondary studies such as summaries of reviews) published in clinical journals or databases.\textsuperscript{[16]} As it is agreed that medical mode has been converted from experiential medicine to evidence medicine in contemporary age, there still exists to be short of unsatisfactory retrieval tools for EBM even if Medline, the most classic of medical MeSH controlled vocabulary, can not focus on very well description and organizations of EBM-related concept of subject\textsuperscript{[17]}. Therefore, in order to establish the CDHR-m program the main content of evidence-based medicine resources should be considered sufficient in the four main elements: structured abstracts, study types, clinical concepts, and the degree of evidence. Japan's EMB-metadate (evidence-based medicine metadata program) has become a good example. In this regard, it is recognized that EBM is the best research evidences based on clinical experiences and patient information. This study for EBM resource used in an integrated search tool across primary
and secondary studies with recognition of EBM specific factors, both studies originate from literatures and reviews published in clinical journals or databases. Japan's EMB-m Search is a comprehensive search for EBM resources (i.e. clinical research literatures), which can identify EBM standard (i.e., the degree of evidence depending on research design) and clinical aspects of the study (e.g., therapy, diagnosis, etiology and prognosis) as well as other subject matters such as patients (with or without diseases) interventions and outcomes. Therefore CDHR-M should learn the views and methods from Japan's EMB-metadate programs. The CDHR-m proposal should be in line with the recognition of primary studies and secondary studies for a comprehensive EBM search tool.

The second factor: (Health level Seven) Standard [18]

HL7 is known as a kind of “medical messages data interchange system”. It was created by American National Standards Institute (ANSI) and accredited by Standards Developing Organizations (SDOs) operating in the healthcare arena. Most SDOs produce standards (sometimes called specifications or protocols) for a particular healthcare domain such as pharmacy, medical devices, imaging or insurance (claims processing) transactions. Health Level Seven’s domain is clinical and administrative data [18]. In recent years, some developed countries like U.S. and European countries have committed to this work, some of them become members of the organization, so did the China's accession in 2000 and become one of the signatory countries of the HL7 standard. HL7 was originally designed to be used for economic services, mainly for the management of medical expenses and hospital information system [19], as a result, it lacks of the contents of the medical literature resources, an important medical information.

Some experts suggested that in view of the HL7’s popularization and application, we should form a national HL7 that not only is under the framework of HL7 standard classification, but also perfect and modify the content of standard classification, coding and content. In addition, as a result of the rapid development of the field of information and medical information services, there is a huge demand for the service targets and an excellent and improved coverage and contents of the health care information systems. It would be the future target that HL7 standards and the medical
literature resources to shared, exchanged and interoperation. So, the CDHR-m project should explore the development of programs for clinical care and medical research to provide more flexible and multi-level medical information related to the elements that are expected to eventually become the localization part of the HL7 standard [20]

The third factor:

Medical assistant information is an important part of health resources. But the common problem of the deficiency of assistant information description both in the medical retrieval databases and the web net resources affects the health information resource of retrieval and utilization. Medical assistant information in fact is closely associated with the body of text [21], such as author information (name, unit, address, telephone no., brief introduction) can supply personal information of author himself; the reference information can indicate that the thesis of the scope and depth about author's, as well as citation situation; besides, the information of Fund supply and resource evaluation (Level publications, impact factor, factor, etc.) can give us the information that the scientific and technological content and quality positioning. Article supporting Comprehensive Medical assistant information can not only help readers quickly and efficiently catch the idea of a content, but also play an important role in information transmission, communication, retrospective retrieval, bibliometrics, statistics [22]. The most significant part of that is to raise the evidence for medical management and decision-making, setting up the Chinese Medical Professional expert's database system originating from author personal information. Meanwhile the fund support information available facilitate to obtain the support types, intensity, distribution (geographical or major) and et al, in order to provide this integrated information content to the Government or the relevant department in accordance with decision-making. Therefore the medical assistant information description will become important factors to consider in CDHR-M.

The fourth factor:

Diversification in the types of medical resources, Network and computer technology has changed the types of medical resources being an unique text form. There are numerous types of multimedias springing up such as the remote medical, major medical network protocol, the medical home,
digital medical images, audio and video, bio-signal graph (ECG, etc.), electronic medical records and other non-text form of the new medical resource types of resources into the medical family. This is the main characteristic of medical resources digitization [23]. In the description and organization of such resources, the application of DC metadata has a strong advantage. Foreign used programs of the three medical metadata: MCM, CISMeF and EBM, expanded property element on DC.type restrictions , the MCM created 35 sub-elements, and CISMeF metadata is on this basis at an additional resource type 20 , which covered all of the network resources types. More than half of the selected words in both above are controlled from the term in the MeSH Thesaurus. EBM metadata resources are all reusable both MCM and CISMeF types of DC metadata description of the resources element analysis can be seen to describe the type of figure of the importance of resources.

**Conclusion:**

The CDHR-m is aiming to create DC metadata model designed to look for new method of Chinese health resources organization and retrieval. Currently it is on the development stage and it is only a point of view. There is a lot of work needed to be done and tried for establishment and improvements. The Government decision-making body of the leading organization of the medical experts is required to research, select, appraise on the metadata element model of CDHR-m, in order to provide a scientific basis for promulgating the standards of CDHR medical metadata.

**references**


3. Yang QY. Mata-data in the network environment. Journal of


2004; (13) 5:57-59.
