Managing a Data Dictionary - Retired

Save to myBoK

This practice brief has been updated. See the latest version here. This version is made available for historical purposes only.

In recent years, the American Recovery and Reinvestment Act, numerous health IT initiatives, and the growth of health information exchange (HIE) have increased the healthcare industry's focus on data management and data use. Currently many organizations store data in multiple health information systems that are disparate-meaning the data within each system stand alone and are not interoperable. The data may or may not be collected consistently.

This lack of data consistency can create challenges for data comparison and reporting for initiatives such as those outlined above and can lead to errors in data use.

Accurate and reliable data are integral to the many health IT initiatives currently under way. According to the International Organization for Standardization (ISO):

> The increased use of data processing and electronic data interchange heavily relies on accurate, reliable, controllable, and verifiable data recorded in databases. One of the prerequisites for a correct and proper use and interpretation of data is that both users and owners of data have a common understanding of the meaning and descriptive characteristics (e.g., representation) of that data. To guarantee this shared view, a number of basic attributes has to be defined.[1]

A data dictionary is one tool organizations can use to help ensure data accuracy.

This practice brief describes common data inconsistencies found within healthcare organizations' systems and defines the data dictionary and its associated data management challenges. It also outlines best practices for maintaining data integrity, including the HIM professional's role.

Common Data Inconsistencies

In many organizations data are stored in different databases and may be of inconsistent quality. Issues such as variable naming conventions, definitions, field length, and element values can all lead to misuse or misinterpretation of data in reporting. The following examples illustrate common data inconsistencies.

Inconsistent naming conventions. The date of the patient's admission is referred to differently in different systems: "Date of Admission" in the patient management module within the EHR, "Admit Date" in the fetal monitoring system, and "Admission Date" in the cardiology database. The unique patient identifier is referred to as a "Medical Record Number" in the patient management system, "Patient Record Identifier" in the operating room system, "A number" (a moniker leftover from a legacy system from 25 years ago), and "Enterprise Master Patient Identifier" in the catheterization lab system.

Inconsistent definitions. The patient access module defines date of admission as the date on which an inpatient or day surgery visit occurs; the trauma registry system defines it as the date on which the trauma patient enters the operating room. Pediatric age is defined as age less than or equal to 13 in the EHR module, while the pediatric disease registry defines pediatric age as below the age of 18. In the bed board system, a nursing unit may be defined as 5W or 5 West. Within the scheduling system, unique locations are defined as short procedure unit or SPU, x-ray, or radiology.

Varying field length for same data element. The field length for a patient's last name is 50 characters in the patient management module and 25 characters in the cancer registry system. The medical record number in the patient management system is 16 characters long, while the cancer registry system maintains a 13-character length for the medical record number.

Varied element values. The patient's gender is captured as M, F, or U in the patient access module, while the gender is captured as Male, Female, or Other in the peripheral vascular lab database.

Inconsistencies in data definitions can lead to innaccurate data use and health data reporting and can potentially affect the quality of care.

Data Dictionary Defined

A data dictionary provides a descriptive list of names, definitions, and attributes of data elements to be captured in an information system or database. It describes the definitions or the expected meaning and acceptable representation of data for use within a defined context of data elements within a data set. It also provides metadata or information about data.

The metadata may include other attributes or characteristics such as length of data element, data type (e.g., alphanumeric, numeric, date, special symbols), data frequency (mandatory or not), allowable values or constraints, originating source system, data owner, data entry date, and when the data element is no longer collected. All systems-administration, financial, and clinical-require data definitions.

The goal is to achieve consistently defined and standardized data.

A simple dictionary can be managed in a spreadsheet or table; a complex dictionary may require a data management program application. AHIMA's "Health Data Analysis Toolkit," available in the AHIMA Body of Knowledge at www.ahima.org, includes an example of a data dictionary.

Data dictionaries should be accessible to data analysts and those authorized users in the organization or enterprise who manage data, use data to manage their work, contribute data to other internal or external systems, and external audit organizations conducting assessments of information system capabilities. The ability to edit the dictionaries, however, should be limited to system administrators.†

A data dictionary can be consulted to understand a data element's meaning and provenance. It is a dynamic document that must be updated as data collection requirements change. The dictionary acts as a resource when reviewing results of reports generated from a data system. It serves as an important tool during data sharing, exchange, or integration purposes.

While data dictionaries are useful for the consistent collection of data, it is imperative that the data are managed and validated for accuracy in reporting.

Why Data Standards Matter

Data standards are an integral component to an organization's data dictionary. Organizations should align the entries in their data dictionary with current data standards to ensure they are in compliance.

Standards play an important role in healthcare. Without standards, the steps toward interoperable HIE might never have been taken. Standards also have enhanced organizational leaders' ability to interpret their data for patient care, business, research, and comparative performance improvement reporting activities.

The American National Standards Institute (ANSI) governs all standards development organizations in the United States. It uses a consensus process to ensure all interested parties associated with standards are involved in their development.

The federal government has established an initial set of standards to support HIE for meaningful use, but adoption of national standards has not become widespread.[2] As HIE continues to increase, healthcare organizations will need to properly identify their data elements for appropriate transmission.

The Benefits of a Data Dictionary

A data dictionary promotes data integrity by supporting the adoption and use of consistent data elements and terminology within health IT systems. By adopting a data dictionary, organizations can improve the reliability,

dependability, and trustworthiness of data use.

An established data dictionary can provide organizations and enterprises many benefits, including:

- Improved data quality
- Improved trust in data integrity
- Improved documentation and control
- Reduced data redundancy
- Reuse of data
- Consistency in data use
- Easier data analysis
- Improved decision making based on better data
- Simpler programming
- Enforcement of standards[1]

A data dictionary promotes clearer understanding of data elements; helps users find information; promotes more efficient use and reuse of information; and promotes better data management.

Note

1. Department of Health and Human Services. "Health Information Technology: Initial Set of Standards, Implementation Specifications, and Certification Criteria for Electronic Health Record Technology." *Federal Register*, July 28, 2010. http://federalregister.gov/a/2010-17210.

Best Practices for Maintaining Data Integrity

Decisions are only as good as the data on which they are based. The data dictionary is the foundational document for maintaining the integrity of an organization's data. A detailed and exacting process is required to create a data dictionary.

A data dictionary is a dynamic document that is evaluated as data needs change or grow. Managing an organization's data and those who enter it is an ongoing challenge requiring active administration and oversight.

The following best practices help organizations maintain their data dictionaries and data integrity.

Know the data. Organizations should define the metadata required of their health information systems and identify implications on technology decisions. The "Data Quality Attributes Grid," in the online version of the September 2007 practice brief "HIM Principles in Health Information Exchange," provides a guide for defining data and their attributes.

When possible, organizations should design the data collection system well in advance of system implementation. This will allow for thoughtful design to identify data elements needed to achieve the purpose of the collection.

Organizations should not collect data simply because they can. Irrelevant data become distractions during the analysis and decision-making processes. Irrelevant or unnecessary data add hidden, unnecessary costs throughout their life cycle.

Organizations should use the "collect once, use many" rule for data collection. "Using and reusing health data for multiple purposes can maximize efficiency [and] minimize discrepancies and errors caused by multiple data entry processes..."[3]

Organizations should consider transitioning to a core data service model where the key or common data are centralized and can be accessed by many, thus reducing the incidence of potential introduction of error in the collection process.

Organizations should understand the data's importance before making changes. They should define and document a data dictionary for each system and understand what data are currently collected, why they are collected, and how they are used. They should research what impact a change to the system would have on the data.

Steps for Maintaining Data Integrity

Managing an organization's data and those who enter it is an ongoing challenge requiring active administration and oversight. The following best practices help organizations maintain their data dictionaries and data integrity:

- Know the data
- Map the data across all systems
- Develop a data quality management process that includes ongoing maintenance and review of the data dictionary
- Comply with regulations and standards
- Ensure accuracy of data collection and reporting
- Establish change management policies and procedures
- Develop active and ongoing user education and training

Map the data across all systems. Exchanging information among systems within an organization and with outside organizations is vital to conducting the business of medicine. Data exchange has gained new importance in light of the meaningful use program.

Organizations should ensure all data uses are consistent by mapping each element across each system and facility and resolving any discrepancies.† An organization should not assume that other organizations do what it does with a data element of the same name.

Organizations should also identify what data are required for HIE participation and any local requirements for coding and reimbursement. Stages 1 and 2 of the meaningful use program define what data will be shared (data capture and data sharing).

Many of the meaningful use requirements are built around production of aggregate data from multiple systems. This makes it imperative that organizations ensure consistency of data across systems.

The key to achieving meaningful use success is effective data management and mapping; understanding and effective implementation of vocabulary standards; and alignment with terminologies and classifications.[4]

Develop a data quality management process that includes ongoing maintenance and review of the data dictionary.[5] To ensure data consistency and accuracy across an organization, the process must be under the direction of an enterprise data quality steering committee. It should include four key components:†

- the purpose for which the data are collected;
- the processes by which data are collected and changes tracked;
- the processes and systems used to archive data and data journals; and
- the process of translating data into information utilized for an application.

Examples of data quality management activities include:

- Frequently reviewing and validating data dictionary content by checking the data quality of clinician entries to ensure proper application and use of data
- Reviewing documentation for errors based on poor techniques such as pulling information forward in the EHR through copy and paste that was not verified or validated by the clinician
- Ensuring overall record integrity among enterprise systems as well as across organizations through periodic review and audit of actual practices

Comply with regulations and standards. Standards are critical because they are the basis for data exchange and interoperability. To ensure compliance, it is essential that all data collected be compared against current state and federal regulations and accreditation agencies (e.g., the Joint Commission) when developing new data fields or performing routine updates.

Ensure accuracy of data collection and reporting. Data reports must be validated to ensure the accuracy of the information produced.† Examples of questions to ask when reviewing data include: Are outliers based on accurate data or are they the result of end user error? Are errors related to a single end user or are they systemic?

If the data reveal apparent inaccuracies, the organization may need to review its data collection process to ensure it is correct and being followed correctly.

Establish change management policies and procedures. Organizations should develop a formal change management process through which all changes to data dictionaries are coordinated. Change management policies and procedures will help organizations prevent disruption of other systems that interact with that particular application.†

Implementing a process for changes, modifications, or deletions to the data dictionary will also ensure consistency in interpretation and version control, if multiple iterations exist or in the event of staff turnover.

Develop active and ongoing user education and training. Organizations must institute an active and ongoing education and training program for all staff who collect, use, analyze, or interact with data on any level. Ongoing education is critical to maintaining a high-quality data dictionary.† Staff turnover and changing data requirements and demands necessitate continuous training.

Organizations should ensure:

- Staff receives education and training so that data capture is consistent across the organization.
- Staff understands the ability to change a data dictionary must be coordinated through the proper change management procedures. For example, the HIM department would be able to change a data element format only in coordination with IT security provisions.
- Funding for maintenance and oversight of the data dictionary and data quality management processes.
- Each employee takes ownership of data integrity and understands how his or her actions affect it.

Whether a data field is being added to an interface or a patient is being registered for the first time, all staff should have a clear understanding of the data definitions and values and the implications of inaccurate data entry. The references at the end of this practice brief can be used to support ongoing education and training efforts.

HIM's Role and Responsibility

As information managers, HIM professionals play an important role in ensuring data integrity. They promote the importance of data quality, and they understand the healthcare record's many functions and the data quality management model's characteristics.

Regardless of the work setting, HIM professionals must be actively involved in software selection and management processes. They must take an active role in defining attributes of prospective applications, as well as maintaining a data integrity program. Many HIM professionals are also responsible for data management and report writing.

When additional services are added to the facility or a field is proposed in the EHR, it is critical to involve an HIM professional responsible for maintaining the data dictionary or have the decisions approved by the data quality steering committee to ensure the impacts are clearly understood.†

In many organizations, the process may be referred to as data administration. Data administration may be defined as the "analysis, classification and maintenance of an organization's data and data relationships. It includes the development of data models and data dictionaries…"[6]

Many organizations have identified the data (or resource) administrator as an IT role. However, this role is a natural progression for an HIM professional working for or with IT to define, manage, and coordinate data dictionaries. The data administrator role is typically included as a part of the overarching data governance program.

Responsibilities of data administrators related to maintaining data dictionaries include:

- Identifying and promoting clear and valid definitions for enterprise data
- Identifying and further defining required validation rules to be applied when capturing enterprise data
- Assessing and resolving data integrity issues (quality, timeliness, accuracy, completeness) and cost-effectiveness issues
- Leading training and educational activities of end users to promote best practices in data collection and use

The foundation established above supports a data administrator's larger responsibilities, which include:

- Ensuring that data structures meet the needs of the various users of the data and implementing established data management practices across the enterprise
- Ensuring sound business decisions when implementing new applications that will access and manage enterprise data and ensuring that proper data management practices are not violated
- Addressing data access and security issues while facilitating the sharing and use of the data across the enterprise

Notes

1. International Organization for Standardization. "Information Technology Parts 1–6 (2nd Edition)." 2004. www.iso.org.
2. Department of Education, Student Aid. "Enterprise Data Dictionary Standards." April 2007. http://federalstudentaid.ed.gov/static/gw/docs/ciolibrary/ECONOPS_Docs/EDM-EnterpriseDataDictionaryStandards.pdf.
3. AHIMA. "Data Mapping Best Practices." *Journal of AHIMA* 82, no. 4 (Apr. 2011): 46–52.
4. Ulmer, Stephen E., and Jan C. Fuller. "Understanding the Meaningful Use Vocabulary Standards." *Journal of AHIMA* 81, no. 11 (Nov.–Dec. 2010): 48–49.
5. AHIMA. "Data Quality Management Model." June 1998. Available in the AHIMA Body of Knowledge at www.ahima.org.
6. Brunson, Duffie. "Data Quality and Data Governance: The Basics." February 15, 2005. www.b-eye-network.com/view/630.

Resources

All resources available in the AHIMA Body of Knowledge at www.ahima.org.

AHIMA. "Data Quality Attributes Grid" in "HIM Principles in Health Information Exchange" (online version). September 2007.

AHIMA. "Health Data Analysis Toolkit." 2011.

AHIMA e-HIM Work Group on EHR Data Content. "Guidelines for Developing a Data Dictionary." *Journal of AHIMA* 77, no. 2 (Feb. 2006).

AHIMA. "Leadership Model: Data Content Standards."

AHIMA. "Accountable Care: Implications for Managing Health Information." 2011.

Birnbaum, Cassi. "One-stop Shop: An HIM Department's Journey to Centralize Core Data Services." *Journal of AHIMA* 78, no. 8 (Sept. 2007).

Clark, Jill. "Tools for Data Analysis: New Toolkit Provides Resources for Health Data Analysts." *Journal of AHIMA* 82, no. 2 (Feb. 2011).

Prepared by

Jill Clark, MBA, RHIA
Barbara Demster
C. Jeanne Solberg, MA, RHIA

Acknowledgments

Cecilia Backman, MBA, RHIA, CPHQ
Jan Barsohpy, RHIT
Joan Croft, RHIT
Linda Darvill, RHIT
Angela K. Dinh, MHA, RHIA, CHPS
Patience Hoag, RHIT, CCS, CCS-P, CHCA, CPHQ
Crystal K. Kallem, RHIA, CPHQ
Priscilla Komara, MBA, RHIA
Jennifer McCollum, RHIA, CCS
Monna Nabers, MBA, RHIA
Sandra Nunn, MA, RHIA, CHP
Cathy Price, RHIT
Laura J. Rizzo, MHA, RHIA
Allison F. Viola, MBA, RHIA
Diana Warner, MS, RHIA, CHPS, FAHIMA
Lou Ann Wiedemann, MS, RHIA, FAHIMA, CPEHR

The information contained in this practice brief reflects the consensus opinion of the professionals who developed it. It has not been validated through scientific research.

† Indicates an AHIMA best practice. Best practices are available in the AHIMA Compendium, http://compendium.ahima.org.

Article citation:
AHIMA. "Managing a Data Dictionary - Retired" *Journal of AHIMA* 83, no.1 (January 2012): 48-52.

Driving the Power of Knowledge