



USING YOUR INFORMATION

By Peter Aiken

METADATA – THE NUMBER ONE ISSUE FOR THE INTELLIGENCE COMMUNITY.

This article describes metadata and the reasons that I consider metadata management to be the number one issue facing the intelligence community. On these pages, I will address topics of importance to the intelligence community such as: What is metadata? Why is it important? Why has metadata been a difficult topic to address in the past? The answers to these questions should make it easier for intelligence community professionals to convince management to plan, budget and invest resources in metadata management.

METADATA – AN ILLUSTRATION

I begin many talks on metadata management with a story to illustrate the key role that metadata plays in organizational data management. I tell the group that they are a group of Air Force officers who have hired me as a contractor. The job for which I was hired was to produce a list of the lengths of all of the runways around the world. This information is considered important enough to put on the personal digital assistants of all Air Force pilots so that in the event that they should need a runway, they could quickly and easily locate a runway of suitable length to land their aircraft in the event of an emergency or other unforeseen circumstances.

I (the contractor) am turning over to the Air Force, my work product – a CD ROM containing the lengths of all runway lengths in the world. The audience's job, is to ask me any questions about the product before I take my payment and close the contract. I ask the group if there is any information they need about the data in the CD ROM that will make it more useful to them. Careful audience members will begin to quickly identify a series of questions that include the following:

- Knowing I use a Macintosh: Can your CD ROM be played on a Windows PC? (Answer: Yes.)
- What format is the data stored in? (Answer: Microsoft Excel.)
- What version of Excel? (Answer: Office 2004.)

Other questions often include:

- Where did the data come from? (Answer: Personal observation from CIA field agents.)
- How accurate is the data? (Answer: All runway lengths were double-checked and are 100 percent accurate.)
- How current is the data – i.e., when was it collected? (Answer: Fall of 2003.)

Most groups will not think to ask one very important question:

- What units of measurement were used to represent the runway lengths? (Answer: meters.)

Obviously knowing the answer to this last question can mean the difference between life and death to a pilot who needs 5000 meters to land a plane. These questions and answers all represent metadata about the runway length data that I collected for the Air Force. As you can see, there is a good deal more that must be learned about the data (the runway lengths) before it can be usefully employed by the Air Force.

WHAT IS METADATA?

So the first question to be addressed is: what is metadata? Unfortunately the data management community has not yet agreed upon a standard definition of metadata. That is unfortunate and makes us look foolish in the eyes of management. When asked, most will reply that metadata is data about data and this is a good starting point. Unfortunately, busy managers who are not familiar with data management have been confused by the term metadata. Others have suggested that a good explanation is that metadata is data resource data. While technically correct, metadata is additional data that permits data (the runway lengths) to be used more effectively as a resource by its consumers, I have not found management to like this explanation any more than 'data about data'.

However, when I explain that metadata isn't a type of data, but a use of data, they seem to better understand what I'm describing. Metadata is data that is used to better understand one or more data items. (I shouldn't need to point out here that data is a combination of facts paired with specific meanings.) Metadata as a use of data means that (correctly) one person's data is another person's metadata. This last is fully compatible with the International Standards Organization's definition of metadata as: "the information and documentation which makes data sets understandable and sharable for users" [ISO, 1996].



Metadata is data that is used to better understand one or more data items



We in the intelligence community need metadata because it makes the data that we collect, analyze and process more useful to the consumers of our work products. If the runway lengths were in inches, were collected in 1901, or were only 10 percent accurate, the Air Force would be well within its rights to not pay me for my CD ROM.

WHY IS IT IMPORTANT? (MORE THAN JUST THE FACTS!)

Coincident with being asked to contribute this article, I finished reading the book *Secret Empire: Eisenhower, CIA, and the Hidden Story of America's Space Espionage*. It tells the story of the real dilemma that President Eisenhower was faced with during the Cold War. While he didn't want to overspend militarily, the launch of Sputnik, statements from the Soviet Premier Nikita Khrushchev, and constant harassment from Congressional Democrats about the feared bomber and missile gaps, led the public to fear that the United States was falling behind the Soviet Union militarily. What Eisenhower lacked was metadata to back up his belief that the US continued to outpace the Soviets in the arms race. He lacked metadata to back up his beliefs and this disquieted him so much that he initiated photographic surveillance of the Soviet Union using spy planes such as the U2 and space-based reconnaissance satellites. Nether technology existed prior to his administration.



Is meta data taking off?

According to the American Heritage English Dictionary, the prefix meta- is derived from a definition meaning 'beyond, transcending, being more comprehensive, and existing at a higher state of development'. Metadata is critical to the intelligence community because without it, we would be unable to report anything other than raw data. Knowing when data was collected, how it was collected, by whom it was collected, are just a few metadata items that specialists use when analyzing intelligence findings. Data provided by known liars and other disreputable sources are necessarily given less weight than data from more reliable sources. Metadata is what enables us to infer a value of specific data items.

Without metadata, organizations are left with just raw facts. Data provided with associated metadata acquires more detailed and comprehensive meaning and becomes valuable beyond its immediate community of uses. Just as President Eisenhower preferred to have facts to back up his assertions of US arm superiority, good analysts prefer data with its associated metadata when producing their estimates.

WHY HAS IT BEEN DIFFICULT?

Aside from the obvious difficulties explaining the concept of metadata to those in charge of funding metadata management initiatives, it has been difficult to implement metadata management schemes because of inherent technological limitations as well as lack of disciplined metadata management engineering practices. Most

data management applications contain only rudimentary metadata management capabilities and those are primarily oriented at maintaining system specific metadata. Our research at the Institute for Data Research indicates that only one-in-ten commercial organizations is even attempting to implement well engineered metadata management schemes. This research finding tends to excuse the report that almost two years after the 9/11 terrorist attacks in New York and Washington DC, the government lacked even a plan to harmonize the metadata of the dozen or so 'watch lists' developed by different intelligence agencies to track undesirable persons movements into and out of the US. The demands of today's metadata management technologies differ substantially from those required in the past in a number of dimensions.

THE METADATA USERS

Modern data analysts have grown to expect that metadata is part of the data that they analyze. Further, the users of metadata encompass a broader range of personnel including: system contractors; developers; architects; managers; and end users. Key to effective metadata engineering is to satisfy differing user information needs with the same core set of metadata.

METADATA BREADTH

Modern metadata is more encompassing than traditional documentation forms. Metadata now needs to develop to be used

interactively, supporting real time fact capture, organization, support and presentation functions.

METADATA INTEGRATION

Metadata documents data from integration perspectives. Metadata integration with other metadata is high. It is related formally using organizationally standard meanings. Overall metadata value generally increases as additional types are integrated. Integration also permits homogeneous access technology and programmatic control over the physical data using logical manipulation forms such as SQL, graphical browsers, object technologies, etc.

METADATA DEPLOYMENT

Metadata users desire metadata access at workstations when working individually and with high-resolution projection capabilities when working with groups. They want to manipulate and analyze the metadata, to aid when performing implementation tasks. High volume metadata is typically implemented using commercial tool kits with: online browsing capabilities; a central repository; interactive graphical analysis tools; one to many distribution via server and LAN/Intranets. While accurate, traditional paper/three ring binder-based documentation is more desirable than no documentation, integrated hypermedia-based metadata provides useful hypertext links to connect individual metadata.

MAKING IT REAL

How can intelligence community professionals convince management to plan, budget and invest in resources for metadata management?

Modern metadata engineering is the means of formally supporting the extraction of the metadata into data management systems and using the resulting repository capabilities available to inform subsequent activities. Metadata users invest in it to better manage and maintain their systems. Metadata engineering results in systems that satisfy metadata user requirements, typically providing access to requested metadata in a fashion supporting the rapid cognitive momentum required of these analysts. Metadata engineering is not a new technique, but instead a formalization of the system fact management practice. Organizational implementation of formalized metadata engineering increases with organizational knowledge/abilities and situation dictates.

Metadata management can benefit from the same sort of techniques that data engineers use when organizing and structuring business information to understand and meet the requirements of a given business function. Metadata engineering can also benefit from leverage achieved understanding a relatively large amount of information by modeling and managing a relatively small amount of metadata. Maintained at the granular level,



How can intelligence community professionals convince management to plan, budget and invest in resources for metadata management?



metadata is variously combined to respond to user needs. From data engineering, we understand how:

- Different combinations of facts can satisfy diverse user needs (for example, contrast operational data uses with warehoused data uses);
- To effectively share the commonly understood metadata within organizations and with business partners; and
- Systems tend to evolve at a rate that is faster than traditional documentation.

These conditions dictate that metadata must be managed as a 'living document' with easy access to shared, re-configurable metadata.

One final note, the advent of extensible markup language (XML) will prove to be a major boon to the intelligence community. Using XML, we will be able integrate data in a manner that is an order of magnitude less expensive than previously available methods. XML has already been adopted by large segments of the business community to help them to trim the 20-40 percent of total IT expenditures that accrue to data and metadata integration tasks. ■