

The DDI: Striking a balance to meet the needs of the data community

Discussants:

Wendy Thomas, University of Minnesota

Peter Joftis, ICPSR

Emiel Kaper, Statistics Netherlands

Marcus Schommler, IZ

Humphrey Southall, University of Portsmouth

^aWith the achievements of the DDI, codebooks can now be created in a uniform, highly structured format that is easily and precisely searchable on the Web, that lends itself well to simultaneous use of multiple datasets, and that will significantly improve the content and usability of metadata. Further, this specification may have far-reaching implications for improvement of the entire process of data collection, data dissemination, and data analysis. -- DDI homepage <<http://www.icpsr.umich.edu/ddi/>>

Data Community

- Data producers
- Data archivists (preservation)
- Data librarians (discovery)
- Software developers
- Disseminators
- Users

∕ Developing controlled vocabularies for as many attributes as possible, again to facilitate machine processing

∕ Extending the DTD to encompass elements describing the CAI process

∕ Adding elements to the DTD to document aggregate and time series data

∕ Extending the DTD to handle complex file types

∕ Enabling the DTD to document "families" of datasets as well as different but related datasets

∕ Adding crosswalks to other bibliographic schemes like Dublin Core, GILS, and MARC

∕ Incorporating standards for metadata about spatial data

- selected items from DDI: Future Directions

Current Applications

- Discovery software (general, higher level)
- Discovery software (data contents)
 - NESSTAR, GRETA, FERRET, CA Counts, ILSES
- Extraction tools
 - NESSTAR, FERRET, CA Counts, ILSES (FASTER, GRETA, NHGIS)
- Analysis tools
 - NESSTAR, FERRET (FASTER?, NHGIS?)
- Questionnaire / Survey development
 - ISSP DataWizard

Observations

- Data processing and analysis software seems to be driving current DDI development
- Focus of initiatives is on sections 3.0 and 4.0 (file and data description)
- Problems of dealing with the dimensions of time and geography seem to be tabled
- General / High level discovery tools and interoperability with other forms of information seem to be minor or side issues

Questions to Panel

- What are some plans or ideas for exploiting the contents of the DDI codebook?
- What features are needed to expand the possibilities for automatic content processing?
- What features should be addressed within the DDI structure and what should be in external descriptions?
- How would these expansions affect the ability of the DDI to meet the needs of the data community?