Access to microdata at Statistics Denmark – Data security in practice

Research Services

DANMARKS STATISTIK

Data security –the background

- Statistics Denmark (DST) provides access to microdata via Research Service because it benefits our society
- Data is confidential and therefore there are rules for accessing and managing data on the research machines that scientists must know before accessing data
- The requirements for security and confidentiality are the same as before, but we have sharpened the focus on compliance with the rules
  - Increased focus on the output researchers send home
  - Tightened sanctions for research environments that break the rules
    - Closing down institutions that send home microdata
    - Increased response of detailed tables send home
  - Follow-up on who is using the system

Data security –general rules

- Individual data is never handed out. They are behind DST's walls, separated from DST's production system
- Only access to data via secure encrypted connection
- Access only through an authorized institution
- Formal agreements must be signed
- Data provided by "need to know" principle
- Data anonymized - unique key for each project
- "Id" and password is strictly personal (and must not be passed over to others)

Data security –general rules

- All work with data takes place on the research servers at DST. Aggregated output can only be send home via a mail server, where all output is logged
- **It is allowed to send home aggregated results but not individual data (microdata)**
- FSE takes samples of output checking that data security rules are followed

What can be send home?

- All forms of:
  - Analytical results
  - Aggregate tables
  - Charts etc.

- **where identification of any individual and/or company is impossible**

- All cells in tables, etc. should contain at least 3 observations
  - A minimum requirement that researchers can increase

What must not be send home?

1. Micordata/ individual data

- Individual data most never be sent home. They must remain at DST.
- What is microdata?
  - Information specific on individuals / companies. This means any individual "records" or data sets with information on the individual level
  - Key Variable as person-numbers, CVR-numbers, address codes, etc. is always regard as individual data and must never be sent home, also if they are anonymized.

What must not be send home?

- Sending home microdata is regarded as a serious data breach. If it happens, we close down data access for the researcher that has committed the breach and for all researchers from the institution that owns the project from where the breach has happened.

- Be careful with program files and logs
  - There are logs that lists a sample of microdata
  - Programming at the individual level contains microdata

What must not be send home- examples - Microdata!

PNR	KØN	INDKOMST	BOPÆL	ALDER	PNRB
1	M	1500000	København	40	15
2	M	1500	Aarhus	45	22
3	K	1000	Bornholm	50	37
4	M	500	Bornholm	50	65
4	M	250	København	40	87
5	M	150	Aarhus	20	
6	K	150	Odense	25	
7	M	750	Herning	35	19
8	K	500	København	60	74
9	M	250	Odense	65	74
10	K	150	Aarhus	20	
11	K	50	Bornholm	15	
12	M	0	København	10	

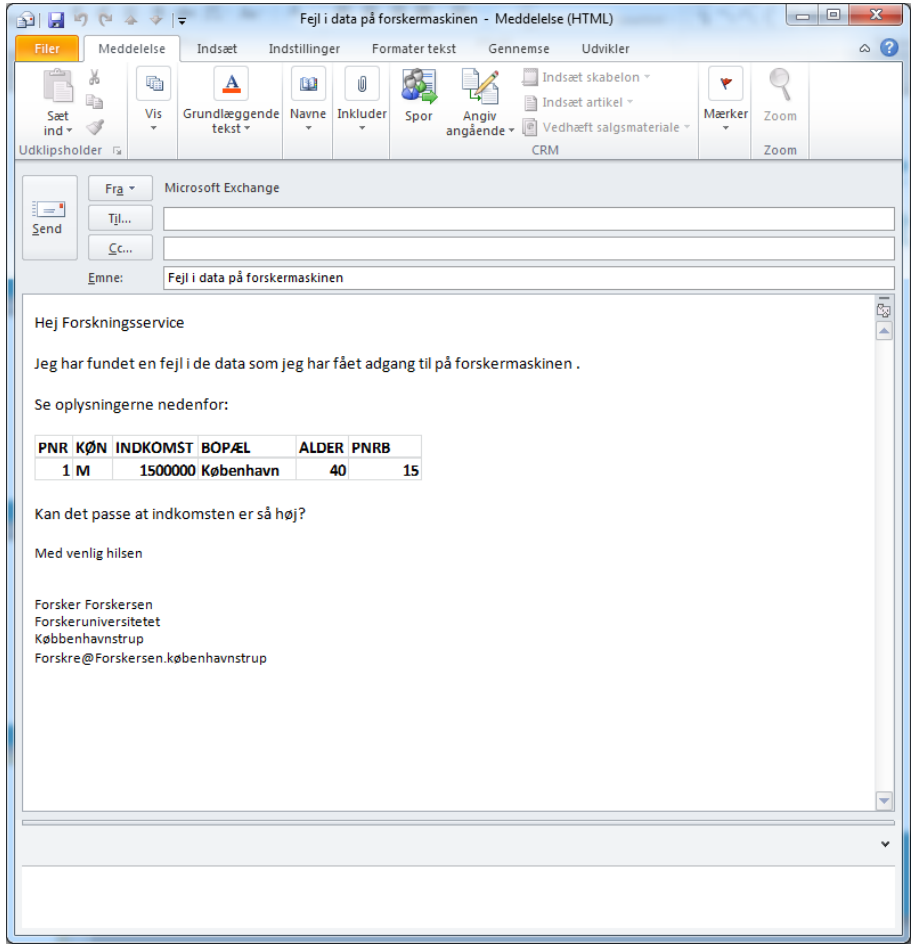
What must not be send home-examples - Microdata!

KØN	INDKOMST	BOPÆL	ALDER
M	1500000	København	40
M	1500	Aarhus	45
K	1000	Bornholm	50
M	500	Bornholm	50
M	250	København	40
M	150	Aarhus	20
K	150	Odense	25
M	750	Herning	35
K	500	København	60
M	250	Odense	65
K	150	Aarhus	20
K	50	Bornholm	15
M	0	København	10

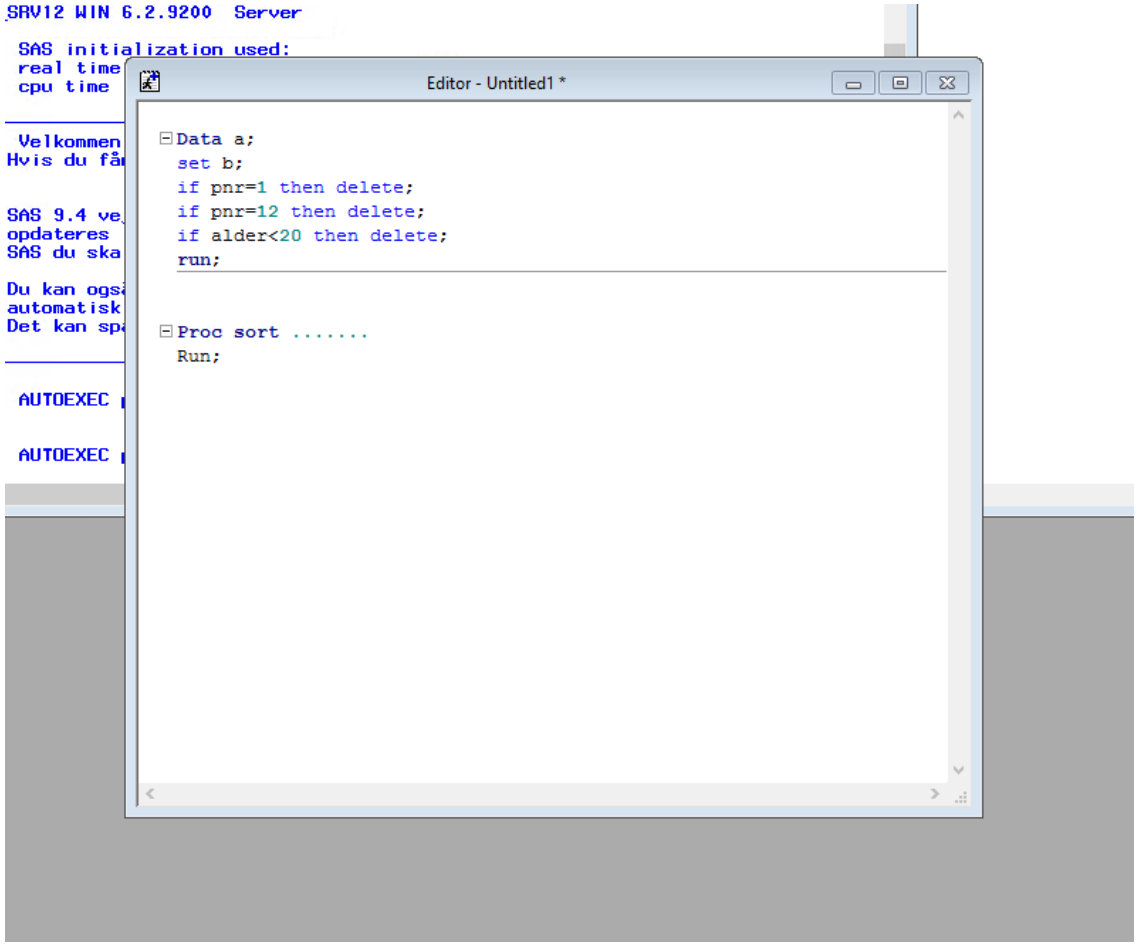
What must not be send home- examples - Microdata!

PNR
1
2
3
4
4
5
6
7
8
9
10
11
12

What must not be send home- examples - Microdata- mails, containing microdata !-



What must not be send home- examples – programs containing microdata !



What should not be send home?

2. Analytical results, aggregated tables, etc.

- Analytical results, aggregate tables, figures, etc. where there are a risk of identifying individuals or companies. That is, tables, etc. with few observations in each cell (less than 3)
- We consider sending home aggregated data with few observations in each cell less serious than in the case of sending home microdata
- There may be situations where it is OK to send home cells containing less than 3 obs. (Avg./max./min.)
- Encourage the researchers to use their common sense. If there a risk of identification – also with more than 3 obs. in the cells - then the output should not be send home

What should not be send home?

- We respond when detailed tables are send home, for example, if there are many cells with one observation - but - we **don't** close the institution down if we see that data has been aggregated
- Here we normally contact the institution and advice/inform them about the rules

If we discover a breach (microdata send home)

- We close access from all researchers from the institution that owns the project where data breach has occurred
  - also if it is an affiliated researcher from another institution who has committed the breach or a foreign researcher
- We contact the institution (researcher and responsible)
  - Institution must send a report of the data breach
  - Plan for future prevention of data breach
- Statistics Denmark's directors if informed and decide the sanction
- Institution is notified of the sanction

What can the institution do if they commit a data breach?

- Notify always as soon as possible Statistics Denmark about the breach
- Describe in the mail when data is sent home, and how many files (the extent of the breach).
- Access to micro data will be closed and we need a report of the breach and a plan for future prevention
- Informed quickly about a data breach is considered as a mitigating factor and can reduce the time the institution is closed down

Good advice for researcher

- For the researcher, who has access to microdata:

- 1. Always check all files to be sent home
  - Always contact the Research Services in case of doubt
- 2. Limit files to be send home to a minimum
  - a. Is it, for example, necessary to send home programs and logs? - stored safely at DST
  - b. Is it necessary send home preliminary results??

Good advice for the responsible of the institution

- For the responsible of the institution:

- 1. Prepare a plan to ensure that all researchers are aware of the security rules for access to data through the research arrangement
  - a. Including both the last incoming researcher and the foreign researcher
  - b. Have a contact person in the environment case of doubt
- 2. All researchers knows the procedure in case of a data breach
- 3. All researchers knows the security ”traps” involved in working with microdata – e.g. that standard output can contain microdata
- 4. Make control measures when sending home output – should another researcher also check the output?
- 5. Make sure that access is regularly updated so access is closed for researchers that don’t need it anymore

New system for sending home output in use

- Limited number of file-types can be send home
- Can close for the individual researchers access to send home output
- Automatic control of files checking for micodata - in test
- This year new research machine with word processing programs, so the report can also be written on the research machine

Output control – a demonstration

- Receive approx. 5000 output files files each week
- Check a random sample
- Each employee responsible for a file format