

The Data Documentation Initiative (DDI)

Joachim Wackerow
GESIS
Mannheim, Germany
joachim.wackerow@gesis.org

Keywords: Data Document Initiative, DDI, metadata, datasets

The Data Documentation Initiative (DDI)⁷⁰ is an international effort to establish an XML-based standard for the compilation, presentation, and exchange of documentation for datasets in the social and behavioral sciences. The most recent version 3.0 of the DDI supports a rich and structured set of metadata elements that not only fully informs a potential data analyst about a given dataset but also facilitates computer processing of the data.⁷¹ Moreover, data producers will find that by adopting the DDI standard they can produce better and more complete documentation as a natural step in designing and fielding computer-assisted interviewing.

DDI 3.0 embraces the full life cycle of the data from conception, through development of the data collection instrument, collection and cleaning of data, production of data products, distribution, preservation, and reuse or analysis of the data. DDI 3.0 is designed to facilitate sharing schemes for concepts, questions, coding, and variables within organizations or throughout the social science research community. Comparison through direct inheritance as in the case of comparison-by-design or through the mapping of items like variables or categories allow capture of the harmonization processes used in creating integrated files in an uniform and machine-actionable way. DDI 3.0 is providing the structural support needed to facilitate comparative survey work in a way that was previously unavailable in an open, non-proprietary system.

A specific DDI module allows for the capture and expression of native Dublin Core elements (DCMES), used either as references or as descriptions of a particular set of metadata. This module uses the simple Dublin Core namespace represented as XML Schema following the guidelines for implementing Dublin Core in XML. In DDI, the Dublin Core is not used as the primary citation mechanism – this module is included to support applications which understand the Dublin Core XML, but which do not understand DDI. This module is used wherever citations are permitted within DDI 3.0 (like citations of a study description or of other material).

DDI 3.0 is aligned with other metadata standards as well: with SDMX (time-series data) for exchanging aggregate data, with ISO/IEC 11179 (metadata registry) for building data registries such as question, variable, and concept banks, and with FGDC and ISO 19115 (geographic standards) for supporting GIS users.

DDI 3.0 is described in a conceptual model which is also expressed in the Universal Modeling Language (UML). Modular XML Schemas are derived from the conceptual model. Many elements support computer processing – that is, it will go beyond being “human readable”, and move toward the goal of being “machine-actionable”. The final release of DDI 3.0 has been published on April 28th 2008. The standard was developed by the DDI Alliance, an international group encompassing data archives and research institutions from several countries in Western Europe and North America.

Earlier versions of DDI provide examples of institutions and applications: the Inter-university Consortium for Political and Social Research (ICPSR) Data Catalog, the Council of European Social Science Data Services (CESSDA) Data Portal, the Dataverse Network, the International Household Survey Network (IHSN), NESSTAR Software for publishing data on the Web and online analysis, and the Microdata Management Toolkit (by the World Bank Data Group for IHSN).

⁷⁰ www.ddialliance.org

⁷¹ <http://www.ddialliance.org/ddi3/>