

Linking SHARE with Administrative Register Data in Denmark

Paul Bingley

Outline

- Registration & legislation
- Register data for research use
 - What is available
 - How to obtain permissions
 - How to work with the data
- Linking survey with register data
 - Our experience with a SHARE linkage
 - Some generalizations
- How does the information compare
 - Income
 - Schooling
 - Retirement

Registration & Legislation

- 1968 Central Person Registration (CPR)
 - Unique 10 digits for each resident
 - First used as key in 1970 Census
 - Since 1980 throughout public administration
- 1987 Public Administration Act
 - A public authority can impose the duty of non-disclosure on non-public servants
 - Disclosure/anonymity/confidentiality with breach punishable by imprisonment
 - Statistics Denmark thereby allows researcher access
- 2000 Act on Processing of Personal Data
 - Duty of notification to data protection agency before collection/use
 - Approves disclosure of data for scientific purposes
 - Applies whether using a register or collecting a survey
 - Private individuals & public agencies can process CPRs for scientific work
 - Subject's right of access to personal data does not apply for scientific work

Statistics Denmark - documentation

- “*high quality data*” + subsets of many registers + ministries have full sets + other registers here & there + surveys + own data
- www.dst.dk > products & services > data for research > documentation > high quality variables
- employment / *population* / personal finance / transfer programs / workplaces / business accounts / schooling / health & justice / housing
- *Families* / fertility / population / households / migration / foreigners
- # *children* / partner ID / X-family ID / X-family status / X-family type
- Name / descriptive name / period / breaks / quality assurer / general description / detailed description / population / distribution plot / values

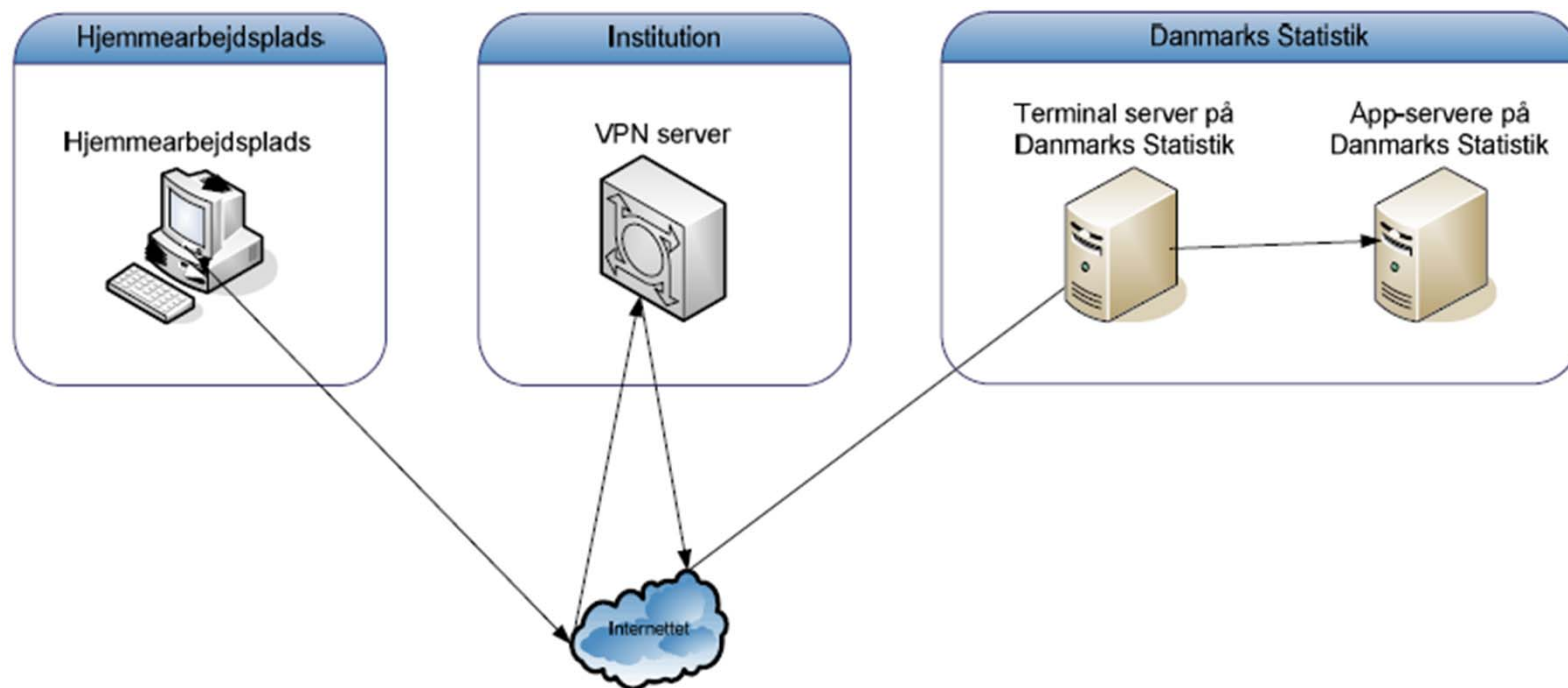
Data access conditions

- 1986 Statistics Denmark on-site access for researchers
 - “Need to know” principle for access to anonymized data based on a project description
 - Trade-off of # variables & sample size
 - Researcher signs confidentiality agreement
- 1995 Statistics Denmark Århus office opens
- 2001 Own workplace access once the analysis environment authorized
- 2003 Access by Danes from abroad via Danish analysis environment authorization
- 2007 Foreigner access from abroad via the same
- 2012 status – 200+ environments & 1500+ researchers

Obtaining permission

- Work for / with an approved analysis environment
- Join a project / write a project description
- Design / negotiate / purchase dataset construction
- Sign a confidentiality agreement
- Install VPN client + receive text message or security token look-up
- Remote desktop access with "printout" via email

Remote Desktop Access from home



Computing

- 7 Servers - STATA1 + Gauss + SPSS + SAS*3 + whatever
- STATA1 Hardware
 - # processors: 4
 - frequency 2.29 GHz
 - # cores: 4
 - memory: 512 GB
- Operating System
 - Windows Server® 2008 R2 Enterprise,
 - Microsoft Windows Server, Version 6.1 (7601, Service Pack 1)
 - x64-based (64-bit)
- Software
 - STATA/MP4 11.2 for Windows 64-bit x86-64
 - STATA/MP4 12.1 for Windows 64-bit x86-64
 - GNU Emacs, version 23.1.1
 - Stat/Transfer (grafisk), Version 11.2.2010.0305
 - WinZip®, version 11.0 (7313)

Non-DST registers

- National Board of Health (SST)
 - Other registers & greater detail
 - More detailed project description required
 - Will export CPRs
 - Prescription drug purchase & psychiatry
- Outside DST & SST
 - Not covered by their data protection agency umbrella agreement
 - Separate data protection agency approval needed

Our SHARE linkage experience

- Waves #1 & #2 with DST registers until 2006
 - Extended initial data protection permission with linkage to labour market research register & national hospital register
 - SFI survey / CentER data / DST – SHARE IDs / CPRs
 - Validity study extension of existing DST project
- SHARE-life & wave #4 with registers until 2011
 - SFI umbrella agreement coverage
 - New validity study – now with cancer & prescriptions
- www.sfi.dk/cssr - centre for survey & survey/register data

Validity study

- Bingley & Martinello (2012)
- Individual linked records measurement error
- Link 1670 of 1707 respondents from 2004 wave
- Schooling
 - SHARE overestimates % with long schooling
 - Especially for those with short schooling & high income
- Labour market status is precise
- Gross household income
 - On average not different in SHARE & registers
 - Classical measurement error