

What Mapping and Modeling Means to the HIM Professional

by Patricia S. Wilson, RT(R), CPC, PMP

Objective

The processes of data mapping and concept modeling are required to help meet the goal of interoperability for an electronic health record (EHR). Interoperability, as defined by the Institute of Electrical and Electronics Engineers, is “the ability of two or more systems or components to exchange information and to use the information that has been exchanged.”¹ This paper explains how these processes contribute to the exchange of information between multiple health information systems. In it, I will discuss the differences and similarities between data mapping and coding for reimbursement. The effect upon HIM professionals will be profound as their job functions either shift to performing data mapping and concept modeling, or as they see the results of these processes upon their current roles.

Discussion

EHR Brings Changes to Health Information Management

In November 2005, AHIMA and the American Medical Informatics Association (AMIA) held a joint summit to discuss work force challenges related to the EHR and the national health information infrastructure (NHII). Their findings were published in April 2006 in the document “Building the Work Force for Health Information Transformation.”² The publication makes recommendations for preparing healthcare professionals, especially those involved in HIM, to understand and to assume the tasks in implementing an EHR and the infrastructure required to support it.

The coding of data is a familiar subject to the HIM professional. The process of coding to ICD-9-CM for diagnosis and procedures became a necessary HIM function in the 1970s when that classification system was first introduced. The function was elevated to a specialized skill in the mid-1980s when ICD-9-CM became an essential part of reimbursement. The migration of the responsibility of the HIM professional beyond the assignment of diagnostic and procedural codes is a natural progression, as technology and electronic health information are embraced by the industry.³ A new focus for the HIM professional would be data mapping, which is the process of matching a given term from a data source to a term in a target set of data that contains the same information but is identified with a different name.⁴

Interoperability

Lack of interoperability was cited by the American College of Medical Informatics as the most significant technical issue preventing the adoption of an electronic health record within the United States.⁵ In June 2000, the National Committee on Vital Health Statistics (NCVHS) presented its findings on uniform data standards for patient medical record information, as required by the Health Insurance Portability and Accountability Act (HIPAA) of 1996. In the report, interoperability is defined as “the

ability of one computer system to exchange data with another computer system.”⁶ The report goes on to state that in order to achieve interoperability, uniform data standards are required to address data comparability and data quality, and that standard messaging formats and standard terminologies will achieve this goal.

The report further defines three levels of interoperability as they relate to data flow within an electronic health record: basic, functional, and semantic. Basic interoperability is the ability to send and receive a message from one computer to another. Functional interoperability is achieved when there is a common message syntax that allows computers exchanging data to interpret the message format, but not the meaning of the message. Semantic interoperability means that information within the message can be interpreted when exchanged between computers. The mechanism for providing the commonality of meaning is standardized codification of data.

Code Sets and Standards

The codification of data discussed in the 2000 NCVHS report is not coding as the HIM professional currently understands it. At this time, the HIM professional primarily uses ICD-9-CM, CPT[®], and HCPCS to code data, which group codes for the purpose of reimbursement. These systems classify terms by predefined parameters, usually for the purpose of reimbursement or population studies. These code sets are likely to remain as standards or be replaced by comparable or more specific code sets and groupers, such as ICD-10-CM, ICD-10-PCS, and APR-DRG[™].

Other functions within the EHR will require more granular terms than those used for reimbursement in classification systems and groupers. The International Organization for Standardization (ISO) standard 1087 defines terminology as a set of terms representing the system of concepts of a particular subject field. Terminologies are often tailored to the particular area of interest such as pharmacy, radiology, laboratory, and clinical management. A medical institution often develops many of these terminologies internally and specific to its needs and uses—for example, a chargemaster for order entry can be considered a locally defined terminology. Some other specific types of terminology in a medical institution can be dictated by outside influences, but are often driven by the needs and desires of clinicians and end users within the institution. For example, some of the more commonly used commercial terminologies include the Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT[®]), National Drug Codes (NDC), and Logical Observation Identifiers Names and Codes (LOINC[®]). There are many different types of code sets because there is no single classification system or terminology that meets the needs of all operations within a medical institution.

The Healthcare Information Technology Standards Panel (HITSP) is sponsored by the American National Standards Institute (ANSI) to bring cooperation between the public and private sectors for establishing standards to support interoperability for local, regional, and national health information networks. In June 2006, they published a list of proposed standards to promote a nationwide HIT infrastructure. The primary HITSP terminology standards are

- ICD-9-CM
- CPT[®]
- HCPCS
- ASTM E1239-04—Standard Practice for Description of Reservation/Registration-Admission, Discharge, Transfer
- ICD-10-CM
- ICD-10-PCS
- HL7 V2. X (for messaging)
- HL7 V3 Clinical Document Architecture (CDA) for text reports
- LOINC[®]
- SNOMED CT[®]
- National Council for Prescription Drug Programs (NCPDP) for pharmacy
- National Drug File Reference Terminology (NDFRT)/RxNorm for formulary⁷

This list of proposed standard terminologies highlights the point that no single standard can provide all of the information needed within an EHR. In October 2006, HITSP further delineated requirements in the areas of biosurveillance, EHRs, and consumer empowerment with a focus on registration and medication history document content.⁸ As HITSP continues its work, additional components will be published producing a set of documents outlining the steps to take toward interoperability.

The Health Insurance Portability and Accountability Act of 1996 required the adoption of standards for the electronic transmission of specific administrative transactions. There are many government committees and industry organizations involved in analyzing what standards are best for the nation. The set of standards to be used by government agencies will be named by the Consolidated Health Informatics (CHI) initiative. There are numerous standard-developing organizations, such as Health Level 7 (HL7) and the American Society for Testing and Materials (ASTM). The standardization of messaging format has largely been accomplished, and many standard terminologies are being used within EHRs such as SNOMED CT[®] and LOINC[®]. In the October 2005 issue of the *Journal of AHIMA*, the Board of Directors of that organization published a position statement calling for “widespread adoption and implementation of SNOMED-CT as a standard clinical terminology in order to facilitate a national health information network and the interoperable exchange of health information between standard electronic health records.”⁹ It is highly unlikely that a single terminology would be selected for use throughout the industry, either by legislation or regulation. Regardless of what standard terminologies should be adopted, an obstacle to semantic interoperability would still be present, due to the existence of historical data already coded to different terminologies requiring data mapping to the new or existing standard terminology.

The Mapping Process

There needs to be a process of cross-referencing different terminologies, primarily for the purpose of interoperability. For example, as the United States healthcare community looks forward to the adoption of ICD-10-CM and ICD-10-PCS, there will need to be a cross-referencing between the current terms in ICD-9-CM and the new classification systems. This cross-referencing process is referred to as mapping. Lee Min Lau and James Campbell define mapping as the process of creating one-way links between concepts and terms for specific purposes, often involving patient, administrative, or interface contexts.¹⁰

Mapping is performed for various reasons and between many different vocabularies, terminologies, and classification systems. An easy-to-understand (although not necessarily consistently simple to perform) type of mapping would be based on synonymy, also known as equivalence. Equivalence maps are as the name implies—a term from the source terminology maps to a term in the target terminology because they mean exactly the same thing. An example of an equivalence map using ICD-9-CM as the source classification system to ICD-10-CM as the target classification system would be as follows:

040.82 Toxic shock syndrome A48.3 Toxic shock syndrome

The different types of rules that may be applicable for mapping, based on the granularity of the term, and attributes associated with the term will depend on the way in which the outcome of the map will be used.

Correct mapping requires a complete understanding of how data will be used. A function requirement, guideline, or use case is required to ensure that maps will be consistent across the application of all data. Use cases have become a common methodology for establishing maps. A use case “defines the intended use, audience, and shared understanding of the target and source—key to development of a useful and reproducible map.”¹¹ As a framework for mapping, the use cases should be comprehensive in outlining the requirements. It is likely that the use cases will change as the maps are completed, and as much information as possible should be gathered and documented at the beginning to provide a strong basis for the maps.

Questions that should be asked when developing use cases are as follows:

- How will the mapped data be used? Will it be used for reimbursement, research, outcome measurements, or public health studies?
- How will the data be transmitted and to what systems?
- Will the data be used to provide a selectable item within a user application, such as a drop-down list?
- Will prompts for medical alerts such as possible drug interactions or allergies be implemented based on the map?
- Will the mapped data be categorized, classified, or grouped into other data sets?
- Will the data be stored? If yes, how will it be stored?
- Where will the mapped data reside?
- How will the maps be performed? Will they be done using automated tools or manual review, or a combination of both?

Building upon the framework of the use case is the application of heuristics, which are the rules and requirements outlined for the purpose of the mapped data.¹² Heuristics also outline the criteria and represent the requirements of the use case. They assure that the maps are understandable, reproducible, and useful. Examples of heuristics include the following:

- Criteria for inclusions or exclusions
- Appropriate interpretation of “Not Elsewhere Classified” or “Not Otherwise Specified”
- Scenarios in which a source term has relationships to multiple target terms and what types of relationships are appropriate

In mapping, an important heuristic is the relationship between the terms from the source to the target—it could be one to many, many to one, or many to many. A one-to-many relationship implies that a single term from the source classification can be mapped to multiple target terms. A many-to-one relationship implies that multiple terms from the source classification can be mapped to the same target term. Many-to-many mapping means both one-to-many and many-to-one relationships happen at the same time between the source and the target terminologies. A many-to-many map is difficult to implement and should be highly scrutinized before including as part of the use case. On the whole, many-to-one mapping is more manageable than one to many for the purpose of interoperability, because “translating” multiple different source terms to the same target term does not require an additional decision once mapping has been completed, whereas for one-to-many mappings, an additional decision has to be made to select which of the many target terms should be the “translation.”

Another consideration would be whether the relationship mapping needs to be bidirectional, that is, requiring not only mapping from source to target term, but also from target back to source. In anticipation of the adoption of ICD-10-CM for use within the United States, it is quite likely that this will be needed. I assert that it is most likely that not only a map between ICD-9-CM and ICD-10-CM will be performed but also a complete map of ICD-10-CM to ICD-9-CM will be required in order to accommodate several possible use cases that have yet to be defined or documented.

Once the use cases and heuristics have been defined, mapping can begin. Mapping the volume of terms in ICD-9-CM to ICD-10-CM and ICD-10-PCS manually would be daunting, so automated tools should be used to assist with the task. There are many commercial companies that provide mapping tools and services to individuals and organizations. Medical institutions and healthcare organizations may develop their own tools to accommodate their particular needs and use cases. Commercial companies and medical informatics groups have published many studies and papers discussing the feasibility and effectiveness of automated mapping tools.^{13, 14, 15} The effectiveness of the algorithms used by each tool depends on the use cases and heuristics that can be programmed into the software. The level of confidence in automated mapping can vary from application to application. In one study performed at by

the Department of Medical Informatics at the University of Freiburg, Germany, the authors used a vector space text retrieval method. This involves assigning a number to textual terms or documents to assist in the process of comparison. The result of unambiguous mapping from ICD-9 to ICD-10 was 64 percent.¹⁶ The efficiency and accuracy of mapping tools, regardless of the method of achieving the map, depends entirely on the ability of the tool to apply the heuristics in a meaningful and consistent manner. To date, I am not aware of any mapping tool that is able to completely perform automated mapping. Manual review is required to a varying extent to map the portions that failed automated mapping and to validate the output of automated mapping. The defined heuristics need to be reproducible in order to validate mapping, regardless of whether it was accomplished by automated or manual means.

Similarities between Mapping and Coding

Many of the principles of mapping are similar to those used in the coding process. Both data mappers and coders are familiar with the structure and use of terminologies. All terminologies have a code and description. Most classification systems, such as ICD-9-CM, ICD-10-CM, and ICD-10-PCS, have a coding structure where the code itself has implied meaning. Other terminologies have moved away from “smart” codes, that is, not assigning codes based on a set format. The code structure has no importance for mappers and has limited importance for coders, with the exception of possibly assisting with memorization of frequently used terms associated with the codes. For example, a coder using ICD-9-CM is inclined to remember code 250.00 and use it in speech or notation instead of writing out the actual description of the term, which is “Diabetes mellitus without mention of complication, type II or unspecified type, not stated as uncontrolled.”¹⁷

All terminologies have a structure of some sort. Even a terminology that is considered “flat” has a structure because the commonality of the terms is implied by the title of the terminology. Many terminologies will usually have a single term under which similar terms are organized, or the structure is implied based on the grouping of similar terms and codes. For example, in the structure of CPT[®], codes beginning with a seven are indicated as being radiology terms; and codes beginning with a 78 are further subdivided into nuclear medicine.¹⁸ Most classification systems have a strict single hierarchical structure in which the term with the highest level of specificity inherits the attributes of the terms directly above it in the hierarchy (see example below). This type of hierarchy uses a parent-child relationship. Terminologies built specifically for a medical institution will have some sort of primary structure or classification—for example, a chargemaster will indicate what charges are pertinent to a given specialty or area within a hospital. The items on a chargemaster are likely to be grouped together based on the department or revenue center. Below is an example of a parent-child relationship from ICD-9-CM, illustrating how the term with the highest level of specificity inherits the attributes of the terms directly above it in the hierarchy.

```

815 Fracture of metacarpal bone(s)
  815.0 Closed
    815.00 Metacarpal bone(s), site unspecified
  
```

Most clinical terminologies, whether built commercially or unique to a medical institution, will have a poly-hierarchical structure. This allows for a single term to have multiple parents. Many terminologies also allow relationships other than parent-child between terms. Although coders do not encounter multiple relationships and hierarchies during the process of coding, it is the personal experience of me and my fellow coder-turned-mapper colleagues that the concept tends to come easily to them when they are exposed to mapping principles. To expand on the example of “closed fracture of the metacarpal bones, site unspecified” a coder will understand the logic of the term having parent-child relationships to “fracture,” “injury to hand,” and “closed injuries.”

Coders must follow heuristics as a part of coding process. The rules and regulations set forth by the Centers for Medicare and Medicaid Services (CMS) or other payers are specific heuristics for coding. The instructions at the front of publications describing ICD-9-CM or other classification systems set out the guidelines for coding and abstracting. The coding applications have embedded within them the coding rules that are pertinent to the purpose, such as a coding product or tool that alerts a coder when a

diagnosis is gender specific. A similar type of alert will occur when mapping heuristics are embedded within mapping tools. Although coders may not formally be aware of them, use cases have certainly been developed by regulatory agencies when they were in the process of establishing coding rules. This exposure to coding rules and regulations make coders comfortable with following mapping heuristics and establishing guidelines when none exist.¹⁹

Differences between Coding and Mapping

As mentioned above, the structure of terminologies can differ greatly. This difference can play a key role in the development of use cases and heuristics for mapping. Although inherited attributes, as explained above as a parent-child relationship, are important to understand when coding, the straightforward nature of a single hierarchy makes coding rules just as straightforward. In addition, most of the heuristics for coding are regulated by government or other regulatory agencies. There are numerous classes, conferences, symposiums, and books written on the principles of coding, and coders can attend formal courses and receive certification and college degrees in coding.

Data mapping is emerging as a specialized skill within the healthcare field. Most data mappers learn their trade while on the job. To date, there is no certification process or college degree for mapping. The heuristics for data mapping are not regulated at this time. They are developed by the organization performing the mapping based on the use cases established at the time of need. For example, an organization may have the need to map all of their pharmacy and supply data to a new product and database in order to implement a computerized provider order entry system. Another example would be the need for mapping when two organizations merge into a single system and data structure. According to an article in the October 2006 *Journal of AHIMA*, titled “The HIM Impact on EHRs,” the HIM professional is likely to be involved in this transition process anywhere from 12 to 69 percent of the time, depending on their level of involvement.²⁰ Thus, the field of mapping provides growth and challenge for those individuals willing to take on the task.²¹

Some of the terminologies that often require mapping contain information that is not usually required for coding. A coder who is presented with this mapping task may feel a bit overwhelmed by the presentation of terms in an unfamiliar format. For example, a mapper who has the task of mapping a LOINC[®] term to SNOMED CT[®] may see something like this:

LOINC Code	LOINC Attributes
INTRAVASCULAR MEAN:PRES:PT:ARTERIAL SYSTEM:QN:	

SNOMED CT [®] Concept ID	Preferred Term
6707001	Mean blood pressure (observable entity)

This is a common mapping scenario, which requires the mapper to have a detailed understanding of the structure and content of the terminologies being mapped. Research and study are needed in order to differentiate between the two structures and understand how they relate.

The biggest area of difference between mapping and coding is the context and the outcome of the two processes. The process of coding and abstracting involves direct access to patient information. The outcome is the assignment of a code or codes based on clinical, demographic, and other information in a patient record. Mapping does not involve any patient information. The map is performed between source and target terminologies based solely on use case and heuristics. No assumptions can be made when mapping, unless the assumption is clearly defined in the heuristics. This is probably the biggest hurdle in moving from coding to mapping. A coder who has been formally trained with coding rules often has a problem setting aside those coding-specific guidelines in order to map.

For example, a coder may be reviewing a patient’s medical record to select a code for the diagnosis of stress incontinence. The coder will look at the patient’s gender in order to decide which code to choose: 788.32, Stress incontinence, male or, 625.6 Stress incontinence, female. A mapper, however, cannot look at a patient record to ascertain the gender. When presented with the term “stress incontinence” for

mapping, the mapper will need to rely solely on the heuristics to decide if a map exists. If no rule has been developed based on the use case, the mapper will map to a similar, non-gender-specific term in the target terminology, or indicate that no map can be found if the target terms are all tied to gender.

Mapping and Modeling

A term that is considered to be a “no map” introduces an additional terminology function called modeling. Modeling is the building of a term in the target terminology in order to provide a map for the source term. Not all mapping use cases and heuristics allow for modeling. Modeling in many terminologies requires special allowances from the organization managing or owning the target terminology. SNOMED CT[®] has made provisions for modeling within its structure by using specially assigned extensions to allow an institution to model terms unique to their enterprise.²² Modeling will often require additional heuristics not used by conventional mapping processes in order to ensure that all relationships and appropriate displays are created in a manner consistent with the development of the target terminology.

Mapping and a Data Dictionary

A work group assembled by AHIMA on EHR data content defined a data dictionary as “a descriptive list of the names (also called representations or displays) and definitions of data elements to be collected in an information system or database. The purpose of the data dictionary is to standardize definitions and ensure consistency of use.”²³

A data dictionary should support encoding, exchange, and comparison of clinical data between independent computer systems. It should provide a structure to enable decision support, reporting, and standard queries. Its goal should be to support standardization of clinical data within a medical institution or across multiple enterprises, by incorporation of industry-standard terminologies.²⁴

Maintaining a data dictionary often requires mapping; vice versa, mapping enterprise data requires a data dictionary. One development model is to create and maintain a data dictionary by mapping multiple industry-standard terminologies to an internal terminology reference acting as the core.²⁵ This process is resource intensive, requiring the investment of people, time, and tools. Experts in standards and terminologies are required in order to build and maintain the content of a data dictionary. The HIM professional is often the person who is called upon to perform this work.²⁶

Conclusion

Many coders may find their role changing to data mapper and modeler as medical institutions and other healthcare organizations work toward interoperability. Coders will be enlisted for these tasks because of their exposure to clinical data and their background in classification systems. The process of coding requires a person who is detail oriented and excellent at establishing and following heuristics, which are essential skills for a good data mapper and modeler.²⁷ Even though there is a natural progression from coding to mapping, there are significant differences in the two processes, and mapping and modeling require specific training. There have been informal discussions within AHIMA regarding a need for a nationally recognized certification process to be developed to provide training and recognize the specialized skills necessary to perform data mapping and modeling. Although nothing has been published or formalized at this time, HIM professionals should look forward to expanding their capabilities as this burgeoning field emerges into national recognition.

Patricia S. Wilson, RT(R), CPC, PMP, is the Team Lead in the Terminology Consulting Services and Healthcare Data Dictionary[®] department at 3M Health Information Systems in Salt Lake City, UT.

Notes

1. Institute of Electrical and Electronics Engineers. *IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries*. New York, NY: Institute of Electrical and Electronics Engineers, 1990.
2. "Building the Work Force for Health Information Transformation." Supplement to the *Journal of AHIMA* 77, no. 4 (April 2006): 1-14.
3. Johns, M. "A Crystal Ball for Coding." *Journal of AHIMA* 71, no. 1 (January 2000): 26-33.
4. McBride, S., R. Glider, R. Davis, S. Fenton. "Data Mapping." *Journal of AHIMA* 77, no. 2 (February 2006): 44-48. [expanded online edition]
5. "Factors and Forces Affecting the EHR System Adoption: Report of a 2004 ACMI Discussion." *Journal of AMIA* 12, no.1 (Jan/Feb 2005): 8-12.
6. National Committee on Vital Health Statistics. "Report on Uniform Data Standards for Patient Medical Record Information." July 6, 2000. Available at <http://www.ncvhs.hhs.gov/hipaa000706.pdf>.
7. Health Information Technology Standards Panel. "Healthcare Information Technology Standards Panel Technical Committees: Selected Standards." Version 2.0, (June 29, 2006): 1-24.
8. Health Information Technology Standards Panel "HITSP Interoperability Specification: Registration and Medication History Document Content Component." Version 1.2, (October 20, 2006): 1-73.
9. American Health Information Management Association. "Implementation of SNOMED-CT Needed to Facilitate Interoperable Exchange of Health Information." *Journal of AHIMA* 76, no. 9 (October 2005): 30-32.
10. Lau, L.M., J.R. Campbell. "Putting Standards to Work: Vocabulary Implementation in the Real World." *Proceedings of the 2002 AMIA Annual Symposium*. Washington, DC: American Medical Informatics Association, November 9, 2002.
11. Campbell, J.R., M. Imel. "The Function of Rule-Based Mapping within Integrated Terminology Management." AHIMA's 77th National Convention and Exhibit Proceedings, October 2005.
12. McBride, S, R. Glider, R. Davis, S. Fenton. "Data Mapping."
13. Nachimuthu S K, Woolstenhulme, RD. "Generalizability of Hybrid Search Algorithms to Map Multiple Biomedical Vocabulary Domains." AMIA 2006 Proceedings. 2006:1042.
14. Shakib, S.C., K.B. Poon, L.M. Lau. "Tools and Processes to Improve Data Mapping Accuracy and Reliability." Proceedings of the 11th World Congress on Medical Informatics. 2004: 1858.
15. Lau, L.M., K. Johnson, K. Monson, S.H. Lam, S.M. Huff. "A Method for the Automated Mapping of Laboratory Results to LOINC." AMIA 2000 Proceedings. 2000: 472-476.
16. Schulz, S., A. Zaiss, R. Brunner, D. Spinner, D.R. Klar. "Conversion Problems Concerning Automated Mapping from ICD-10 to ICD-9." *Methods of Information in Medicine* no. 37 (September 1998): 254-259.
17. American Medical Association. "International Classification of Diseases, ICD-9-CM 2007". Volumes 1, 2, and 3. Ninth Revision, Clinical Modification, 2007.
18. American Medical Association. "Current Procedural Terminology CPT® 2007." Professional Edition, 2007.
19. McBride, S., R. Glider, R. Davis, S. Fenton. "Data Mapping."

20. Fenton, S., M. Amatayaku, M. Work. "The HIM Impact on EHRs." *Journal of AHIMA* 77, no. 9 (October 2006): 36-40.
21. Giannangelo, K. "Establishing Professional Development Goals." *Journal of AHIMA* 77, no. 9 (October 2006): 98-99.
22. College of American Pathologists. "Introducing Extensions (SNOMED CT Complement)." *SNOMED Clinical Terms[®] Technical Reference Guide*. Northfield, IL: College of American Pathologists, (January 2006): 61-67.
23. Adam, A., D. Barley, et al. "Guidelines for Developing a Data Dictionary." Practice Brief. *Journal of AHIMA* 77, no. 2 (February 2006): 64A-64D.
24. 3M Health Information Systems. "Making Sense of the Data: Using a Medical Data Dictionary to Integrate, Share, and Understand Clinical Data." White Paper. (August 2005): 1-16.
25. Adam, A., D. Barley, et al. "Guidelines for Developing a Data Dictionary."
26. Fenton, S., M. Amatayaku, M. Work. "The HIM Impact on EHRs."
27. McBride, S., R. Glider, R. Davis, S. Fenton. "Data Mapping."