December 7 (Fri.), 2012
13:00～16:30

Social Science Data Preservation and Dissemination in East Asia

Moderator: Shin Arita
Opening Remarks: Hiroshi Ishida

Data Preservation and Dissemination at SSJDA (Social Science Japan Data Archive): Yukio Maeda

Implementing Data Documentation Initiative (DDI) in SSJDA: Yuki Yonekura

Data Archiving at KOSSDA (Korea Social Science Data Archive): Acquisition, Preservation, Documentation, and Dissemination: Joon Han

The Practice of the Survey Research Data Archive in Taiwan: Ruoh-rong Yu

The University of Tokyo
Akamon General Research bldg.
5F (room 549)

Language: English
Inquiry: ssjda@iss.u-tokyo.ac.jp
Data Preservation and Dissemination at Social Science Japan Data Archive

Yukio Maeda
Center for Social Research and Data Archives
Institute of Social Science
University of Tokyo

Table of Contents

I. Historical background and overview
II. The state of data preservation and dissemination in Japan
III. Social Science Japan Data Archive
   1. Institutional framework
   2. Some statistics
   3. Staff members
   4. Data curation policy
   5. Processing
   6. Actual dissemination
   7. Service through Internet
   8. International activities
   9. Education and training
  10. Future Issues
I. Historical background and overview

### Historical background

- In Japanese universities, it is difficult to get a funding and position for a permanent staff member other than professors or purely administrative staff.
- A few initiatives between 1960 and 1990 but none of them materialized.
- No full-time archivist at SSJDA until 2007
- A vicious circle (of distrust)
  - No institutionalized archive -> No trust and budget -> No datasets preserved and disseminated -> No secondary users -> Low productivity -> No institutionalized archive
### Various data organizations in Japan (1)

1. The small but renowned publisher (Bokutaku Ltd.) set up a branch to sell election survey datasets in 1991. Leviathan Data Bank (LDB).
   - Each data set costs $200~$500.
   - LDB only handles datasets in political science but still holds the largest collection today (49 datasets).
   - The datasets have been used extensively among political scientists in Japan.

2. The Social and Opinion Research Database Project (SORD) at Sapporo Gakuin University started its data archive service late in the 1990s.
   - SORD originally started as the database project for the member of the Japanese Sociological Association in 1990.
   - It holds a small number of datasets (15 datasets in 2012) and attracts some users (52 data requests between 2007 and 2012).

### Various data organizations in Japan (2)

3. JEDI system at Hyogo University of Teacher Education
   - [http://essrc.hyogo-u.ac.jp/jedi/?locale=en](http://essrc.hyogo-u.ac.jp/jedi/?locale=en)
   - Archive data from educational research and provides advice research in the filed of education.
   - It currently holds around 150 data sets.

4. RUDA (Rikkyo University Data Archive)
   - [https://ruda.rikkyo.ac.jp/dspace/](https://ruda.rikkyo.ac.jp/dspace/)
   - The most recent organization established for data preservation and dissemination. Since April 2011.
   - Archive survey data in social science, mostly from sociology. It currently holds 29 datasets.
   - Please see its broschure for details.
II. The state of data preservation and dissemination in Japan

The number of publicly available academic survey data for secondary analysis has been increasing.

Many social scientists in young generation now assume that secondary analysis of survey data is an important repertoire of their research strategies.

- It is expected that those young researchers deliver their data to a data archive when they obtain a fund to conduct their own survey.

Opinion data from news agencies are hard to come but a few of them are now available from SSJDA.
Data policy

- Neither the Ministry of Education, Culture, Sports, Science and Technology (MEXT) nor the Science Council of Japan requires the researcher to make their data available for secondary analysis. It is completely up to the judgment of principal investigators (PIs).
- No standard for metadata content and data format.
- No recommendation issued by academic or professional associations.
- Some PIs set up their own websites for downloading. The long-term commitment for preservation and dissemination is unclear.

Official statistics

- New Statistical Act was enacted in 2007.
- The legal framework is set for academic use of official micro data.
- The National Statistics Center (NSTAC) provides official micro data sets for academic users. Its collection is now expanding rapidly.
- Long time and many papers may be necessary to have an access to official micro data files for secondary analysis. A few universities cooperate with NSTAC to facilitate the use of official micro data.
Official Statistics (2)

- Universities with Satellite Facilities
  a. Hitotsubashi University
  b. Hosei University
  c. Kobe University
  d. Institute of Statistical Mathematics

- Available Data sets
  a. Labor Force Survey
  b. Housing and Land Survey
  c. National Survey of Family Income and Expenditure
  d. Employment Status Survey
  e. Survey on Time Use and Leisure Activities

III. SSJDA
1. Institutional Framework

- SSJDA was established in 1998. The first “large-scale” data archive in Japan.
- It is now a part of the Center for Social Research and Data Archives at Institute of Social Science, University of Tokyo.
- The Center is designated as “Joint Usage and Research Center” by the MEXT (2009 ~ ).
- It has been the hub organization for the ICPSR Japanese National Membership since 1999.
- IFDO member since 2001 (only one in East Asia at this moment).

2. Some statistics

<table>
<thead>
<tr>
<th>Year</th>
<th>New release</th>
<th>The total number of datasets</th>
<th>The total number of applications</th>
<th>The total number of researchers</th>
<th>The number of datasets provided</th>
<th>The number of articles published</th>
</tr>
</thead>
<tbody>
<tr>
<td>1998</td>
<td>276</td>
<td>276</td>
<td>11</td>
<td>14</td>
<td>22</td>
<td>2</td>
</tr>
<tr>
<td>1999</td>
<td>91</td>
<td>367</td>
<td>20</td>
<td>33</td>
<td>72</td>
<td>4</td>
</tr>
<tr>
<td>2000</td>
<td>42</td>
<td>409</td>
<td>51</td>
<td>74</td>
<td>147</td>
<td>7</td>
</tr>
<tr>
<td>2001</td>
<td>50</td>
<td>459</td>
<td>114</td>
<td>278</td>
<td>545</td>
<td>22</td>
</tr>
<tr>
<td>2002</td>
<td>40</td>
<td>499</td>
<td>222</td>
<td>489</td>
<td>796</td>
<td>35</td>
</tr>
<tr>
<td>2003</td>
<td>77</td>
<td>576</td>
<td>224</td>
<td>613</td>
<td>1134</td>
<td>26</td>
</tr>
<tr>
<td>2004</td>
<td>101</td>
<td>677</td>
<td>343</td>
<td>748</td>
<td>1641</td>
<td>68</td>
</tr>
<tr>
<td>2005</td>
<td>95</td>
<td>772</td>
<td>335</td>
<td>1,036</td>
<td>2527</td>
<td>75</td>
</tr>
<tr>
<td>2006</td>
<td>185</td>
<td>957</td>
<td>366</td>
<td>944</td>
<td>1577</td>
<td>49</td>
</tr>
<tr>
<td>2007</td>
<td>208</td>
<td>1171</td>
<td>483</td>
<td>1,208</td>
<td>1851</td>
<td>86</td>
</tr>
<tr>
<td>2008</td>
<td>56</td>
<td>1227</td>
<td>481</td>
<td>1,434</td>
<td>2034</td>
<td>113</td>
</tr>
<tr>
<td>2009</td>
<td>66</td>
<td>1293</td>
<td>606</td>
<td>1,756</td>
<td>2784</td>
<td>116</td>
</tr>
<tr>
<td>2010</td>
<td>61</td>
<td>1354</td>
<td>666</td>
<td>1,354</td>
<td>1630</td>
<td>135</td>
</tr>
<tr>
<td>2011</td>
<td>70</td>
<td>1424</td>
<td>774</td>
<td>2,093</td>
<td>1812</td>
<td>186</td>
</tr>
</tbody>
</table>
3. Staff members

- The complicated (or chaotic one may say) organizational structure makes it difficult to count the number of people working at SSJDA.
- Director + Several faculty members
  - In addition to SSJDA, each has his/her own teaching and research obligations.
- A few research associates (post doctoral level)
  - Responsible for daily duties of SSJDA such as handling data requests, maintaining the user database, and website.
  - Engaged in developing tools for data management and dissemination.
- One data archive manager (full-time archivist)
  - Responsible for data cleaning and documentation
- Several staff members of varying time commitments
  - Documentation, clerical works, IT support

4. Data curation policy

- We accept survey data from researchers, government agencies, labor unions, and commercial firms with research division.
- The most of the data we receive are from probability sampling.
- We accept non-probability data such as quota sampling and internet surveys as long as it has research value.
- We do not curate aggregate data, qualitative data, and non-numeric data.
- No plan to process official statistics.
5. Processing (1)

- Checking documentation
  - Whether the data files accompany sufficient metadata such as questionnaire and reports.
  - Whether the report prepared by PIs can be photocopied without infringing copyrights.

- Checking data files
  - Whether the frequencies produced from the data files are consistent with those printed in the report.
  - Removing personal identifiers and sensitive information
  - Converting data into the format for preservation and dissemination.

5. Processing (2)

- Producing metadata for preservation and dissemination.
  - Abstract (sampling design, response rate, number of observations, PIs, and etc.)
  - Preparing questionnaire and frequency tables.

- Double-checking consistencies between the data file and the metadata produced at SSJDA.

- Finally, the data and metadata have to be approved by the PIs before release.
5. Processing (3)

### Abstract

<table>
<thead>
<tr>
<th>Survey Number</th>
<th>0142</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survey Title</td>
<td>Surveys of New Start-ups, 1998</td>
</tr>
<tr>
<td>Depositor</td>
<td>The Research Institute of National Life Finance Corporation (NLIFC)</td>
</tr>
<tr>
<td>Restriction of Use</td>
<td>Apply to SSJ/ Data Archive. NLIFC's approval required</td>
</tr>
</tbody>
</table>

### Summary

Changes in the industrial structure and consumption behavior have brought a structural recession in the older industries and declining shipping deficits in oil and iron. Moreover, as the employment condition has rapidly deteriorated, more and more people have felt the need for new employment opportunities. Consequently, the NLIFC Institute has conducted surveys to clarify the current status of new start-ups from various perspectives, and since 1991 the Institute has been publishing an annual white paper to share the findings. This 1998 survey reveals that new start-ups have become more frequent and new start-ups with employment have increased in number. The survey also offers a detailed analysis of the timing of entrepreneurship concerning the idea of starting new businesses and factors behind such timing as well as the ways of taking issues as in the initial period of business.

### Universe

Companies which were, among others, NLIFC-financed between April and September, 1997, either less than one year old or before opening of the offices.

### Sample Size

6,468 companies, of which 1,661 (24.1%) returned total responses.

### Time Period

August, 1998

### Spatial Unit

All over Japan

### Sampling Procedure

See Subjects above.

### Data Collection Method

Investigator: NLIFC Research Institute

### Sponsor(Funding)

NLIFC Research Institute (July, 1990), Shinko-Kogyo Asakura (Deputy), Akihisa Hase (The 1990 White Paper on New Start-ups) / Chuou-Kogyo Tetsudo Zentaro (The Business and Small Business Research Center, Inc.)

6. Actual dissemination

#### <Classic Method>

- **How to obtain a data file from SSJDA**
  1. One needs to fill out an application form and mail it with a self-addressed envelope with postage.
  2. The form (paper) is forwarded to the PIs to obtain his/her permission for one year use by the applicant.
  3. Data and metadata are burned on the CD and mailed to the applicant.
  4. After one year, s/he has to fill out an report form and submit it with publications (if available). Then, s/he either has to destroy the data files or apply for an extension.

*Every step is conducted manually!!*
6. Actual dissemination (continued)

- Why is the rule so strict?
  - The rule was set when SSJDA started. We tried to be as strict as possible to gain trust from potential data depositors.
  - As the number of data requests grew rapidly, this “stringent” procedure is now exempted for widely used academic surveys such as the Japanese General Social Survey (JGSS) and Social Stratification and Mobility Survey (SSM).
  - In order to cope with increasing demands, SSJDA set up direct download system for heavily used datasets (next slide).

7. Service through Internet

- The “direct” download system was introduced in April, 2009. [https://ssjda.iss.u-tokyo.ac.jp/Direct/](https://ssjda.iss.u-tokyo.ac.jp/Direct/)
  - Actually, still human being working behind!

- The on-line data analysis system (using SPSS web) started working in 2005.
  - This system will be discontinued in the future due to mechanical reasons.

- We started Nesstar service this fall and are working on metadata editor based on DDI. Yonekura will present this topic in depth later.
8. International Activities

- Cooperation with data archives in the US and Europe.
  - Through the framework of ICPSR and IFDO
  - Inviting guests from social science data archives overseas such as GESIS in Germany.

- Cooperation with the data archives in the East Asian Region
  - ISS held the international symposium titled “Data Archives and Their Prospects in East Asia: Collaborative Data Collection, Dissemination, and Data Sharing” in 2008.
  - Mutual visits between SRDA at Academia Sinica in Taiwan.
  - Visiting KOSSDA in 2012.

- The panel survey project within the Center also tries to enhance cooperative activities within East Asia.
9. Education and training

- The center holds two-week spring seminar in statistics in every March. Its courses are, for example:
  - Introduction and intermediate courses on survey data analysis using SPSS.
  - Regression analysis using SPSS.
  - Statistics using R.
- Every year we offer one or two data confrontation seminars.
  - We pick the data available through SSJDA and invite the PI as a visiting professor for a year.
  - We solicit participants for 8-10 months data analysis seminar in which young researchers and graduate students analyze the data while being supervised by the PIs.
  - A few books have been published from the past seminars.

The books from the past data confrontation seminars
10. Future issues

- Expanding the capability of data acquisition, preservation, and dissemination.
- Promoting cooperation with the other data organizations within Japan.
- Enhancing cooperation and mutual data use within East Asia.

Thanks for listening and for your patience.

Social Science Japan Data Archive
Institute of Social Science
University of Tokyo
Implementing Data Documentation Initiative (DDI) at SSJDA

Yuki Yonekura
Institute of Social Science, The University of Tokyo

Data Documentation Initiative (DDI)

- An international standard for describing survey metadata
- DDI Alliance (http://www.ddialliance.org/) provides
  - XML schema, DTD
  - Useful tools based on DDI
- DDI-Codebook (ver. 2.x) and DDI-Lifecycle (ver. 3.x) are in use
  - DDI-Codebook: For codebook description. The latest version is 2.5
  - DDI-Lifecycle: Record metadata along with survey process – Data lifecycle model. Current version is 3.1

Data Lifecycle (DDI Alliance, 2009)
Data sharing, dissemination and metadata

- Metadata is used for
  - Finding data
    - Search data by topic, methodology, question, variable etc...
  - Understanding data
    - Data quality
    - Survey population
    - Question information
    - Variable definition

Benefits of DDI

- International standard
  - Metadata contents can be controlled
  - Facilitate data sharing with data archives
- Provide rich metadata
  - More than 500 elements are available (DDI-Lifecycle)
- Machine readable metadata
  - Efficient search (Question Bank, Variable Bank, etc...)
  - Associate micro data with metadata easily
    - Good feature for online analysis
- XML documents can be transformed to several documents
  - Questionnaire
  - CAPI, CATI script
  - Web page
  - Codebook
  - Setup syntax for various statistical package
Our current use of DDI

• Nesstar
  – View metadata online
  – Online analysis
    • Cross tabulation
    • Correlation analysis
    • Regression Analysis

• Easy DDI Organizer (EDO)
  – Metadata editor
  – Survey planning tool

Nesstar System
https://nesstar.iss.u-tokyo.ac.jp/webview

• You can
  – View metadata online
  – Analyze data online
    • Cross tabulation
    • Correlation analysis
    • Regression analysis

• Three datasets are available
  – Japanese General Social Survey - 2008 (Japanese ver, English ver)
  – Japanese Life Course Panel Survey (wave 1)
• We started operational test from Nov. 2012
Potential use of Nesstar for data sharing and dissemination in East Asia

• Integrated data catalogue with Nesstar might be possible
  – CESSDA (Council of European Social Science Data Archives) offers this service
  – We have the same platform to disseminate data now
  – Easier access to East Asian data

---

CESSDA Catalogue

One can search the holdings of 16 data archives at once through this integrated catalogue.
http://www.cessda.org/accessing/catalogue/index.html

---

The CESSDA Catalogue

- Provides a seamless interface to datasets from social science data archives across Europe
- May be searched via a free text search from the top search box
- Proceed via the options in the left hand side menu
- Can be viewed in any of five languages. The default language is dependent on the regional setting of the computer user
- The language can be switched at any time by selecting the relevant language from the drop down list above. Please note: changing this language will take the user back to the CESSDA catalogue home page

- On the free text search:
  1. "Enter search term"
  2. "Search" (for the entire catalogue)

---

CESSDA Data Publishers

- 1600 datasets available
- 400 (25%): SOR (25%)
- 018 (18%): SO (18%)
- 019 (19%): EURO (19%)
- 020 (20%): OECD (20%)
- 021 (21%): EU (21%)
- 022 (22%): UN (22%)
- 023 (23%): UNESCO (23%)
- 024 (24%): ILO (24%)
- 025 (25%): UNESCO (25%)
- 026 (26%): ILO (26%)
- 027 (27%): UNESCO (27%)
- 028 (28%): ILO (28%)
- 029 (29%): UNESCO (29%)
- 030 (30%): ILO (30%)
- 031 (31%): UNESCO (31%)
- 032 (32%): ILO (32%)
- 033 (33%): UNESCO (33%)
- 034 (34%): ILO (34%)
- 035 (35%): UNESCO (35%)
- 036 (36%): ILO (36%)
- 037 (37%): UNESCO (37%)
- 038 (38%): ILO (38%)
- 039 (39%): UNESCO (39%)
- 040 (40%): ILO (40%)
- 041 (41%): UNESCO (41%)
- 042 (42%): ILO (42%)
- 043 (43%): UNESCO (43%)
- 044 (44%): ILO (44%)
- 045 (45%): UNESCO (45%)
- 046 (46%): ILO (46%)
- 047 (47%): UNESCO (47%)
- 048 (48%): ILO (48%)
- 049 (49%): UNESCO (49%)
- 050 (50%): ILO (50%)
- 051 (51%): UNESCO (51%)
- 052 (52%): ILO (52%)
- 053 (53%): UNESCO (53%)
- 054 (54%): ILO (54%)
- 055 (55%): UNESCO (55%)
- 056 (56%): ILO (56%)
- 057 (57%): UNESCO (57%)
- 058 (58%): ILO (58%)
- 059 (59%): UNESCO (59%)
- 060 (60%): ILO (60%)
- 061 (61%): UNESCO (61%)
- 062 (62%): ILO (62%)
- 063 (63%): UNESCO (63%)
- 064 (64%): ILO (64%)
- 065 (65%): UNESCO (65%)
- 066 (66%): ILO (66%)
- 067 (67%): UNESCO (67%)
- 068 (68%): ILO (68%)
- 069 (69%): UNESCO (69%)
- 070 (70%): ILO (70%)
- 071 (71%): UNESCO (71%)
- 072 (72%): ILO (72%)
- 073 (73%): UNESCO (73%)
- 074 (74%): ILO (74%)
- 075 (75%): UNESCO (75%)
- 076 (76%): ILO (76%)
- 077 (77%): UNESCO (77%)
- 078 (78%): ILO (78%)
- 079 (79%): UNESCO (79%)
- 080 (80%): ILO (80%)
- 081 (81%): UNESCO (81%)
- 082 (82%): ILO (82%)
- 083 (83%): UNESCO (83%)
- 084 (84%): ILO (84%)
- 085 (85%): UNESCO (85%)
- 086 (86%): ILO (86%)
- 087 (87%): UNESCO (87%)
- 088 (88%): ILO (88%)
- 089 (89%): UNESCO (89%)
- 090 (90%): ILO (90%)
- 091 (91%): UNESCO (91%)
- 092 (92%): ILO (92%)
- 093 (93%): UNESCO (93%)
- 094 (94%): ILO (94%)
- 095 (95%): UNESCO (95%)
- 096 (96%): ILO (96%)
- 097 (97%): UNESCO (97%)
- 098 (98%): ILO (98%)
- 099 (99%): UNESCO (99%)
- 100 (100%): ILO (100%)

---

The CESSDA Catalogue

- Provides a seamless interface to datasets from social science data archives across Europe
- May be searched via a free text search from the top search box
- Proceed via the options in the left hand side menu
- Can be viewed in any of five languages. The default language is dependent on the regional setting of the computer user
- The language can be switched at any time by selecting the relevant language from the drop down list above. Please note: changing this language will take the user back to the CESSDA catalogue home page

- On the free text search:
  1. "Enter search term"
  2. "Search" (for the entire catalogue)
Easy DDI Organizer (EDO)

- Desktop application for managing metadata based on DDI-Lifecycle (ver. 3.1)
- Functions
  - Edit metadata
  - Import metadata from DDI document
  - Import variable information from data file of statistical package (SPSS)
  - export several documents (codebook/questionnaire/setup syntax for statistical packages/DDI)

Functions of EDO- Editing metadata

<table>
<thead>
<tr>
<th>Metadata Item</th>
<th>Corresponding module of the DDI3.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data lifecycle event record</td>
<td>Record events from starting study to data publishing</td>
</tr>
<tr>
<td>Study description</td>
<td>Back ground information of the study such as purpose, abstract, coverage</td>
</tr>
<tr>
<td>Data collection</td>
<td>Sampling procedure, survey method</td>
</tr>
<tr>
<td>Question design</td>
<td>concepts, question item, response domain, category, code</td>
</tr>
<tr>
<td>Logical/Physical data information</td>
<td>Variable information, Record Layout</td>
</tr>
<tr>
<td>Comparison</td>
<td>comparison of concept/variable between studies</td>
</tr>
</tbody>
</table>
Functions of EDO- Exporting Questionnaire

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 never read books.
   1. About one
   2. About two
   3. About three
   4. Four or more
   5. No answer

Q3. How often do you read the newspaper?
   1. Almost every day
   2. Several times a week
   3. Once a week or less
   4. Never
   5. 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. Ordering for information
   2. Internet shopping
   3. Internet banking
   4. Sending email
   5. No answer

Q5. Sending email by a cellular phone (including PHS)
   1. 4 times a day or more
   2. 3 times a day
   3. 2 times a day
   4. 1 time a day
   5. Several times a week
   6. Once a week or less
   7. Never
   8. No answer

Q6. Sending email by a PC
   1. 5 times a day or more
   2. 4 times a day
   3. 3 times a day
   4. 2 times a day
   5. Once a week or less
   6. Never

Functions of EDO- Exporting Codebook

Japanese General Social Surveys (JGSS-2008)
Codebook

Index
1. Abstract
2. Mode of Data Collection
   2.1. Overview
   2.2. Sampling Procedure
   2.3. Data Collection Phase Period
3.4. Mode of Data Collection
5. Validation

12

20
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

Our next steps

2012
- Nesstar
  - Continue operational test
- Easy DDI Organizer
  - Function improvement
  - User friendly question sequence manager
  - Questionnaire template
  - Function expansion
  - Import / export DDI-Codebook
  - Manage information about related materials

2013
- Nesstar
  - Start full-scale operation
  - Expand available data
- Easy DDI Organizer
  - Publish as an open source software
Acknowledgement

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

• I appreciate cooperation of Mr. Takeshi Iwanami, the president of epoch Ltd and Mr. Masaru Tsuda

• I also thank all project members and staffs of SSJDA

Thank you for your attention!
Any questions?

Contact:
Yuki Yonekura: yyonekura@iss.u-tokyo.ac.jp
Organization and Activities of KOSSDA

December 7, 2012

About KOSSDA

- Started as a non-profit social science library in 1983
- Added survey data archive in 2003
- Integrated into digital data and literature archive in 2006
- Provide integrated online service of data and literature on Korea
Major Activities

- Data and Library Services
- Research
- Education and Training
### Data Acquisition

- Both qualitative and quantitative data
- Acquired from both individuals and organizations
- Contact individual researchers either through collaboration with NRF or by analyzing journal articles
- Agreement with 98 research institutes (private, university and government), research companies, civic organizations

### Data Archiving

- Assessing data quality
- Data cleaning & error correction
- Metadata construction
- Protecting confidential information
- Uploading data and meta-data
### Data Holdings

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Acquisition</td>
<td>660</td>
<td>238</td>
<td>175</td>
<td>177</td>
<td>321</td>
<td>372</td>
<td>1,943</td>
</tr>
<tr>
<td>Archiving</td>
<td>458</td>
<td>262</td>
<td>144</td>
<td>135</td>
<td>142</td>
<td>209</td>
<td>1,350</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Qualitative Data</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acquisition</td>
<td>10</td>
<td>14</td>
<td>24</td>
<td>10</td>
<td>15</td>
<td>35</td>
<td>108</td>
</tr>
<tr>
<td>Archiving</td>
<td>25</td>
<td>28</td>
<td>23</td>
<td>20</td>
<td>20</td>
<td>28</td>
<td>144</td>
</tr>
</tbody>
</table>
Survey Data

Korean General Social Survey (10)
ISSP Annual Topical Module Survey (11)
Gallup Omnibus Survey (22)
Values & Attitudes (44)
Population (16)
Family & Gender (31)
Elderly (17)
Youth (105)
Education (46)
Inequality (20)
Labour & Employment (113)
Migrant Workers (25)
Labour Relations (26)
Organizational Management (50)

Business & Technology (18)
Industrial Safety & Health (42)
Media & Communication (65)
Cultural & Leisure Activities (63)
Deviance & Crime (100)
Juvenile Delinquency (29)
Criminal Justice & Judicial System (93)
Politics & Public Opinion (123)
Public Administration (73)
Social Order & Social Cohesion (22)
Health & Welfare (64)
Quality of Life (27)
Community & Urban Studies (72)
Others (24)

Statistical Tables and Qualitative Data

Social Trends (689)
Regional Statistics (58)

Economic Growth 1910-1945 (95)
Intl Comparative Statistics (ISSP) (469)

Population (16)
Local Community (40)
Political and Social Movement (30)
Poverty (12)

Overseas Korean (5)
Migrant Workers (5)
Social Research Documents (5)
Others (25)
Data Search

- Open to public
- Search & browse based on
  - subject
  - thesaurus
  - keyword
  - variable
Subject Areas for browse

Politics, Public Administration, and Law
Economy and Business Management
Society and Culture
Social Problems, Issues, and Welfare
Psychology and Education
Regional Studies

Data Download & Analysis

- Restricted to members
- Membership include individual and institutional
- Online application and download
- Online analysis of data using NESSTAR
- Online preview of qualitative data
English Service

- Free access for foreign researchers in 88 academic institutions provided with support from Korea Foundation
- Translation of 96 survey data (including questionnaire codebook) and 763 statistical tables has been completed
- Metadata for 1040 survey data and 30 qualitative data has been translated

Download and User statistic

<table>
<thead>
<tr>
<th>Registered Users</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Member</td>
<td>57</td>
<td>638</td>
<td>1,449</td>
<td>1,290</td>
<td>1,569</td>
<td>1,717</td>
<td>6,720</td>
</tr>
<tr>
<td>Non-member</td>
<td>572</td>
<td>524</td>
<td>364</td>
<td>220</td>
<td>360</td>
<td>312</td>
<td>2,352</td>
</tr>
<tr>
<td>Total</td>
<td>629</td>
<td>1,162</td>
<td>1,813</td>
<td>1,510</td>
<td>1,929</td>
<td>2,029</td>
<td>9,072</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data download</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>20'11</th>
<th>2012</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Downloads</td>
<td>58</td>
<td>134</td>
<td>729</td>
<td>955</td>
<td>1,412</td>
<td>1,139</td>
<td>4,427</td>
</tr>
<tr>
<td>No. of Data</td>
<td>167</td>
<td>290</td>
<td>1,210</td>
<td>1,862</td>
<td>2,435</td>
<td>2,275</td>
<td>8,239</td>
</tr>
</tbody>
</table>
Library holdings

- Books and articles from research on Korea
- Holds 52,000 research report, 41,000 books, 2,600 electronic and 1,500 printed academic journals
- Documentary delivery service provided to registered members

Research

- System development for data preservation and archiving
- Revision of Korean “social indicator” system under the contract with bureau of statistics
- Development and operation of data acquisition and dissemination system for Korea Research Foundation
Education and Training

- Summer/Winter workshop on data analysis for researchers and students (10 days)
- Spring/Fall short course on advanced statistical method (4 weekends)
- Introductory lecture on using KOSSDA system for data search

Organization and Governance

- Board of trustees
- Consortium of data depositing organization
- Executive committee
- Staff: 3 directors and 8 staff members in the division of planning & management, data development, and research & education
Recent development: K-Poll

- Modeled after I-Poll service of Roper center
- Expanding into new area of statistical information media
- Tables & graphs showing serial trend and international comparison of statistical information on hot issues
- Searching variable summary across various data sets
Introduction to Korea Social Science Data Archive

December 7, 2012

Goal: professionalization and standardization

Methods: literature reviews and case surveys

Background for Developing a Data Archiving System
Roles of data archives

- Data identification and collection
- Systematic management and long-term preservation of data
- Data dissemination
- Provision of information and knowledge related to data

Why archiving?

- To minimize the efforts and costs for data management and preservation
- To reuse the accumulated data
- To provide comprehensive information on data

Data archiving workflows

Source: IFDO (International Federation of Data Organizations for the social Sciences)
Research Project 2 - Digital Archiving

- Born digital vs. digitize
- Long-term preservation
- OAIS reference model
- Data curation
- Life-cycle model

Research Project 3 - Examples of Major Data Archives

- ICPSR (Inter-University Consortium for political and Social Research)
  http://www.icpsr.umich.edu/ICPSR/

- GESIS (German Social Science Infrastructure Services)
  http://www.gesis.org/

- ESDS (Economic and Social Data Service)
  http://www.esds.ac.uk/

- Roper Center
  http://www.ropercenter.uconn.edu/
### Research Project 4: Professionalization and Standardization (1)

#### International standards for metadata schema
- DDI (Data Documentation Initiative)
- EAD (Encoded Archival Description)
- ISAD(G) (General International Standard of Archival Description)
- MARC (Machine Readable Cataloging)
- DC (Dublin core)

#### Online analysis tools
- NESSTAR (Networked Social Science Tools and Resources)
- SDA (Survey Documentation and Analysis)
- Various commercial statistical packages (SPSS and domestic SW)

#### Subject service
- Social Science Subject Thesaurus
  - For example, UKDA HASSET, UNESCO Thesaurus, and Parliamentary Library Standard Thesaurus
- Social Science Variable Database

#### Other
- Data processing and preservation
- Data use and deposit

---

<table>
<thead>
<tr>
<th>Data preservation standards</th>
<th>ILOES (International Standard for Bibliographic Description)</th>
<th>ISADACDH (International Standard for Archival Description General)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data content standards</td>
<td>- Record for Archival Description (RAD)</td>
<td>- Record for Archival Description (RAD)</td>
</tr>
<tr>
<td></td>
<td>- Functional Record (FAR)</td>
<td>- Functional Record (FAR)</td>
</tr>
<tr>
<td></td>
<td>- Data Description Initiative (DDI)</td>
<td>- Data Description Initiative (DDI)</td>
</tr>
<tr>
<td></td>
<td>- Standard Format for Digital Objects (SFDI)</td>
<td>- Standard Format for Digital Objects (SFDI)</td>
</tr>
<tr>
<td>Data structure standards</td>
<td>- MARC (Machine Readable Cataloging)</td>
<td>- MARC (Machine Readable Cataloging)</td>
</tr>
<tr>
<td></td>
<td>- EAD (Encoded Archival Description)</td>
<td>- EAD (Encoded Archival Description)</td>
</tr>
<tr>
<td></td>
<td>- ISAD(G) (General International Standard of Archival Description)</td>
<td>- ISAD(G) (General International Standard of Archival Description)</td>
</tr>
<tr>
<td>Data value standards</td>
<td>- Medical Subject Headings (MSH)</td>
<td>- Medical Subject Headings (MSH)</td>
</tr>
<tr>
<td></td>
<td>- Learning of Objects (LOOM)</td>
<td>- Learning of Objects (LOOM)</td>
</tr>
<tr>
<td>Data models</td>
<td>- Virtual Research Association (VRA) Core Categories</td>
<td>- Virtual Research Association (VRA) Core Categories</td>
</tr>
</tbody>
</table>

Source: Metadata demystified, published by the National Information Standards Organization (NISO); Introduction to metadata: pathways to digital information, by the Getty Institute.
DDI (Data Documentation Initiative)

- DDI is an international standard for describing social science data which facilitates the sharing and comparison of data collections.
- Developed into two specifications: DDI-L (DDI-Lifecycle) and DDI-C (DDI-Codebook)

KOSSDA applied DDI 2.1 version

**Application of Metadata Standards for Data Management (1)**

DDI (Data Documentation Initiative)

- DDI is an international standard for describing social science data which facilitates the sharing and comparison of data collections.
- Developed into two specifications: DDI-L (DDI-Lifecycle) and DDI-C (DDI-Codebook)

KOSSDA applied DDI 2.1 version
DDI 2 Structure and Compatibility

- **KOSSDA**
  - **Application of Metadata Standards for Data Management (2)**
  - **Source:** DDI (Data Documentation Initiative)

  - Developed jointly by Norwegian Social Science Data Services (NSD), UK Data Archive (UKDA), and Danish Data Archive (DDA). It can search, analyze, and download data on the web without using a separate data analysis program.
  - It can provide DDI-based metadata
  - Current version: 4.0
  - **KOSSDA applied 3.5 version**

Introducing Online Statistical Analysis Tools for Data Search

**NESSTAR (Networked Social Science Tools and Resources)**

- Developed jointly by Norwegian Social Science Data Services (NSD), UK Data Archive (UKDA), and Danish Data Archive (DDA). It can search, analyze, and download data on the web without using a separate data analysis program.
- It can provide DDI-based metadata
- Current version: 4.0
- **KOSSDA applied 3.5 version**
OAIS (Open Archival Information System) Reference Model

- Developed by CCSDS (Consultative Committee for Space Data Systems) in NASA
- Used as an international reference for developing preservation metadata
- Functional digital archiving model for digital preservation systems

![ICPSR OAIS functional model application example](image)

Initiating the Development of Data Archiving system

- Supplementing an existing system:
  - Literature management
  - User management
  - Use, deposit

- Developing a new system:
  - Data management
  - Online statistical analysis
  - Variable management
  - Thesaurus

- 2007: KORENG website opened
- 2008
Developing Management Systems (1)

- Metadata Management System

Developing Management Systems (2)

- Online Statistical Analysis and Variables Management System
Developing Management Systems (3)

**Thesaurus management system**

- Integrated search for data and literature
- Browse by subject area
- Thesaurus search
- Survey data variable search
- Frequency analysis
- Cross tables and graph creation
- Correlation & regression analysis
- Qualitative data preview
- Data order & download
- Data deposit
- Document delivery service
- MY KOSSDA
Survey Data Variable Search

Developing User Systems (6)

KOSSDA
Survey Data Variable Search

Developing User Systems (7)

KOSSDA
Online data analysis
Developing User Systems (8)

Data Download

- Register
- Search
- Order
- (Approve)
- Download

Procedure for data download

Future Development Plan

Data curation
Life-cycle model
Upgrade

DDI
NESSTAR
Etc.

K-Poll service
KOSSDA
Open API

Long-term preservation

KOSSDA
Data Download

For research
For education
For browse

Procedure for data download
Thank you for your attention!

http://www.kossda.or.kr
The Practice of
The Survey Research Data
Archive in Taiwan

Ruoh-rong Yu
Research Fellow and Executive Director
Center for Survey Research
Research Center for Humanities and Social Sciences
Academia Sinica, TAIWAN

Development of Data Archives
in Taiwan
Data Archives in Taiwan: SRDA

- Survey Research Data Archive (SRDA)
  - Center for Survey Research (CSR), established in 1993, originally aimed to conduct surveys for researchers at Academia Sinica.
  - In 1994, CSR constructed the SRDA to preserve and disseminate the survey data collected.
  - Now has the largest collection of survey data in Taiwan.

Data Archives in Taiwan: Government

Some government organizations have their own archives, for instance,

- Directorate General of Budget, Accounting and Statistics (DGBAS), established since 1931
  - Sample surveys and Censuses conducted by DGBAS

- Collaboration Center for Health Information Application, established since 2011
  [http://biostat.cmu.edu.tw/~nhip/indexn.html](http://biostat.cmu.edu.tw/~nhip/indexn.html)
  - National Health Insurance administrative records
  - Registration records of Taiwan aborigines
Data Archives in Taiwan: Survey Projects

- Some large-scale survey projects maintain their own data archives, for example,
  - Taiwan Social Change Survey (TSCS)
  - Taiwan Integrated Postsecondary Education Database
    https://www.cher.ntnu.edu.tw/
  - Taiwan Election and Democratization Survey
    http://www.tedsnet.org/cubekm2/front/bin/home.phtml

Data Archives in Taiwan: Brief Summary

- In general, government organizations are more conservative with data dissemination than academic institutes or researchers
- Survey data, either conducted by governments or academic institutes, can be accessed more easily than administrative records
- Most academic survey data can be accessed easily without cost, while government archives usually require application and fee
- General attitudes toward data opening are getting more and more positive in Taiwan society
Data Policies in Taiwan

Data Policies: Government

- Open data is a hot issue for Taiwan government in recent years
- Besides central government, some local governments have setup open data platforms
- Collaboration Center for Health Information Application is important as the first formal data archive of registration and administrative records
- However, Personal Information Protection Act implemented since this October has strict regulation on collecting, processing and using of personal information
Data Policies: Research Funders and Associations

- National Science Council (NSC)
  - Require the survey projects subsidized by the Council should deposit their data in SRDA within a certain number of years
  - However, the requirement is not mandatory
- Academic associations of social sciences
  - Some associations have setup their ethical codes in recent years
  - These ethical codes are primarily related to data collection and data usage, instead of data dissemination

Introduction to Survey Research Data Archive (SRDA)
Survey Research Data Archive (SRDA)

- Established in 1994 by Center for Survey Research (CSR)
- CSR is one of the five research centers at the Research Center for Humanities and Social Sciences (RCHSS), Academia Sinica
- Academia Sinica
  - As the nation's premier research institution, is directly responsible to the President of Taiwan
  - It supports research in a wide variety of disciplines, ranging from sciences to humanities and social sciences
Current Staff of CSR

- Researchers
  - 5 research fellows
  - 2 adjunct research fellows
- Full-time workers
  - 25 full-time employees, with budget from Academia Sinica
  - 2 full-time research assistants, under research projects subsidized by National Science Council
- Part-time workers or outsourcing manpower
CSR Staff involved in Data Division

- Supervising researcher
- Head of division
- Working staff
  - 9 full-time workers
  - Within the 9 workers, 2 are under the NSC research projects
Data Acquisition and Processing of SRDA

Data Acquisition

- Mainly sample surveys
  - Surveys implemented by CSR
  - Surveys subsidized by NSC
  - Surveys conducted by government
  - Donations from researchers and research institutes
- Censuses
  - Administrative district-level statistics
  - Collaborated with Center for GIS at RCHSS
    - Demographic Statistics since 1946
    - Urban and Regional Development Statistics since 1975
- In-house processed data
  - Merged adjacent Manpower Utilization Surveys
  - Combined Taiwan Social Change Surveys
Sources of Government Surveys

- Directorate General of Budget, Accounting and Statistics
- Ministry of the Interior
- Ministry of Transportation and Communication
- Ministry of Education
- Department of Health
- Council of Labor Affairs
- Research, Development and Evaluation Commission
- Judicial Yuan

~ Hope to persuade more government organizations to authorize SRDA to disseminate their data

Role of SRDA

Academic Surveys → SRDA → Data Users

Government Surveys → SRDA → Data Users
Data Processing within SRDA

- Check for risk of disclosure, and decide whether to release or not
- For releasable data,
  - Clean data and delete sensitive information
  - Prepare required metadata
  - Create data files with different formats (e.g., SAS, SPSS, STATA)
- Downloadable files include raw data (e.g., SPSS, STATA, SAS), questionnaire, codebook, user's guide, etc.
Number of Datasets Archived by SRDA

Based on the statistics from November 2012,

- Academic sample surveys
  - Acquired data: 1,379
  - Released data: 1,112
- Government sample surveys
  - Acquired data: 447
  - Released data: 402
- Government censuses
  - Acquired & released: 33

Academic Survey Data: By Fields

- Sociology 25%
- Political Science 8%
- Economics 9%
- Management 9%
- Others 6%
- Public Health 3%
- Regional Studies 10%
- Education 11%
- Art 2%
- Psychology 7%
- Others 6%
Government Survey Data: By Organizations

- Department of Budget, Accounting and Statistics (DGBAS) 68%
- Ministry of Interior (MOI) 9%
- Council of Labor Affairs 3%
- Research, Development and Evaluation Commission 6%
- Department of Health 2%
- Ministry of Transportation and Communication 4%
- Judicial Yuan 1%
- Censuses of DGBAS 7%
- Censuses of MOI 0.01%

Data Dissimilation of SRDA

Survey Research Data Archive, Center for Survey Research. RCHSS, Academia Sinica
Membership

- Two-year membership
  - Academic researchers, college faculty, researchers from government organizations
- One-year membership
  - Graduate or college students, research assistants at colleges or research institutes
- Academia Sinica membership
  - Research fellows at Academia Sinica
- Free membership

Data Access Channels

By confidentiality of data,
- Data directly downloadable from SRDA website
- Data with restricted access
  - Access channels
    - On-site access
    - Remote access
  - Existing restricted data
    - Taiwan Education Panel Survey (TEPS)
    - Taiwan Youth Project (TYP)
Application Procedure of Data Usage

- **Non-restricted academic surveys**
  - Once a member logs in, can download data from website

- **Non-restricted government surveys**
  - A member has to submit a brief research plan first
  - Submitted documents reviewed by working staff
  - Once approved, can download data from website

- **Restricted surveys**
  - A member has to submit a filled application form
  - Submitted documents reviewed by researchers
  - Once approved, a set of username & password is provided to the applicant
  - Qualified users can use data via restricted access

---

**On-site Access**

Facilities at on-site room
Remote Access

Internet  \[\rightarrow\]  VPN  \[\rightarrow\]  Intranet

- Users
- Server with Statistical Softwares Available

Recent Data Usage Situation

- At the end of this October,
  - There were 1,334 effective members
- From last July till this October,
  - Accumulated number of downloads for non-restricted data is 13,345
- From last June till this September,
  - Accumulated number of restricted-data users is 365
Services Provided on SRDA Website

User-friendly Interface

- Questionnaires and other metadata can be downloaded by both members and non-members
- Non-restricted data can be easily downloaded by members
- On-line statistical analyses can be conducted via Nesstar, with some advanced functions restricted to members
- Survey data or question items can be searched by multiple ways
- Publications related to specific data can be browsed or searched
Search Functions with SRDA

- For academic surveys, users can
  - Search specific data by PI’s name, project title, abstract, or keywords
  - Search specific data or question item by “research concepts”
  - Browse the list of data of a specific research discipline
- For government surveys, users can
  - Search specific survey data

Search tools on the SRDA website

Can search specific data by the survey project’s PI, title, abstract, keywords, etc.
Search tools on the SRDA website

Can search specific data or question item by “research concepts”

Can browse the list of data of a specific research discipline
Search tools on the SRDA website

Can search specific survey data conducted by government

Publications Based on SRDA Data

- How is the list of publication constructed?
  - For any member who had been access to a dataset, SRDA staff shall send a brief letter for possible publications
  - SRDA staff would send periodic questionnaires to members for suggestions and publications
  - SRDA staff would actively search for publications of some vital surveys
Some Advanced Tasks in Progress

- Develop videos of specific data or analytical tool for potential users or new members
- Adopt information security management system (ISMS), and has acquired ISO 27001 since May 2010
- Re-process some vital datasets to make them more user-friendly
- Construct English-version website and data files

Promotion Activities
Workshops

- In recent years, we held workshops on topics:
  - Data processing and management
  - Confidential data
  - Panel data
  - Usage of specific survey data
  - Analytical methods
  - Analytical tools

Relevant Publications

- Journal of Survey Research
  - Biannual academic journal
  - On survey methodology and data application
- SRDA Newsletter
  - Published quarterly
- SRDA E-digest
  - Published biweekly
International Activities

International activities in recent years,
- Invite researchers from ICPSR and Netherlands Bureau of Statistics as workshop speakers
- Keep contact with some survey and data centers in China (e.g., Beijing University, Renmin University, Chinese Academy of Social Sciences) and other areas
- Attend IASSIST Annual Conference
- Attend CODATA and other international conferences

Welcome to visit the Survey Research Data Archive (SRDA) website!
https://srda.sinica.edu.tw/