METADATA AND NICS: Joys, Sorrows and Payoffs

Introduction to Metadata

What is metadata? The catchy definition is "data about data." This phrase, however, fails to signify metadata's importance for data collectors and users. Metadata is a summary of information about the form and content of a resource. Good metadata makes information stored in databases transparent, allowing data users to analyze data with confidence and know whether they can combine disparate datasets. The success of NICS depends on the widespread development and implementation of metadata standards and automation tools to facilitate better data sharing across data communities.

While metadata has been a widely accepted formal term only since the 1990s, the word has been in use since the 1960s and the concept is even older. Any organizational scheme that seeks to identify data or to relate multiple data sets relies on metadata. For instance, library card catalogues tag books with a coding system that allow users to locate a book based on title, author, publisher, or another defining feature. The catalogue, therefore, contains the metadata of each book it stores, allowing for easy retrieval.

Metadata has grown in its practical significance through the digital revolution, and we rely on it for most of our common daily tasks. With the Internet today serving as a primary means for the spread of information, detailed and accurate metadata helps us to assess a given data set for its value to a particular purpose. Similarly, web users expect to be able to search for an undetermined data set based only on a few key search terms. Without metadata codes, Outlook, for example, wouldn't know how to sort email by date, by sender, or by subject. Google wouldn't be able to find relevant pages based on a few key words, let alone know how to distinguish an image from a text document.

The need for accurate and well-documented statistical metadata is critical in running smooth data collection operations and contributing to informed dissemination and

analysis of data results. In "The Role of Metadata in Statistics,"[1] Cathryn Dippo of the Bureau of Labor Statistics explains: "Metadata descriptions go beyond the pure form and contents of data. Metadata are also used to describe administrative facts about data, like who created them, and when. Such metadata may facilitate efficient searching and locating of data. Other types of metadata describe the processes behind the data, how the data were collected and processed, before they were communicated or stored in a database. An operational description of the data collection process behind the data (including e.g., questions asked of respondents) is often more useful than an abstract definition of the 'ideal' concept behind the data."

We use statistical metadata to understand the content of the statistical data we are analyzing and to understand its limitations and possibilities for integration with other information. Statistical metadata allows us to confidently decide whether we can relate data from one study with data from another. When data users know and understand the source and timeliness of the data, and method by which it was collected, they are better informed of its strengths and weaknesses, and, therefore, the limitations and possibilities for its uses for their particular purposes.

The value of providing detailed metadata is becoming more widely-known among data collectors and users. There is, however, a lack of standards for metadata content and of formatting tools for the automation of detailed statistical metadata from disparate data sets. Current metadata automation tools are geared more towards the limited information librarians use and do not take into account differences between metadata for surveys and metadata for administrative records that have been released as statistical files. Furthermore, many data collectors who recognize the need to provide metadata simply do not know how to start the process, let alone maintain metadata over time.

At present, data users across the country struggle to compile and integrate data sets to help them reveal patterns and trends. Metadata facilitates the process by which people

[1] Dippo, C. (2000) The Role of Metadata in Statistics, *Proceedings of the 2nd International Conference on Establishment Surveys*, 909-918, American Statistical Association, Washington.

make connections between data they want and data that already exists; it also provides users with a measure of assurance as to whether the data they are looking at will meet their needs.

Metadata and NICS: Joys, Sorrows and Payoffs seeks to bridge these knowledge gaps and bring together statistical data collectors, disseminators, tool developers, and users in a shared recognition of metadata's importance. NICS understands that data communities approach metadata with different perspectives and various levels of experience. *Metadata and NICS* has been planned with this in mind, to inform those who deal with statistical data in any form and to bring people together under the shared goal of improving metadata standards for the benefit of data users across the country.