

# Data Documentation Initiative (DDI) Tag Document

The Data Documentation Initiative is an international effort to establish a standard for technical documentation describing social science data which is written in XML. The tagged structure enables computer processing of the information. Although, DDI is currently a specification and not a formal ISO standard, it is an important goal of the initiative.

Previously, searches across a range of data repositories took time and effort due to a lack of consistency in metadata formatting. However, within DDI-compliant repositories, metadata and datasets use internationally recognized documentation schema, and this allows for the detailed and structured searching of information. Various documentation elements of a survey are contained within 'tags'; each tag is standardized to contain specific contextual information about the data and metadata. For example, if a user is conducting a search for *adult literacy*, instead of searching through bodies of text or individual lists of studies, text in the tag <keyword> (2.2.1.1), within DDI-compliant repositories, will quickly reveal all studies containing this type of data. Or, for instance, if a user wants to search for a specific *producer* of data, the tag <producer> (2.1.3.1.) contains this information and a list of relevant datasets will be displayed on one search screen.

The overall objective of this reference document was to have certain tags from the DDI Tag Library that were deemed necessary for Statistics Canada and the Data Liberation Initiative (DLI) community to implement when marking up DDI compliant files. Also, this document will help in establishing best practices to ensure a consensus on the interpretation of the DDI standards, thereby files will be marked-up in a similar fashion so that information can be easily shared. This document may evolve in the future as new tags may be added or removed as needed.

The tag document is divided into five sections: Document Description, Study Description, Data Files Description, Variables Description and Other Materials. Each section has a brief overview, a list of the selected tags, along with a concise description of each tag with pertinent examples which will further assist in the understanding and interpretation of the tag. It is important to note that the Document Description section refers to the person/organization that is marking up the file. Whereby, the Study Description section is information about the data collection. Some tags are repetitive but they take on different meanings depending on the context and the location within their section. Even though many of the tags are listed as Optional, the tags listed within this document are the ones which are recommended to be completed when marking up files. To facilitate the process in marking up Statistics Canada surveys, a suggested source of where to find the information is included in the necessary tags.

Although, the Data Liberation is distributing this document it is important to recognize the contribution of the former members of the CANDDI Group for all their ground work; also, for sharing their knowledge and expertise in the preparation of this document. Special thanks to Michelle Edwards from Guelph University, Jane Fry from Carleton University and Irene Wong from the Research Data Centre in Alberta University for their advice and comments as this document was put together. Even though this document may not meet the needs of everyone within the DLI community, is our hope that it will be a good reference point.

For a more complete listing of all the DDI section numbers, you can refer to the ICSPR site at:  
<http://www.icpsr.umich.edu/DDI/index.html>