

Development of a Taxonomy for Health Information Technology

Brian E. Dixon, M.P.A.^{a,b}, Atif Zafar, M.D.^{a,b,c}, Julie J. McGowan, Ph.D.^{a,b,c}

^aAHRQ National Resource Center for Health Information Technology, Maryland

^bRegenstrief Institute, Inc., Indiana

^cSchool of Medicine, Indiana University, Indiana

Abstract

Taxonomies provide schemas to help classify entities and define the relationships between them. Early computing enabled the development of ontologies and Medical Subject Headings (MeSH), the first modern classification of medical terminology as applied to medical literature. Later developments, such as MEDLINE, expanded MeSH to include a number of medical informatics terms. However, a lack of specificity in MeSH and other existing informatics taxonomies for terminology used to describe the growing field of health information technology (health IT) created the need for the development of a specialized taxonomy. Experts associated with the Agency for Healthcare Research and Quality's (AHRQ's) National Resource Center for Health Information Technology (NRC) created and evaluated a taxonomy for health IT, to enable users of a public health IT Web site to efficiently identify resources within an online, searchable repository.

Keywords:

controlled vocabulary, internet, information systems, medical informatics applications, computer communications networks

Introduction

Taxonomy is derived from two Greek words: taxis, meaning order or arrangement, and nomos, meaning law. Common definitions include a focus on the laws of classification and natural history, including the relationships among entities organized in a hierarchical structure [1][2].

While taxonomies have historically been applied to a variety of different content areas, the focus remained on the classification of those areas and the hierarchical structure that explained vertical relationships. During the twentieth century, computers enabled the development of ontologies, data models that represent concepts and facilitate understanding of relationships, particularly in the definition of ambiguous concepts and their concomitant associations.

The science of taxonomy and its related ontologies enabled the development of Medical Subject Headings (MeSH) and its classification structure (e.g., tree), the first true classification of medical terminology as applied to the world's medical literature. Its genesis was in the early work

of John Shaw Billings and the publication of *Index Medicus* in 1879. However, the first major expansion of the terminology did not come until 1958, recognizing the specialization of the profession of medicine [3].

The creation of the specialized taxonomy for medicine enhanced the ability of users to retrieve information from printed materials. However, without the use of a computer to facilitate the application of the classification schema, albeit through punched cards and high-end photographic equipment to print the newly reconstituted *Index Medicus*, the foundation for the current MESH would not have been laid [4].

The computerized MEDLINE followed over a decade later, with its MESH expanded to include a number of tangential concepts. The vocabulary and its hierarchical relationships have been developed over time and have incorporated a number of medical informatics terms.

However, there is a lack of specificity of concepts and granularity in the health information technology (health IT) area, particularly as it relates to current practices in electronic health record (EHR) implementation. The MESH vocabulary also lacks terminology to encompass dynamically growing areas like health information exchange and consumer health informatics.

In trying to develop a searchable knowledge core for the Agency for Healthcare Research and Quality's (AHRQ's) National Resource Center for Health Information Technology (NRC) Web site, a major first step was the creation of a taxonomy to define the interrelationships among concepts around health IT planning, implementation, and evaluation.

Because health IT bridges medical informatics and hospital and clinical practice organizations, it became imperative in the development of a taxonomy to understand the scientific and sociological issues impacting health IT. MESH offered some concepts, but the major sources of the taxonomy development came from divergent vocabularies. This supports the premise that a good taxonomy must be developed on the basic principles of a hierarchical and interrelated classification schema, searchable within a Web environment.

Materials and methods

In March 2005, a panel of experts with a wide array of experiences in implementing health IT systems was convened to design a national, searchable Web site for persons interested in the implementation and evaluation of health information technologies. These experts, drawn from medical informaticians across the United States, agreed that a basic system of categorization would be useful to organize content into a searchable knowledge base. The experts also expressed a desire for the taxonomy to be used as a method for educating visitors about common terminology used in the field of health IT.

The panel began design of the taxonomy using a structured brainstorming exercise. The brainstorming involved thinking through the various types of content the Web site would feature, which stakeholders the Web site would target, and what topics best matched the goals of the Web site designers, stakeholders, and potential users. A preliminary confluence of ideas was slowly organized into logical buckets by the end of the brainstorming session. The buckets top level terms and the terms contained therein branches as used in most controlled vocabularies comprised the first draft of the taxonomy.

The panel next looked to a number of existing taxonomies to refine each top level term and its branches. The EHR Mind Map, created by the Healthcare Information Management Systems Society (HIMSS), provided some useful terms, but it was found to be narrowly focused on EHR systems, neglecting technologies such as telemedicine and health information exchange [5]. The panel also considered the ACMs Computing Classification System (CCS), but it consists primarily of computer science terms (e.g., programming languages, mathematical theory, and data structures) whereas the target audience for the Web site is clinical and health administration leaders engaged in health IT implementation activities [6]. Finally, the panel considered MESH but found it too lacked sufficient granularity to cover the complete technical (e.g., EHR, HIE, telemedicine) and sociological (e.g., adoption, process redesign, project management) spectrums that encompass the field of health IT [7].

To create a taxonomy for health IT, the panel created a draft using terms from the three major taxonomies it examined. The group then organized the list into clusters of related terms under broad categories, and it began to search for additional terms that would fit into its categories. The organized list became version 1.0 of the taxonomy.

Since the panels original meeting, the taxonomy has been expanded and refined. A first round of edits came after a second expert review was conducted by experts associated with the NRC who could not attend the original expert panel meeting. A second round of edits came when two of the authors (Dixon and Zafar) used the taxonomy to classify journal articles and other resource documents in the online library the taxonomy was developed to support. A third set of modifications was made when two NRC experts (Dixon and Samarth) applied the taxonomy to the

more than 120 health IT grants and contracts awarded by AHRQ within the last two years (FY04 and FY05).

The taxonomy was also evaluated on two subsequent occasions to ensure comprehensiveness and usability. A formal review of the taxonomy was conducted in late 2005 by the large panel of experts on staff with the NRC. The recommendations of the panel were implemented as part of the development version of the AHRQ health IT Web site.

In early 2006, a second type of evaluation took the form of usability testing, a critical component in software and Web development. Health care and IT professionals from leading non-profit and government organizations were asked to use an online, interactive version of the taxonomy to search for health IT resources. Testing was conducted at the National Library of Medicine (NLM) usability lab in Washington, D.C. User experiences were observed, recorded, and analyzed by NLM and NRC staff.

Based on recorded user experiences, the taxonomy and its appearance on the Web site were altered. For example, a minor category EHR Systems was renamed Implementation of Health IT after some users confused it with the EHR minor category. Users also reported that they did not associate the term clinical decision support with the EHR minor category, so it was renamed Electronic Health. Other users commented on the taxonomys appearance, noting that they would prefer to see terms such as electronic health records and health information exchange as entry points (e.g., major categories) to the taxonomy. Instead of altering the major dimensions of the taxonomy to accommodate user requests, prevalent terminology was summarized together under its parent branch for display on the Knowledge Library main page.

Results

The taxonomy is organized into six major and 28 minor categories (high-level terminology) with two additional sub-levels for controlled vocabulary terminology (discrete, classifiable concepts) and a dictionary for controlled vocabulary terminology synonyms. Based on the results of the usability testing, a subset of the 400+ terms and synonyms were selected to drive terminology browsing of the online resources.

When users click on a term, a search is performed in the background, and results are displayed for the user to browse through. Resources from the core collection, those tagged by an NRC expert, are promoted to the top of the result list. Partner contributions, online resources aggregated from a variety of health IT-focused organizations, fill out the result list and are sorted based on the prevalence of the selected term within the resource.

Of the 6872 online resources available through the AHRQ health IT Web site, 702 (10.2%) reside in the core collection. Table 1 presents the major and minor categories along with the number of core collection resources belonging to them.

Table 1 - Taxonomy for health IT and volume of core collection resources

I. Organizational Strategy (171)	A. Financial (18)
	B. Planning (44)
	C. Process Change (18)
	D. Implementation of Health IT (77)
	E. Policy (14)
II. Technology (170)	
	A. Mobile (3)
	B. Infrastructure (1)
	C. Security (2)
	D. Standards (34)
	E. Electronic Health (57)
	F. Telehealth (9)
	G. Health Information Exchange (51)
III. Value (227)	
	A. Research (76)
	B. Evaluation Outcomes (151)
IV. Laws and Regulations (35)	
	A. Sample Legal Documents (2)
	B. Privacy (16)
	C. Security (7)
	D. Government (10)
V. Organizations (54)	
	A. Professional Societies (12)
	B. Payers (8)
	C. Governmental (14)
	D. Nonprofit Organizations (15)
	E. Magazines (5)

VI. Operations (35)	
	A. Governance (14)
	B. Project Management (12)
	C. Systems (5)
	D. Dissemination (4)

Organizational strategy

Technology is implemented in order to solve problems and improve processes. Strategies designed to address specific problems and processes will ensure the successful implementation of technology within organizations.

- Financial strategies create initial and sustained revenue for the deployment and maintenance of technology.
- Strategic planning helps technology deployment meet the needs of an organization and ensure that technology solutions adhere to the mission and vision of an organization.
- Strategies for process change give implementers tools for addressing the human component of technology adoption.
- Implementation of health IT strategies provide roadmaps to help organizations procure and adopt technologies. In health care, these strategies should be crafted to ensure interoperability and the protection of privacy for personal health information.
- Policy strategies address legal issues, such as federal and state regulations related to the sharing of patient data across state lines. Strategic policies can also include those designed to ensure technology meets certification requirements and accreditation criteria.

Technology

Technology in this taxonomy is broadly defined, and its definition includes individual components such as hardware (e.g., PDA, tablet PC) and complete systems (e.g. laboratory information system - LIS, computerized physician order entry - CPOE).

- Mobile computing is important to health care since the workforce is by-and-large mobile, traveling from room-to-room or clinic-to-hospital.
- Infrastructure is a critical component of all technology projects as it describes how data is communicated.
- Security is a major priority in health care as we strive to provide the right amount of data to the appropriate group of individuals providing care to a patient.
- Standards enable interoperability, the ability for two disparate information systems to seamlessly share data. Many standards are available in the market, however, so distinguishing between them becomes important when exchanging information.

- **Electronic Health** is a broad concept that encompasses complete systems (e.g., PACS, EMR, CPOE), functions (e.g., clinical messaging, results reporting), and components (e.g., knowledge base, dashboard, rules engine).
- **Telehealth** describes technologies deployed that allow care to be provided remotely, including remote monitoring of patients in their homes.
- **Health Information Exchange (HIE)** involves the exchange of patient data between providers and disparate clinical information systems. HIE includes various labels (e.g., LHIE, NHIE, NHIN), policy elements (e.g., governance, data sharing agreements), technical components (e.g., architecture, interface engine), and applications (e.g., research, public health reporting).
- **Consumer Health** encompasses technologies that engage patients in the delivery and management of their health.

Value

Evaluation is a core discipline in health IT as the tolerance for mistakes is extremely small. Research and outcomes are also important to demonstrate return-on-investment (ROI) and IT system impacts on patient safety and the quality of delivered care.

- **Research** includes study designs, data analysis techniques, evidence based medicine (EBM) practices, institutional research board (IRB) policies, and grant writing resources.
- **Evaluation outcomes** focuses on results of published studies and outcomes of non-academic lessons learned by professionals in the field in a variety of areas (e.g., safety, quality, ROI, patient satisfaction, etc.).

Laws and standards

Laws and standards affect the development, implementation, and adoption of health IT.

- **Sample Legal Documents** includes a variety of legal documents developed by health care organizations across the nation.
- **Privacy** includes information on protected health information laws and standards, such as HIPAA.
- **Security** contains resources on administrative, physical, and technical security as well as security audits.
- **Government** includes resources on state and federal regulations as well as government standards related to the national health information network (NHIN).

Organizations

There are many entities involved in the field of health IT, and each organization has an agenda and a role to play in development and adoption of technology in health care. It is important to understand these and the ability to tell the difference between entities.

- **Professional societies** are member organizations which advocate on behalf of their membership, advance professional ethics, and strive to set standards for the industry.

- **Payers** include those organizations that reimburse for health care expenditures.
- **Governmental** organizations include state and federal agencies that support health IT initiatives.
- **Nonprofit organizations** are independent entities who seek to educate, develop, promote, and research the advancement of technologies for use in health care.
- **Magazines** include news and trade publications which seek to report on the health IT industry.

Operations

Management of operations is critical to the prolonged use of technologies in health care.

- **Governance** includes organizational policies related to human and technological resources.
- **Project management** consists of resources to help manage the implementation and maintenance of health information technologies.
- **Systems operation** includes resources to help with day-to-day maintenance and operation of health IT systems.
- **Dissemination** is crucial primarily to those conducting research in the field of health IT. However, sharing best practices and successful implementations can be strategic for many organizations.

The taxonomy covers the broad spectrum of topics discussed in the field of health IT. The controlled vocabulary terms are too numerous to publish here, but many of them can be found on the AHRQ health IT Web site, <http://healthit.ahrq.gov>. User feedback will contribute to the taxonomy's future development.

Discussion

The need for a taxonomy for health IT was driven by the creation of a Web site containing knowledge-based resources to support health IT planning, implementation, and evaluation. No single source existed which could provide a contextual vocabulary with the breadth and depth necessary for efficient and effective access of the online resources. The process to create such a taxonomy involved panels of health IT experts identifying and refining terms, hierarchies, and cross-references. The taxonomy was then evaluated for thoroughness and usability.

The formal reviews and usability testing emphasized the need for robust synonymy within any taxonomy. For example, Electronic Health is the term chosen by our experts for a minor category under Technology. Synonymy is critical to ensure that documents which use terms such as electronic medical records and computer-based patient records will be appropriately selected and ranked when searches are performed against the online library.

Usability testing revealed a strong need for the development of inter-relationships between orthogonal concepts. We found that users not only searched for documents related to computerized physician order entry, but they desired CPOE related documents in specific practice settings (e.g., ambulatory, inpatient). We further found that users desired to distinguish between documents related to

small practices and large medical centers. Based on this feedback, we intend to develop interrelated concepts to address these needs.

Conclusion

While health information technology embodies a set of concepts familiar to medical informaticians and others in the health IT field, its concepts are foreign to many in the health care field and present barriers to effective planning, implementation, and evaluation of electronic health record systems and other health information technologies. Because of the need to provide a national resource for health IT information, and to make this knowledge repository readily accessible, a health IT taxonomy was developed through a series of iterative processes.

The health IT taxonomy provides a hierarchical and interrelated classification schema for use by those engaged in the development, implementation, and evaluation of health information technologies. Its use on the AHRQ health IT Web site enables users to efficiently access the resources created and aggregated by NRC staff. Application of the taxonomy to other NRC activities and external knowledge stores is planned, and we believe it will create better understanding of common terms and their interrelations.

Just as the field of information technology is rapidly changing, the health IT taxonomy must remain dynamic. In addition to continued application of it to other components of the AHRQ health IT web site, additional evaluation is planned. This taxonomy is offered to anyone creating a searchable knowledge repository, and we hope that future development will become a living process with contributions from all who use it.

Acknowledgements

This work was supported by the Agency for Healthcare Research and Quality (AHRQ) as part of the AHRQ National Resource Center for Health Information Technologies (Contract No. 290-04-0016) project to build and maintain a free, public Web site for those interested in the

implementation and evaluation of health IT. We would also like to thank the following NRC colleagues who contributed greatly to the development of the taxonomy: Davis Bu, MD, Center for Information Technology Leadership, Mark Frisse, MD, MBA, Vanderbilt Center for Better Health, Anita Samarth, ASTECH Consulting, and Jack Starmer, MD, Vanderbilt Center for Better Health.

References

- [1] Bruno D, Richmond H. The truth about taxonomies. *Info Mgmt J*. 2003 Mar/Apr: 44-53.
- [2] The Columbia Encyclopedia. 6th Ed. [homepage on the Internet] Classification; c2001-05 [cited 2006 Mar 9]. Available from: <http://www.bartleby.com/65/cl/classifi.html>.
- [3] National Library of Medicine. Introduction and history. *Bull Med Libr Assoc*. 1961;49Pt2:1-7.
- [4] National Library of Medicine. Operational history. *Bull Med Libr Assoc*. 1961;49Pt2:64-71.
- [5] Healthcare Information Management Systems Society [homepage on the Internet]. EHR mind map: a comprehensive resource guide. c2003 [updated 2004 Jan 21; cited 2006 Mar 3] Available from: <http://www.himss.org/content/mindmaps/EHR/index.htm>.
- [6] Association for Computing Machinery [homepage on the Internet]. The ACM Computing Classification System [1998 Version]. c2006 [cited 2006 Aug 11] Available from: <http://www.acm.org/class/1998/>.
- [7] MeSH Browser [homepage on the Internet]. U.S. National Library of Medicine [2006 MeSH]. [cited 2006 Sep 5] Available from: <http://www.nlm.nih.gov/mesh/MBrowser.html>

Address for Correspondence

Brian E. Dixon, M.P.A.
Regenstrief Institute, Inc.
410 West 10th Street, Suite 2000
Indianapolis, IN 46202
U.S.A.
(317) 423-5582
bdixon@regenstrief.org

Copyright of *Studies in Health Technology & Informatics* is the property of IOS Press and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.