Data Mapping and Its Impact on Data Integrity
AUTHORS AND CONTRIBUTORS:
Linda Hyde, RHIA; Theresa Rihanek, MHA, RHIA, CCS; Terry Santana-Johnson, RHIT, CDIP, CCS, CCS-P; Rita Scichilone, MHSA, RHIA, CCS, CCS-P; Cortnie Simmons, MHA, RHIA, CCS; Jane Beth Turner, RHIA; Wendy Zumar, MA, RHIA, CCS

ACKNOWLEDGEMENTS:
Sue Bowman, MJ, RHIA, CCS, FAHIMA; June Bronnert, RHIA, CCS, CCS-P; Marlisa Coloso, RHIA, CCS; Angie Comfort, RHIT, CDIP, CCS; Katherine Downing, MA, RHIA, CHP, PMP; Lesley Kadlec, MA, RHIA; Cheryll Rogers, RHIA, CDIP, CCS, CTR; Joanne Romasko, RHIA, CPC, CHDA; Rayna Scott, RHIA, CHDA, MS

EDITOR:
Anne Zender

DESIGN:
Maria Sitelis

Representing more than 71,000 specially educated health information management professionals in the United States and around the world, the American Health Information Management Association is committed to promoting and advocating for high quality research, best practices, and effective standards in health information and to actively contributing to the development and advancement of health information professionals worldwide. AHIMA's enduring goal is quality healthcare through quality information.

© 2013
INTRODUCTION

The current rise in data mapping projects is the result of the need to link disparate electronic data systems in a rapidly changing environment. Mapping projects are valuable in a variety of situations where data elements from one code or data set are compared to another set and evaluated for equivalence of meaning to accomplish a defined “use case.” Code sets related to health information functions include CPT and its modifiers, ICD-9-CM, ICD-10-CM/PCS, and HCPCS level II, as well as LOINC, Rx NORM, and SNOMED CT®. Additionally, quality measures such as those used by the Agency for Healthcare Research and Quality and National Quality Forum are frequently linked to these code sets and may require internal mapping to ensure accurate measurement. Data mapping is not limited to just these code sets; there are many different types of maps in the healthcare realm.

The increased demands for data sharing and interoperability, especially across different practice settings and different classification systems, increase reliance on data mapping tools and techniques. The use of these tools requires frequent integrity checks. Understanding the role and context of data maps, as well as their strengths and weaknesses, is essential in ensuring the reliability of the data entries derived from maps. Data mapping tasks may be as simple as matching a provider’s administrative codes for disposition to an external standard such as UB-04 or taking a more complex clinical condition and creating a standard representation across different standardized representations. In all situations employing maps, processes and guidelines must be clearly defined and documentation prepared to explain how the map was created, tested, and performing correctly for its intended use case.

Thinking through the risks and unintended consequences of map use is mandatory before any projects are planned. Careful consideration of liability is required before any mapping between disparate sources is attempted for healthcare situations.

This thought leadership paper explores the relationship of data mapping and data integrity assurance by providing guidance to avoid adverse outcomes involving the use of maps.

Information management projects are data-centric projects. Data mapping requires knowledge of information technology, the data sets being mapped, and project management, which often requires a diverse team approach. Health information management (HIM) professionals are frequently involved in planning and developing data maps that utilize code sets, as there is a need for competency in both the source and target systems involved in the mapping projects. The source is the origin of the map or the data set from which one is mapping. The target is the data set in which one is attempting to find equivalence or define the relationship. Figure 1 depicts the typical mapping process.

All installed and active software maps of any kind must be identified, inventoried, and checked for validity. Poorly designed and out-of-date mappings create significant data integrity problems in health information systems. Undetected errors in data maps have the potential to introduce many problems, including the filing of false claims to insurers, delivering the wrong information for patient care and/or quality measures, or causing a breach in patient privacy.

Figure 1 Mapping Workflow
EXAMPLES OF DATA INTEGRITY INVOLVING MAPPING

Ensuring data integrity when data maps are present requires diligence to ensure:

- Trustworthiness of mapped information over its life cycle and use
- Integrity is accomplished by validation and regular “checkups” of the data flow and map performance.
- Mapped entries reflect current content through the entire workflow, consistent with the use case for the map in both primary and secondary uses of the same data.
- Integrity is achieved by monitoring the mapping results specific to the use case and how those results are repurposed for other uses.
- Maintain the intended content of a validated map, in case the mapped data structure changes through data processing. Ongoing vigilance is required to test results of the map.
- The map provides uniform, reliable, complete, and unchanged semantic output through data stewardship protocols.

DANGERS OF INAPPROPRIATE MAPPING

When evaluating data integrity issues involving mapping, it’s critical to understand the elements of all the code sets or data sets being mapped. The strengths and weaknesses of each source or code set, their intended use, and how the map is created are all important to building a successful data map. Using a map for a purpose not intended or misunderstanding the construct of the source and target can lead to incomplete, incorrect, and inappropriate maps. The following examples illustrate several key issues related to inappropriate mapping practices:

- SNOMED CT is a comprehensive clinical terminology that contains content for both human and veterinary medicine. SNOMED CT hierarchies are complex and require careful consideration in developing heuristics and guidelines. The structure supports terms within the hierarchy commonly referred to as a “parent” or “child.” The definition of a subtype is a (SNOMED CT) concept that has a direct subtype relationship to a specified concept.2 SNOMED CT also includes (subtype) descendants that are all subtypes of a concept, including subtypes of subtypes. For example, if a concept has four children, then descendants are those children plus all the concepts descended from those four children.
  
  - When exporting concepts for a map using SNOMED CT, it is critical to ensure the correct reference set is used to exclude terms exclusively used for non-human concepts when working on human-only content.
  
  - For example, when there are child terms for a selected parent concept, all children should be reviewed for inclusion in the map. Children of the term “aneurysm” include both acquired and congenital aneurysms, but the map may only be intended to identify acquired conditions.
  
  - When deploying maps for health record use, the integrity of results will be compromised if the source system entry does not represent valid health record content precisely (that is, semantic match). Data maps cannot provide context or inference of content or integrity. It’s important to build in steps to ensure appropriate structure and validation of the map before it is used, and to test in the “live” environment to ensure there are no errors in the data processing cycle.
• When developing maps for health records or clinical data management, take care to ensure reliability of the results. Maps should not be used as a substitute for health record encoding. Maps are created without context and information available to assign codes for health records as a source for administrative purposes including reimbursement for care. For example, the term “allergy to penicillin” can be appropriately coded when the context of the health record reveals this is a history or status rather than an acute allergic reaction. However, in mapping, this same contextual information is not available, which requires guidance for the user to confirm results.

• Creating maps internally and using unqualified personnel in map development compromises integrity of results. Use skilled personnel familiar with data mapping requirements, limitations, and pitfalls to ensure reliable results.

• Using maps for healthcare claims can lead to compliance issues or even allegations of fraud if the map results in incorrect code submission linked to health insurance billing for reimbursement.

Exploring the Importance of Business Rules and Map Heuristics

Data mapping is a key component of requirements analysis. This analysis needs to be completed by a subject matter expert early in the project or after the mapping has been completed but not yet deployed. It’s essential to carefully plan and assess requirements, since mapping projects require time, skills, and effort that may be costly. Data maps should not be attempted if there is a more direct process available for the use case. In data mapping, planning time spent to ensure integrity of the map performance can be provided on the front-end or back-end project planning.

All maps require an investment of time and resources, and some mapping projects may require more than others. When the source data and the target data are similar in structure with a high percentage of exact matches in content and meaning, the time needed for validation will be minimal. However, when the source and target do not result in an exact match, time must be spent to determine which of the choices are appropriate based on the map’s use case. Unless the mapping is a one-to-one match from source to target, decisions must be made to meet the intent of the map.

The adoption of business rules adds another tier to systems that automate business processes. For mapping projects, this is often referred to as map heuristics. If there is a one-to-one match from source to target information, automated systems are able to perform the mapping activity with little human intervention. If this is not the case, a commitment of time must be applied to the mapping process to ensure the end result meets the map’s intent and conforms to the project’s use case. Subject matter experts managing or reviewing the map must be well versed concerning source data with guidance on what the target data should be.

For example, specific expertise in the use of Current Procedural Terminology may be required to understand how to map codes correctly for their intended use in professional fee billing. For this reason, optimal mapping results require a comprehensive understanding of how the mapped data will be used. A functional requirement, guideline, or use case is required to ensure maps will be consistent across the application of all data requirements. Mapping heuristics are “rule of thumb” guidance that provide rules for how to map from source to target in a consistent manner. Detailed instructions are provided so consistency is ensured between map developers throughout the project. Every mapping project must have clear instructions to make the map results “understandable, reproducible, and reliable.”

Along with a well-documented use case describing the need for a map and its intended purpose, heuristics are essential for mapping success. It is imperative that decisions made regarding business rules and map heuristics are clearly documented so evidence is available describing why each decision was made and by whom. This serves as an audit trail of decisions made during the mapping process.
Mapping is almost always a resource-intensive project requiring hands-on review and considerable knowledge about the source and target. Human intervention is necessary for mapping design and validation of map outcomes. Mapping tools may assist in the process by providing varying degrees of automation. Manual review is required, to a varying extent, to map the portions that failed automated mapping and to validate the results of automated mapping. The defined map development rules must be reproducible in order to validate mapping(s), regardless of whether it was accomplished by automated or manual means. Assumptions may be used with caution in data mapping when they are fully defined and included in the documentation of the use case and can be relied on to produce results without data integrity errors. For example, the ICD-10-CM/PCS to ICD-9-CM reimbursement mappings assume the use case is for inpatient admissions and reimbursement.¹

**REVIEWING VENDOR MAPS**

For all managers of health information, it is important to identify all existing applications that are using maps of any kind for health information use, including those involving clinical terminology systems as a source or target. The vendor should be able to provide a list of maps that transpose or translate one data set to another being used for review. HIM professionals should be aware of and monitor these, since there is significant potential for error if the vendor is not keenly aware of the content and how it is designed to be used. This documentation should include a data dictionary to provide a complete definition and the intended use of each data component. The data dictionary will ensure the interpreted meaning of the data element is correct. For example, a field for “Provider ID” could have many different definitions such as billing identification number, national provider identifier, social security number, and so forth.

End users should create an acceptance testing process to verify the map as part of their decision to purchase the product. Such maps may be as simple as admit, discharge, and transfer information mapped into products, or as complex as provider maps that determine under which provider a bill may be submitted. Validity testing protects against mapping errors that threaten the integrity of the mapped results.

Linking data from one coded data system to another is rarely perfect and even subtle differences in the meaning can cause integrity problems. When the source and target are not a semantic match, the problems begin when a word or term has more than one meeting and the context is not available to the user and assumptions are made.

**MAINTENANCE PROCESS OF MAPS CANNOT BE IGNORED**

The source and target systems for any map are expected to change over time. In some cases, these systems have periodic, regularly scheduled updates such as ICD-9-CM or ICD-10-CM, or release dates for SNOMED CT. If a map is created between ICD-10-CM and SNOMED CT, which both have different update schedules, this makes the maintenance schedule more complex and costly, but the maintenance is necessary to keep the map functioning properly. This maintenance process not only needs to accommodate the addition of new codes, but also the retirement of existing codes, changes in the code description, or any other change. When revisions or deletions are made in the code sets and these codes are used in a map, versioning is needed so that historical data that uses this map is not adversely affected. Versioning provides a reference to ensure that maps that were current in a given time period are still valid going forward, but replaced with current maps for the current time period.
MAPPING PROCESS AND TOOLS AFFECTING DATA INTEGRITY

The challenges are numerous in electronic systems and workflows. These include:

- Drop-down pick lists for recording clinical facts generated from maps, which may result in inappropriate or less specific selections
- Computer-assisted encoding software automation submitting codes that misrepresent the facts of the encounter
- Use of templates developed from mapped data, which may exaggerate the clinical facts or omit important input
- Workflows involving maps resulting in the submission of non-specific or inappropriate codes or data elements
- Forgetting to update maps when the source and target systems have mapped items changed or discontinued
- Automated mapping producing errors without detection
- Interface engines that employ maps that are not functioning properly
- Default responses used within rule-based maps resulting in inaccurate results
- Duplicate data entries if software applications are employing maps and manual data entry is submitted through a separate process
- System override capabilities cancel the use of the map for health records and the ability to copy information from one source and paste it into another affects the integrity of the map’s target

Awareness of these common areas of concern should bring additional scrutiny to the data mapping planning process. For example, a billing system designed to facilitate claims submission may have certain items mapped incorrectly for deployment through an interface, causing the host system to house the correct data, with the “mapped to” results returning an incorrect value. Another possible issue may arise when a source field is mapped to the wrong field (as opposed to the wrong value) in the destination system. For all these reasons, careful design of the map and validity of the setup is required.

Data integrity includes the ability to depend on finding the accurate data and being able to trust its reliability. Having the data stored once increases reliability and reduces the opportunity for conflicting results. Data dictionaries assist in ensuring that data is requested once and that results are consistent.

VALIDITY TESTING FOR DATA INTEGRITY ASSURANCE

While testing is an important component of development of any application or system, it must continue after the application is released for use. Validity testing for data maps needs to occur on a regular basis in the production environment to verify the map is still being used for its intended purpose and is consistently delivering accurate mapping results. This can be accomplished using a random sampling basis or by focusing on the maps in search of high-volume data to watch for data integrity problems. A monthly validation report on a field-by-field basis which includes the benchmark is helpful to quickly identify any potential new mapping issues that have developed. For example, if the field “Referring Physician” is typically populated 83 percent of the time and this month’s validation report shows there is a decrease to 70 percent, this is a red flag requiring research to identify the source prior to using the data. Whenever possible, validity testing should be conducted in the “live” environment to ensure end-to-end integrity of the content.
RECOMMENDATIONS FOR BEST PRACTICES TO AVOID DATA MAPPING ERRORS

Any data integrity process is an integral part of a data stewardship program. Data stewardship involves administering the organization’s data assets to ensure ongoing usability, accessibility and quality. Recommendations to avoid data mapping errors include:

- Document the map heuristics and standing business rules surrounding the development of each map.
  - These rules would include well developed use cases for each map, the identification of applications that would utilize the map, and documentation to explain how mapping rules are created and deployed in the workflow.

- Create a program and process to test the validity and reproducibility of the map.
  - A testing program should cover the development process of the map and any associated tools used from map development to end-user acceptance testing and approval.

- Create and implement a maintenance program.
  - Source and target code sets for any map may be subject to change due to periodic updates, removal from source or target data sets, discontinuation, or major version changes from either end of the map.

Notes


References

AHIMA. “Putting the ICD-10-CM/PCS GEMs into Practice.” Updated May 2013. Available in the AHIMA Body of Knowledge at ahima.org.

