Metadata Offers Roadmap to Structured Data

Save to myBoK

By Diane Dolezel, MSCS, RHIA

Database metadata provides a valuable roadmap for locating the structured data stored in healthcare databases. However, many electronic health information management (e-HIM) professionals do not possess the necessary basic understandings of where to find the metadata and how to use it after it is located. This article discusses how to find the metadata, describes the contents of typical systems and user metadata repositories, and explores the relationship of metadata to data quality in the context of e-HIM development.

A single facility often maintains multiple database repositories to handle the flood of data from clinical applications, mobile monitoring devices, user interfaces, and electronic health records (EHRs). This proliferation of stored data makes it difficult for healthcare professionals to know where to find the data they need. Additionally, e-HIM developers are challenged to create high-quality interfaces referencing the correct data field definitions. The National Information Standards Organization (NISO) endorses the use of metadata to locate database data, and to support interoperability.[1] Moreover, the Office of the National Coordinator for Health IT (ONC) recommends using metadata to increase data integrity for health data exchange.[2]

The process of understanding the metadata roadmap unfolds with these five steps:

1. Find the metadata catalogs
2. Explore system metadata
3. Examine user metadata
4. Survey organization-wide metadata

Find the Metadata Catalogs

Imagine you are an e-HIM project manager for a data integration project for a large hospital. You must work with the project developers to design a database interface that will accept input from a web-based form. To maintain data quality, the user's form data must be validated against the database definitions before the user is allowed to submit the form. You know that there are metadata system catalogs that describe each row and column (i.e., field) in the database tables. However, you do not know where to find them or how to obtain access.

Explore System Metadata

First, consider metadata creation. When the database administrator installs the database software on the server, the installation process creates the database management system (DBMS) metadata catalogs automatically. These system catalogs, which are actually a set of tables stored in the database, are updated automatically by the DBMS when changes occur, like creating a new table or user.[6] However, metadata catalogs are accessible only to database administrators and super users.

To get access for the integration project, the project manager or project lead should meet with a database administrator (DBA) and explain what data they needed access to, and whether the developers will need to view the data or make changes to the data (i.e., update or insert new data). The DBA will then query the system catalog to locate the information and will provide appropriate access as well as documentation on the table definitions.

A copy of the metadata definitions for the project objects, such as tables or procedures, should be distributed to all developers. It should be referenced designing the user's web form validation. This will help guarantee that interfaces are sending correctly validated data to the database.

For example, if the gender field is defined as "M" or "F", then these codes must be used to validate the web form so that a form input value of a "1" or "0" would be not be submitted. Moreover, the user would get an error message with examples of

the correct data format for that input field.

Typical information in a DBMS systems catalog is:[7]

- User names, roles, privileges
- Names of all tables, columns, indexes, constraints
- Names of user created procedures or indexes
- Audit trail of updates, edits, deletions

Oracle and IBM DB2 are well-established corporate databases. For a visual example of the IBM database, visit the developerWorks website at www.ibm.com/developerworks/data/library/techarticle/dm-0411melnyk/.[8] The Oracle system catalog interface has a similar design.

Examine User Metadata and Survey Organization-wide Metadata

As developers work on the project they will create test tables and load them with test data in the development database in order to test the forms. Those data definitions will appear in the users' data catalog, and in the systems catalog. For example, if the developer creates a test table "Patient," in an Oracle test database, then the Oracle DBMS catalog tables ALL_TABLES, and ALL_TAB_COLUMNS would be updated.[9] This is similar to what happens when a person creates a new file on their computer and the directory listing is updated.

To ensure that the developer's code works in the production systems, it is essential that the development database tables have the same column definitions, primary keys, indexes, etc., as the production tables. Specifically, the development area should match the production area as closely as possible.

This can be achieved efficiently by having the developers, or system designers, refer to their copy of the system metadata definitions when creating objects in their development database. Failing to take this action substantially increases the likelihood of problems when the web screen goes into production. A portion of the ALL_TAB_COLUMNS catalog table that lists all tables or views accessible to a user is available for review above.[10] In many settings, the user will have permission to view this table.

Lastly, an e-HIM developer or designer may be working on a small project integral to health information management (HIM), like tracking HIM employee credentials that will not be implemented in the company's production area. In this case, there is a relational database called Microsoft Access that is commonly used for this type of small-scale development. It has many benefits, such as a system catalog. It includes a rich tool set for querying and creating objects, and the data stored in the system is easily exportable for upload to Oracle or DB2, should that be desired later.

Table 2 [...] shows a view of the system catalog for the table Patient in a Microsoft Access database, created on the computer desktop of the author of this article. The field names are listed (i.e., ptmrn, ptfname), and the data types are still the default type of text. The primary key, which is ptmrn, is of data type autonumber, indicating that a new incremented number will be associated with a record when it is added to the database.

At the bottom of the page under the General tab, one can see information about the ptmrn column, which is the field name selected at the top. For example, the ptmrn column is a long integer, with an index, and no duplicates are allowed.

Table 1: ALL_TAB_COLUMNS Catalog Table Example

This table provides a list of all the tables, columns, and other database objects that are available to the current user.

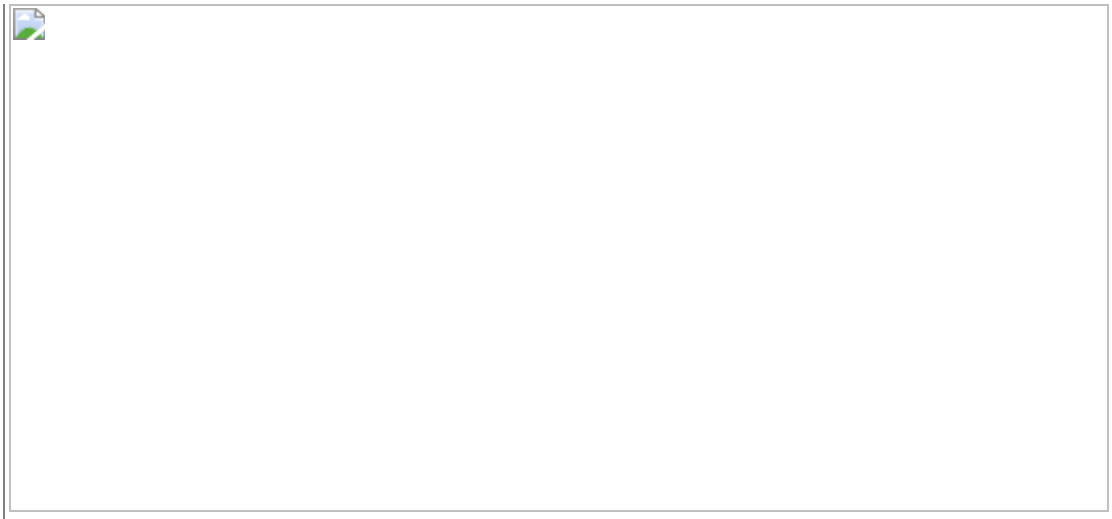
Column	Datatype	NULL ALLOWED	Description

OWNER	VARCHAR2(128)	No	User name of owner of the table, view, or cluster
TABLE_NAME	VARCHAR2(128)	No	Table, view, or cluster name
COLUMN_NAME	VARCHAR2(128)	No	Column name
DATA_TYPE	VARCHAR2(30)	Yes	Datatype of the column
DATA_LENGTH	NUMBER(10)	Yes	Length of the column in bytes
DATA_PRECISION	NUMBER(10)	Yes	Decimal precision for NUMERIC and DECIMAL datatype; binary precision for FLOAT, REAL, and DOUBLE datatype; NULL for all other datatypes
DATA_SCALE	NUMBER(10)	Yes	Digits to the right of decimal point in a NUMERIC or DECIMAL
NULLABLE	VARCHAR2(1)	Yes	Indicates if the column allows NULLs. Value is N, if there is a NOT NULL constraint on the column or if the column is part of a primary key

Source: Oracle. “A.1.11 ALL_TAB_COLUMNS.” https://docs.oracle.com/html/B10100_01/wncat.htm#i631886.

Table 2: Example of System Catalog for Table Patient in Microsoft Access Database

This table is created automatically by the database management system when the user creates a new table.



Learn to Navigate Metadata

Metadata is useful for finding structured data in a database. HIM professionals should learn how to navigate the metadata, or system catalogs, because they provide information about the data in the database that is needed for designing and developing electronic interfaces. Metadata also helps maintain data quality in a database.

Notes

[1] National Information Standards Organization. *Understanding Metadata*. Bethesda, MD: NISO Press, 2004. www.niso.org/publications/press/UnderstandingMetadata.pdf.

[2] Office of the National Coordinator for Health Information Technology. "Metadata Standards to Support Nationwide Electronic Health Information Exchange." *Federal Register* 76, no. 153 (Aug. 9, 2011): 48769–76. www.gpo.gov/fdsys/pkg/FR-2011-08-09/pdf/2011-20219.pdf.

[3] Ibid.

[4] Viola, Allison and Shefali Mookencherry. "Metadata and Meaningful Use." *Journal of AHIMA* 83, no. 2 (February 2012): 32-38.

[5] Milstead, J. and S. Feldman. "Metadata cataloging by any other name." *Online* 23, no. 1 (1999): 24.

[6] Ibid.

[7] Oracle. "Oracle9i Lite Developer's Guide for Windows CE/Pocket PC, Release 5.0.2, System Catalog Views." 2002. http://docs.oracle.com/html/B10100_01/wncat.htm.

[8] Melnyk, Roman. "DB2 Basics: How to get useful information from the DB2 UDB system catalog." *developerWorks*. November 2, 2004. www.ibm.com/developerworks/data/library/techarticle/dm-0411melnyk/.

[9] Oracle. "Oracle9i Lite Developer's Guide for Windows CE/Pocket PC Release 5.0.2, System Catalog Views, A.1.11 ALL_TAB_COLUMNS." 2002. https://docs.oracle.com/html/B10100_01/wncat.htm#i631886.

[10] Ibid.

Diane Dolezel (dd30@txstate.edu) is an assistant professor in the health information management department at Texas State University, based in San Marcos, TX.

Article citation:
Dolezel, Diane. "Metadata Offers Roadmap to Structured Data" *Journal of AHIMA* 86, no.2 (February 2015): 44-46.