

ISO/IEC JTC 1/SC 32 N 1261

Date: 2005-04-05

REPLACES: --

<p style="text-align: center;">ISO/IEC JTC 1/SC 32</p> <p style="text-align: center;">Data Management and Interchange</p> <p style="text-align: center;">Secretariat: United States of America (ANSI) Administered by Farance, Inc. on behalf of ANSI</p>

DOCUMENT TYPE	Other Document (Open)
TITLE	Data Semantics and Metadata Standards
SOURCE	Dan Gillman (WG 2)
PROJECT NUMBER	
STATUS	Extract from ISO Focus of February 2005
REFERENCES	
ACTION ID.	FYI
REQUESTED ACTION	
DUE DATE	
Number of Pages	3
LANGUAGE USED	English
DISTRIBUTION	P & L Members SC Chair WG Conveners and Secretaries

Douglas Mann, Secretariat, ISO/IEC JTC 1/SC 32

Farance, Inc *, 360 Pelissier Lake Road, Marquette, MI 49855-9678, United States of America

Telephone: +1 906-249-9275; E-mail: MannD@battelle.org

available from the JTC 1/SC 32 WebSite <http://staging.itc1sc32.org/>

*Farance, Inc. administers the ISO/IEC JTC 1/SC 32 Secretariat on behalf of ANSI

Data semantics and metadata standards



by Dan Gillman, information scientist, Bureau of Labor Statistics, USA

Almost every organization manages data. No matter whether the organization is public or private, for or not for profit, data are a major corporate resource and the problems of describing those data are similar. Descriptions are essential for locating, understanding and using data, whether the users are inside or outside the organization.

Metadata are the descriptions of data or other resources. Here, we use the term “description” in a general sense. For example, the name of

a study through which some data are collected does not describe the data, strictly speaking, although it is useful for locating the data of interest. On the other hand, the definition of a code used to represent some data is very descriptive. Both are metadata.

The meaning of data

Semantics is the study of meaning, so the term “semantics of data” refers to the meaning of the data. Without getting into a deep philosophical discussion, meanings are part of the information that data convey. Information is represented by (other) data, and this other data is metadata for the original data.

IT:
**Success
and
progress**

Traditionally, the metadata for databases and files are developed individually, without reference to similar data in other sources. Even when metadata exist, they are often incomplete or incompatible across systems. As a result, the data contained in these databases and files are poorly understood. In addition, the metadata often disappear after the data reach the end of the business life-cycle.

There are now techniques for representing the semantics of data independently of the system in which they are stored, the model used to organize them, the source from which they came and the time when they were created. These semantics and other descriptors are stored as metadata.

Managing metadata

The International Standard ISO/IEC 11179, *Metadata registries (MDR)* addresses the management of metadata needed to understand, locate and manipulate data. The purposes of the standard are to:

- establish standard descriptors for data;
- promote shared understanding of data across organizational elements and among organizations;
- facilitate interchange, re-use, harmonization and standardization of data and components of data over time, space and applications, both within and across organizations; and
- support management of the components of data.

ISO/IEC 11179 is a six-part standard as follows:

Part 1: Framework – presents an overview of the entire standard and a description of the basic concepts and principles.

Part 2: Classification – describes how to manage a classification scheme in a metadata registry.

Part 3: Registry metamodel and basic attributes – provides the basic framework, in the form of a conceptual model, for a metadata registry.

Part 4: Formulation of data definitions – establishes rules for forming quality definitions for data elements and their components.

Part 5: Naming and identification principles – explains how to document naming conventions for data elements and their components.

Part 6: Registration – specifies the roles and basic procedures for the registration process in an ISO/IEC 11179 metadata registry.

The goals of registration

A metadata registry that conforms to ISO/IEC 11179 is denoted MDR, this being a database of metadata that supports the functionality of registration. Registration accomplishes three main goals: identification, provenance and monitoring quality. Identification is accomplished by assigning a unique identifier (within the registry) to each object registered there. Provenance addresses the source of the metadata for the object described. Monitoring quality ensures that the metadata does the job it is designed to do.

About the author



Dan Gillman is an information scientist employed by the US Bureau of Labor Statistics in its Office of Survey Methods Research. His main research interests are metadata, metadata stand-

ards and automatic text classification. At the Bureau of Labor Statistics, he is responsible for instituting metadata management principles throughout the agency. Gillman has published and presented numerous papers on metadata and metadata standards in professional forums, he chairs the UNECE/Eurostat/OECD Work Group on Statistical Metadata, and he chairs the US ANSI-accredited metadata standards committee, INCITS/L8.

An MDR manages the semantics of data. Understanding data is fundamental to all aspects of its use so the underlying model for an MDR is designed to capture all the basic components of the semantics of data, independent of any application or area of subject matter.

“ISO/IEC 11179 defines the attributes which guarantee that both users and owners of data have a common understanding of the meaning and descriptive characteristics of that data.”

In ISO/IEC 11179 the basic container for data is called a data element. It may exist purely as an abstraction or in some application system. In either case the description of a data element is the same in ISO/IEC 11179. Data element descriptions have both semantic and representational components, and the semantics are further divided into contextual and symbolic types.

Contextual semantics are described by the data element concept (DEC). The DEC describes the set of objects for which data are collected and a characteristic of those objects being measured, for example the marital status (characteristic) of US adults (set of objects).

Symbolic semantics

Symbolic semantics are described by the conceptual domain (CD), which is a set of categories, not necessarily finite. Each category has a definition and these correspond to the classes of a partition of the set of objects determined by a characteristic of those objects. For instance, commonly recognized categories for the marital status of US adults are single, married, divorced and widowed. Each adult belongs to one category.

The representational component is about the permitted values that a data element may use, with each value corre-

sponding to one of the categories in the CD. The set of these permitted values is called a value domain (VD). In the case of marital status, for example, the values S, M, D and W correspond to the categories defined above.

The semantic and representational components are described through attributes contained in the conceptual model of a metadata registry as specified in Part 3 of ISO/IEC 11179. A metadata registry that conforms to ISO/IEC 11179 can describe a wide variety of data. The attributes and relationships associated with a particular object in an MDR give that object meaning, although the depth of this meaning is limited because it is impossible to describe an object fully.

In addition, since ISO/IEC 11179 is subject-matter area independent, an MDR does not contain specific subject-matter attributes that could add meaning to an object. For instance, a data element is further described by knowing the data collection procedure, if one exists. This information, however, is beyond the scope of ISO/IEC 11179.

A descriptive framework for a shared view

Interestingly, the attributes described in Part 3 are themselves data elements and they can be registered in an ISO/IEC 11179 metadata registry. There are two main consequences to this: first, the metadata registry can describe itself; but secondly, metadata layers or levels are not defined in ISO/IEC 11179, which is a general descriptive framework for any kind of data but does not address other data management needs that are beyond its scope and should be addressed elsewhere.

The increased use of data processing and electronic data interchange relies heavily on quality data. One of the prerequisites for this is that both users and owners of data have a common understanding of the meaning and descriptive characteristics of that data. ISO/IEC 11179 defines the attributes that guarantee this shared view. ■