
DST SECURITY AND USE

VER 1.4

DST SECURITY AND USE

- ▶ Introduction to the use of DST
- ▶ Application for access
- ▶ **THE PERSONAL ACCESS**
- ▶ How do I log on?
- ▶ How do I download information?
- ▶ **WHAT ARE THE RULES CONCERNING DOWNLOADING OF INFORMATION?**
- ▶ Breach, sanctions and consequences

INTRODUCTION TO USE OF DST

- ▶ **Danmarks Statistik (DST) - www.dst.dk**
- ▶ Researchers can obtain access to information in registries with personal data at DST
- ▶ Researchers can also obtain access to information from other registries via DST e.g.:
 - Seruminstittet (SSI) -> **Health Data Board** – All merge of variables requires permission!
 - National Registry of Patients (NRP)
 - The joint unit for quality development for general practitioners and the regions (DAK-E)
 - Department of Clinical Epidemiology (LABKA)
 - and others
- ▶ Requires specific permissions
- ▶ Special access through **DST researcher computers**

APPLICATION FOR ACCESS

- ▶ **If you have not completed this course, there is no access!**
- ▶ The Head of Department is responsible for ALL agreements made with DST
- ▶ Project description
- ▶ Permissions from at least the Danish Data Protection Agency
- ▶ Additional permissions from external registries (not DST data)
- ▶ Applications to DST:
 - Number from Danish Data Protection Agency, project description, protocol etc.
 - <http://www.dst.dk/da/TilSalg/Forskningsservice>
- ▶ **You can obtain help to getting started from the Data Management Group**

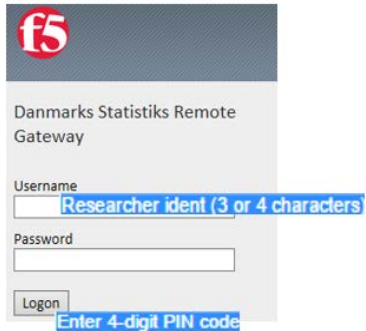
PERSONAL ACCESS

- ▶ **When you have obtained an access, it is personal!**
- Nobody else can use your access
- Nobody is allowed to look over your shoulder
- You are not permitted to show your data to others
- ▶ **Data confidentiality**

**Individual-level data are confidential, see
The Public Administration Act § 27, section 3
the Penal Code §152.**

HOW DO I LOG ON?

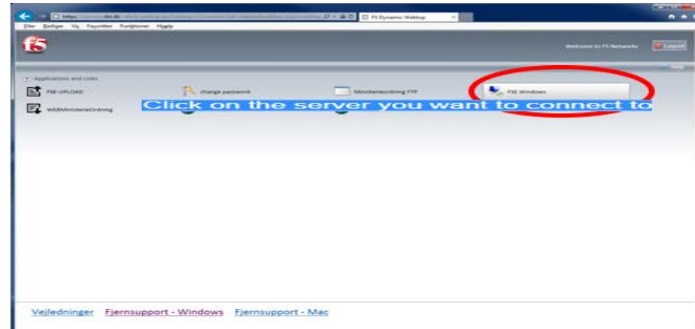
- ▶ When permissions are obtained, you will be registered as user at **DST researcher service**
- ▶ This gives you access to DST research computers
- but only concerning your own project.



The screenshot shows the login interface for the f5 service. It features the f5 logo at the top left. Below it, the text "Danmarks Statistiks Remote Gateway" is displayed. There are two input fields: "Username" and "Password". A blue callout box points to the Username field with the text "Researcher ident (3 or 4 characters)". Below the Password field, another blue callout box says "Enter 4-digit PIN code". A "Logon" button is located at the bottom left.



This screenshot shows the next step in the login process. It features the f5 logo at the top left. The main text reads "Enter your one-time password." Below this is a password input field with a blue callout box that says "SMS with a one-time password". A "Logon" button is positioned at the bottom left.



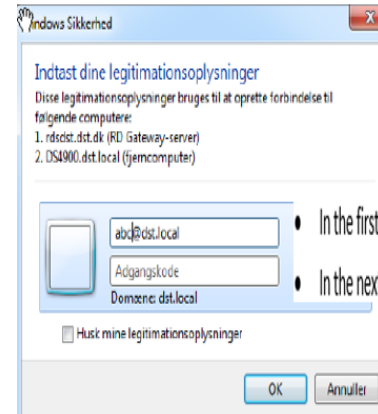
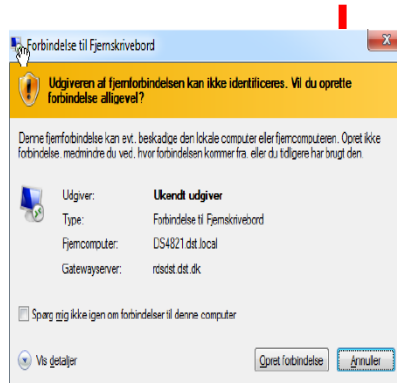
HOW DO I LOG ON?

- ▶ After this slide we are almost half way through the login procedure...
- this is called **data security!**

Windows

- Klik på den server du vil forbindes til

Herefter vil der nederst fremkomme nedenstående i browseren



HOW DO I LOG ON (WINDOWS)?

Once you have completed your work on the server, log on to this by clicking on Logout



HOW DO I LOG ON (MAC)?

The instructions can be obtained here:

<http://http://dst.dk/ext/2418449283/0/forskning/Forbindelse-til-forskerservere-i-Danmarks-Statistik-Mac-brugere---pdf>

HOW DO I DOWNLOAD INFORMATION?

- ▶ How do I download information?
- ▶ **This is where problems can occur...**
- ▶ and **NO**, you cannot download information!
- ▶ That is why you work at DST research computers.
- ▶ That is why you have to apply for all permissions.
- ▶ That is why you are personally responsible.

WHAT ARE THE RULES?

- ▶ **Quote from DST "Download of files from research computers.pdf":**
- ▶ *1. Files may not contain **identifiable data**, i.e. data containing single records or data where companies or individuals may be identified.*
- ▶ *2. The file may not contain e.g. **one variable from one observation** from a set of raw data no matter the content of this variable.*
- ▶ *3. Counts, tables, output, operation logs, programmes and graphic **files** may be transferred if the data does not contain identifiable data.*

WHAT ARE THE RULES?

- ▶ **Rule no. 1 "Identifiable data"**
- ▶ No complete records, or an observation, can be transferred from a set of raw data
- ▶ Neither a reduced record with information that can identify a person or a company.
- ▶ Example: A violation could be e.g. information on: "A company with a turnover of 20 billion DKK in the Municipality of Bjerringbro". In this instance two variables are enough – turnover and municipality. Everybody knows it is about the company Grundfos.

WHAT ARE THE RULES?

- ▶ **Rule no. 2. "one variable from one observation "**
- ▶ It actually states that not a single number from a set of raw data must be downloaded!
- ▶ Regardless of the value of the number
- ▶ I.e. just one number fra a set of raw data.
- ▶ **You really have to be careful here!**
- ▶ Example: In the example with Grundfos we had the value of one variable of "more than 20 billion DKK". If we had the exact figure for the turnover from raw data, e.g. 22.6 billion DKK, which was the turnover in 2012, then this one number would be a violation of the rules!

WHAT ARE THE RULES?

- ▶ **Rule no. 3 "File types"**
- ▶ It sounds as if the types of files mentioned can be downloaded
- if they do not contain identifiable data.
- ▶ **However, be careful!**
- ▶ In a do-file, it may say "drop if pnr == 1234" - UPS! A value! **A violation!**
- ▶ Log-files from Stata may contain far more information than you would think!
- and more information than you have seen on the screen!

WHAT ARE THE RULES?

- ▶ **Rule no. 3 "File types"**
- ▶ It could be tempting to plot a variable and download the plot to get an overview.
- ▶ In this case significant values outside the ?? plagemage of many values contain identifiable dataa!
- ▶ It could e.g. be an item plot of the turnover for all companies in Bjerringbro, where one large company stands out from the others – as one point!
- and in this way it becomes identifiable data!

WHAT IS PERMITTED THEN?

- ▶ You are permitted to download **aggregated data**:
 - I.e. numbers in data tables have to be an average of 5 numbers/cells? In raw data.
 - I.e. **graphical** items cannot be identifiable down to at most 5 cells? On average in raw data.
 - **Do-files** must be cleaned for data which is identifiable down to less than or at most 5 cells? in raw data.
 - **Log-files** must be cleaned for data which is identifiable down to less than an average of 5 cells? In raw data.
 - Or data which **in another way** can identify a person or a company.

WHAT IS PERMITTED THEN?

- ▶ Can I download Min, Max, Percentiles and Median values?
- ▶ No, you cannot!
(written answer from DST)

- ▶ May I refer to "how many" if the number is below 3 and all are below 80% of the sum.
(previously 5)
- ▶ No, you cannot!
(written answer from DST)

WHAT IS PERMITTED THEN?

- ▶ Demands to **special variables**:
- ▶ Never download pnr numbers and similar.
 - The identifialbe civil registration numbers called pnr, recnum, famid etc.
- ▶ Never download key numbers
 - i.e. database keys, as record key values or observation index.
- ▶ In other words, it is more than difficult to download information.

WHAT IS ALLOWED THEN?



In other words, you can only download finished

Results and documentation

- if they comply with the demands mentioned previously!

BE CAREFUL WITH LOG-FILES

- ▶ If you e.g. use STATA, be careful with commands such as:

```
. codebook, . codebook, compact  
. summarize  
. list
```

These commands may write **individual cell values?, min and max** in the log file!

- ▶ Stop Log: `. log close`
- ▶ Start Log: `. Log using mylogfile.log`

OTHER PROBLEMS

Other problems:

- ▶ If e.g. your supervisor is not approved, he/she must not look over your shoulder!
- ▶ Do not make work copies or bring identifiable data to meetings.
- ▶ Only work in the closed world of DST researcher computers.
- ▶ Never e.g. zip the information you download.
- ▶ Never download "large amounts" of files at the same time.

What you download is controlled by automatic scanners – and spot checks!

GOOD SOFTWARE PRACTICE

We share the DST server resources!

- ▶ STATA allocates a lot of RAM
- ▶ SAS allocates a lot of disc space
- ▶ If you start the programmes more than once, you use more licenses!
- ▶ If you start the programmes more than once, you use more resources!
- ▶ Remember to close the programmes!
- ▶ Temporary files are only deleted when you log-off!

Others can work better when you are not logged on!

BREACH, SANCTIONS AND CONSEQUENCES

- ▶ **ONE** violation shuts **EVERYBODY** down at the Department for an indefinite length of time
 - to open again requires the board in DST to meet

- ▶ The **individual** person responsible for this will also get:
 - First time: 3 months quarantine
 - Second time: 12 months quarantine
 - Third time: Permanent exclusion via PH.

- ▶ Conscious actions to acquire data:
 - **Report to the police - see the Public Administration Act § 27, section 3 and the Penal code §152**

THANK YOU FOR YOUR PARTICIPATION

- ▶ If you have any doubts, please contact:
- ▶ DST researcher services
- ▶ Or get counselling as provided by the Data Management Group.



AARHUS
UNIVERSITET