

Metadata for Evidence Based Medicine resources

Yukiko Sakai

Hiyoshi Media Center, Keio University

yukiko.sakai@adst.keio.ac.jp

Abstract

A new metadata element set based on Dublin Core Metadata Element Set (DC) and Administrative Container Core (A-Core) was proposed for Evidence Based Medicine (EBM) sources after reviewing metadata elements and contents of current EBM sources and medical metadata for Internet resources. The metadata schema was designed to provide a common format for existing primary and secondary studies; further for Internet resources as prospective sources. An enhanced DC.Description element can store structured abstracts of primary studies in primary and secondary studies of clinical research; A-Core elements are used for indexers or creators of metadata for primary studies. Two encoding schemes were suggested as EBM qualifiers for the DC.Subject element to distinguish important factors of EBM practices: the degree of evidence and focuses of clinical perspectives such as therapy, diagnosis, prognosis, and etiology. An additional feature of this metadata schema is in distinction of a variety of "types" (e.g., study type, resource type, format, genre)

Keywords: Evidence Based Medicine, Metadata, Dublin Core, Administrative Core, Resource Description

1. Introduction

Evidence Based Medicine (EBM) is the integration of the best research evidence with clinical expertise and patient values [1]. An important process of EBM practices is finding strong evidence supported by clinical studies (i.e., both primary studies and secondary studies such as summaries or reviews) published in clinical journals or databases.

In order to find strong evidence, a comprehensive search for EBM sources (i.e.,

clinical study literatures) that identifies EBM criterion (i.e., the degree of evidence depending on research design) and clinical aspects of the study (e.g., therapy, diagnosis, etiology and prognosis) in addition to other subject matters such as patients (with or without diseases), interventions, and outcomes should be conducted.

Some difficulties in the search for EBM sources using existing tools have been pointed out. For primary studies, high retrieval is a problem when searching against existing large bibliographic databases such as MEDLINE [2]. Unlike the search for secondary studies that consist of pre-selected primary studies (based on the criteria for EBM), primary studies should be filtered out with EBM specific factors (i.e., research design and clinical aspects) using special techniques. Another difficulty is the shortage of tools for searching multiple secondary resources or an integrated search mechanism for primary and secondary sources.

An additional problem that the author would add is the lack of established tools for searching Internet resources that are excluded in EBM sources so far due to their questionable quality [3]. They too have possibility of becoming EBM sources in the near future.

2. Scope of the study

The purpose of this study is to propose a metadata¹ schema for EBM resources used in an integrated search tool across primary and secondary studies with a recognition of EBM specific factors (i.e., the degree of evidence and focuses of clinical perspectives).

¹ Metadata represents the content, structure, and logistical information of any information objects including electronic resources. Metadata is used for data discovery and control of data. (Hudgins, Agnew, Brown, 1999 [4]).

First, existing metadata contents and elements of current EBM sources and medical resources on the Internet were analyzed to reveal the strengths and weaknesses of those metadata for EBM searching. Metadata from the MEDLINE database was examined as the most comprehensive bibliographic database used in searches for primary studies. Metadata from the *ACP Journal Club* articles was also examined as an example of secondary study journals or databases. For resources on the Internet, Medical Core Metadata (MCM) and CISMef (catalog et index des sites medicaux francophones) metadata were reviewed.

Second, an EBM metadata schema focusing on relevance, comprehensiveness and EBM specific subjects was proposed based on this review

3. Review of existing metadata

3.1. MEDLINE

MEDLINE is one of the largest index and abstract databases of medical journal articles which contains over 11 million references maintained by the U.S. National Library of Medicine [5].

The strength of MEDLINE for searching medical literature is its use of a sophisticated controlled vocabulary called the Medical Subject Headings (MeSH). MeSH represents clinical concepts well (e.g., patients, interventions, and outcomes), but is weak in searching EBM sources. Two important factors for EBM (i.e., research design and clinical aspects) are not represented well using MeSH, because controlled vocabulary for research design and clinical aspects of EBM are not sufficient and scattered throughout several components of MeSH. Some study types, such as those representing research designs relating to the degree of evidence (e.g., “randomized controlled trials” as the best study in therapy) are listed as PUBLICATION TYPE (PT) [6] in MeSH. PT also contains other types such as genre or format (e.g., lecture notes, pictorial works), while concepts on study types are also found in other components in MeSH (i.e., main headings [7] and check tags [8]). Concepts representing clinical perspectives are split into subheadings (e.g., therapy, diagnosis) [9] and main headings (e.g., Prognosis, Morbidity, Mortality, Recurrence).

This shortage results in the need for complex search strategies, so called EBM hedges or EBM

filters, to increase precision and recall for EBM materials [10].

3.2. ACP journal club

The *ACP journal club* is the first secondary journal focusing on EBM. It focuses on the synopsis of a single study and reviews of multiple related studies. This journal has been published both in print and online

The structure of the metadata schema used by the ACP journal is similar to that of bibliographic databases for primary studies except for elements for comments, since the secondary studies should hold metadata of primary study(ies) as a structured abstract(s). This journal is inconsistent in that some labels vary among different clinical perspectives of a study (e.g., “intervention” for therapy, “description of test” for diagnosis) even in the same journal.

A strength of the metadata for secondary studies in the *ACP journal club* is that study types related to the degree of evidence are easily recognized in a DESIGN element in the primary study abstracts. Clinical perspectives (e.g., therapy, diagnosis) are also entitled, while MeSH is not used for keywords.

It should be noted that such secondary journals are unlikely to be indexed in major bibliographic databases (e.g., MEDLINE) and the metadata schema varies in each secondary journal. This inconsistency may be an obstacle for comprehensive and consistent searching in EBM sources.

Ovid Evidence Based Medicine Reviews (Ovid EBMR) ‘s multifile search feature [11] for MEDLINE and several secondary journal databases using a full text search is one of the solutions for a meta search of primary and secondary studies. Searching by full text instead of utilizing sophisticated controlled vocabularies - such as MeSH; however, may cause other problems in terms of comprehensiveness and consistency in the search result.

3.3. MCM and CISMef

MCM (Medical Core Metadata) was proposed by a project team of Oregon Health Sciences University in 1998 to enhance access to medical resources on the Internet [12].

CISMef (catalog et index des sites medicaux francophones) metadata schema has been used as an online catalog and index of health Internet resources for French-speaking health professionals

and consumers operated by Rouen University Hospital in France [13].

Both of these metadata schemas are based on the Dublin Core Metadata Element Set whose goal is to define the minimum set of metadata elements necessary for the representation of a variety of resources, including Internet documents, to provide interoperability among information systems [14]

MeSH, which is used in both schemas for subject representation, must provide consistent access to EBM sources even on the Internet. Furthermore, MeSH in character strings and the identifier of Unified Medical Language System (UMLS) Metathesaurus [15] in MCM is useful for enhanced search systems involving other vocabularies.

The weakness of using MeSH in both schemas for searching EBM sources is that they omit most study types from PUBLICATION TYPE of the MeSH controlled vocabulary when they created an encoding scheme for DC.Type element. Instead, they introduced different resource types (e.g., home pages, meeting and conference announcements, teaching material) and creators (e.g., hospital department, scientific society) to focus on Internet resources. Other kinds of types (e.g., format, genre) are forced into the DC.Type element (e.g., database, directory, textbook in MCM; text, registry, congresses in CISMef) which is also problematic.

4. Proposal for EBM metadata

This section proposes a metadata schema for EBM resources based on the fifteen elements of the Dublin Core Metadata Element Set (DC) and the eight elements of the Admin-Core: Administrative Container Metadata(A-Core) as shown in Table 1.

A feature of this metadata schema is the possibility of holding primary and secondary studies in a similar way.

DC was chosen as the international metadata schema standard for Internet-based resources because, its interoperability was considered an important factor for an integrated search mechanism among EBM sources based on disparate metadata schemas. A-Core, designed to describe the provenance, management or administration (e.g., who, what, when) of other sets of descriptive metadata [16] was chosen to store information about the creator of metadata for primary studies including both the indexers and the

creators of abstracts of primary studies used in the secondary studies.

Table 1. Outline of EBM metadata

Element	Encoding Scheme	Element Refinement	Element (continued)
DC.Title	MeSH EBMC ^{*1} EBMS ^{*2}	Abstract Objective Design Setting Participants Analyzed factors OutcomeMeasures Results Conclusion Commentary	DC.Publisher
DC.Creator DC.Subject			DC.Contributor DC.Date DC.Type DC.Format DC.Identifier DC.Source DC.Language DC.Relation DC.Coverage DC.Rights AC.Name AC.Activity AC.Email AC.Contact AC.Date AC.DateRange AC.Rights AC.Location

*1 Evidence Based Medicine Clinical Perspectives

*2 Evidence Based Medicine Study Type

The following are the details of enhanced or refined DC elements for EBM sources with a note about conversion from existing sources. The remaining elements simply follow the definition put forth by the Dublin Core Metadata Element Set [14] and the A-Core [16] as shown in the sample data in the Appendix.

- *DC.Subject*

DC.Subject prefers to hold MeSH as an Element Encoding Scheme as per the DC Qualifiers recommendation [17] to provide general clinical concepts such as patient, intervention and outcomes for EBM sources. The syntax of the element is string + UMLS identifiers for human readability and the machine mechanisms involving other medical vocabularies following the suggestion of the MCM metadata.

EBMC is an encoding scheme listing clinical perspectives (e.g., therapy, prognosis, diagnosis, or etiology). EBMS is another scheme listing study types. These schemes contribute to providing two important factors for EBM: the degree of evidence and focuses of clinical perspectives. Both schemes are applied to represent concepts derived from primary studies even in the secondary studies.

Subheadings of MeSH pertaining to clinical perspectives may be converted to EBMC; study types in MeSH-PT go to EBMS.

- *DC.Description*

DC.Description holds an abstract of primary studies in primary and secondary studies under the ABSTRACT refinement. (DC Qualifiers [17]). For a structured abstract, further refinements as shown in the sample data are possible in order to specify clinical studies.

An additional refinement for DC.Description is COMMENTARY that holds commentary in secondary studies.

Research design shown in secondary study articles can be converted to the encoding scheme for study types (i.e., EBMS); entitled clinical perspectives go to another encoding scheme (i.e., EBMC) as well.

- *DC.Type*

DC.Type should be defined as the resource type as stated in the original definition of the DC.Type element [14]. Types of genre or the mean of genre and format in MeSH-PT (e.g., journal article) or resource types in DC.Type of MCM and CISMef can be converted to DC.Type in EBM metadata.

- *DC.Format*

DC.Format contains the physical manifestation, which equals the original definition of DC.Format [14]. Some types in MCM's encoding scheme for DC.Type may fall into this element.

5. Conclusion

After reviewing the metadata of current EBM sources and medical metadata for Internet resources, an EBM metadata element set based on the DC and the A-Core was proposed. This new schema is designed for the mutual description of metadata for primary and secondary studies as well as Internet resources. The features of this metadata schema are: the capability of providing structured abstracts of primary studies; a suggestion of two additional encoding schemes in addition to MeSH for important factors of EBM practice (i.e., study types and clinical perspectives); and the possibility of the usage of this schema for Internet resources.

Further studies are needed to examine the usefulness of the EBM metadata by reviewing the schemas of more complex EBM sources such as the *Cochrane Systematic Reviews* [16]. Studies should also examine implementing this schema for the conversion of metadata from existing sources and its effects on search performance.

Acknowledgement

This paper is based on the author's term project for "Metadata architectures and applications" at the School of Information and Library Science, the University of North Carolina at Chapel Hill (UNC-CH) in Fall 2000. The author would like to express gratitude toward the course instructor Dr.

Jane Greenberg who encouraged her throughout this work. The author is also grateful to the instructors of "Evidence based medicine and the medical librarian" in Spring 2001 who helped her to understand the concept of EBM, and her colleagues in the Health Sciences Library of UNC-CH for their constructive suggestions.

References

- [1] Sackett, DL.; Straus, SE.; Richardson, WS.; Rosenberg, W.; Haynes, RB *Evidence-based medicine: how to practice and teach EBM*. 2nd ed. Edinburgh: Churchill Livingstone, 2000.
- [2] McKibbin, A.; Wilczynski, N.; Hayward, RS. The medical literature as a resource for health care practice. *Journal of the American Society for Information Science*. 46: 737-43, 1995.
- [3] McKibbin, A. *PDQ: evidence-based principles and practice*. Hamilton, Ontario: B.C. Decker, 1999.
- [4] Hudgins, J.; Agnew, G.; Brown, E. *Getting mileage out of metadata: applications for the library*. Chicago: American Library Association, 1999.
- [5] National Library of Medicine. *Fact sheet: MEDLINE*, 2000. [online] Available: <http://www.nlm.nih.gov/pubs/factsheets/medline.html>
- [6] National Library of Medicine. Medical Subject Headings. Publication types: genre terms. 2000. [online] Available: <http://www.nlm.nih.gov/mesh/pubtypesg.htm>
- [7] National Library of Medicine. Medical Subject Headings. 2001. [online] <http://www.nlm.nih.gov/mesh/meshhome.html>
- [8] National Library of Medicine. Medical Subject Headings. Features of the MeSH vocabulary, 2000. [online] Available: <http://www.nlm.nih.gov/mesh/features.html>
- [9] National Library of Medicine. Medical Subject Headings. Topical subheadings with scope note, abbreviations, and allowable categories. 2000. [online]. Available: <http://www.nlm.nih.gov/mesh/topsubscope.html>
- [10] Haynes, RB. Developing optimal search strategies for detecting clinically sound studies in MEDLINE. *Journal of the American Medical Informatics Association*, 1: 447-458, 1994.
- [11] Ovid. *Evidence-based medicine reviews, 2001?* [online] Available: <http://www.ovid.com/products/clinical/ebmr.cfm>
- [12] Malet G.; Munoz F.; Appleyard, R.; Hersh, WR. A model for enhancing Internet medical document retrieval with "medical core metadata." *Journal of the American Medical Informatics Association*. 6: 183-208, 1999.
- [13] Darmoni, SJ.; Thirion, B. A standard metadata scheme for health resources. *Journal of the American Medical Informatics Association*, 7(1): 108-109, 2000.

- [14] Dublin Core Metadata Element Set, Version 1.1: reference description. July 2, 1999. [online] Available: <http://dublincore.org/documents/1999/07/02/dces/>
- [15] National Library of Medicine. *Fact sheet: UMLS Metathesaurus*, 2001. [online] Available: <http://www.nlm.nih.gov/pubs/factsheets/umlsmeta.html>
- [16] Iannella, R.; Campbell, D. The A-Core: metadata about content metadata, 1999. [online] Available: <http://metadata.net/admin/draft-iannella-admin-01.txt>
- [17] Dublin Core Qualifiers (DCMI recommendation). July 11, 2000. [online] Available: <http://dublincore.org/documents/2000/07/11/dcmes-qualifiers/>

Appendix. Sample EBM metadata

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/TR/2000/CR-rdf-schema-20000327"
  xmlns:dc="http://purl.org/DC/documents/rec-dces-19990702.htm"
  xmlns:dcq="http://www.loc.gov/marc/dcqualif.htm"
  xmlns:ebmq="http://hc.lib.keio.ac.jp/~yukiko/ebm/qualifier.htm"
  xmlns:ac="http://purl.org/dc/agent/1.0/"
  xmlns:mesh="http://nlm.nih.gov/mesh/MBrowser.htm"
  xmlns:ebmc="http://hc.lib.keio.ac.jp/~yukiko/ebm/clinicaperspective.htm"
  xmlns:ebms="http://hc.lib.keio.ac.jp/~yukiko/ebm/studytype.htm"
  xmlns:ISO3166="http://www.din.de/gremien/nas/nabd/iso3166ma/codlstp1/index.html"
  xmlns:ISO639-2="http://lweb.loc.gov/standards/iso639-2/langhome.html">
  <rdf:Description about="http://www.bmj.com/cgi/content/full/314/7095/1655">
  <dc:title>Alcohol consumption and cognitive performance in a random sample of Australian soldiers who served ...</dc:title>
  <dc:title>
  <dcq:alternative>Alcohol consumption was not associated with cognitive performance or brain atrophy in ...</dcq:alternative>
  </dc:title>
  <dc:creator>Dent, OF</dc:creator>
  <dc:creator>Sulway, MR</dc:creator>
  <dc:creator>Broe, GA</dc:creator>
  <dc:creator>
  <dcq:corporateName>Department of Sociology, Australian National University</dcq:corporateName>
  <dcq:corporateName.address>Canberra, ACT0200, Australia</dcq:corporateName.address>
  </dc:creator>
  <dc:subject>
  <dcq:subjectScheme>EBMC</dcq:subjectScheme>
  <rdf:value>Etiology</rdf:value>
  </dc:subject>
  <dc:subject>
  <dcq:subjectScheme>EBMS</dcq:subjectScheme>
  <rdf:value>Cohort Study</rdf:value>
  </dc:subject>
  <dc:subject>
  <dcq:subjectScheme>MeSH</dcq:subjectScheme>
  <rdf:value>Alcohol-induced Disorders#CUI0266194#SUI0911544/physiopathology#CUI0005144#SUI0003155 ...</rdf:value>
  <rdf:value>Pick Disease of the Brain#CUI3100954#SUI2210599
  /etiology#CUI0001335#SUI0001966/Australia#CUI0399555#SUI9155376</rdf:value>
  <rdf:value>Aged#CUI4588012#SUI0344805</rdf:value>
  <rdf:value>Men#CUI0087524#SUI0092464</rdf:value>
  </dc:subject>
  <dc:description>
  <ebmq:objective>To determine the association between daily alcohol consumption and ...</ebmq:objective>
  <ebmq:design>Cohort Study</ebmq:design>
  <ebmq:setting>Community-based study in Sydney</ebmq:setting>
  <ebmq:participants>342 men (mean 63 y) who wer veterans of ...</ebmq:participants>
  <ebmq:analyzedfactors>Self-reported average daily alcohol consumption was ...</ebmq:analyzedfactors>
  <ebmq:outcomemeasures>Cognitive performance was assessed ...</ebmq:outcomemeasures>
  <ebmq:results>In 1982, 85% of the men drank alcohol once/wk...</ebmq:results>
  <ebmq:conclusion>No association was found between alcohol consumption and cognitive performance...</ebmq:conclusion>
  <ebmq:commentary>Heavy drinking and alcoholism have been linked to cognitive disorders, but ...</ebmq:commentary>
  </dc:description>
  <dc:type>Journal Article</dc:type>
  <dc:identifier>http://www.bmj.com/cgi/content/full/314/7095/1655</dc:identifier>
  <dc:source>BMJ. 1997 Jun 7; 314:1655-7</dc:source>
  <dc:source>The study funding from: Australian Department of Veteran's Affairs and Australian National Health and...</dc:source>
  <dc:language>eng</dc:language>
  <dc:relation>
  <dcq:references>1.Reid MC, Anderson, PA. Med Clin North Am. 1997; 81:99-1016.</dcq:references>
  <dcq:references>2.Hendrie HC, Gao S, Hall Ks, Hui Sl, Unverszagt FW. J Am Geriatr Soc. 1996; 44:11158-65.</dcq:references>
  <dcq:references>3.Christian JC, Reed T, Carmelli D, et al. Stud Alcohol. 1995; 56:414-6.</dcq:references>
  <dcq:references>4.Hebert LE, Schrr PA, Beckett LA, et al. Am J Epidemiol. 1993; 137:881-91.</dcq:references>
  </dc:relation>
  <dc:contributor>Fones, TV</dc:contributor>
  <ac:date>1998-01</ac:date>
  <ac:location>Evidence-based medicine. 4(1), 26</ac:location>
  </rdf:Description>
</rdf:RDF>
```