



# Managing a Data Dictionary (2016 update)

*This update supplants the 2012 practice brief “[Managing a Data Dictionary](#).”*

As healthcare organizations move to the electronic environment, the large volume and complexity of data collected is growing at unprecedented rates. While access to data has the potential to enhance decision-making and benefit the organizations and their patients, this can only be achieved if the organization understands the data, follows industry standards for formatting that data, and maintains it in a reliable manner. Challenges to the reliability of data originate within organizations that may store information in multiple systems and in mergers of healthcare organizations that need to combine and consolidate their data and systems and in the growth of health information exchange (HIE)—which demands interoperability.

The growth in available data creates opportunities for more complex data analysis and application of big data tools and principles. However, the lack of data consistency can create challenges for data comparison and reporting, which can ultimately lead to errors in data use.

Accurate and reliable data are integral to many health IT initiatives currently under way. (For a discussion of the characteristics of data quality, see the [AHIMA Data Quality Management Model](#).) A data dictionary is one tool organizations can use to help ensure data accuracy. According to the [International Organization for Standardization \(ISO\)](#):<sup>1</sup>

The increased use of data processing and electronic data interchange heavily relies on accurate, reliable, controllable, and verifiable data. 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.

This Practice Brief defines the data dictionary, describes common data inconsistencies found within healthcare organizations' systems, and discusses associated data management challenges. It also outlines best practices for maintaining data integrity, including the health information management (HIM) professional's role.

## Data Dictionary Defined

A data dictionary is a crucial component in a relational database because it 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, which provides information about other data.

The metadata may include descriptive information including keywords or author information, structural information such as the sequence of data objects, or administrative information including ownership or date collected. For most relational database management systems (RDBMS), the database management system software needs the data dictionary to access the data within a database. In non-relational databases, database structures can be less consistent as data is not stored in tables with defined data types and relationships. Non-relational database models vary in structure but may include data arrays or graph forms. Fields in individual records can vary as the analyst

creates each record from the data pool on demand.

The goal is to achieve consistently defined and standardized data. Health data analysis depends on the consistency and reliability of the data and the analyst's ability to understand the attributes of the 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](#)" includes an example of a [data dictionary](#).<sup>2</sup>

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.

A data dictionary can be referenced to understand a data element's meaning and origin. 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 is managed and validated for accuracy in reporting.

For details on how to develop a data dictionary, refer to the AHIMA Practice Brief "[Guidelines for Developing a Data Dictionary](#)."

### **Additional Elements of a Data Dictionary**

A data dictionary could include interoperability rules on when and how data is shared (i.e., making it available to the portal, making it available to auto-fax to the primary care provider, or adding a "confidential document" type—to hold from sharing based on security).

### **Common Data Inconsistencies**

Data inconsistency occurs when there is conflicting value in the same data element. In many organizations data is stored in different databases or in different formats, resulting in data inconsistency. This practice creates unreliable information and compromises the overall data quality. Inconsistencies in data definitions can lead to inaccurate data use and health data reporting and can potentially affect the quality of care.

Some common 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.** Name conventions or identifiers occur differently in different systems. For example, the date of the patient's admission can be referred to as "Date of Admission" in the patient management module within the electronic health record (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 name conventions or identifiers are the same, but definitions are different. For example, 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 data element values.** Same data elements are displayed in different values. For example, 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.

## Why Data Standards Matter

Data standards are an integral component to an enterprise's data dictionaries. As a part of an overall information governance structure, enterprises should align the entries in their data dictionaries with current data standards to ensure they are in compliance with regulatory, legal, risk, environmental, and operational requirements.<sup>3</sup>

Data standards play an important role in patient safety. In the article entitled “HIM Functions in Healthcare Quality and Patient Safety,” the authors stated, “Patient safety and compliance issues represent a major factor in data integration. Through the use of clinical decision support and electronic documentation, healthcare-associated infections, falls, and other negative healthcare-associated events can be more quickly identified, tracked, monitored, and eliminated.”<sup>4</sup>

Data standards play an equally important role in interoperability. Data standards are necessary for interoperability as words have different meanings in different systems. The [American National Standards Institute \(ANSI\)](#) governs standards development organizations in the United States. ANSI's use of a consensus process ensures all interested parties associated with particular standards are involved with the development of those standards. The [federal government](#) has established a set of standards to support HIE for what it defines as “meaningful use” within the requirements of the “meaningful use” EHR Incentive Program.

As HIE continues to increase, healthcare organizations will need to properly identify their data elements for appropriate transmission and apply the proper [Information Governance principles](#) to ensure that the standards they adhere to properly address the organization's needs for that data. HIM professionals must maintain responsibility for monitoring the quality of documentation while working collaboratively with other members of the healthcare team to maintain the clinical accuracy and completeness of the data. These efforts are key to identifying system and process problems within the realm of patient safety and quality of care.<sup>5</sup>

### 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
- | Ability to reuse data
- | Consistency in data use
- | Easier data analysis
- | Improved decision making based on better data
- | Simpler programming
- | Enforcement of standards<sup>6</sup>

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.

## 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 AHIMA Practice Brief "[HIM Principles in Health Information Exchange](#)" provides a guide for defining data and their attributes.<sup>7</sup>

When possible, organizations should design the data collection system well in advance of any new technological 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."<sup>8</sup>

Organizations should consider transitioning to a core data service model where the key or common data elements 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 is currently collected, why it is collected, and how it is

used. They should research what impact a change to the system would have on the data. Information governance practices for change control should be implemented. These change control practices will minimize impact to data integrity.<sup>9</sup>

### **Steps for Maintaining Data Integrity**

Managing an organization's data and those who enter it is an ongoing challenge requiring active administration and oversight. Change management is key, and access controls are also an additional consideration. The following best practices help organizations maintain their data dictionaries and data integrity as a part of an overall data governance initiative:

- | Know the data and its source
- | Map the data across all systems
- | Develop a data quality management process that includes ongoing maintenance and review of the data dictionary
- | Comply with all regulations and industry 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 use data elements of the same name in the same manner.

Organizations should also identify what data is required for HIE participation and any local requirements for coding and reimbursement. Stages 1 and 2 of the [meaningful use](#) program defined what data is 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.<sup>10</sup>

**Develop a data quality management process** that includes ongoing maintenance and review of the data dictionary.<sup>11</sup> To ensure data consistency and accuracy across an organization, the process should be under the direction of an enterprise data quality steering committee which should include representatives from IT, HIM, and risk/legal departments. It should include five key components:

1. The purpose for which the data is collected
2. The processes by which data are collected and changes tracked
3. The processes and systems used to archive data and data journals

4. The process of translating data into information utilized for an application
5. The process for determining who can change data during its lifecycle

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
- | Follow data elements throughout their lifecycle to ensure inappropriate changes are not made
- | Assess for incorrect data mapping

**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 (i.e., [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 should institute an active and ongoing education and training program for all staff that are responsible to 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, which should be documented and the records maintained as required by regulatory guidance or internal policies.

**Incorporate Data Integrity in Data Governance Initiatives.** Data integrity is a key component in ensuring effective governance of data. According to Nunn (2009), "[Data governance](#) is the set of policies and procedures that determine the who, how, and why of data management within the organization. Strong data governance supports compliance and legal efforts by organizing data for retrieval and retention, especially over the long term."<sup>12</sup> Lack of data integrity impedes workforce performance and creates medical and legal risks when necessary data or records cannot be located or if they are inaccurate or incomplete. Information governance initiatives should elaborate on organizational change control practices. Standard practice for change should be implemented. These change control practices will minimize impact to data integrity.

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.
- | Data integrity must be incorporated into the overall data governance initiatives for the organization. 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**

In healthcare organizations, numerous stakeholders and business units are involved in the collection, maintenance, and use of data. The HIM professional's role is unique in that they promote the importance of data quality for patient safety and quality improvement, and they understand the healthcare record's many functions and the data quality management model's characteristics. As information managers, HIM professionals must work with all stakeholders to establish the data dictionary and protect the integrity of the organization's data.

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. HIM professionals bring to the data management discussion knowledge of interoperability standards for health information exchange, data requirements for registries and public reporting, and quality management and business reporting needs.

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..."<sup>13</sup>

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
- | Monitor targeted data elements to proactively identify data integrity issues

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

- | Ensuring data structures meet the needs of the various data users 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 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](http://www.iso.org).

[2] AHIMA. "[Health Data Analysis Toolkit](#)." 2014.

[3] Cohasset Associates and AHIMA. 2014 Information Governance in Healthcare Whitepaper. [www.ahima.org/~media/AHIMA/Files/HIM-Trends/IG\\_Benchmarking.ashx](http://www.ahima.org/~media/AHIMA/Files/HIM-Trends/IG_Benchmarking.ashx).

[4] AHIMA. "[HIM Functions in Healthcare Quality and Patient Safety](#)." *Journal of AHIMA* 82, no.8 (August 2011): 42-45.

[5] Department of Education, Student Aid. "[Enterprise Data Dictionary Standards](#)." April 2007.

[6] 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.

[7] AHIMA e-HIM Work Group on HIM in Health Information Exchange. "[HIM Principles in Health Information Exchange](#)." *Journal of AHIMA* 78, no.8 (September 2007): [expanded online version].

[8] AHIMA. [Information Governance Principles for Healthcare](#). AHIMA thanks ARMA International for use of the following in adapting and creating materials for healthcare industry use in IG adoption: Generally Accepted Recordkeeping Principles® and the Information Governance Maturity Model. [www.arma.org/principles](http://www.arma.org/principles). ARMA International 2013.

[9] "Data Mapping Best Practices." *Journal of AHIMA* 82, no. 4 (Apr. 2011): 46–52.

[10] Ulmer, Stephen E. and Jan C. Fuller. "Understanding the Meaningful Use Vocabulary Standards." *Journal of AHIMA* 81, no. 11 (Nov.–Dec. 2010): 48–49.

[11] Davoudi, Sion et al. "Data Quality Management Model (2015 Update)." *Journal of AHIMA* 86, no.10 (October 2015): expanded web version.

[12] Nunn, Sandra L.. "[Driving Compliance through Data Governance](#)." *Journal of AHIMA* 80, no.3 (March 2009): 50-51.

[13] Brunson, Duffie. "[Data Quality and Data Governance: The Basics](#)." February 15, 2005.



# Resources

AHIMA. "[Accountable Care: Implications for Managing Health Information.](#)" 2011.

AHIMA. "Data Quality Management Model." 2015.

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

AHIMA. "[Leadership Model: Data Content Standards.](#)"

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

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

Centers for Medicare and Medicaid Services. "[Principles for Accelerating Health Information Exchange.](#)"

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

## Prepared by (2016 update)

Sion Davoudi  
Jill Flanigan, RHIT  
Shannon Houser, PhD, MPH, RHIA, FAHIMA  
Lesley Kadlec, MA, RHIA, CHDA  
Annessa Kirby  
Daniel VanSlyke, RHIA, CHDA  
Annemarie Wendicke, CHDA, MPH

## Acknowledgments (2016 update)

Maria Barbetta, RHIA  
Aurae Beidler  
Suzanne Goodell, MBA, RHIA  
Maribeth Hernan, MA, RHIA, CHP  
Linda Howard, CCS  
Beth Liette, MS, RHIA  
Jennifer McCollum, RHIA, CCS  
Lori McNeil Tolley, M.Ed., RHIA  
Laurie Miller, RHIT, CCS-P  
Sharon Slivochka, RHIA  
Nicole Van Andel  
Amanda Wickard, RHIA, MBA  
Holly Woemmel, MA, RHIA, CHPS

## **Original Authors**

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

## **Original 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

---

Article citation:

AHIMA Practice Brief. "Managing a Data Dictionary (2016 update)" (Updated December 2016)