
DST SIKKERHED OG ANVENDELSE

VERSION 2.7 – 2024 – KURSUS FOR MEDARBEJDER VED INSTITUT FOR FOLKESUNDHED VED AARHUS UNIVERSITET – UNDER AUTORISATION 50

DST SIKKERHED OG ANVENDELSE

- ▶ HVAD ER REGLERNE FOR NEDTAGNING AF INFORMATIONER?
- ▶ Eksempler og værktøjer
- ▶ Kontrol af data før hjemtagning
- ▶ Brud, sanktioner og konsekvenser
- ▶ Hvordan hjemtager jeg analyseresultater?
- ▶ Introduktion til brug af DST
- ▶ Ansøgning om adgang til DST
- ▶ DEN PERSONLIGE ADGANG
- ▶ Hvordan kommer jeg på?

HVAD SIGER REGLERNE?

- ▶ Citeret fra DST "Hjemsendelse-af-filer-fra-forskermaskiner.pdf":
 - ▶ 1. Filerne må ikke indeholde *identificerbare data*, dvs. data der indeholder enkeltrekords, eller hvor virksomheder eller personer direkte eller på anden måde er identificerbare.
 - ▶ 2. Filen må derfor fx ikke indeholde *én variabel fra én observation* fra et rådatasæt, lige meget hvad denne variabel indeholder.
 - ▶ 3. Optællinger, tabeller, output, kørselslogs, programmer og grafik *filer*, må gerne hjemsendes. Alt sammen under forudsætning af, at de ikke indeholder identificerbare data.

HVAD SIGER REGLERNE?

Regel nr. 1. "Identificerbare data"

- ▶ Ingen hele records/rækker, altså **en observation**, må hjemtages fra et rådatasæt.
- ▶ Eller en reduceret record/række med informationer, som kan identificere en person eller et firma direkte eller indirekte.
- ▶ **Eksempel:** En overtrædelse kunne f.eks. være en information, der siger: "Et firma med en omsæt på over 20 mia. kr. i Bjerringbro". Her er to variable nok, *omsætning og by*. Så ved alle, at det i dette eksempel drejer sig om Grundfos.

HVAD SIGER REGLERNE?

Regel nr. 2. "*én variabel fra én observation*"

- ▶ Der står faktisk; at ikke så meget som **én celle**værdi fra et rådatasæt må hjemtages!
- ▶ Uanset, hvad værdien af cellen er.
- ▶ Dvs. bare ét tal fra ét rådatasæt!
- ▶ **Her skal man virkelig passe på!**
- ▶ Eksempel: I eksemplet med Grundfos, havde vi den ene variables værdi sat til "over 20 mia. kr.". Hvis vi nu havde det **præcise tal** for omsætning fra rådata, f.eks. 22,6 mia. kr., som omsætningen var i 2012, så ville dette ene tal være en overtrædelse af reglerne!

HVAD SIGER REGLERNE?

Regel nr. 3. "Filtyper"

- ▶ Dette lyder som om, at de nævnte filtyper gerne må hjemsendes
- ▶ Men, her skal man også passe på!
- ▶ Tilladte filtyper til hjemtagning
- ▶ Forudsat at de ikke indeholder Overtrædelser af Regel nr. 1 og Regel nr. 2



Filtype	Filekstension
Excel	xls xlsx xlsxm xlm xml
Grafik	eps png wmf tif jpg gif emf jpeg svg bmp ppt pptx pptm tiff
HTML	htm html mht
Programkode	sas r do doh ado
Logs	Log
SAS	lst sas7bdat
SPSS	sps sav spv
Stata	dta smcl
Tekst	tab txt
Latex	Tex
PDF	pdf
Word	doc docx rtf

HVAD SIGER REGLERNE?

Eksempler på ulovligt indhold i filer:

- ▶ I en do-fil kan der f.eks. stå "*drop if pnr == 1234*", hvilket er en **databasenøgle værdi**
- ▶ **Tællinger** som viser at der indgår mindre end 5 personer
- ▶ Specifikke **datoer og lokationer** for events
- ▶ **Logfiler** fra Stata kan indeholde langt mere information, end man lige tænker over!
– og mere information, end man har set på skærmen!
- ▶ **Plot af en variabel** (signifikant værdier, "haler" med få personer, begge akser! osv)
- ▶ **Signifikant værdi** (en uden for plagemage værdier, indeholder identificerbar data!)
- ▶ **Tekstvariable** med unikke oplysninger

HVAD SIGER REGLERNE?

- ▶ Krav til **særlige variable**:
- ▶ Der må aldrig nedtages pnr-numre og lign.
 - De af identificerbare cpr-numre kaldet pnr, recnum, famid etc.
- ▶ Der må aldrig nedtages nøgle-numre
 - Dvs. databasenøgler (keys), som nøgleværdier eller observations indeks.
- ▶ **Det er med andre ord mere end vanskeligt at hjemtage informationer!**

HVAD MÅ MAN SÅ?

Der må hjemtages **aggregeret data**:

- Dvs. at tal i **datatabeller** mindst skal være et gennemsnit af 5 celler i rådata.
- Dvs. at **grafiske** punkter ikke må være identificerbare ned til mindre end gennemsnit af 5 celler i rådata.
- **Datoer** skal angives som måned, ugenummer, relativ tid etc.
- **Do-filer** skal være rensed for data, der er identificerbare ned til mindre end gennemsnit af 5 celler i rådata.
- **Log-filer** skal være rensed for data, der er identificerbare ned til mindre end gennemsnit af 5 celler i rådata.
- Eller at data ikke **på anden måde** kan identificere en person eller en virksomhed.
- **Kaplan-Meier Plots** skal bruge en relativ tid, f.eks. "antal dage/uger sinde start" som x-akse.

HVAD MÅ MAN SÅ?

- ▶ Må jeg hjemtage **Min, Max, Percentiler og Median** værdier?

Nej, det må du ikke!
(Skriftlig svar fra DST)

- ▶ I særlige tilfælde accepteres et gennemsnit af 3 celleværdier (i stedet for 5), hvis:
Alle 3 værdier hver i sær udgør mindre end 80% af summen af de 3 værdier?

I særlige tilfælde ja – Men, vi har valgt mindre end 5 som udgangspunkt!
- dette skyldes, at vi ellers skal kontrollere ovenstående ibm. hver eneste værdi.

PSEUDO PERCENTILES

- ▶ Niels Henrik Bruun har lavet nogle værktøjer, som ligger på **DST og VDI**
- ▶ Her regnes **min, max, medianer, persentiler...** som et gennemsnit af 5 værdier omkring den oprindelige værdi
- ▶ Ligeledes ligger der værktøjer til at lave base-table, hvor værdier der er mindre end gennemsnittet af 5 automatisk er fjernet.
- ▶ http://www.bruunisejs.dk/StataHacks/Datamanagement/pseudo_percentiles/pseudo_percentiles/#pseudo-percentiles

HAV STYR PÅ DINE PLOTS!

Vi har en vinder!



Mulige udfald

Individuelle Personer

Alle værdier kommer
fra 1 person !!!

Scores



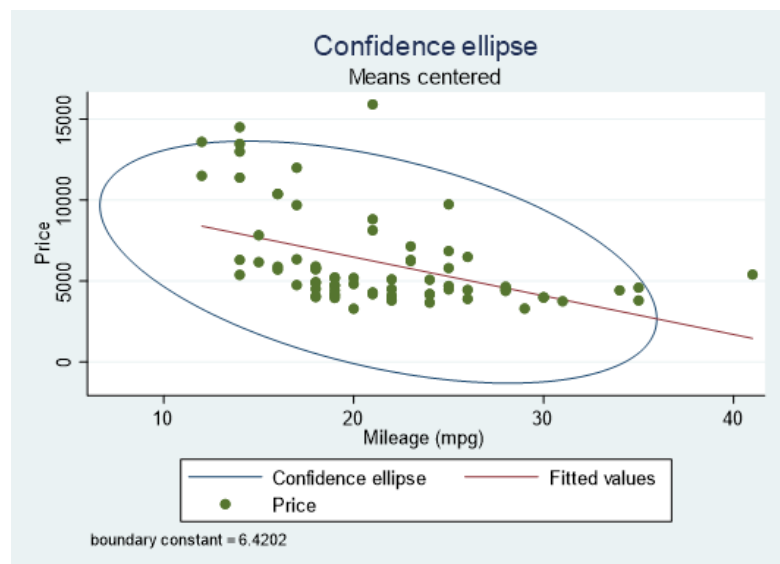
AARHUS
UNIVERSITET

NIELS-SØREN BØGH
DST SIKKERHED OG ANVENDELSE

DST SIKKERHED OG ANVENDELSE
28. JANUAR 2020

HAV STYR PÅ DINE PLOTS!

► Scatter Plots – no go!



Install this feature:

<http://www.haghigh.com/statistics/stata-blog/stata-programming/download/ellip.html>

It is done like this (search and install): *findit gr32_1*

installs ellip

Then it can be done like this, on the test database in Stata about cars (Auto.dta):

cls

sysuse auto, clear

ellip price mpg, means constant(pt2) plot(lfit price mpg)

ellip price mpg, means constant(pt2) plot(lfit price mpg || scatter price mpg)

Du må ikke plotte punkterne, da de er mikrodata!

Men du kan lave en statistisk repræsentation, der giver den samme informationer som et scatter-plot.

MISSINGS?

- ▶ Krav til **særlige værdier**:
- ▶ **Missings** [.]
- ▶ Missings, altså manglede værdier, bliver på DST også betragtet som en værdi
- ▶ Der må derfor heller ikke rapporteres et antal missings, mindre end 5
- ▶ Hvis antallet af rækker med missing er <5 skal disse tages ud af undersøgelsen
- ▶ *Tip: Nogle vælger derfor at erstatte få missings med gennemsnittet eller medianen af de øvrige værdier.*

PAS PÅ MED LOG-FILER

- ▶ Hvis du f.eks. bruger STATA, så pas på med kommandoer som:

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

Disse kommandoer kan skrive **enkelte celleværdier, min og max** ned i logfilen!

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

ANDRE PROBLEMATIKKER

Andre problematikker:

Kommunalt ansatte snagede i 300 personers følsomme oplysninger

To ansatte i Horsens Kommune er blevet afskediget, og sagen er meldt til Datatilsynet og politiet

- ▶ Hvis f.eks. din vejleder ikke er godkendt på projektet, så må de ikke kikke med!
- ▶ Lad være med at lave arbejdskopier eller medbringe identificerbare data til møder.
- ▶ Arbejd kun inde i den lukkede verden på DSTs Forskermaskiner.
- ▶ Du må aldrig f.eks. zippe det, du nedtager.
- ▶ Du må aldrig nedtage "store mængder" filer af gangen.

Det, du nedtager, bliver kontrolleret af automatiske scanner – samt stikprøver!

HVORDAN HJEMTAGER JEG ANALYSERESULTATER?

- ▶ Hvordan hjemtager jeg **analyseresultater** fra DST?
- ▶ Det er her alle problemerne kan opstå...
- ▶ Og **NEJ**, du må ikke "bare" nedtage informationer!
- ▶ Det er derfor, at du arbejder på DSTs forskermaskiner.
- ▶ Det er derfor, at du skulle søge om alle tilladelserne.
- ▶ Det er derfor, at du er personligt ansvarlig.
- ▶ Der er derfor, at Datamanagement skal lave en kontrol



Spørg
PH Hjemtagningen
ph_hjemtagning@ph.au.dk

POLITIK OM KONTROL AF HJEMTAGELSER

Inden du hjemsender informationer på IFS

- ▶ Skal du henvende dig til Datamanagement, så vi kan lave en kontrol af dine information!
- ▶ Derefter kan den projektansvarlige lave selve hjemtagningen
- ▶ Mange brugere har ikke en download adgang og kan derfor ikke hjemtage
- ▶ Tag kun den information hjem du vil publicere
- ▶ Hjemtagning er altid en risiko!

ph_hjemtagning@ph.au.dk

- ▶ \\..\Projektmappe\Hjemtagning\2017-09-27\mine_filer

http://ph.medarbejdere.au.dk/fileadmin/ph/Interne_politikker/ProcedurerForskerservice_040915.pdf

HVAD MÅ MAN SÅ?



Med andre ord, må du kun nedtage færdigbehandlede og kontrollerede

Aggregerede resultater og dokumentation

- såfremt at de overholder førnævnte krav!

Vi hjemtager aldrig mere end højest nødvendigt!

NOTE

NOTE – RULES on DATA REPATRIATION from VDI

It's NOT allowed to:

- save data to other locations than your CryptShare
- e-mail, upload or transfer data out from the VDI
- copy, cut or paste data from VDI
- print, make screen dumps or take photos of the screen
- make hand copies of data from the screen
- take out any microdata - like rows, cell- or key values

Data repartitions MUST be checked and allowed by Datamanagement:

- PH_hjemtagning@ph.au.dk

Data can leave the system in aggregated form ONLY:

- at least the average of 5 persons
- no min, max, medias or percentiles
- no significant points in plots
- no micro data in axis in plots
- no missing values below at least 5 missing values per variable

Put data for repartitions in the folder:

..\hjemtagning\<Date (yyyy_MM_dd)>\<your files>

GOD SOFTWARE SKIK

Vi er fælles om at dele DST-Serverens Ressourcer!

- ▶ STATA allokerer meget RAM
- ▶ SAS allokerer meget Disk Plads
- ▶ Starter du programmerne mere end en gang, så bruger du flere licenser!
- ▶ Starter du programmerne mere end en gang, så bruger du flere ressourcer!
- ▶ Husk at lukke programmerne!
- ▶ Midlertidige-filer slettes først når du logger af!

Så kan de andre arbejde bedre, når du ikke er på!

BRUD, SANKTIONER OG KONSEKVENSER

- ▶ **ÉN** forseelse lukker **ALLE** på instituttet ned på ubestemt tid
 - en genåbning kræver et direktionsmøde i DST
- ▶ Derudover får **den enkelte**, der var årsagen:
 - 1. gang: 3 mdr. karantæne
 - 2. gang: 12 mdr. karantæne
 - 3. gang: Permanent udelukkelse via PH.
- ▶ Bevidste handlinger for at tilegne sig data:
 - **Politianmeldelse, jf. Forvaltningslovens § 27, stk. 3 og Straffelovens §152**

DANMARKS STATISTIK

VERSION 2.7 – 2024 – KURSUS FOR MEDARBEJDER VED INSTITUT FOR FOLKESUNDHED VED AARHUS UNIVERSITET – UNDER AUTORISATION 50

INTRODUKTION TIL BRUG AF DST

- ▶ **Danmarks Statistik (DST)** - www.dst.dk
- ▶ Særlig adgang via tjenesten **DSTs Forskermaskiner** - <https://remote.dst.dk>
- ▶ Forskere kan her få adgang til informationer i personfølsomme registre hos DST:
Registre som: Befolkning (BEF), Uddannelse (UDDA), Landspatientregistret (LPR), Sygesikring (SSSY), Død (DOD), Indkomst (FAIK) m.f
- ▶ Forskere kan også få adgang til informationer fra andre registre via DST.
Registre som: Sundhedsdatastyrelsen (LSR, LMDB), Landspatientregisteret (LPR), De praktiserende lægers og regionernes fælles enhed for kvalitetsudvikling (DAK-E), Klinisk Epidemiologisk Afdeling (LABKA) m.f.
- ▶ Institut for folkesundhed AU og CIRRAU har selv projektdatabaser – *kan være en genvej!*
- ▶ Adgang til registrene kræver ansøgning og godkendelse



ANSØGNING OM ADGANG

- ▶ Har du ikke gennemført dette kursus, så er der ingen adgang!
 - ▶ Institutlederen er ansvarlig for ALLE aftaler indgået med DST
 - ▶ Tilladelser fra mindst Datatilsynet (AU's fortegnelse)
 - ▶ Derudover tilladelser fra evt. eksterne registre (ikke DST data)
 - ▶ Tilknytnings aftale med IFS
-
- ▶ Ansøgninger til DST via DDV:
 - Datatilsynsnummer, Projektbeskrivelse, Protokol m.m.
 - <http://www.dst.dk/da/TilSalg/Forskningservice>

ANSØGNING OM ADGANG

Procedure for adgang til Danmarks Statistik-projekter/forskeraftaler under autorisation 50: Institut for Folkesundhed

1. Ved oprettelse af projekt på Danmarks Statistik (DST) kontakt datamanager Marianne Pedersen (mp@ph.au.dk) eller datamanager Lone Fredslund (lf@ph.au.dk).

Der udfyldes en projektindstilling på Danmarks Datavindue (DDV)

2. Datamanagement sørger for, at projektet indsendes til DST. De sørger for underskrift fra institutleder med kopi til [datamanagement](#).

3. Oprettelser af nye forskere på eksisterende projekter sker ved udfyldelse af [skemaet for intern procedure for adgang til at arbejde på Danmarks Statistik](#).

4. Forsker modtager login fra DST.

<https://ph.medarbejdere.au.dk/it-sikkerhed-web-og-telefoni/danmarks-statistik-dst-forskningservice/>

DEN PERSONLIGE ADGANG

- ▶ **Når du har fået DST-adgang, så er den personlig!**
 - Ingen andre må bruge din adgang
 - Ingen andre må se dig over skulderen
 - Du må ikke vise dine data til andre
- ▶ **Datafortrolighed**

Husk!
Du må kun lave det med
data, som du har søgt om
tilladelse til!

**Individdata er fortrolige, jf.
Forvaltningslovens § 27, stk. 3
Straffelovens §152.
EU Dataforordning fra 25. maj 2018**

EU DATAFORORDNING

Den nye EU Dataforordning fra 25. maj 2018 (GDPR)

- ▶ Den nye danske persondatalov gælder også på DST
- ▶ Underretning om overtrædelser til DST
- ▶ Underretningspligt om overtrædelser til Datatilsynet
- ▶ Underretningspligt om overtrædelser til de omfattede personer
- ▶ Mulighed for bøder for overtrædelse – ikke fastsat endnu
- ▶ Erstatning til krænkede personer, ca. 1000 kr. pr. person
- ▶ Hertil sanktioner fra DST og evt. politianmeldelse.



Udover GDPR
har DST egne
regler for arbejde
med persondata!

VEJLEDNINGER PÅ DST

- ▶ <https://www.dst.dk/da/TilSalg/Forskningservice/brug-af-forskermaskiner>
- ▶ <https://www.dst.dk/da/TilSalg/Forskningservice/hjemtagelse-af-analyseresultater>

Brug af forskermaskiner

Her findes vejledninger til opsætning, logon og brug af Danmarks Statistiks forskermaskiner:

Danske vejledninger

- ⬇ Forbindelse til forskerservere i Danmarks Statistik - Windows brugere
- ⬇ Forbindelse til forskerservere i Danmarks Statistik - Mac brugere
- ⬇ Forbindelse til Ministerieordningen i Danmarks Statistik - Windows brugere
- ⬇ Forbindelse til Ministerieordningen i Danmarks Statistik - Mac brugere
- ⬇ Vejledning til skift af password
- ⬇ Vejledning til reset af password
- ⬇ Nødvendige opdateringer af Microsoft RDP Klient på Windows7
- ⬇ Forbindelse til Danmarks Statistiks forskerservere fra udlandet
- ⬇ Arbejde på Danmarks Statistiks forskerservere
- ⬇ Adgang til Statistikbanken og Danmarks Statistiks hjemmeside fra forskerservere
- ⬇ Vejledning til brug af SAS formater i Danmarks Statistik
- ⬇ Vejledning til upload af data
- ⬇ Vejledning til hurtig genskabelse af filer
- ⬇ Hardware og software på Danmarks Statistiks forskermaskiner

Hjemtagelse af analyseresultater

Det er vigtigt, at du følger Danmarks Statistiks hjemtagelsesregler. Overtræder du reglerne, risikerer du, at hele din institution bliver midlertidigt lukket ude af ordningen. Har du ved en fejl hjemtaget mikrodata, skal du straks orientere Danmarks Statistik ved at sende en mail til forskningsservice@dst.dk. Det er dit ansvar at sætte dig ind i Danmarks Statistiks datasikkerhedsregler.

⬇ [Datasikkerhedsregler under mikrodataordningerne herunder regler for hjemsendelse af analyseresultater og sanktioner ved databrud](#)

Du må kun hjemtage via Danmarks Statistiks hjemtagelsesværktøj.

Vejledning til hjemtagelse og download af filer



Hvad må ikke hjemtages

Ingen form for mikrodata må hjemtages, hverken via hjemtagelsesværktøjet eller på andre måder fx ved at skrive oplysninger fra skærmen, tage et skærmpoint eller et billede af skærmen med mobiltelefonen. Det gælder også de mikrodata, som du selv har indsendt i forbindelse med et konkret projekt.

Nedenfor er en liste over, hvad der ikke må hjemtages. Listen er ikke udtømmende.

- Tabeller med under tre observationer må ikke hjemtages
- Datasæt – eller dele af datasæt – med oplysninger på individ-/virksomhedsniveau, hvilket kan være baggrundsoplysninger som indkomst, uddannelse og socioøkonomisk status eller beregnede variabler på individniveau, pseudonymiserede nøglevariabler som personnumre, CVR-numre, arbejdsstedsnumre, adressekoder mv.
- Filformater, der ikke er godkendt til hjemtagelse

Tilladte filtyper



HVORDAN KOMMER JEG PÅ DST (WINDOWS)?

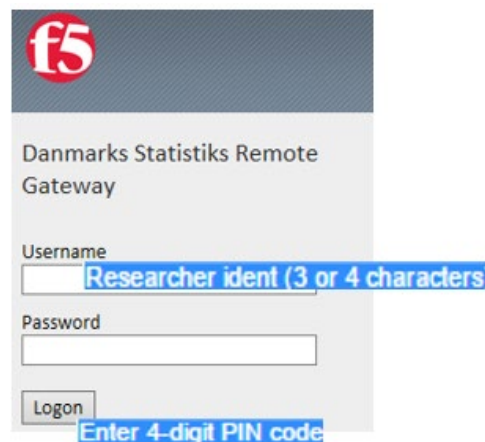
- ▶ Når tilladelserne er på plads, bliver du oprettet som bruger på **DST Forskerservice**
- ▶ Dette giver dig adgang til DSTs Forskermaskiner
- men kun til dit eget projekt.

Linket til vejledningen er:

<http://dst.dk/ext/2115449284/0/forskning/Forbindelse-til-forskerservere-i-Danmarks-Statistik-Windows-brugere--pdf>

I browser indtast adressen

<https://remote.dst.dk>



HVORDAN KOMMER JEG PÅ DST (WINDOWS)?

1. Overblik over projekter (DDV)

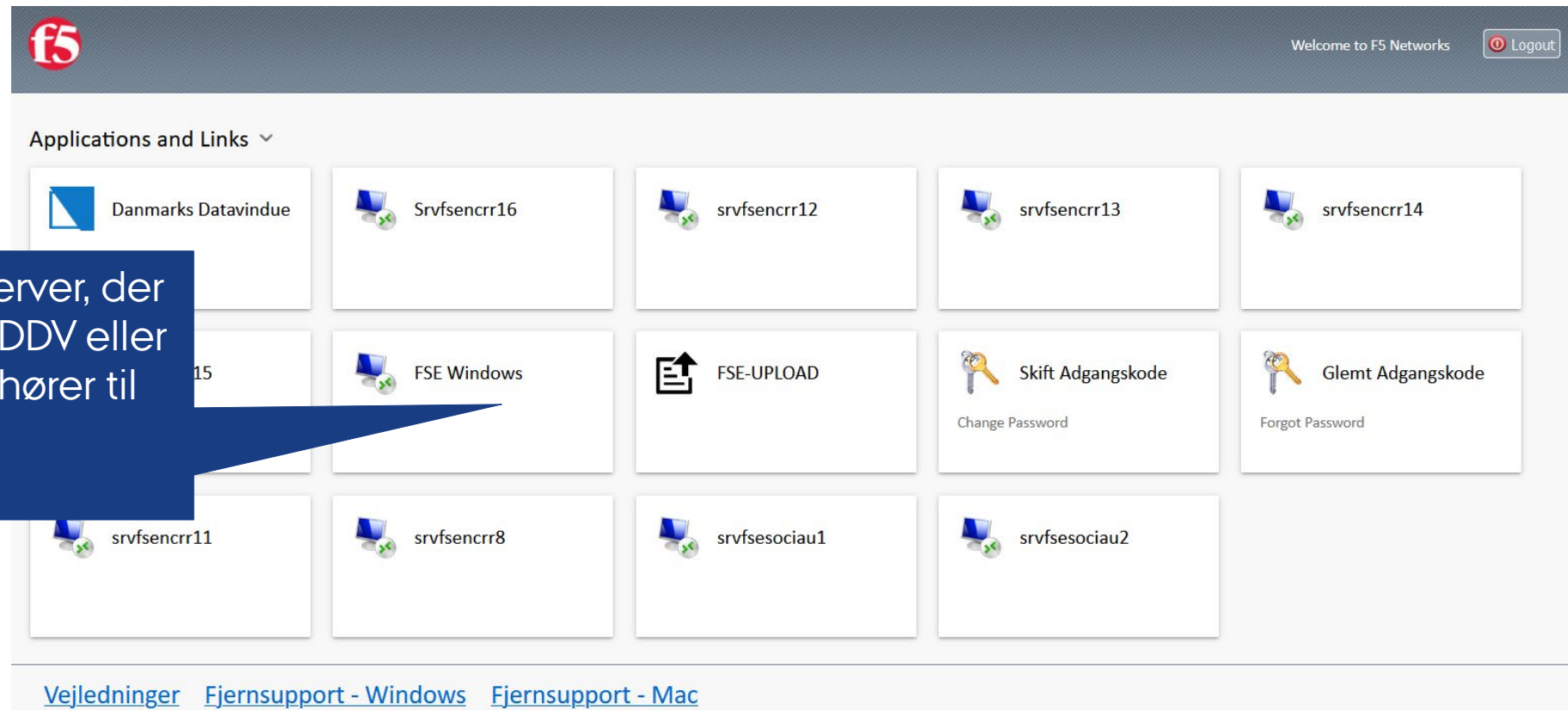
4. Log ud

3. Aflever data til DST

2. Vælg dit password

The screenshot shows the F5 Networks dashboard. At the top left is the 'f5' logo. To the right of the logo is the text 'Welcome to F5 Networks' and a 'Logout' button. Below the header is a section titled 'Applications and Links'. It contains a grid of 15 tiles. The first tile is 'Danmarks Datavindue', which is circled in red. The second row contains 'srvfsencrr15', 'FSE Windows', 'FSE-UPLOAD' (circled in red), 'Skift Adgangskode' (circled in red), and 'Glemt Adgangskode' (circled in red). The third row contains 'srvfsencrr11', 'srvfsencrr8', 'srvfsesociau1', and 'srvfsesociau2'. At the bottom of the dashboard are links for 'Vejledninger', 'Fjernsupport - Windows', and 'Fjernsupport - Mac'.

HVORDAN KOMMER JEG PÅ DST (WINDOWS)?



Navnet på den server, der skal bruges står i DDV eller i den e-mail, der hører til projektet. F.eks. "FSE Windows"

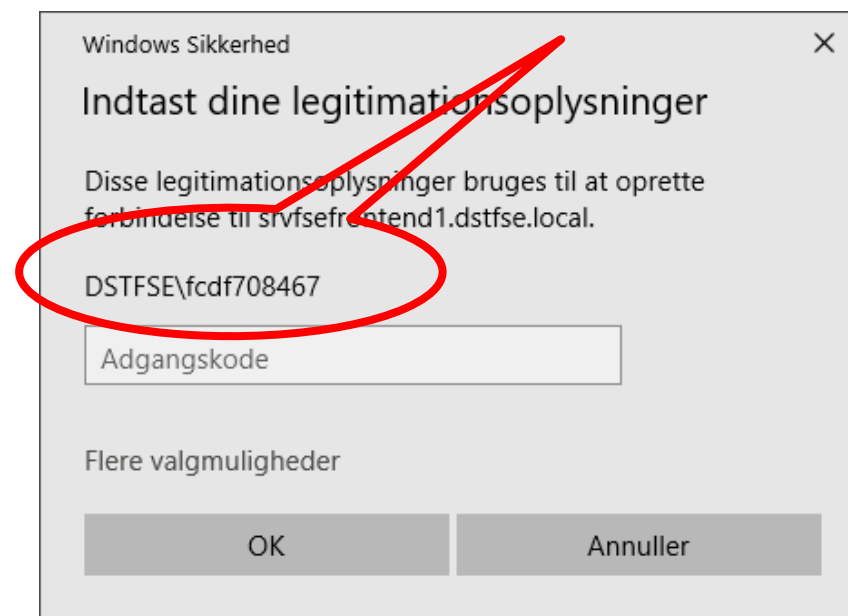
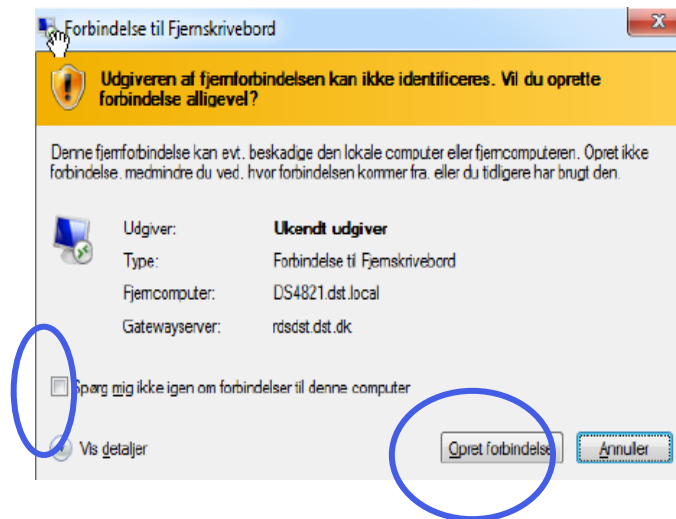
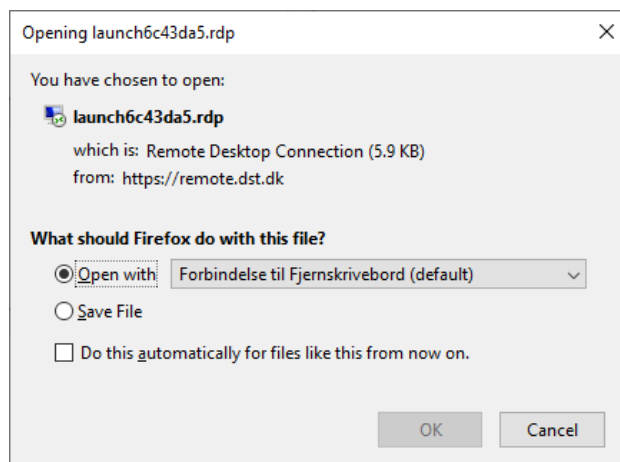
HVORDAN KOMMER JEG PÅ DST (WINDOWS)?

- ▶ Efter denne slide, så er vi næsten halvvejs i login proceduren....
 - dette kaldes **datasikkerhed**!

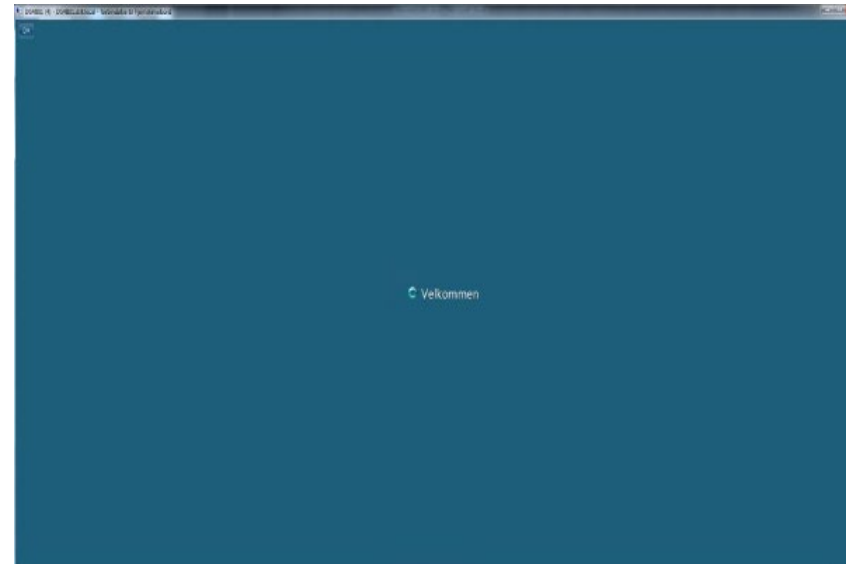
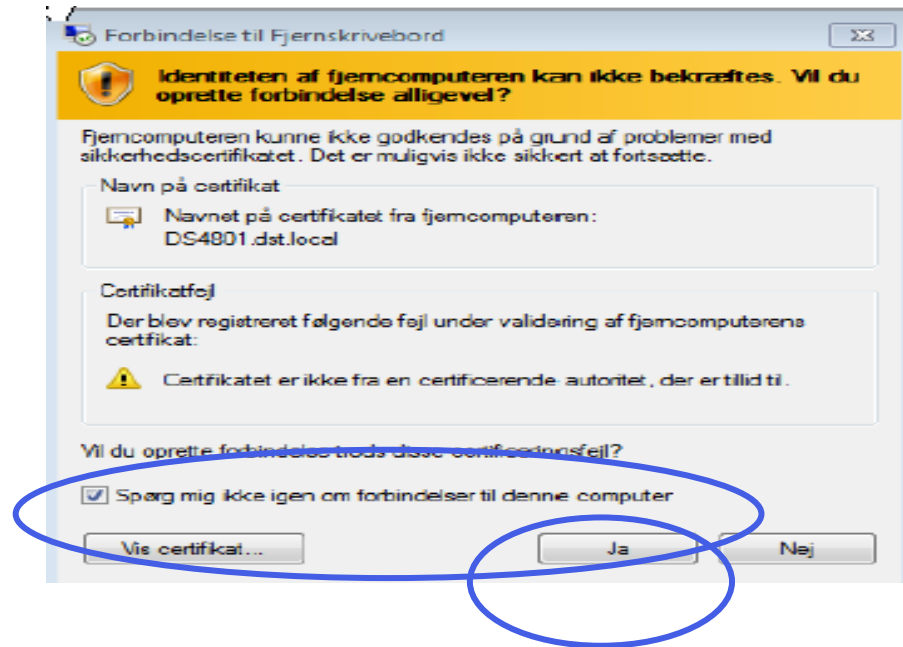
Domain er DSTFSE

Brugernavn er ident + projektnr.

Windows



HVORDAN KOMMER JEG PÅ DST (WINDOWS)?



HVOR ER MINE DATA?

På DST er der to slags data **RAW** og **WORK**:

\\70nnnn\RAW

- ▶ er de rådata man har fået udleveret.
- ▶ De er i SAS-format og er skrivebeskyttet

\\70nnnn\WORK er arbejdsdata.

- ▶ Dette er rådata konverteret fra SAS til e.g. Stata eller R
- ▶ Eller programkode, beregnede data, grafiske fremstillinger osv.
- ▶ Der må/kan kun hjemtages data fra WORK (efter kontrol fra Datamanagement)



SAS convertering
til Stata, R osv

TAK FOR JERES DELTAGELSE

Er du i tvivl, så kontakt:

- ▶ **DST Forskerservice** - 39 17 31 30 eller forskningsservice@dst.dk
- ▶ Eller få vejledning hos **Datamanagement på IFS**

Marianne Pedersen mp@ph.au.dk

Lone Fredslund lf@ph.au.dk

Niels-Søren Bøgh nsb@ph.au.dk



AARHUS
UNIVERSITET