

## **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 implications 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 and combine disparate datasets with confidence. NICS is particularly interested in the implementation of metadata and the widespread development of metadata standards to facilitate better data sharing across data communities.

While metadata as a formal term has only been around since the 1990s, the word has been in use since the 1960s and the concept is even older. Any organizational schema that seeks to identify data or to relate data sets relies on metadata. Library card catalogues tag books with a unique 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 the focal point for the spread of information, detailed and accurate metadata helps us to assess a given data set for its value to a particular project. 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 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 word document.

The need for accurate and well-documented statistical metadata is critical in running smooth data collection operations and ensuring accurate dissemination and analysis of

data results. In her paper “The Role of Metadata in Statistics,”<sup>1</sup> Cathryn S. 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 to 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. In this vein, statistical metadata allows us to confidently 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 of its use.

The implications of providing detailed metadata are becoming more widely-known among data collectors and users. Nevertheless, there is lack of metadata formatting standards, thus preventing widespread order or continuity of metadata organization among disparate data sets and data repositories. Furthermore, there are knowledge gaps among organizations and individuals that deal with data, and many data collectors who recognize the need to provide metadata simply don’t know how to do this or where to start. Local governments across the country struggle to compile accurate data sets that will help them reveal local patterns and trends. Metadata facilitates the process by which people make connections between data they want and data that already exists; it also provides users with a measure of confidence that the data they are looking at will meet their needs.

---

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

*Metadata and NICS: Joys, Sorrows and Payoffs* seeks to bridge these knowledge gaps and join vast communities of statistical data collectors, disseminators, and users via their 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 communities across the country.