Hannele Keckman-Koivuniemi and
Mari Kleemola:
Data Processing in FSD:
CHALLENGES IN A NEW ARCHIVE

IASSIST2003
Ottawa, May 2003

fsd@uta.fi
http://www.fsd.uta.fi
General

• FSD established in 1999
• About 400 archived studies
• One third of them are international
• FSD receives circa 100 new studies annually
Pre-Requirements

- Data belongs to some area in the field of social sciences
- Aspects of copyright and ownership are clear, no legal impediments
- Original purpose, information content and technical properties make data suitable for archiving
Beginning the Process

- Depositor delivers:
  - Machine-readable data (mostly SPSS or Excel)
  - Supplementary material (questionnaire, coding instructions, collection procedure, publications)
  - Completed and signed deposit forms (material description and material deposit agreement)
Working on the Data (1)

- FSD processes all studies INTENSIVELY in the same way
  - time-consuming
- Objective: Content of the archived data corresponds closely to the original questionnaire
Working on the Data (2)

- Data are checked by using SPSS and preserved in SPSS portable format:
  - Data and identification information (labels) in the same file
  - Converting possible in the future
  - Customers use it, too

- All modifications are documented in the SPSS syntax
Checking the Data (1)

- FSD makes a copy of the original data
- Number of variables and cases are reviewed
- FSD checks frequencies to verify variables and values - both valid and invalid
Checking the Data (2)

- Variables are renamed (q1, q2_1, q2_2...) and labels are created corresponding to questionnaire
- Filter variables are examined
- Some variables may have to be dropped or added
Data Protection

- Depositors are supposed to delete personal data before delivery
- Sometimes FSD stores materials which contain personal data
- FSD anonymizes the data according to its own and depositor’s instructions
Versions

- FSD aims to process datasets once and for all
- Some of them have to be reprocessed
- First final: version 1.0
- Cosmetic changes: version 1.1
- Significant changes: version 2.0
Documentation

- Finally the dataset is described
- FSD uses DDI
- Metadata are both in Finnish and English, codebooks in Finnish
  - Available on the Internet
- NESSTAR, Madiera
Challenges for the Future (1)

• Introducing processing levels
  – what is the minimum level that enables preservation

• Getting supplementary documentation
  – lack of sufficient documentation delays the archiving process

• Version control
  – what to do with the original dataset and how to document changes
Challenges for the Future (2)

- CATI and CAPI versus survey
- Downloading the data through online-system
- Long term preservation of electronic documents
  - we are not alone here!
Contact Information

Finnish Social Science
Data Archive

FIN-33014
UNIVERSITY OF TAMPERE
FINLAND

Tel. +358 3 215 8519
Fax +358 3 215 8520

http://www.fsd.uta.fi/

hannele.keckman-koivuniemi@uta.fi
mari.kleemola@uta.fi