A review of NSD's access to social science research data from Statistics Norway

Trond Pedersen

Norwegian Social Science Data Services

Gothenburg, 16. – 17. September 2010

Introduction

- Statistics Norway (SN):
  - Administers a vast amount of national registers and survey data
  - World leading with respect to the amount of register data
  - Obvious source for research data
- NSD has had a long and good relationship with SN since the 70's
- The data transfers regulated by contract
- NSD acts as one of SN's tools for dispersion of research data

NSDs main purpose

- Work for better access to research data for students and researchers
- Contribute to better knowledge on what kind of accessible data is out there (through NSD's web pages and the Nesstar tool)
- Be of service to students and researchers in their process of data collection

Statistics Norway as a data provider

- The data quantity and quality makes SN an ideal source for access to research data
- A lot of obstacles to overcome
  - Formalities
  - Main task is production of statistics, not serving researchers
  - Time-consuming
  - Economical issues
- Researchers, especially those connected with research institutes, typically have a longer time horizon, are familiar with the searching procedures, and can afford the prices SN charge => not a problem
- The obstacles incompatible with students needs (in a hurry, not able to pay for data, little knowledge about formalities) => turn to NSD

NSD compared to SN

NSD has a complementary role to that of SN

- Has close dialog with the academies, and good knowledge about their needs
- Generally more focus on the needs of researchers and students (SN has more focus on statistics production)
- Offer data free of charge (unlike SN)
- Efficient data delivery (may not be the case for SN)
- The typical SN client is a researcher, while NSD serves both researchers and students

Cooperation on 3 types of data

- Large surveys (LFS, Living Conditions Survey etc)
- Regional data
- Register data

Survey data

- Serving of researchers with survey data has been delegated to NSD by SN since the 70's
- Established routines
- Remote access to files through data server
- Large amount of data: Most of the important surveys produced by SN, including the complete collection of standard LFS and the Living Conditions Surveys
- Access to anonymized micro data requires filling out a web form
- Direct downloading or online analysis applies to some data sets
- Users: Limited to researchers and students
- Statistics: Ca. 500 surveys delivered per year

Regional data

- Tailor made tables have been regularly transferred from SN since the early 70's

- Organized in NSD's Regional Database

- Frequently update of data

- Aggregate level: Municipality and urban district

- 320 000 variables, and time series covering the last 240 years (covering censuses, labour market, health, social security benefits etc)

- Direct access by logging in to NSD's database since 2001 (requires registration)

- 1500-2500 downloads per year

- Users: Limited to employees and students at the academies, and for research purposes only

- Little flexibility when adapting to the researchers needs – could be better if NSD were aloud to make own adaptations on micro data

Register data

- Good access until recent years
  - KIRUT
  - FD-Trygd
- Change in practice has led to poor accessibility especially amongst students
- Currently only two kinds of register data in NSD's possession, of which none of them comes from SN:
  - Kontantstøttedatabasen – the Cash Benefit for Parents of Infants Database

    (Source: NAV - the Norwegian Labour and Welfare Administration)
  - Fastlegedatabasen – the Family Doctor Database (Source: the Norwegian Directorate of Health)

The KIRUT period

- KIRUT – "Client Streams in to, within and out of the Welfare System and the Labour Market"
    - Established in the early 90's
    - Covers clients streams in and out of a wide range of social security benefit arrangements
    - 10% of the population aged 16-67
    - 1989-1996
    - A unique and important data source for social security benefit research
- Advantages:
    - Rich data (many variables, longitudinal data, large population sample)
    - Free of charge
    - Established routines
    - Well documented

The KIRUT period II

- Statistics:
  - KIRUT-data delivered to approximately 100 research projects from the beginning to 2000
  - 1/3 student projects (master's degree)
- No more updates after 1996
  - Redundant due to SN's new events database, FD-Trygd
  - No longer willing to serve KIRUT with new updates
  - The use of KIRUT continued, despite the lack of updates: During the period January-March 2000, NSD executed on average one order per week

The FD-Trygd period

- FD-Trygd – SN's events database
- A further development of KIRUT
- Established in 1999
- A large longitudinal events database covering the streams in to, out of and within most of the Norwegian social security benefit arrangements
- Covers the whole population
- Time period: 1992-
- Input data: A large number of administrative registers, mostly from NAV (the Norwegian Labour and Welfare Administration)
- Managed and maintained by SN
- Free of charge for many years (subsidised by NFR – the Norwegian Research Council)

The FD-Trygd period II

- Relatively good cooperation:
  - NSD aided researchers and students in the application process
  - SN delivered anonymized data to through NSD
  - NSD received un-anonymized data for adaptation and anonymization before delivery in some cases
- Statistics
  - 2001: 6 data applications received by NSD
  - 2002: 12 data applications received by NSD

Years with poor register data availability

- Due to the success of FD-Trygd, SN's delivery capacity became stretched to the limit around 2005 => longer waiting to get data access (NSD experienced waiting time up to 5-6 months
- Tried to come around the problem by delegating data adaptation and anonymization to NSD in 2006
- Eventually, SN realised that the NFR subsidies were no longer in accordance with the resources spent
- SN's solution: Payment introduced in 2007!
- To get around the economical and delivery time issues, students started reusing data across research projects
  - 5 requests in 2008, one in 2009 and two so far in 2010
  - Not optimal, but the best option available
  - Free of charge
  - Quick access to data

The prosess of bettering the register data service

- NSD took the initiative to build up a new data service, based upon the KIRUT-modell, and direct negotiations started in April 2008

- Informal discussions had been going on for some time, prior to 2008 (from February 2006)

- A draft of data specification were formed the same year, in cooperation with an interdisciplinary group of researchers at the Bergen University

- Agreement to go for a 20% random sample of SN's FD-Trygd, and more or less most of the central variables and social security benefit areas in the database

- Long period of meetings back and forth between NSD and SN followed, with the leaderships highly involved

- Mixed signals have been sent, and compromises have been made to secure the anonymity of the data

The prosess of bettering the register data service II

A new data service finally started to take form at the beginning of 2010

- Agreement on data specifications
- Agreement on how to take care of privacy protection (anonymization procedures)
- Agreed to transfer un-anonymized micro data, so that NSD can adapt, anonymize and supply students (and researchers)
- Possible through a data management contract with specific anonymization instructions attached
- However, SN has to make some adjustments to secure a certain level of anonymity: Data on income, wealth, place of residence etc has to be censured before the data transfer to NSD

The prosess of bettering the register data service III

Last development:

- NSD given data access via a remote server
- Data will soon be ready, but for internal use only (control, documentation etc)
- NSD can not supply students and researchers with anonymized micro data until a final contract is ready
- The last obstacle is the finalization of the anonymization routines
- Data supply to students and researchers may be possible within a couple of months

Summary

- Long and good relationship with SN
- "Rome not built in a day"
- When all the formalities and routines are in order, NSD will be able to serve particularly students in an excellent way (good data quality, efficient delivery, and free of charge)
- Things are looking good:
  - NFR are focused on bettering the data availability and have in this respect announced large funding for suitable projects
  - NSD and SN currently working on a joint application - the goal of the project is to establish an efficient data delivery system