DST SECURITY AND APPLICATION

NIELS-SØREN BØGH
DST SIKKERHED OG ANVENDELSE

AARHUS
UNIVERSITET

DST SECURITY AND APPLICATION

▶ Introduction to using DST

▶ Application for access

▶ THE PERSONAL ACCESS

▶ How do I log in?

▶ How to download information from DST?

▶ WHAT ARE THE RULES FOR DOWNLOAD OF INFORMATION?

▶ Violations, sanctions and consequences

AARHUS
UNIVERSITET

NIELS-SØREN BØGH
DST SIKKERHED OG ANVENDELSE

DST SIKKERHED OG ANVENDELSE
10. DECEMBER 2014

INTRODUCTION TO USE OF DST

▶ **Statistics Denmark (DST) -** www.dst.dk

▶ Researchers can obtain access to information in sensitive personal records in DST

▶ Researchers can also access information from other registers through DST, eg .:

- Serum Institute (SSI)
- National Patient Register (NPR)
- They practitioner doctors and regions common unit for quality development. (DAK-E)
- Clinical Epidemiology Chamber (LABKA)
- Etc.

▶ Requires special permits

▶ Special access via a service called **DST's Researcher Machines**

APPLICATION FOR ACCESS

▶ If you do not complete this course, then there is **NO** access!

▶ Søren K. Kjærgaard is responsible for ALL agreements concluded with DST

▶ Project Description

▶ Permits from at least **The Danish Data Protection Agency** (Datatilsynet)

▶ In addition, you need permits from external registers also (not DST data)

▶ Applications for DST:

· Datatilsyn Number, Project Description, Protocol etc.

· http://www.dst.dk/da/TilSalg/Forskningsservice

▶ **You can get help to get started from the Data Management Group**

THE PERSONAL ACCESS

▶ **Once you have got access then it is personal!**

▪ No one else can use your access

▪ No one else can look over your shoulder

▪ Do not show your data to others

▶ **Data Privacy**

Individual Data is confidential pursuant.
The Administration Law § 27.3
The Criminal Law §152.

AARHUS
UNIVERSITET

NIELS-SØREN BØGH
DST SIKKERHED OG ANVENDELSE

DST SIKKERHED OG ANVENDELSE
10. DECEMBER 2014

HOW DO I LOG IN?

▶ Once authorizations are in place, then you have an account on **DST Research Services**

▶ This gives you access to DSTs Researcher Machines
 - But only for your own project.
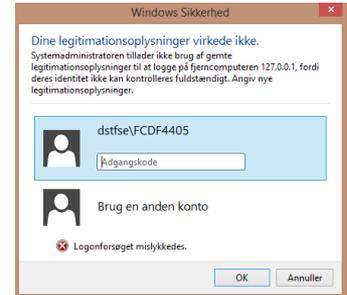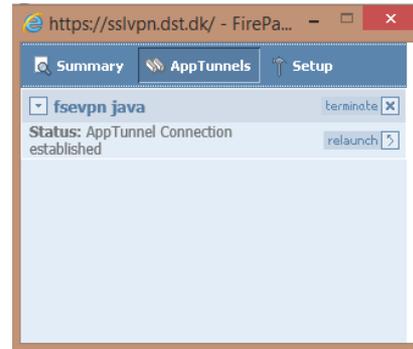
HOW DO I LOG IN?

▶ After this slide, we are almost halfway in the login procedure ....
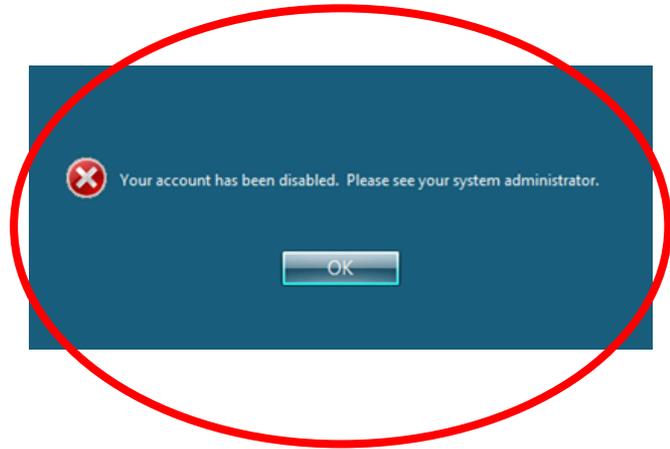- This is called **Data Security!**

AARHUS
UNIVERSITET

NIELS-SØREN BØGH
DST SIKKERHED OG ANVENDELSE

DST SIKKERHED OG ANVENDELSE
10. DECEMBER 2014

HOW DO I LOG IN?

▶ Here should we have talked about the rest of login procedure ...

▶ However, one of the Department of Public Health employees have in the meantime **unproved** violated **one small spot** in the DST rules!

▶ This closed the **ENTIRE** Institute access to DST down **indefinitely!**

▶ **These rules must you hear about today!**

AARHUS
UNIVERSITET

NIELS-SØREN BØGH
DST SIKKERHED OG ANVENDELSE

DST SIKKERHED OG ANVENDELSE
10. DECEMBER 2014

HOW TO DOWNLOAD INFORMATION?

▶ How do I download information?

▶ **This is where all the problems can occur ...**

▶ And **NO** you must not take down the information!

▶ That's why you're working on DSTS researcher machines.

▶ That's why you should apply for all permits.

▶ It is, therefore, that you are personally responsible.

AARHUS
UNIVERSITET

NIELS-SØREN BØGH
DST SIKKERHED OG ANVENDELSE

DST SIKKERHED OG ANVENDELSE
10. DECEMBER 2014

WHAT SAYS THE RULES?

▶ **Quoted from DST "Return-of-files-from-forskermaskiner.pdf":**

▶ *1. Files must not contain **identifiable data**, e.g. data containg single record from raw data or information from which companies or persons otherwise are identifiable.*

▶ *2. Therefore the file can not contain **one variable from one observation** from raw data, no matter whatever this variable contains.*

▶ *3. Counts, Tables, Outputs, Logs, Programs and Graphics **Files** are allowed to be downloaded. All under the condition that they do not contain identifiable data as described above.*

WHAT SAY THE RULES?

▶ **Rule # 1. "Identifiable Data"**

▶ No single records (an observation) must bee taken home from raw data.

▶ Or a reduced record with information that can identify an individual or a company.

▶ Example: A violation could, for instance be an information that says: "A company with a turnover of over 20 billion. DKr. in the city of Bjerringbro". Here are two variables enough; turnover and city. Then all can locate Grundfos as the case for this example.

AARHUS
UNIVERSITET

NIELS-SØREN BØGH
DST SIKKERHED OG ANVENDELSE

DST SIKKERHED OG ANVENDELSE
10. DECEMBER 2014

WHAT SAY THE RULES?

▶ **Rule # 2. "***One variable from one observation* "

▶ There is actually; that not even one single value of a raw data cell may bee downloaded!

▶ Whatever the value of the cell are.

▶ That is just one number from a raw dataset.

▶ **Here you have to be really careful!**

▶ Example: In the example with Grundfos, we had one variables value set to "more than 20 billion. kr. ". If we had the exact figures for the turnover from raw data, e.g. 22.6 billion. kr., as turnover was in 2012, then a single figure will be a violation of the rules!

AARHUS
UNIVERSITET

NIELS-SØREN BØGH
DST SIKKERHED OG ANVENDELSE

DST SIKKERHED OG ANVENDELSE
10. DECEMBER 2014

WHAT SAY THE RULES?

▶ **Rule # 3. "*File Types* "**

▶ This sounds like that these file types are allowed to download
- if they do not contain identifiable data.

▶ **But, here you have to be careful!**

▶ In a STATA do-file one line can stand "**drop if pnr == 1234** "
- UPS! A cell value! **A violation!**

▶ Logs from STATA can contain much more information than you think!
 - And more information than you have seen on the screen!

AARHUS
UNIVERSITET

NIELS-SØREN BØGH
DST SIKKERHED OG ANVENDELSE

DST SIKKERHED OG ANVENDELSE
10. DECEMBER 2014

WHAT SAY THE RULES?

▶ **Rule # 3. "***File Types* **"**

▶ One could easily be tempted to plot a variable, and take home the plot, to get an overview.

▶ Here some significant values can act as identifiable data!

▶ This could, for instance be a point-plot of the turnover of all companies in Bjerringbro, where a large company stands out significant from all the others - **one point!**
- and thus becomes the identifiable data!

AARHUS
UNIVERSITET

NIELS-SØREN BØGH
DST SIKKERHED OG ANVENDELSE

DST SIKKERHED OG ANVENDELSE
10. DECEMBER 2014

WHAT SHOULD YOU DO?

▶ Downloads must contain **aggregated data** only:

▪ That is, **data cells** has to to be at least an average of 5 cells in the raw data.

▪ That is, **graphic** points may not be identifiable down to less than the average 5 cells in raw data.

▪ **Do-files** must be cleaned of identifiable data down to less than the average of 5 cells in raw data.

▪ **Log-files** must be cleared of identifiable data down to less than the average of 5 cells in raw data.

▪ Or that the data **otherwise** can identify a person or a company indirectly.

AARHUS
UNIVERSITET

NIELS-SØREN BØGH
DST SIKKERHED OG ANVENDELSE

DST SIKKERHED OG ANVENDELSE
10. DECEMBER 2014

WHAT SHOULD YOU DO?

▶ Can I download **Min, Max and Median** values?

▶ No, do not!
(Written answers from DST)

▶ May I refer to the "how many" if there is less than 5?

▶ No, do not!
(Written answers from DST)

AARHUS
UNIVERSITET

NIELS-SØREN BØGH
DST SIKKERHED OG ANVENDELSE

DST SIKKERHED OG ANVENDELSE
10. DECEMBER 2014

WHAT SHOULD YOU DO?

▶ Requirements for **special variables**:

▶ Never download pnr's and the like.
▪ Those are "none" identifiable social security numbers called pnr, recnum, famid etc.

▶ Never download record key numbers
▪ That database keys, record key values or observation index.

▶ In other words, it is more than difficult to download information legally.

WHAT SHOULD YOU DO?

▶

In other words, you must only take down only

Aggregated Results and Documentation

- Provided that they are comply within the aforesaid requirements!

AARHUS
UNIVERSITET

NIELS-SØREN BØGH
DST SIKKERHED OG ANVENDELSE

DST SIKKERHED OG ANVENDELSE
10. DECEMBER 2014

BE CAREFUL WITH LOG-FILES

▶ For example. using STATA, so be careful with commands like:

  ”codebook”
  ”codebook, compact”
  ”summarize”
  ”list”

  These commands can write **individual cell values, min and max** into the log-file!

▶ Stop Log: log close
▶ Start Log: log using mylogfile.txt

AARHUS
UNIVERSITET

NIELS-SØREN BØGH
DST SIKKERHED OG ANVENDELSE

DST SIKKERHED OG ANVENDELSE
10. DECEMBER 2014

OTHER ISSUES

Other problems you run into:

▶ If for example your supervisor is not approved for the project, so they may not look with!

▶ Do not make copies for external work or bring identifiable data to meetings.

▶ Work only within the closed world of DST's Researcher Machines.

▶ Do never zip (compress) the data you download - It is not allowed.

▶ Never download "large quantities" of files at a time.

What you download are controlled by automatic scanners - and samples are inspected!

AARHUS
UNIVERSITET

NIELS-SØREN BØGH
DST SIKKERHED OG ANVENDELSE

DST SIKKERHED OG ANVENDELSE
10. DECEMBER 2014

VIOLATIONS, SANCTIONS AND CONSEQUENCES

▶ **ONE VIOLATION** will close the entire department down indefinitely
- Reopening requires a board meeting in DST

▶ **Besides this the violator will get:**

▪ 1st time: 3 months. quarantine

▪ 2nd time: 12 months. quarantine

▪ 3rd time: Permanent exclusion from the agreement.

▶ Deliberate actions to acquire indentifiable data will cause:

▪ **Police Involvement due to Administration Law § 27.3 and The Criminal Law §152**

AARHUS
UNIVERSITET

NIELS-SØREN BØGH
DST SIKKERHED OG ANVENDELSE

DST SIKKERHED OG ANVENDELSE
10. DECEMBER 2014

THANK YOU FOR YOUR PARTICIPATION

▶ If in doubt, contact:

▶ DST Research Services

▶ Or get guidance from the Data Management Group.

AARHUS
UNIVERSITET

NIELS-SØREN BØGH
DST SIKKERHED OG ANVENDELSE

DST SIKKERHED OG ANVENDELSE
10. DECEMBER 2014