# Making Information Visible, Accessible, and Understandable: Meta-Data and Registries

Clay Robinson

*Department of Defense Chief Information Officer Office of Information Policy*

*The term* metadata *is often misused and misunderstood. It is important to understand the categories, multiple meanings, and value of using metadata to improve the interoperability, discovery, and utility of data assets throughout the Department of Defense (DoD). Proper use and understanding of metadata can substantially enhance the utility of data by making it more visible, accessible, and understandable. Expanded use of metadata leads to better-informed decision making, improved management of information, increased return on investment for digital asset production and publishing, improved security management, and more effective information sharing.*

The DoD Net-Centric Data Strategy requires that information assets be tagged with metadata. The concept of metadata can be confusing and many people are unclear how metadata contributes to the mandates of improved discovery, accessibility, and understandability.

There are many reasons to use metadata. First, it improves precision search for specific queries; second, it clarifies context for understanding; third, it allows identification of security classifications/controls. Expanded use of metadata leads to better-informed decision making, improved management of information, increased return on investment for digital asset production and publishing, and improved security management and information sharing. The best metadata provides a rich description of information assets so that a simple search query produces meaningful results in which a user can easily determine the usefulness of the data asset. Good metadata enables users to avoid sorting through many search responses that are not relevant because of context conflicts or file type mismatches, thereby reducing time for decision-making.

In its simplest meaning, *metadata* is information about something. The term metadata, as used in this article, refers to structured definitions that describe the properties of distinct computer data assets. *Metacard* is the term often used to describe the aggregate of metadata about a particular asset similar to the notion of a catalog card in a library. An example of metadata is the description of a music file specifying the creator, the artist that performed the song, the data created, the length of play time, album name, and the genre. Without resource metadata, portable digital music players would not be so popular due to the difficulty in creating and sorting playlists or finding particular songs. Another exam-

ple may be a metacard that contains information regarding an improvised explosive device (IED) event database. The IED metacard may include details such as security classification, geographic locations covered, event type, time, point of contact for access to the data (if not already granted), etc. Metadata is much more than just keyword tags; it provides richer information. Many existing programs and applications automatically produce metadata when data is created. For example, standard commercial word processing applications produce metadata such as title, time stamp, author or creator, and type of file.

> ## "*Metadata can be categorized in numerous ways, but three ... are resource (bibliographic), structural, and semantic.*"

Metadata can be categorized in numerous ways, but three principle categories are resource (bibliographic), structural, and semantic. Resource metadata contributes principally to visibility of an information asset. Resource metadata includes security classification, title, description, creator, publish date, and other attributes. Resource metadata is similar in concept to cards in a library catalog used to locate books. In this case, metadata helps the user locate data or services. The DoD has published the DoD Discovery Metadata Specification (DDMS) (https://meta-data.dod.mil) to define a particular type of resource metadata to support precision search.

Structural metadata is critical to accessibility and *usability*. It includes schemas and models that describe struc-

ture and formatting which are critical to interoperability and the management of databases. Going back to the portable music player example, not all devices play all audio and video file formats. Designation of file format lets a user match the file type to his device. In the case of a warfighter looking for information, he may have a desktop that is limited to the types of files (i.e. Portable Document Format or Power Point) he can view and by knowing file type or size, the user can download accordingly.

Semantic metadata helps with understandability of terms and includes shared vocabularies, taxonomies, and ontologies. Communities of Interest (COIs) usually speak in their own vernacular. Terms often have unique meanings within a given COI's context, and metadata enhances understanding of their terms. As an example, the data element or term *frequency* may relate to radio spectrum in the signals intelligence community, but *frequency* may relate to the periodicity of payments for the finance community. It is unreasonable and unrealistic to have a single meaning across the entire DoD for that term. However, within particular COIs, terms should have specific meanings. Once a user recognizes a term is from a particular community, then she can better relate to the term and understand its meaning and applicability. For several years, the DoD attempted to standardize data elements with a single common meaning across the DoD. Considering the DoD's size and broad set of communities and missions, department-wide data element standardization was not successful. The DoD now recognizes the concept of COIs and is fostering an environment for each COI to describe their vocabularies using metadata.

A number of metadata-related activities are under way throughout the DoD. To promote effective use of metadata,

the DoD has issued the DoD Net Centric Data Strategy Directive 8320.2, <www.dtic.mil/whs/directives/corres/html/832002.htm>, the DDMS, DoD Net-Centric Data Strategy Program, Decision Memorandum III, and other implementing guidance. The Defense Information Systems Agency (DISA) chairs the DoD Metadata Working Group which meets bi-monthly to address a variety of metadata topics. DISA also manages the DoD Metadata Registry and Clearinghouse as well as the COI Directory. The DoD Metadata Registry and Clearinghouse provides software developers access to data technologies to support DoD community mission applications. Through the Metadata Registry and Clearinghouse,

software developers can access registered extensible markup language data and metadata components, database segments, reference data tables, and related metadata information. These data technologies increase the DoD's core capabilities by integrating common data and enterprise data services built from reusable data components. For more information on the referenced items, see <www.dod.mil/cio-nii> and <http://metadata.dod.mil>. For the DoD to successfully operate in a net-centric environment, people must understand metadata. Metadata is a key element of information sharing and interoperability. For further information, see <http://metadata.dod.mil>.◆

**About the Author**

**Clay Robinson** is associate director in the DoD CIO's Office of Information Policy where he has led the development of the DoD's Discovery Metadata Specification, which is a key component of the DoD's Net-Centric Data Strategy. Robinson has a bachelor's degree in economics from Virginia Tech.

**Office of the DoD CIO**
**Information Policy**
**E-mail: clay.robinson@osd.mil**
**Phone: (703) 602-0937**

# LETTER TO THE EDITOR

**Dear CROSSTALK Editor:**
The function point analysis (FPA) described in Ian Brown's article *Controlling Software Acquisition Costs with Function Points and Estimation Tools* implies the estimating tool accepts adjusted function points (AFPs) per International Function Point Users Group (IFPUG) standard 4.2 as input and allows the estimator to perform trade-off analyses to arrive at an acceptable cost and schedule.

The FP count is backfired into equivalent source lines internal to the estimating tool. The AFP provides a single valued input, unless there is a variance associated with the FP count, which will produce a point estimate. The outputs produced in the article are all related to output distributions of cost and schedule. Point inputs produce point outputs. Are we to assume the AFP produces an input with low – most likely – and high FP counts? The article also discusses the use of commercial off-the-shelf (COTS) and reused components as part of the trade-off analysis. The use of these components in the trade-off analysis raises the zero function point problem when dealing with the cost and schedule impact associated with reused system components.

– Dr. Randall Jensen
<randall.jensen@hill.af.mil>

**Dear CROSSTALK Editor:**
In spite of the fact that function points have been around for more than a quarter of a century now, there are still many misconceptions and misunderstandings about function points. Let me address each point in turn.

First, most estimation tools accept unadjusted function points as a sizing input. The tools rely on more targeted parameters such as multiple site development, reuse required, and requirements volatility to calculate estimation adjustments that might have been handled by the general systems characteristics and AFPs before parametric tools were as prevalent as they are today.

Second, function points are but one input into an estimation tool. Other cost drivers, such as personnel capabilities and

experience, development environment, and product requirements are used to tailor the cost estimate to the particular program. Very often these parameters are expressed as ranges – particularly in an acquisition environment where specific information may not be available. For example, the program office may have a minimum Capability Maturity Model® Integration level required for the vendor, which would set a minimum level for some of these parameters. But some vendor may bid that performs well above that level, so the acquisition cost framework should include a range of inputs to account for this possibility. When any of the input parameters are set as ranges, the estimation tool will produce a range of cost and schedule outputs. That being said, Dr. Jensen does bring up an excellent point: the function point count itself may be expressed as a range (low, likely, and high). The acquisition process may be in such an early stage that requirements may not be fully defined, or there may be some uncertainty associated with system functionality. In this case, it is completely appropriate to use a size range to develop the acquisition cost and schedule framework.

Finally, let's talk about the *zero function point problem*. Function points measure software size independent of language, technology, or platform – and that includes COTS and reused components. If I've got a set of requirements that translates into 500 function points, and I decide to use a COTS product to meet half of those requirements, I've still got system that is 500 function points in size. It did not all of a sudden just become 250 function points. I would simply have to model the effort differently in the estimation tool than I would if all requirements would be custom developed. I would need to make sure that I knew how to reflect these differences appropriately in the parametric model. This is why you need an experienced person working with the tool. A fool with a tool is still a fool – these tools are powerful and flexible enough that you can get all kinds of answers out of them, and the trick is understanding if you've got the inputs set up right.

– Ian Brown
<brown_ian@bah.com>