Developing the DDI in Taiwan - the Case of TSSDA at CSR

Alfred Ko-wei Hu
Center for Survey Research
Academia Sinica
TSSDA at the CSR

- TSSDA: Taiwan Social Science Data Archive
- CSR: The Center for Survey Research at Academia Sinica

A latecomer for survey data archive

A PC-based data archive, almost no mainframe experience in data archiving
TSSDA at the CSR

- **Technology**
  - Little mainframe facility support
  - Internet applications is limited

- **Software**
  - Mostly stand alone PC software
  - Never have Oracle experience

- **Data Storage and Internet**

- **Personnel**
  - Database management
  - Internet applications
Benefits of the DDI Approach: As Specified by the DDI Alliance

- Interoperability
- Richer content
- Single document - multiple purposes.
- On-line subsetting and analysis
- Precision in searching.

Background
DDI Objectives
DDI in Taiwan
DDI Processing
Lessons
Challenges ahead
Benefits of the DDI Approach: As Specified by the DDI Alliance

- Interoperability
- Richer content
- Single document - multiple purposes.
- On-line subsetting and analysis
- Precision in searching.

Hardware and software sophisticated archive with enough internet support
Benefits of the DDI Approach:
Generally viewed by the CSR

- **Interoperability**
  
  Enhance information Integration both textual and numerical information

- **Richer content**

- **Single document - multiple purposes.**
  
  Enhance efficiency in file management

- **On-line subsetting and analysis**
  
  expand potential users of survey data

- **Precision in searching.**
  
  In need of constructing Thesaurus database

- **Basis for English web pages**
Benefits of the DDI Approach:
The CSR expectation: by Web features

Background
DDI Objectives
DDI in Taiwan
DDI Processing
Lessons
Challenges ahead

- Study search
- Online analysis
- Literature search
- Study search
- Var. search
- Single variable
- Cross tab
- Trend analysis
- Published work based on survey data
Benefits of the DDI Approach: The CSR expectation: by Collection

Intend to including the following survey data . . . . .

- Surveys of large scale since the late 1970s
- Serial data and panel data
- Surveys with potential use for comparative studies of East Asian societies
Benefits of the DDI Approach: The CSR expectation: by Phases

**Phase One** (Jul-Dec, 2001)
- Studying
  - the DDI and XML
  - the DDI applications
  - the DDI processing

**Phase Two** (2002-2003)
- DDI input and conversion
- Search engine
- Basic online analysis
- Conversion tools

**Phase Three** (2004-2005)
- English Website
- Upgrading online analysis
- Enhance web features
- Server-based archive

**Background**

**DDI Objectives**

**DDI in Taiwan**

**DDI Processing**

**Lessons**

**Challenges ahead**
What have been done?

Taiwan Social Change Survey

1984 – 2001 (28 volumes)

Panel Studies of Family Dynamics in Taiwan

RI - 1999
RII - 2000
RIII - 2001
RIV - 2002
RV - 2003

RI - 2000
RII - 2001
RIII - 2002
RIV - 2003
What will be done in the next?

Contents:
- Add more survey information (e.g. studies of election & democratization)
- Fix information already in place
- Construct English website

Web Features:
- Add new features of online analysis (e.g. comparison of means, comparison of correlations, regression)
- Improve search engine
- Improve web design
General view

Background

DDI Objectives

DDI in Taiwan

DDI Processing

Lessons

Challenges ahead

Construct DDI

Online Analysis

Questionnaire in PDF

DDI Key in

Checking DDI Format

SPSS data format

XML Generator

DDI Variable Information

XSLT to HTML

Study Info.

Variable Info.

My SQL

Web Design and Program Integration

Textual Information

Download

Search engine for study description, variable info. and codebook
Steps of Work Flow

Textual data
- Study description
- Variables and answering items
- Related literature

Codes of variables mapping
- Def. file + Numeric data
- Revised Def. file + Numeric data
- Excel table For trend analysis

Numeric Data
- Def. file + Numeric data
- Revised Def. file + Numeric data
- Excel table For trend analysis

Add-on Contents
- Thesaurus
- Checking and integration
- DDI File

Web technology
- XSLT
- Search engine
- MYSQL
- PHP

Content conversion

Web features
- Codebook
- Study search
- Variable search
- Lit search
- Online analysis
  - Single variable
  - Cross-tab
  - Trend Analysis
  - Correlation
  - Means
  - Regression
  - Subsetting and downloading
    - Whole
    - Subsetting
Taiwan Social Science Data Archive: A Preview

Background

DDI Objectives

DDI in Taiwan

DDI Processing

Lessons

Challenges ahead

Take a look at . . . .

TSSDA

tssda.sinica.edu.tw
DDI Lessons Learned

Data Collection

Cleaning and Checking

DDI Processing

Release

Webpage

Server

Store

INPUT

CHECK

OUTPUT

Traditional Processing

DDI Processing

Problematic Area
Problematic Area: A1

Background

DDI Objectives

DDI in Taiwan

DDI Processing

Lessons

Data Migration

Learning DDI DTD Tags

Conversion Tools

Issues of Variable Codes

Variable recoding for cross-tab and trend analysis

Challenges ahead
Problematic Area: A1

Background

DDI Objectives

DDI in Taiwan

DDI Processing

Lessons

Data Migration

Learning DDI DTD Tags

Conversion Tools

Issues of Variable Codes

Variable recoding for cross-tab and trend analysis

Challenges ahead

Manageable
Problematic Area: A2

Background

DDI Objectives

DDI in Taiwan

DDI Processing

Lessons

Data Checking

- Pick typos
- Information in the right tag
- Well-formed XML document
- Define valid codes

Information Update

- Construction of thesaurus
- DDI and SPSS syntax

Challenges ahead
Problematic Area: A2

Background

DDI Objectives

DDI in Taiwan

DDI Processing

Lessons

Data Checking

- Pick typos
- Information in the right tag
- Well-formed XML document
- Define valid codes

Information Update

- Construction of thesaurus
- DDI and SPSS syntax

Challenges ahead

Manageable
Problematic Area: A3

Integrate DDI and Relational Database

- Issues of Mapping
  - DDI + Numerical data → relational database

- Issues of Uploading
  - DDI → XSLT → HTML
  - Fetching files from MySQL

- Debugging Chinese Big5 codes
Problematic Area: A3

Integrate DDI and Relational Database

Issues of Mapping
DDI + Numerical data → relational database

Issues of Uploading
DDI → XSLT → HTML
Fetching files from MySQL

Debugging Chinese Big5 codes

Learning to be Manageable
Problematic Area: A4

Web Programming outsourcing web applications

Server and Web Management server administrator vs. webmaster

Background

DDI Objectives

DDI in Taiwan

DDI Processing

Lessons

Challenges ahead
Problematic Area: A4

Web Programming
outsourcing web applications

Server and Web Management
server administrator vs. webmaster

Difficult to manage at present
Problematic Area: A5

Balance among The Seven C's of Web Design -- Albert DeSimone, Jr., Dr. Walter B. McRae, and The Georgia Web Group

- Comprehensiveness
- Currentness
- Client-Orientation
- Clarity over Coolness
- Courtesy over Coolness
- Compatibility without Compromise
- Cross-Linking and Validation
Problematic Area: **A5**

**Balance among The Seven C's of Web Design**  -- Albert DeSimone, Jr., Dr. Walter B. McRae, and The Georgia Web Group

- **Comprehensiveness**
- **Currentness**
- **Client-Orientaion**
- **Clarity over Coolness**
- **Courtesy over Coolness**
- **Compatibility without Compromise**
- **Cross-Linking and Validation**

**Image discovery processes**
Conclusion

Traditional Data Management and DDI Processing

In-house work routine or outsourcing

Online analysis and users’ response

Challenges ahead
Conclusion

Challenges ahead

Background

DDI Objectives

DDI in Taiwan

DDI Processing

Lessons

End

Thank You