Feature enhancement of Easy DDI Organizer (EDO)

Yuki Yonekura, Keiichi Sato, Yukio Maeda
Institute of Social Science, The University of Tokyo
Introduction: About SSJDA

• A data archive of social survey in Japan
  – gathering and storing social survey data
    • About 1300 datasets are stored
    • Datasets available in English are few
  – provide datasets to researchers and students
    • 1630 datasets were provided to 1354 researchers in 2010
    • About 60 new datasets are published per year (2008-2010)

SSJDA homepage(ssjda.iss.u-tokyo.ac.jp/en/)
Challenges in metadata preparation at SSJDA

• We organize survey metadata from sources (report, codebook, papers, etc.) provided by the depositors

Two problems

• Time consuming
  – Ask depositors about unclear information
  – Write HTML manually

• Variation in the quality of metadata
  – There are some variations in information recorded in resources provided by the depositors
  – We cannot provide information which is not recorded in the resources
  – There is no standard for recording survey metadata in Japan

• These problems would be reduced if depositors edit metadata of their own data based on certain standard
Easy DDI Organizer (EDO)

- Desktop application for managing metadata based on DDI 3.1 (Lifecycle)
- Following functions were implemented
  - Edit metadata
  - Export DDI 3.1 document
# Functions of EDO- Editing metadata

<table>
<thead>
<tr>
<th>Metadata Item</th>
<th>Corresponding module of the DDI3.1(Lifecycle)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data lifecycle event record</td>
<td>Archive</td>
</tr>
<tr>
<td>Record events from starting study to data publishing</td>
<td>StudyUnit</td>
</tr>
<tr>
<td>Study description</td>
<td>StudyUnit</td>
</tr>
<tr>
<td>Background information of the study such as purpose, abstract, coverage</td>
<td>StudyUnit</td>
</tr>
<tr>
<td>Data collection</td>
<td>DataCollection</td>
</tr>
<tr>
<td>Sampling procedure, survey method</td>
<td>DataCollection</td>
</tr>
<tr>
<td>Question design</td>
<td>ConceptualComponent/DataCollection</td>
</tr>
<tr>
<td>Concepts, question item, response domain, category, code</td>
<td>ConceptualComponent/DataCollection</td>
</tr>
<tr>
<td>Logical/Physical data information</td>
<td>LogicalDataProduct/PhysicalDataProduct/PhysicalInstance</td>
</tr>
<tr>
<td>Variable information, Record Layout</td>
<td>LogicalDataProduct/PhysicalDataProduct/PhysicalInstance</td>
</tr>
<tr>
<td>Comparison</td>
<td>Group/Comparative</td>
</tr>
</tbody>
</table>
New Features of EDO

• Edit question sequence information
• Import
  – Study metadata from DDI document
  – Variable information from SPSS file (.sav)
• Export
  – questionnaire
  – code book
  – Setup syntax for SPSS
Question sequence editor

Type of Item

Item number

Description of Item
Question sequence editor
- Change question sequence
Question sequence editor
- Add conditional branch
Branch setting dialogue

Condition viewer

Condition editor
Importing study metadata from DDI document
Importing DDI document

[Diagram showing a window to select study units and another window to specify which items to import, with options such as Study Description, Data Lifecycle Events, Study Member, Abstract, Coverage, and Funding.]
Abstract

Institute of Regional Studies at Osaka University of Commerce (key institute on the frontiers of academic projects designated by the Ministry of Education) launched JGSS project in collaboration with the Institute of Social Science at the University of Tokyo. JGSS Research Center at Osaka University of Commerce (Joint Usage/Research Center accredited by Minister of Education, Culture, Sports, Science and Technology) has taken over the project.

There have been 2 pilot surveys and 7 main surveys (JGSS-2000, 2001, 2002, 2003, 2005, 2006, 2008) so far, and 9 main surveys will be conducted biennially hereafter. Generally, surveys consist of questions to understand time-series shifts and topical issues. In addition, there have been multiple developments in the surveys. In JGSS-2003, 2 separate self-administered questionnaires were prepared, and special module (social network questions) was set in...
Importing variable information from SPSS file
<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Concept</th>
<th>Variable Label</th>
<th>Question</th>
<th>Universe</th>
<th>Variable Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>IDUSE</td>
<td>Imported Varial</td>
<td>ID Number for</td>
<td></td>
<td></td>
<td>Numeric</td>
</tr>
<tr>
<td>BLOCK</td>
<td>Imported Varial</td>
<td>Regional Block</td>
<td></td>
<td></td>
<td>Category</td>
</tr>
<tr>
<td>PREF</td>
<td>Imported Varial</td>
<td>Prefecture Narr</td>
<td></td>
<td></td>
<td>Category</td>
</tr>
<tr>
<td>SIZE</td>
<td>Imported Varial</td>
<td>Size of Municip</td>
<td></td>
<td></td>
<td>Category</td>
</tr>
<tr>
<td>SIZE2K</td>
<td>Imported Varial</td>
<td>Size of Municip</td>
<td></td>
<td></td>
<td>Category</td>
</tr>
</tbody>
</table>

**Variables Type Detail**

Select From Existing Response Style

<table>
<thead>
<tr>
<th>Type Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Exporting questionnaire

Questionnaire A

<table>
<thead>
<tr>
<th>Item</th>
<th>Number</th>
<th>Question/Statement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question</td>
<td>Q1</td>
<td>Q1: In an average day, about how many hours do you personally watch television? Enter 0 (zero) if you never watch television.</td>
</tr>
<tr>
<td>Question</td>
<td>Q2</td>
<td>Q2: How many books do you read a month on average (comics and magazines excluded)?</td>
</tr>
<tr>
<td>Question</td>
<td>Q3</td>
<td>Q3: How often do you read the newspaper?</td>
</tr>
<tr>
<td>Question</td>
<td>Q4</td>
<td>Q4-2: Have you ever done any of the following by the Internet using a PC or cellular phone? Choose all that apply.</td>
</tr>
<tr>
<td>Question</td>
<td>Q5</td>
<td>Q4-3B: Sending e-mail by a cellular phone (including PHS)</td>
</tr>
<tr>
<td>Question</td>
<td>Q6</td>
<td>Q4-3C: Sending e-mail by a PC</td>
</tr>
<tr>
<td>Question</td>
<td>Q7</td>
<td>Q4-3A: Speaking on a cellular phone (including PHS)</td>
</tr>
<tr>
<td>Question</td>
<td>Q8</td>
<td>Q4-3B: Sending e-mail by a cellular phone (including PHS)</td>
</tr>
<tr>
<td>Question</td>
<td>Q9</td>
<td>Q4-3C: Sending e-mail by a PC</td>
</tr>
<tr>
<td>Question</td>
<td>Q10</td>
<td>Q5: Do you regularly do any exercises or play any sports (walking, swimming, baseball, etc.)?</td>
</tr>
</tbody>
</table>
Q1. In an average day, about how many hours do you personally watch television? Enter 0 (zero) if you never watch television.

Q2. How many books do you read a month on average (comics and magazines excluded)?

0. I rarely read books.
1. About one
2. About two
3. About three
4. Four or more
9. No answer

Q3. How often do you read the newspaper?

1. Almost everyday
2. Several times a week
3. About once a week
4. Less than once a week
5. Never
9. No answer

Q4. Have you ever done any of the following by the Internet using a PC or cellular phone? Choose all that apply.

1. Searching for information
2. Internet shopping
3. Internet banking
4. Making web pages
5. None

Q5. Sending e-mail by a cellular phone (including PHS)

1. 5 times a day or more
2. 3-4 times a day
3. 1-2 times a day
4. Several times a week
5. Once a week or less
6. Never
8. Not applicable
9. No answer

Q6. Sending e-mail by a PC

1. 5 times a day or more
2. 3-4 times a day
3. 1-2 times a day
4. Several times a week
Exporting codebook

[Software interface showing codebook export for Japanese General Social Surveys (JGSS-2006)]

1. Select the codebook to export.
2. Navigate to the export options within the software interface.
3. Choose the codebook feature to export the questionnaire data.

Questionnaire A

<table>
<thead>
<tr>
<th>Item</th>
<th>Number</th>
<th>Question/Statement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question</td>
<td>Q1</td>
<td>Q1: In an average day, about how many hours do you personally watch television? Enter 0 (zero) if you never watch television.</td>
</tr>
<tr>
<td>Question</td>
<td>Q2</td>
<td>Q2: How many books do you read a month on average (comics and magazines excluded)?</td>
</tr>
<tr>
<td>Question</td>
<td>Q3</td>
<td>Q3: How often do you read the newspaper?</td>
</tr>
<tr>
<td>Question</td>
<td>Q4</td>
<td>Q4-2: Have you ever done any of the following by the Internet using a PC or cellular phone? Choose all that apply.</td>
</tr>
<tr>
<td>Question</td>
<td>Q5</td>
<td>Q4-3B: Sending e-mail by a cellular phone (including PHS)</td>
</tr>
<tr>
<td>Question</td>
<td>Q6</td>
<td>Q4-3C: Sending e-mail by a PC</td>
</tr>
<tr>
<td>Question</td>
<td>Q7</td>
<td>Q4-3A: Speaking on a cellular phone (including PHS)</td>
</tr>
<tr>
<td>Question</td>
<td>Q8</td>
<td>Q4-3B: Sending e-mail by a cellular phone (including PHS)</td>
</tr>
<tr>
<td>Question</td>
<td>Q9</td>
<td>Q4-3C: Sending e-mail by a PC</td>
</tr>
<tr>
<td>Question</td>
<td>Q10</td>
<td>Q5: Do you regularly do any exercises or play any sports (walking, swimming, baseball, etc.)?</td>
</tr>
</tbody>
</table>
Index
1. Abstract
  1.1. Study Member
  1.2. Purpose
  1.3. Abstract
  1.4. Funding

2. Mode of Data Collection
  2.1. Universe
  2.2. Sampling Procedure
  2.3. Data Collection Date/Period
  2.4. Mode of Data Collection

3. Variable Summary
1. Abstract

1.1. Study Member

IWAI Noriko Osaka University
TANIOKA Ichiro Osaka University President,
NITTAMichio The University
<未入力><未入力> Central Research Services Inc

1.2. Purpose

JGSS Research Center at Osaka University of Commerce (Joint Usage/Research Center accredited by Minister of Education, Culture, Sports, Science and Technology) has taken over the project.

1.3. Abstract

Institute of Regional Studies at Osaka University of Commerce (key institute on the frontiers of academic projects designated by the Ministry of Education) launched JGSS project in collaboration with the Institute of Social Science at the University of Tokyo. JGSS Research Center at Osaka University of Commerce (Joint Usage/Research Center accredited by Minister of Education, Culture, Sports, Science and Technology) has taken over the project. There have been 2 pilot surveys and 7 main surveys (JGSS-2000, 2001, 2002, 2003, 2005, 2006, 2008) so far, and main surveys will be conducted biennially hereafter. Generally, surveys consist of questions to understand time-series shifts and topical issues. In addition, there have been multiple developments in the surveys. In JGSS-2003, 2 separate self-administered questionnaires were prepared, and special module (social network questions) was set in to one of the questionnaire forms (Form B). Since JGSS-2005, survey questions have been publicly sought from researchers in and outside of Japan, and included in main surveys. Since JGSS-2006, questions for the EASS (East Asian Social Survey) have been incorporated. EASS project is a cross-national survey which incorporates a common module into repeated social survey already implemented in each country and region. It is planned to conduct every two years starting from 2006. The constituent teams that launched the EASS were JGSS, Taiwan Social Change Survey (TSCS), Korean General Social Survey (KGSS), and CGSS (Chinese General Social Survey). The theme for EASS 2006 is “Families in East Asia” and for 2008 is “Culture and Globalization in East Asia”. JGSS-2008 consists of replicating questions, topical questions, publicly-sought questions, and EASS module. Respondents' basic demographics, such as their occupations and family structure, are asked in face-to-face interview, while their daily behavior and attitudes are asked in self-administered questionnaire. There are two forms of self-administered questionnaires. Form A mainly consists of replicating questions and topical questions, and Form B mainly consists of EASS module. Form A and Form B were randomly distributed to a half of the sample. EASS integrated data can be obtained from EASSDA.

1.4. Funding
2. Data Collection

2.1. Universe

2.2. Sampling Procedure
Two-stage stratified random sampling; stratified by regional block (6) and population size (4). 529 sampling points.

2.3. Data Collection Date/Period
2008-12

2.4. Mode of Data Collection
Face to Face Interview

3. Variable Summary

IDUSE ID Number for users (Numeric)
Question Text

BLOCK Regional Block (Category)
Question Text

<table>
<thead>
<tr>
<th>Code</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Hokkaido or Tohoku</td>
</tr>
<tr>
<td>2</td>
<td>Kanto</td>
</tr>
<tr>
<td>3</td>
<td>Chubu</td>
</tr>
<tr>
<td>4</td>
<td>Kinki</td>
</tr>
<tr>
<td>5</td>
<td>Kyushu</td>
</tr>
</tbody>
</table>

PREF Prefecture Name (Category)
Question Text

<table>
<thead>
<tr>
<th>Code</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Hokkaido</td>
</tr>
<tr>
<td>2</td>
<td>Aomori</td>
</tr>
<tr>
<td>3</td>
<td>Iwate</td>
</tr>
<tr>
<td>4</td>
<td>Miyagi</td>
</tr>
<tr>
<td>5</td>
<td>Akita</td>
</tr>
<tr>
<td>6</td>
<td>Yamagata</td>
</tr>
</tbody>
</table>
Exporting setup syntax
get data
/arrangement=Delimited
/delimiters=";"
/variables=IDUSE F BLOCK F PREF F SIZE F SIZE2K F SEXA F AGEB F MARC F FORM F DOBYEAR F XJOB...
variable labels
IDUSE "ID Number for users"
BLOCK "Regional Block"
PREF "Prefecture Name"
SIZE "Size of Municipality"
SIZE2K "Size of Municipality (as of 2000)"
SEX A "Sex"
AGE B "Age"
MARC "Marital Status (Same as DOMARRY)"
FORM "Form of Self-administered Questionnaire"
DOBYEAR "Date of Birth: Year (Christian Era)"
XJOB1WK "Work Status"
tp5unemp "Reason for Not Working"
XWORK "Ever Had Job"
agestpkw "Age of Quitting the Last Job"
XJOBHWK "Work Hours per Week"
szsbhwk "Working Hours of Secondary Job"
TPJOB "Work Arrangement"
TPJOBP "Job Title"
TPJBDP "Dispatch of General Workers"
TPJBESE "Having Employees"
DOCM "Commuting"
SZCMTHRS "Commuting Time: Hours"
szcmtime "Commuting Time: Minutes"
SZCMTTL "Commuting Time: Total"
Our next step

2012
- Function improvement
  - User friendly question sequence manager
  - Questionnaire template
- Function expansion
  - DDI 1/2 import/export

2013
- Beta-version release
Potential use and benefit

Researchers
• planning/conducting surveys easier
• High quality data
• Reduce cost of preparing metadata

Data users
• Richer metadata
• High quality data

Education
• Learn how to plan/conduct surveys

Data archives
• High quality services
Acknowledgement

• This project is funded by Joint usage/ research center project of the Ministry of Education, Culture, Sports, Science and Technology, Japan.

• We appreciate cooperation of Mr. Takeshi Iwanami, the president of epoch Ltd and Mr. Masaru Tsuda.
Thank you for your attention!
Any questions?

Contact
Yuki Yonekura: yyonekura@iss.u-tokyo.ac.jp