

LLNL-JRNL-400269



LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

METADATA REGISTRY, ISO/IEC 11179

R. K. Pon, D. J. Buttler

January 7, 2008

Encyclopedia of Database Systems

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

METADATA REGISTRY, ISO/IEC 11179

Raymond K. Pon

UC Los Angeles, <http://www.cs.ucla.edu/~rpon>

David J. Buttler

Lawrence Livermore National Laboratory, <http://people.llnl.gov/buttler1>

This work (LLNL-JRNL-400269) was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

SYNONYMS

Metadata Repository, MDR

DEFINITION

ISO/IEC-11179 [1] is an international standard that documents the standardization and registration of metadata to make data understandable and shareable. This standardization and registration allows for easier locating, retrieving, and transmitting data from disparate databases. The standard defines the how metadata are conceptually modeled and how they are shared among parties, but does not define how data is physically represented as bits and bytes. The standard consists of six parts. Part 1 [2] provides a high-level overview of the standard and defines the basic element of a metadata registry – a data element. Part 2 [3] defines the procedures for registering classification schemes and classifying administered items in a metadata registry (MDR). Part 3 [4] specifies the structure of an MDR. Part 4 [5] specifies requirements and recommendations for constructing definitions for data and metadata. Part 5 [6] defines how administered items are named and identified. Part 6 [7] defines how administered items are registered and assigned an identifier.

HISTORICAL BACKGROUND

The first edition of the standard was published by the Technical Committee ISO/IEC JTC1, Information Technology Subcommittee 32, Data Management and Interchange, starting in 1994 and finally completing in 2000. The second edition was started in 2004 and was finally completed in 2005. The second edition cancels and replaces the first edition of the standard.

SCIENTIFIC FUNDAMENTALS

Metadata is data that describes other data. A metadata registry is a database of metadata. The database allows for the registration of metadata, which enables the identification, provenance tracking, and quality monitoring of metadata. Identification is accomplished by assigning a unique identifier to each object registered in the registry. Provenance details the source of the metadata and the object described. Monitoring quality ensures that the metadata accomplishes its designed task. An MDR also manages the semantics of data, so that data can be re-used and interchanged. An MDR is organized so that application designers can determine whether a suitable object described in the MDR already exists so that it may be reused instead of developing a new object.

Part 1: Framework

Part 1 introduces the building blocks of the MDR standard: data elements, value domains, data element concepts, conceptual domains, and classification schemes. An MDR is organized as a collection of concepts, which are mental constructs created by a unique combination of characteristics. A concept system is a set of concepts with relations among them. One such concept system that classifies objects is a classification scheme. A classification scheme is organized with some specified structure and is designed for assigning objects to concepts defined within it.

The basic construct in a metadata registry is the data element. A data element consists of a data element concept and a representation. A data element concept (DEC) is a concept that can be represented as a data element described independently of any particular representation. The representation of a data element consists of a value domain, a data-type, units of measure, and a representation class. A data element concept may consist of an object class, which is a set of abstractions in the real world that can be identified with explicit boundaries, and a property, which is a characteristic common to all members of an object class. A value domain is a set of permissible values. Each value domain is a member of the extension of a concept known as the conceptual domain. A conceptual domain is a set of value meanings, which are the associated meanings to values.

An MDR contains metadata describing data constructs. Registering a metadata item makes it a registry item. If the registry item is subject to administration, it is called an administered item. An ISO/IEC 11179 MDR consists of two levels: the conceptual level and the representational level. The conceptual level contains the classes for the data element concept and conceptual domain. The representational level contains classes for data element and value domain.

Part 2: Classification

Part 2 provides a conceptual model for managing concept systems used as classification schemes. Associating an object with a concept from a classification scheme provides additional understanding of the object, comparative information across similar objects, an understanding of an object within the context of a subject matter field, and the ability to identify differences of meaning between similar objects.

Classification schemes are registered in an MDR by recording their attributes, such as those regarding its designation, definition, classification scheme, administration record, reference document, submission, stewardship, registration authority, and registrar.

Part 2 also defines the mechanism for classifying an administered item, which is the assignment of a concept to an object. Objects can also be linked together by relationships linking concepts in the concept system.

Part 3: Registry metamodel and basic attributes

Part 3 describes the basic attributes that are required to describe metadata items and the structure for a metadata registry. The standard uses a metamodel to describe the structure of an MDR. A metamodel is a model that describes other models. The registry metamodel is specified as a

conceptual data model, which describes how relevant information is structured in the real world, and is expressed in the Unified Modeling Language [8].

The registry model is divided into six regions:

- **The administration and identification region:** supports the administrative aspects of administered items in the MDR. This region manages the identification and registration of items submitted to the registry, organizations that have submitted and/or are responsible for items in the registry, supporting documentation, and relationships among administered items. An administered item can be a classification scheme, a conceptual domain, context for an administered item, a data element, a data element concept, an object class, a property, a representation class, and a value domain. An administered item is associated with an administration record, which records administrative information about the administered item in the registry.
- **The naming and definition region:** manages the names and definitions of administered and the contexts for names. Each administered item is named and defined within one or more contexts. A context defines the scope within which the data has meaning, such as a business domain, a subject area, an information system, a data model, or standards document.
- **The classification region:** manages the registration and administration of classification schemes and their constituent classification scheme items. It is also used to classify administered items.
- **The data element concepts region:** maintains information on the concepts upon which the data elements are developed, primarily focusing on semantics.
- **The conceptual and value domain region:** administrates the conceptual domains and value domains.
- **The data element region:** administrates data elements, which provide the formal representations for some information (e.g., a fact, observation, etc.) about an object. Data elements are reusable and shareable representations of data element concepts.

Part 4: Formulation of data definitions

Part 4 specifies the requirements and recommendations for constructing data and metadata definitions. A data definition must be stated in the singular. It also must be a descriptive phrase, containing only commonly understood abbreviations, that state what the concept is (as opposed to what the concept is not). A data definition must also be expressed without embedding definitions of other data. The standard also recommends that a data definition should be concise, precise, and unambiguous when stating the essential meaning of the concept. Additionally, a data definition should be self-contained and be expressed without embedding rationale, functional usage, or procedural information, circular reasoning. Terminology and consistent logical structure for related definitions should also be used.

Part 5: Naming and identification principles

Part 5 defines the naming and identification of the data element concept, the conceptual domain, data element, and value domain. Each administered item has a unique data identifier within the

register of a Registration Authority (RA), which is the organization responsible for an MDR. The international registration data identifier (IRDI) uniquely identifies an administered item globally and consists of a registration authority identifier (RAI), data identifier (DI), and version identifier (VI).

Each administered item has at least one name within a registry of an RA. Each name for an administered item is specified within a context. A naming convention can be used for formulating names. A naming convention may address the scope of the naming convention and the authority that establishes the name. A naming convention may additionally address semantic, syntactic, lexical, and uniqueness rules. Semantic rules govern the existence of the source and content of the terms in a name. Syntactic rules govern the required term order. Lexical rules govern term lists, name length, character set, and language. Uniqueness rules determine whether or not names must be unique.

Part 6: Registration

Part 6 specifies how administered items are registered and assigned an IRDI. Metadata in the MDR is also associated with a registration status, which is a designation of the level of registration or quality of the administered item. There are two types of status categories: lifecycle and documentation. The lifecycle registration status categories address the development and progression of the metadata and the preferences of usage of the administered item. The documentation registration status categories are used when there is no more development in the quality of metadata or use of the administered item.

Each RA establishes its own procedures for the necessary activities of its MDR. Some activities include the submission, progression, harmonization, modification, retirement, and administration of administered items.

KEY APPLICATIONS

The standardization that ISO/IEC 11179 provides enables for the easy sharing of data. For example, many organizations exchange data between computer systems using data integration technologies. For example, in data warehousing schemes, completed transactions must be regularly transferred to separate data warehouses. Exchanges of data can be accomplished more easily if data is defined precisely so that automatic methods can be employed. By having a repository of metadata that describes data, application designers can reuse and share data between computer systems, making the sharing of data easier. ISO/IEC 11179 also simplifies data manipulation by software by enabling the manipulation of data based on characteristics described by the metadata in the registry. This also allows for the development of a data representation model for CASE tools and repositories [9].

There are several organizations that have developed MDRs that comply with ISO/IEC 11179, such as the Australian Institute of Health and Welfare [10], the US Department of Justice [11], the US Environmental Protection Agency [12], the Minnesota Department of Education [13], and the Minnesota Department of Revenue [14]. Currently, there is also an MDR available developed by Data Foundations [15].

CROSS REFERENCES

Metadata

RECOMMENDED READING

- [1] ISO/IEC JTC1 SC32 (2007). ISO/IEC 11179, Information Technology – Metadata registries (MDR), <http://metadata-standards.org/11179/index.html>.
- [2] ISO/IEC JTC1 SC32. (2004): Part 1: Framework. Information Technology – Metadata registries (MDR). 2nd Edition.
- [3] ISO/IEC JTC1 SC32. (2005): Part 2: Classification. Information Technology – Metadata registries (MDR). 2nd Edition.
- [4] ISO/IEC JTC1 SC32. (2003): Part 3: Registry metamodel and basic attributes. Information Technology – Metadata registries (MDR). 2nd Edition.
- [5] ISO/IEC JTC1 SC32. (2004): Part 4: Formulation of data definitions. Information Technology – Metadata registries (MDR). 2nd Edition.
- [6] ISO/IEC JTC1 SC32. (2005): Part 5: Naming and identification principles. Information Technology – Metadata registries (MDR). 2nd Edition.
- [7] ISO/IEC JTC1 SC32. (2005): Part 6: Registration. Information Technology – Metadata registries (MDR). 2nd Edition.
- [8] Object Management Group. (2007). Unified Modeling Language. <http://www.uml.org/>.
- [9] ISO/IEC JTC1 SC32. (1999): Part 1: Framework for the specification and standardization of data elements. Information Technology – Metadata registries (MDR). 1st Edition.
- [10] Australian Institute of Health and Welfare. (2007). Metadata Online Registry (METeOR). <http://meteor.aihw.gov.au/>.
- [11] US Department of Justice. (2007). Global Justice XML Data Model (GJXDM). <http://justicexml.gtri.gatech.edu/>.
- [12] US Environmental Protection Agency. (2007). Environmental Health Registry. <http://www.epa.gov/edr/>.
- [13] Minnesota Department of Education. (2007). Metadata Registry (K-12 Data). <http://education.state.mn.us/mde-dd>.
- [14] Minnesota Department of Revenue. (2007). Property Taxation (Real Estate Transactions), <http://proptax.mdor.state.mn.us/mdr>.
- [15] Data Foundations. (2007). Metadata Registry. http://www.datafoundations.com/solutions/data_registries.shtml

ADDITIONAL DEFINITIONS

Title: Metadata

Byline: Raymond K. Pon and David J. Buttler

Definition: Data that defines and describes other data.

Main Text: Metadata are data and data become metadata when they are used to describe other data. Because metadata are data, metadata can be stored, organized, and retrieved in a database.

Cross-References: Metadata registry, ISO/IEC 11179