ECCMA White Paper

Transparency

It is about the data - the quality of the data

Mr. Peter R. Benson, project leader for ISO 8000

Background:

We live in a data centric environment where the real world is increasingly viewed not first hand but through the data that represents it. Very few of us actually touch, or even see the tangible objects, or participate in the events we report on; instead we rely on the data and we contribute to its quality. Today, more than ever before, our access to data, the ability of our computer applications to use it and the ultimate accuracy of the data determines how we see and interact with the world we live and work in.

On the one hand, complexity and detail can be measured by the quantity of data; a good example of this is where an image comes into focus, or a web page is completed, as more and more data arrives. On the other, the accuracy with which we see the real world can be measured by the quality of the data.

Data is intrinsically simple and can be divided into data that identifies and describes things, master data, and data that describes events, transaction data.

Master data describes and identifies both tangible and intangible things, such as individuals, organizations, locations, tangible goods and intangible services but also processes, laws, rules and regulations. Typically we use identifiers to reference master data as in an airport code, a tax id, a passport number, a vehicle license number, a part number, a serial number and a credit card number; these are references to master data.

Transaction data describes an event such as the completion of a process, a credit card transaction, a purchase, a sale or a transfer. The ability to resolve the references to master data contained in transaction data is an important aspect of the quality of transaction data.

> **Transparency** requires that:
>
> (1) Transaction data accurately <u>identifies</u> who, what, where and when
>
> and
>
> (2) Master data accurately <u>describes</u> who, what and where

ECCMA White Paper

<u>Understanding data quality</u>

Data is defined as the "symbolic representation of something that depends, in part, on its metadata for its meaning," it follows therefore that the quality of the metadata must play an important part in determining data quality. Metadata gives data meaning. For example "50-02-01" is a meaningless string of characters but apply the metadata "Date of Birth" and it becomes meaningful data. To make it unambiguous we need to have a syntax such as CCYY-MM-DD and the associated value upgraded to 1950-02-01.

Good quality metadata comes from a metadata registry or a technical dictionary. This will contain a definition of the concept. For example, the concept: "Date of birth" has a concept definition of: "Year, month and day in which a person or an animal is born". Even better, an open technical dictionary will assign a language independent public domain concept identifier, as for example 0161-1#02-065175#1 in the Electronic Commerce Code Management Association (ECCMA) open technical dictionary (eOTD). This allows the data 0161-1#02-065175#1:1950-02-01 to be rendered as either Date of Birth: February 2, 1950 or Date de naissance: 2 Février 1950.

Using quality metadata from an open technical dictionary creates not only quality data in the sense that it is unambiguous, but it also creates portable data, data that can be easily moved from one application to another and preserved over time independently of software. Finally using pubic domain concept identifiers as metadata protects the intellectual property in the data.

<u>Implementing ISO 8000 the international standard for data quality</u>

ISO 8000 is concerned with the principles of data quality, the characteristics of data that determine its quality, and the processes to ensure data quality. The standard is in several parts that allow it to be implemented for a specific type of data as well as incrementally within the type.

ISO 8000-110:2008 is the foundation standard for master data quality. Master data that is compliant with the standard is portable data that is formatted according to a published syntax and where the metadata is explicit, either included with the data or by reference to an open technical dictionary.

Requesting or requiring that master data is provided in ISO 8000-110:2008 compliant format is not a burden to the data provider. The requirements of ISO 8000-110:2008 are simple; they require no specialized technology or the purchase of any product or service and are within the capability of all companies regardless of their size. [ISO 8000-110:2008 is available from the ANSI eStandards store](#).

ISO 8000-120:2009 is a supplement to ISO 8000-110:2008 that covers master data provenance. The standard is designed to assist in tracking the extraction of data elements through to their original source. Implementation of this standard requires knowledge of database management.

ISO 8000-130:2009 is a supplement to ISO 8000-120:2008 that covers master data accuracy. The standard is designed to assist in tracking claims to accuracy of data elements. Implementation of this standard requires knowledge of database management.

www.eccma.org

❧