NATO Codification System as the foundation for ISO 8000, the International Standard for data quality

Mr. Peter R. Benson

Executive Director and Chief Technical Officer of the Electronic Commerce Code Management Association (ECCMA)

In October 2004 NATO Allied Committee 135 (AC/135) and the Electronic Commerce Code Management Association (ECCMA) signed a memorandum of agreement to work together to promote the NATO Codification System as an international standard. One of the outcomes of this partnership is ISO 8000 the new international standard for data quality.

ISO 8000 is a key standard at the heart of the next generation of high speed and high relevance internet searches that rely on accurate and unambiguous descriptions. The first published part of the new standard is ISO 8000-110 Master data: Exchange of characteristic data: Syntax, semantic encoding, and conformance to data specification. This part of ISO 8000 describes the compliance requirements for the exchange of quality master data, data that describes individuals, organizations, locations, goods, services, rules and regulations.

While ISO 8000-110 describes the requirements for meeting data quality, ISO 22745 describes the tools needed and finally the ECCMA Open Technical Dictionary (eOTD), and the associated XML interchange formats, are the tools themselves that you use to build master data that meets ISO 8000 data quality standards. The eOTD is an industrial version of the Military NATO Codification System (NCS).

While you do you do not need to know the theory behind the eOTD or ISO 8000 to put it into practice, a basic understanding helps.

With in excess of 5,000 man years in its development, spanning over 50 years and over 50 countries, the NATO Codification System (NCS) represents a uniquely reliable system for creating the standardized descriptions that are critical in managing a large distributed inventory and an efficient supply chain.

From design, development, procurement and maintenance through to ultimate disposal; reference numbers in the form of drawing numbers, stock numbers and part numbers are critical components of the supply chain but they are only as good as the quality and accessibility of the underlying master data they reference. Without a solid and systematic method for creating complete and accurate master data records and the consistent descriptions that are derived from good master data, the reference numbers themselves can lead to the proliferation of duplicates that undermine part standardization, competitive purchasing and create bloated inventories that not only tie up substantial capital and are expensive to operate but ultimately slow down the supply chain with frequent stock outages and long lead times. Poor descriptions are a leading cause of failure to achieve readiness goals and to failure to mange obsolescence and diminishing manufacturing sources.

Good quality master data is the key to inventory visibility throughout the supply chain and the key to the new generation of open market electronic procurement systems for both goods and services. Finally good quality master data is also the key to accurate contracts for goods and services.

The fundamental principles of NATO codification System (NCS) are simple but it is the attention to detail and the discipline with which the principles are applied that are the hallmark of its success. These principles are at the core of the ECCMA Open Technical Dictionary and both ISO 22745 and ISO 8000.

In 1594, Shakespeare wrote in *Romeo and Juliet*: "What's in a name? That which we call a rose by any other name would smell as sweet" This is the first principle – everything starts with a name. In the eOTD this is actually a number, the number of a concept or the concept identifier of the class to be precise. This concept identifier will be associated with the NATO Approved Item Name and definition but additionally it may also be associated with other names and definitions taken from many other national or international standards or industry associations. The eOTD is a superset of the NCS effectively linking the NCS vocabulary to the vocabularies used by others. There is no need for anyone to "convert" to a single standard vocabulary but only that they map their vocabulary to the eOTD, this is the process managed by ECCMA.

The principle of unambiguous naming extends beyond item names to all "concepts". ISO 8000 requires that all data labels or metadata be explicit. No more cryptic column headers, field names or xml tags, every data label must be resolvable to a full definition. This is the purpose of the eOTD.

Using the eOTD as a metadata repository, it is easy to build ISO 8000 compliant data using the second principle; that all individuals, organizations, locations, goods and services can be described through a sequence of property value pairs where the property (the metadata) is defined in the eOTD.

Let's take something as simple as a person's name. It is a good example because locally "everyone knows" the convention for naming individuals but as soon as you leave the local area the convention becomes meaningless. The concept of a person's name is actually very complex, you have first names, given names, family names, last names, maiden names, married names not including middle names or initials, then there are prefixes and suffixes. Actually, I expect in the not to distant future, we will all be identified by a number that represents our DNA signature very much like GPS today is replacing complicated and changing physical addresses with a simple globally unique and unchanging number.

While we all take syntax for granted, or at least we expect our computers will deal with this in the background, the issue today is not so much with the data but the accurate labeling of the data and this is where the eOTD has an important role to play in providing a universal tool for open semantic encoding. This leaves us with "conformance to data specification".

ISO 8000-110 recognizes that a very important feature of data quality is its ability to meet the requirements of the end user. What was missing is the ability to express these data requirements, and here again the NATO Codification System held the answer. The first task in building the NATO codification system was to decide on what data was required to describe an item, the Item Identification Guide (IIG). Originally built by literally an army of engineers, the maintenance an evolution of these identification guides are at the heart of the NATO Codification System and this is also where the eOTD differs from the NCS.

The members of NATO and the users of the NCS have very specific data requirements. Because as a group the have agreed to share data, they have also agreed, in principle, to use the same Identification Guides or to collect the same data for all items that share the same item name. In practice however requirements for data differ not only from one organization to another but even within an organization from one business function to another. Just as physical inventory costs money to acquire and to maintain so too does data. It costs money to acquire data and it costs more money to maintain data; collecting data that you do not need is wasteful. Where ISO 8000-110 requires data to conform to data specifications, ISO 22745-30 defines how these data specifications are built and eOTD-i-xml is the XML format for building a data requirements specification using the eOTD.

The simplicity of eOTD-i-xml is by design and in recognition that companies need to build and maintain hundreds, if not a couple of thousand, of these data requirement specifications. A company will typically start with a single global data requirement specification such as a manufacture name, a part number and a 40 character description. By using these data requirement statements a company can measure the quality of their data and to improve it over time by increasing the breath and depth of their data requirement statements.

The eOTD-i-xml format is now being used in a number of codification tools and content building solutions. While recognizing the diversity of the requirements for data eOTD-i-xml represents a universal way to express these requirements in computer sensible form.

The final goal in data quality is the recognition and acknowledgement of the authoritative source and developing an ability to request data as required. This is the purpose of the eOTD-q-xml and eOTD-r-xml exchange formats. Once a company has a data requirement specification it is a simple matter of formatting it as a request (a query) for data validation or for missing data. Ideally the response will also be formatted in XML that references the eOTD or in eOTD-r-xml. As not everyone is familiar with XML there are simple spreadsheet equivalents for both the query and the response. However, the fundamental principle remains the same; you must know what data you need before you ask your supplier for data, asking for specific data for a specific item is much more likely to succeed than asking for "your data".

In the end the quality of our master data depends not so much on our ability to research and abstract data but rather to specify what data we need and to know how to ask for it.

Automating the data supply chain
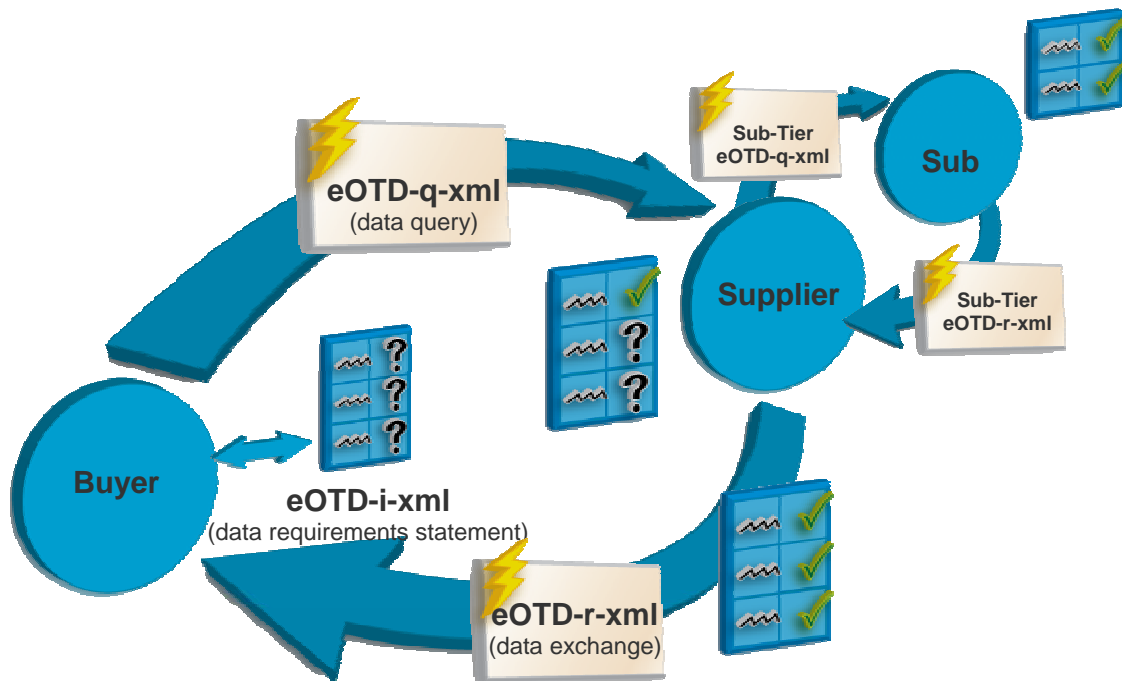
Once we have quality master data then we can start to build quality descriptions; be they short descriptions for an ERP application or longer descriptions for a request for quotation or a purchase order. Creating cryptic and ambiguous free text descriptions is rapidly becoming a thing of the past as software applications that are able to read master data in eOTD-r-xml can render this data into descriptions of any length and for any purpose.

Automating the creation of quality descriptions