

## GROUP SEVEN:

### Metadata: Current Issues in Building Successful Metadata

Participants: Carol Hert (chair), David Barker-Plummer, Tony Stefanidis, Mathew Weaver, AthmanBouguettaya,

The success of digital government initiatives often depends on the ability to combine multiple data sources, share knowledge about those data sources and enable access to them. Metadata are critical for the successful accomplishment of these tasks. Current issues in metadata systems such as appropriate metadata standards and structures, location of metadata, and evaluation of metadata repositories will be discussed.

#### DISCUSSION

The session started with introductions expressing our diverse interests in metadata and metadata usage and then, on to a discussion of the definition of metadata, of which we all seemed to have our own. We had differences in what would be considered metadata, how to model relationships among metadata, data, and other information.

Particular points of interest during this discussion included:

- Rationales and their capture as metadata
- The semantics of metadata within a database and across databases
- Metadata definition must be within a context. Different uses for metadata (for search and retrieval, for rationale recording and use, for integration of databases, etc.)

Agencies have a different idea of metadata. For example, how to do cognitive tests and the rationale behind survey questions, etc. The derived data—called metadata by survey folks at workshop. Statistical people “This column has a variance of X” —think this is metadata, but we do not think that it is. Should there be a distinction between derived and metadata? What if you do an analysis on data? For any survey instrument, there are levels of data. Derived data is below the collected data. Metadata is above the collected data.

Sometimes there is meta-metadata above the metadata.

There are two problems: semantics within the database, and then semantics across databases. The schema is the metadata, so schema integration data is meta-metadata. This problem really has been plaguing us. We have used an ontological approach, but we are not satisfied. The semantics that you chose may not be the semantics that were intended. Agencies that collect data is geared from the producer perspective rather than the user perspective. They sometimes neglect the metadata in process design. Users could be very diverse—a citizen planning a vacation, someone trying to respond to a natural disaster, a real estate agent building a virtual reality of a neighborhood. The question is, are we dealing with structured or unstructured data? The problems are different for each. In structured data (i.e., databases) the schema is already there. In the case of semi-structured data, you still have to build the schema. People feel more comfortable with schemas because you can reason about schemas, query schemas, etc. We are dealing mostly with databases. The problem of metadata itself is not such a big problem—except for understanding the schema. We want to have a uniform interface to interact with the semi-structured and the structured data. Our main problem is to automatically look at the schema and generate the metadata (ontologies).