DDI
Does it have a life beyond IASSIST?

IASSIST/IFDO 2005
Edinburgh

Ernie Boyko
NESSTAR Americas
Ottawa

May, 2005
Outline

- About DDI
- Who is using it?
- Producer issues
- Some Metadata standards
  - DDI
  - ISO 11179
  - SDMX
- Analysis and conclusions
About DDI

- Established in 1995 to create a universally supported metadata standard for the social science community (first meeting in Quebec city at IASSIST 1995)
- Initiated and organised by the Inter-University Consortium for Political and Social Research (ICPSR), Michigan, USA
- Guided by the DDI Alliance
Why <ddi>?

- An XML structure for a codebook to be:
  - manipulated
  - viewed
  - searched, and
  - employed by stat packages

- Involves diverse participants:
  - data producers
  - archives/data centres
  - researchers/users
Who is using it?

- Data libraries and archives
- Some research projects
- Some government projects
- Some data producers
- Builders of dataset collections

The dream is to have data producers do the DDI markup at source... ... but!
Data Producer Issues

- See it as extra work
- Have existing systems and are reluctant to change
- Complex surveys
- Process management
- Are often not interested in archiving

But more importantly, they are concerned about which standard they should use!
What Other Standards?

- ISO/IEC 11179 – Metadata registries
- SDMX - Statistical Data Metadata Exchange
- There are others but they do not come on the data producer radar screen
ISO/IEC 11179 – Metadata registries

- A standard for describing any kind of data and organizing the descriptions in a registry (previously, focus was data elements)  
  Dan Gillman US Bureau Labor Statistics

- Being used by NSO’s to describe data as part of their metadata management strategy

- Enables comparison of data from different sources e.g., classification schemes

- Freely available at [http://www.jtc1.org](http://www.jtc1.org)
ISO/IEC 11179 – Metadata registries

- Not aware of any full implementation of an ISO 11179 model
- Statistics Canada is close with their IMDB project
- A good source for creating codebooks
- See http://www.statcan.ca/english/concepts/index.htm
Definitions, data sources and methods

About
Definitions, data sources and methods

Notices
North American Industry Classification System (NAICS) 2002

Changes to the Standard Geographical Classification 2001

New Surveys

Consultation on classifications
North American Products Classification (NAPCS)

Statistical Activities

- List by subject
- Alphabetical list

Surveys and derived statistics

- List by subject
- Alphabetical list

Questionnaires

- List by subject
- Alphabetical list

Variables

- List by topic
- List by statistical unit
- List by classification domain

Standard classifications

- Industry
- Occupation
- Product
- Geography
- Chart of Accounts: Financial Position and Performance
- Other classifications
  - List by subject
  - Alphabetical list

Information on data quality

- Policy on informing users of data quality and methodology
- Policy on standards
- Quality Assurance
**SDMX**

- To support efficient processes for exchanging data between and among partners
- Common vocabularies
- International code lists
- National Accounts Frameworks
- Balance of Payments Manuals
SDMX

- Bank for International Settlements
- European Central Bank
- Eurostat
- International Monetary Fund
- OECD
- UN Statistical Office
- World bank
SDMX

- Long history of standards and protocols
- Numerous metadata repositories
- Strong focus on consistency and data quality
- See http://www.sdmx.org

Deals only with aggregate data with a major emphasis on time series
DDI Version 3.0

- Conceptual and representation
- Looks beyond data archiving to the complete survey life cycle
- Instrument documentation
- To facilitate more re-use of metadata
- Comparative analysis
- Takes advantage of new XML technologies
Analysis

- Different players with different objectives
- Data elements, data sets, aggregates and time series
- Administrative components
- Consistency and data quality
- Different software and data processing considerations
- Different implementation rates
Conclusions

To quote Karsten Boye Rasmussen

“…..standards must be important,”
Conclusions

To quote Karsten Boye Rasmussen

“…..standards must be important, that is why there are so many to choose from!”
Conclusions

- There is no gold standard for creating metadata and facilitating data and metadata exchange and preservation.
- Different objectives require different standards.
- Organizations will not invest in new tools and standards if they do not provide them with a direct benefit.
Conclusions

To answer the question that I posed at the beginning........

Yes, I think that DDI will have a life well beyond the IASSIST community
Thank you for your attention

Ernie Boyko, Nesstar Americas
boykern@yahoo.com