Strategies of Promoting the Use of Survey Research Data Archive

Meng-Li Yang

Center for Survey Research, RCHSS, Academia Sinica
Background

• Survey Research Data Archive (SRDA) of Center for Survey Research (CSR) archives survey datasets deposited by the academics and government agencies.
• Maintained by the Data Division of CSR.
• Promotion activities: newsletters and workshops on specific popular datasets or data cleaning skills.
Search in SRDA

• Search is separated by sources of data.
• Power of Nesstar is underutilized because most datasets are not moved to it and several of its functions are not open for use due to SRDA administrative reasons.
• Search criteria are limited to descriptions of datasets.
• Metadata of datasets completely rely on depositors to supply.
Survey of SRDA use in Oct. 2011

• Sample: those who ever submitted grant proposals in humanities and social sciences fields to National Science Council (NSC) over the last five years.

• A first survey ever, as an additional small set of questions to a main topic commissioned by the NSC.
Results of the survey-1

• 52% of the 3019 R’s did not hear of SRDA.
• Among the 257 R’s who ever tried but gave up,
  -17%: metadata provided insufficient for effective search;
  -10%: difficult to find needed data with the search criteria.
Results of the survey-2

• Among the 295 one-time members,
  - 17%: did not know how to find what they needed.
  - 16%: could not find what they needed.

Conclusions:
  1. need to promote SRDA.
  2. need to improve its search effectiveness.
To promote SRDA

1. Put video/audio tapes of workshops on SRDA websites.

2. Create demonstrations of small analyses on relatively unfamiliar datasets to answer simple research questions.

3. Create power point files introducing SRDA and research articles using survey data. Send to college faculties and put on the web before the fall semester begins.
To improve search effectiveness-1

1. Revise the abstracts and keywords of datasets.
2. Define concepts for each variable; quality is to be controlled by checking consistencies among items with identical elements of words.
3. Move all datasets to Nesstar to make use of the attributes of variables stored in it.
To improve search effectiveness-2

4. Construct a new search engine where:
   - all datasets can be searched at a time.
   - concepts and words of a question item are added to the search criteria.
   - up to 3 words/concepts can be used to search within a dataset or within a variable.
   - search results can be searched further.
Current progress and evaluation-1

- The work of revising abstracts and keywords was formidable.
  - Need staff with training in social sciences.
- Demonstrations of small analyses are made for seven datasets by my part-time research assistants, but still wait for me to check.
- Introductions of research articles are completed also by my assistant, but about half of them also wait for me to review and revise.
Current progress and evaluation-2

- A data cleaning workshop in March was videotaped and put on the website. More workshops and videotapes are expected.
- Staff needs to learn to appreciate the value of providing such information on the web.
- The moving of datasets to Nesstar is still going on.
Current progress and evaluation-3

- Concept defining is commissioned to experts (researchers, professors).
- Similar variables with seemingly different concepts are found and sent to me for judgments.
- Identical variables in other datasets are given the same, so that fewer and fewer items will need to be given concepts.
Current progress and evaluation-4

• Concepts are completed for about 10 datasets by now.
• Speed is expected to pick up as fewer and fewer items will have to rely on experts.
  - The construction of concepts cannot stop if it is to be useful because new datasets will always come in.
Current progress and evaluation-5

• The new search engine is almost completed. In addition to the old search criteria, we now can either search within datasets or search within items.

• The refined search function is to be completed around July.

- Real data users are needed in the planning of improvement projects for effective changes to take place.
Conclusion

• Staff members with trainings in social sciences are needed in handling data archives; they can help with construction of metadata of datasets and variables.

• Users’ active participation is needed in planning and designing archives in order for the archived datasets to have the highest likelihood of being re-used.