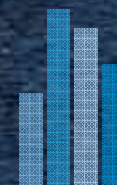


The Danish System for Access to Micro-data

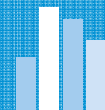
Charlotte Nielsen

Research Services, Statistics Denmark



Outline

- **Rules for researchers' access to micro data**
- **Data available for researcher**
- **Main task for research service unit**
- **Typical progress of a research project**
 - step by step procedures and services seen from a researchers point of view
- **Data security**



Data for research handled by a specialized Research Service Unit

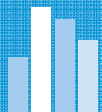
- Staff of 17
 - 13 academics
 - 3 IT specialists
 - 1 secretary



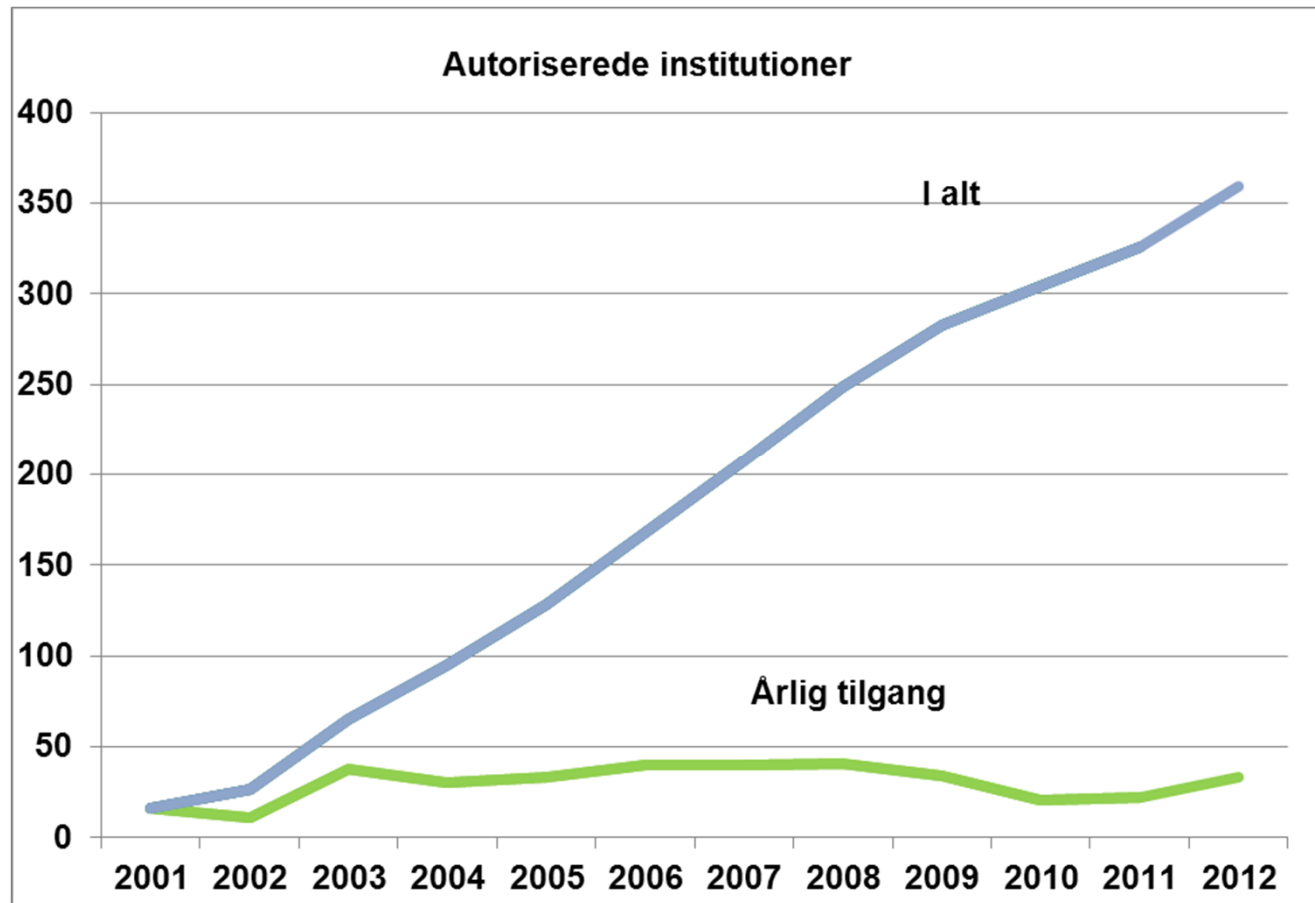
Authorization:

Who can get access to microdata?

- Only researchers or analysts from **authorized institutions**
 - Stable institutions with a responsible manager and with a group of researchers/analysts (Statistic Denmark will not grant authorization to a single person)
 - Public research environments (Universities, hospitals etc.)
 - Private sector
 - Non-governmental organisations
 - Consultancy firms
 - Enterprises
 - Foreign researchers

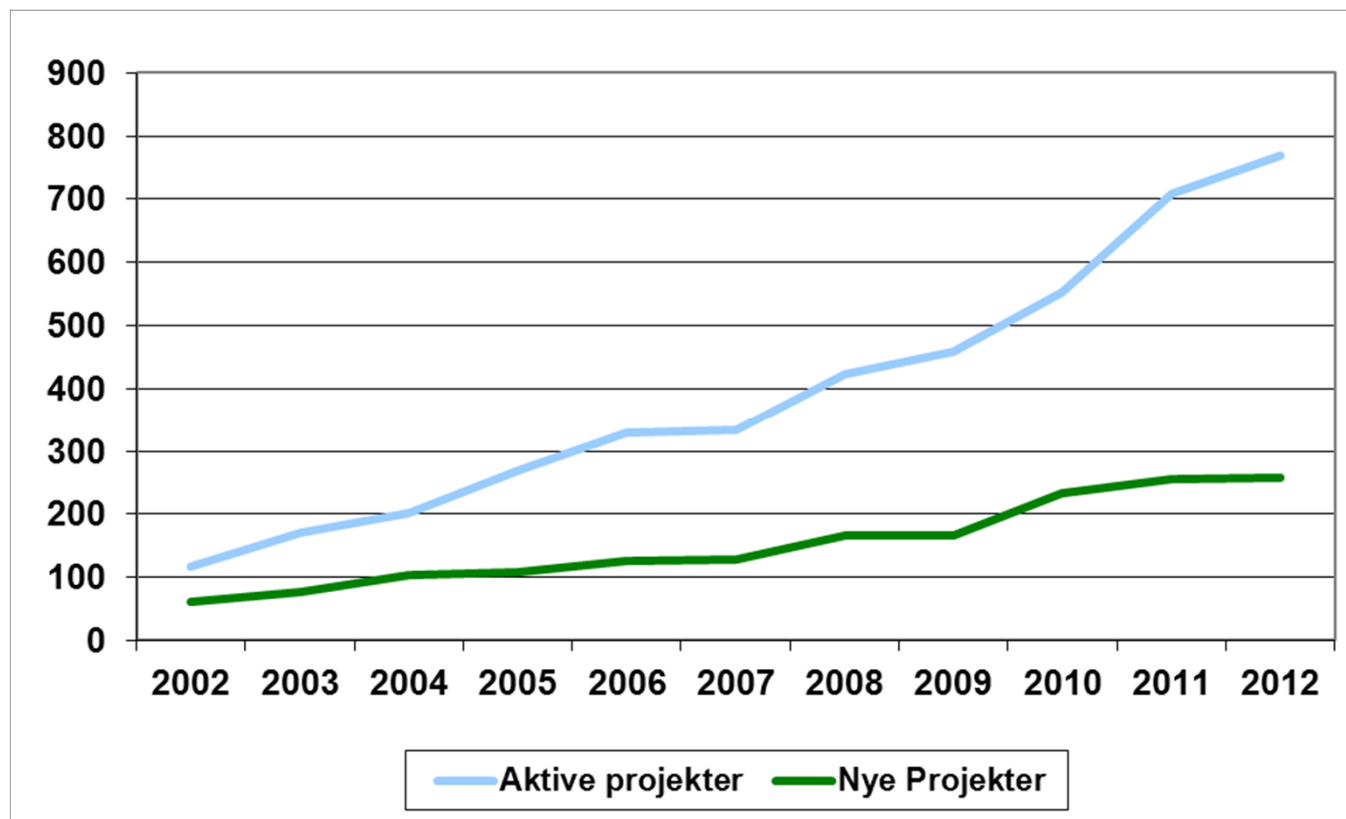


Authorized Institutions - development



359 institutions are authorized to day

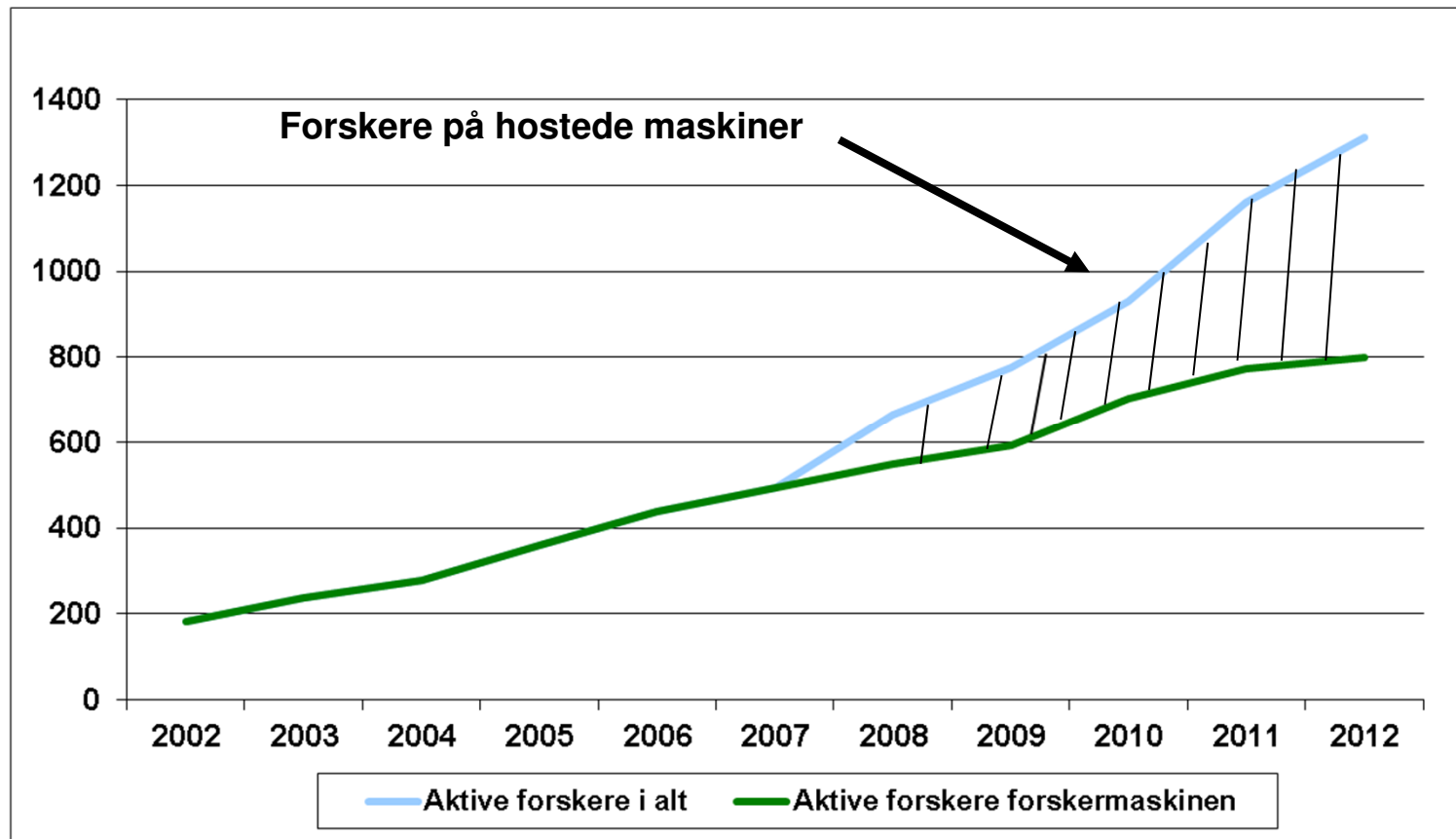
New and active projects



258 new projects in 2012

768 active projects at the research machines

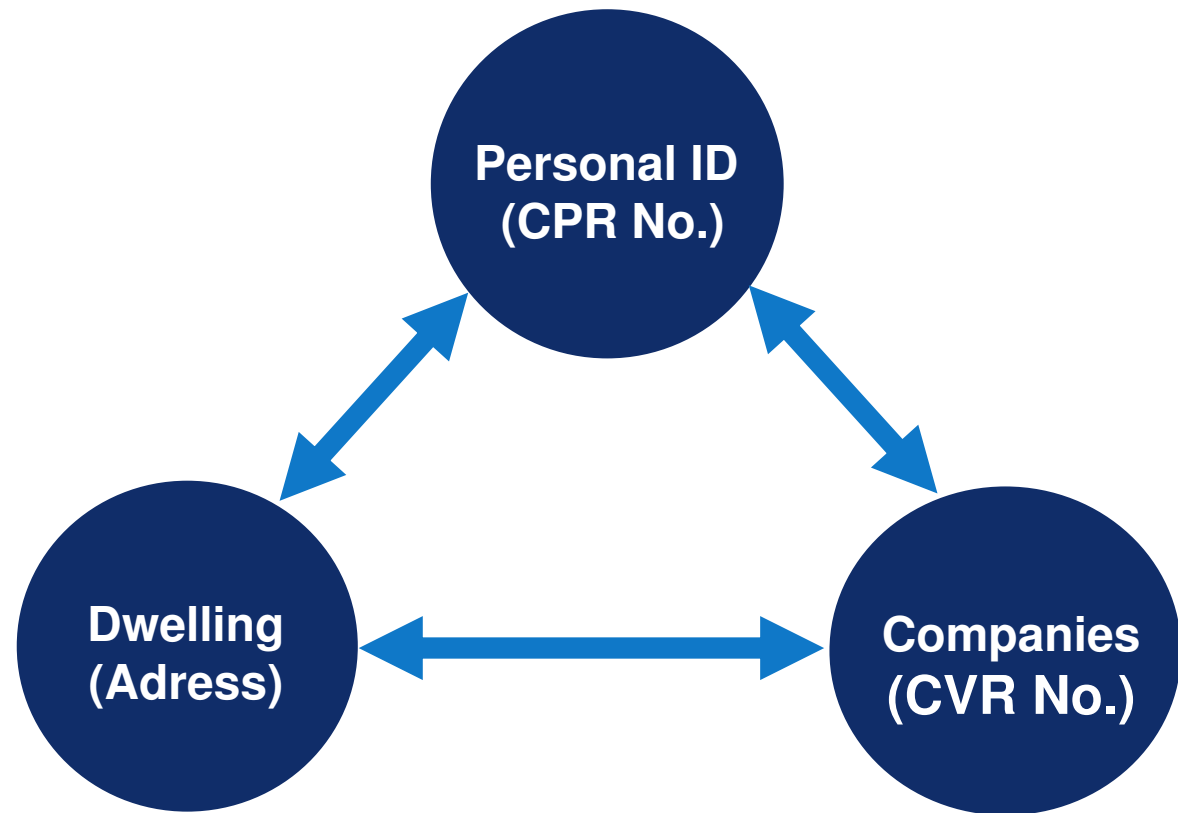
Active researchers on the research machines and the hosted machines



Nearly 1500 active researchers per year

Data available for researchers

- **Data from 250 subject areas – several dates back to the early 1980's**
 - Most data from administrative registers – high quality – cover the entire Danish population
- **HEADLINES**
 - Demographic data
 - Education
 - Labour market
 - Earnings
 - Social conditions
 - Health
 - Justice
 - Income
 - Business enterprises
 - Housing
 - Environment
 - External trade
 - Data from other sources
e.g. own surveys



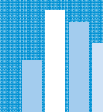
Demographic data

- **Demographic data from 1980 onwards**
 - Immigration and emigration
 - Internal migration within the country
 - Marriages and divorces
 - Births and deaths (including causes of death)
 - Family formation
 - References between parents and children

**Information from the Central Population Register (CPR)
and SSI (Earlier the National Board of Health)**

Health data

- **National Register of Patients**
 - Hospitalizations, diagnoses, treatment etc. from 1977 and onwards
- **Register of Medicinal Product Statistics**
 - Prescription medicines
- **Health insurance**
 - Visits to medical doctors, dentists and other health provider 1992-2012



Labour, earnings and income

- Register-based labour force 1980-2012
- Commuting
- Personnel and family income statistics 1980 -2012
- Salary 1980 and onwards
- Unemployment 1980 and onwards
- Public transfer benefits 1984 - 2012

Education

- **Highest completed level of education 1981- 2013**
- **Full-time educations (1973-2012)**
- **Courses and adult educations**

Business sector

- **Enterprises in general**
 - Employment in businesses
 - Purchases and sales by enterprises
 - General enterprise statistics
 - Enterprise groups
 - Business demography
- **IDA – Integrated Database for Labour Market Research**
 - IDA linkages (persons - workplaces – employees) 1981-2012

Main tasks for Research Services Unit

- Administrates authorizations, contracts etc.
- Approve research projects – “Need to know” principle
- Collect micro data from the other units in Statistic Denmark
- Advise researchers both in data security and content and quality of data – in cooperation with other units at Statistic Denmark
- Extract data for research projects
- Give on-line access to microdata from the researchers own work stations
- Check output files from the research servers

Typical progress of the project

Step 1: Authorization

- **Submission of an authorization application by the research Institution**
 - Description of the research environment
 - Designation of responsible manager of the authorization
(Often the Head of the Department)
- **An authorization contract between Statistics Denmark and the research Institution/environment is signed**
 - by The National Statistician - Jørgen Elmeskov and the responsible research manger

Appendix 2

AUTHORISATION

Statistics Denmark hereby grants

[organisation]

Authorisation for

Remote electronic access to selected datasets at Statistics Denmark

Remote access via the Internet is subject to the following terms:

Step 1: Authorization – important points

- The person signing this agreement on behalf of the research/analysis environment undertakes personally to **supervise** or to appoint a person to supervise that the provisions of this agreement are observed.
- The researcher **may not, directly or indirectly, download the dataset** or any datasets derived there from. Furthermore, **individual records may not be printed**, and all output must be aggregated to an extent that eliminates any risk of direct or indirect identification of persons or enterprises. The researcher may not attempt to make such identification.

Step 2 – Project approval

- The project description is received
 - Description of purpose
 - Population to be studied (**Need to know**)
 - DST registers included in the study (**Need to know**)
 - External data (survey data or data from external registers)
 - Period data is needed for
 - Whom will have access to data
- If there are external data, permission must be granted by the Danish Data Protection Agency (responsibility of the researcher)



Step 3 – Selection of specific registers

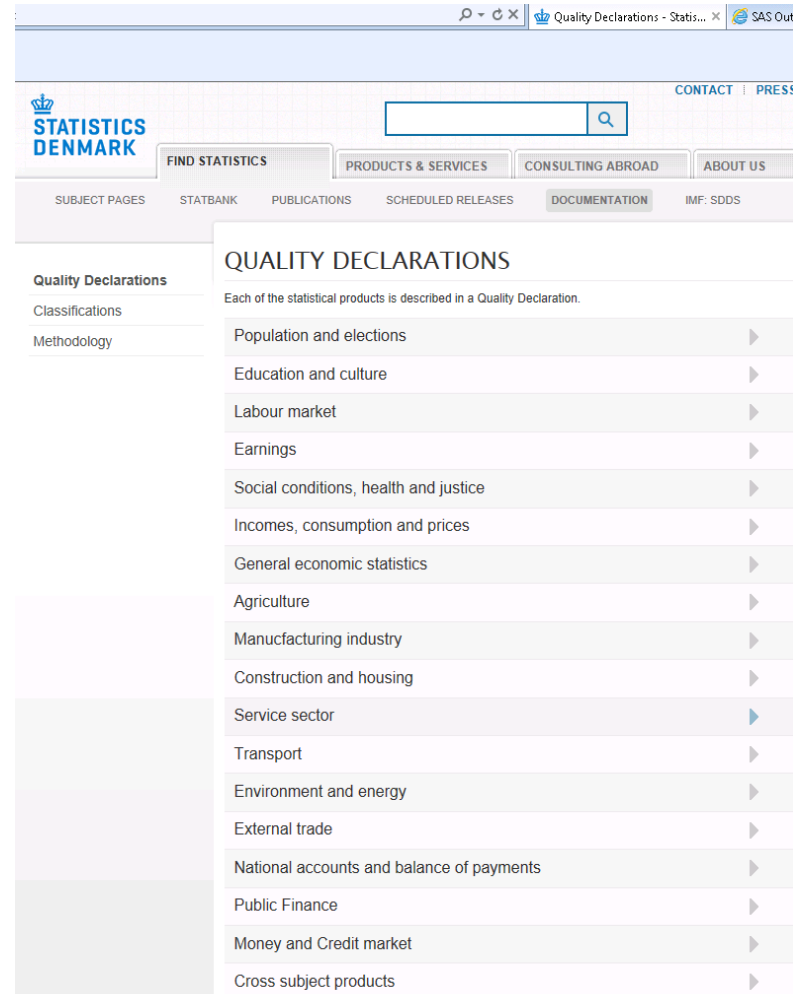
177 registers (on Internet)

Register forkortelse	Register titel	Ansvarlig	Første år	Seneste år	Antal datasæt	Statustidspunkt	opdatering
ADOP	Adoptioner fra CPR	KAE	1988	2009	22	ÅRET	ÅRET
AEFB	Eldredokumentation forebyggende hjemmebesøg	CHJ	2008	2012	5	ÅRET	ÅRET MAJ-JUNI ÅRET EFTER REFERENCEÅR
AEFV	Eldredokumentation visiteret hjemmehjælp frit valg	CHJ	2008	2012	5	ÅRET	ÅRET MAJ-JUNI ÅRET EFTER REFERENCEÅR
AEPB	Eldredokumentation visiteret hjemmehjælp plejebolig	CHJ	2008	2012	5	ÅRET	ÅRET MAJ-JUNI ÅRET EFTER REFERENCEÅR
AEPI	Eldredokumentation borgere på plejehjem bolig (imputeret)	CHJ	2008	2012	5	ÅRET	
AETR	Eldredokumentation genoptræning vedligeholdelsestræning	CHJ	2008	2012	5	ÅRET	ÅRET MAJ-JUNI ÅRET EFTER REFERENCEÅR
AKM	Arbejdsklassifikationsmodulet	CLR	1976	2005	30	ÅRET	Opdateres ikke længere - data findes i indkomst
AMFO	Amfora	CLN	1994	2006	13	ÅRET	ÅRET UDGAET - ERSTATTES AF PUOB
BANT	Antal beskæftigede på virksomheden	CLR	2006	2008	3	ÅRET	ÅRET MARTS 2 ÅR EFTER REFERENCEÅR (DATA FOR ÅRET 2009 KOMMER I MAR
BBR	Bygge- og boligregister	CLR	2005	2013	9	1. JAN I ÅRET	1. JAN I ÅRET JUNI ÅRET EFTER REFERENCEÅRET (2009 DATA FORVENTES JUNI 2010)
BC	Beskæftigede CRAM	CLN	1985	2007	23	ÅRET	ÅRET
BEF	Befolkningen (år)	KAE	1986	2013	28	1. JAN I ÅRET	1. JAN I ÅRET
BEFF	Befolkning flygtninge	KAE	2011	2011	1	1. JAN I ÅRET	1. JAN I ÅRET
BEFK	Befolkningen (kv.)	KAE	2013	2013	1	HVERT KVARTAL	HVERT KVARTAL OPDATERET TIL OG MED 2. KVARTAL 2013
BEFO	Befolkningens Oprindelse	JHS	2013	2013	1		
BES	Landspatientregistret ambulante besøg	KAE	1994	2011	18		
BFL	Detaljeret lønmodtagerdata fra e-Indkomst	CLN	2012	2012	1	KUMULERET	OPDATERET TIL OG MED 2. KVARTAL 2012
BFLA	Aggregerede lønmodtagerdata på brancher	CLN	2012	2012	1	KUMULERET	DATA TIL OG MED 4. KVARTAL 2012
BFLB	Aggregerede lønmodtagerdata på kommune	CLN	2012	2012	1	KUMULERET	DATA TIL OG MED 4. KVARTAL 2012
BFLC	Aggregerede lønmodtagerdata på sektor branche region	CLN	2012	2012	1	KUMULERET	DATA TIL OG MED 4. KVARTAL 2012
BOBO	Børnetilskud barn	KAE	2012	2012	1	ÅRET	
BOL	Bolitællingen	clr	1981	2004	24	1. JAN. I ÅRET	1. JAN. I ÅRET UDGAET - HISTORISK
BOPA	Bopælsændringer	KAE	2012	2012	1	ÅRET	ÅRET
BOSA	Boligsikring	JKP	1984	2011	28	DEC. I ÅRET	DEC. I ÅRET DATA FOR 2012 FORVENTES I 3. KVT. 2013
BOTI	Børnetilskud tilskudsmodtager	KAE	2012	2012	1		
BOYA	Bolitælling	JKP	1984	2011	28	DEC. I ÅRET	DEC. I ÅRET DATA FOR 2012 FORVENTES I 3. KVT. 2013
BOYD	Tilskud og ydelser (børnefamilieydelse)	KAE	1995	2009	15	året	året Maj måned året efter referenceår
BUA	Børn og unge anbragte	CLN	2011	2011	1		OPDATERES ÅRLIGT

Step 3 - Selection of specific registers

Quality declarations (on Internet)

- **Declarations of Contents – general description of the registers (persons responsible, etc.)**
- <http://www.dst.dk/en/Statistik/dokumentation/Declarations.aspx>



Step 4 - Selection of specific variables

Overview

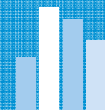
		År																											
BEF - Befolkningen (år)		1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Navn	Label																												
ADRDATO	Adressedato	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
AEGTE_ID	Ægtefælle ID	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
ALDER	Alder pr. 1. januar.																							2008	2009	2010	2011	2012	2013
BOPIKOM	Bopæl i kommunen	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
CIVST	Civilstand	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
CIV_VFRA	Civilstandsdato	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
EFALLE	Personnummer E-fælle	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
FAMILIE_ID	Familiens identificerende nummer.	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
FAMILIE_TYPE	Familietype	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
FAR_ID	Faderens Person Id	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
FKIRK	Folketællingsmarkering	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
FM_MARK	Forældremarkering	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
FOEDREG_KODE	Kode for personens fødselsregistreringssted.	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
FOED_DAG	Fødselsdato	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
GENERATION	Kategorisering af 2. og 3. generationsindvandrere	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
HUSTYPE	Husstandstype	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
IE_TYPE	Indvandrere, efterkommere, personer med dansk opr.	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
KOEN	Køn	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
KOM	Kommunekode	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
MOR_ID	Moderens Person Id	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
OPR_LAND	Oprindelsesland	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
PLADS	Familiestatus	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
PNR	Personnummer	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
REG	Region	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
STATSB	Statsborgerskab	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013

Step 4 - Selection of specific variables

Documentation

- High quality documentation of e.g. 600 selected variables:
- Systematic Documentation:
 - Detailed variable description, including data breaches
 - Validity (period) /population
 - Tables and graphs
 - Formats
 - Appendices with special interest

<http://www.dst.dk/da/TilSalg/Forskningsservice/Dokumentation/hojekvalitetsvariable.aspx>



Step 5 – Data section protocol

- **Data extraction protocol**

<i>Education</i>	EDUCATION		
	<u>From UDDA</u>	<u>Description</u>	<u>Year</u>
	PNR	Personnel identification number	1981-2012
	HFUDD	Highest education completed	1981-2012
	HFAFGTP	Time of completed	1981-2012
	IGUDD	Education in progress	1981-2012
	HF_KILDE	Data source	1981-2012

Further information on variable documentation:

<http://www.dst.dk/da/TilSalg/Forskningsservice/Dokumentation/hoejkvalitetsvariable/hoejst-fuldfoerte-uddannelse.aspx>

Step 6 – Price is calculated and a contract is signed

- A price is calculated ~150 EURO per hour is charged for extraction of data (January 2013)
- 10 hours of free initial advisory services for new projects managed by public research and analysis environment
- Additional charge for disk storage in the case of large-scale projects and log-on charge for private companies
- A contract stating price and expected time of delivery is sent



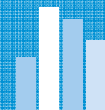
Step 7 – Data are extracted



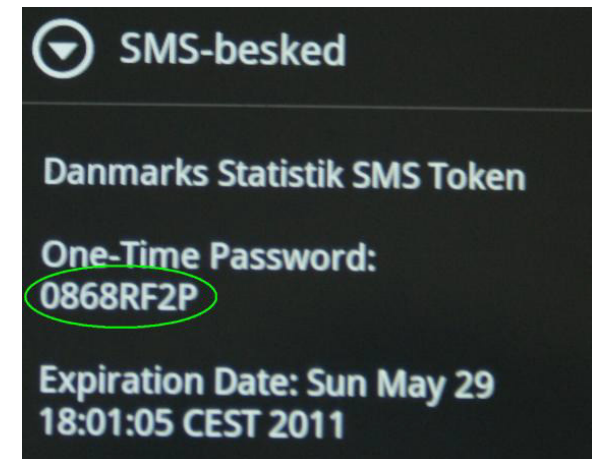
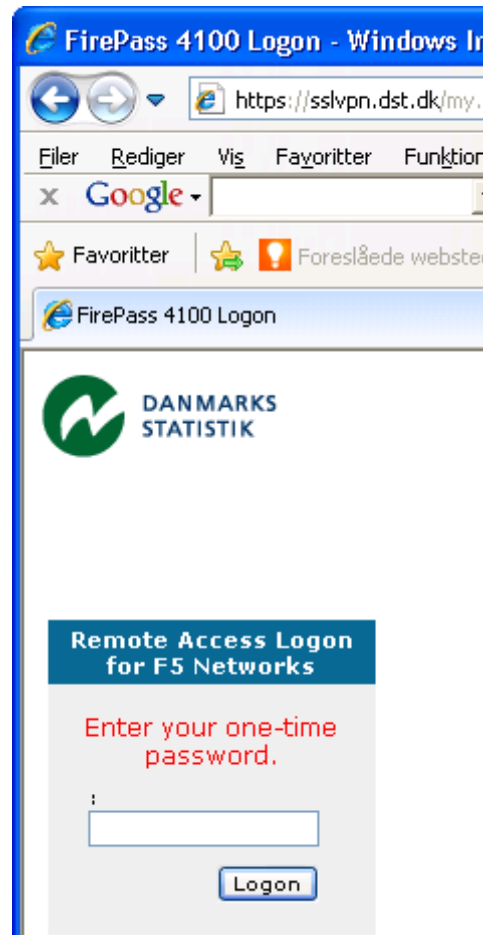
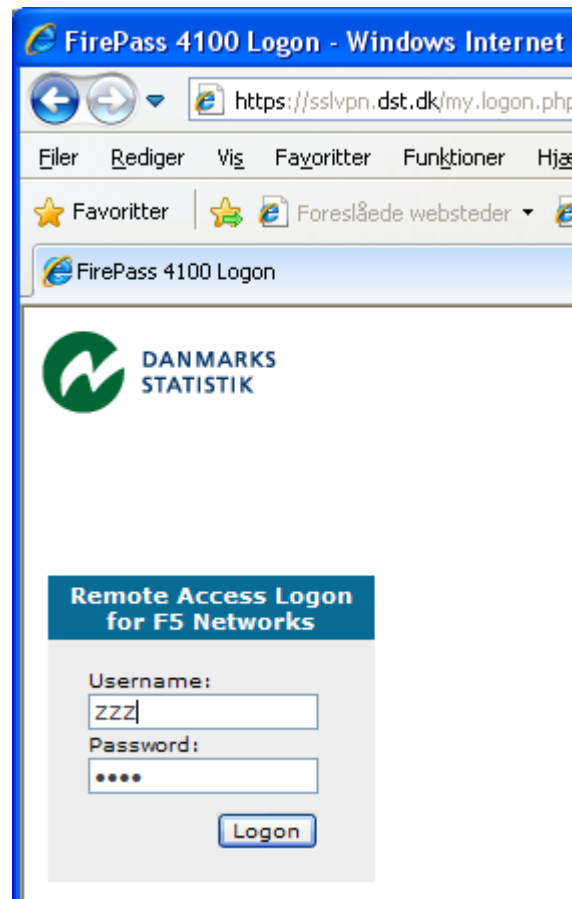
- Data anonymized by means of a project-specific key before data are entered into the researcher computer (all identifying variables such as CPR numbers, addresses etc. are replaced by project specific random numbers)
- Usually one file pr. register and year (data not merged)
- Data placed on a special server separated from the production network

Step 8- Agreement is signed by researcher

- **The researcher signs an agreement on access to microdata**
 - States that all work on micro data must take place on the research server
 - No attempts to identify people or enterprises must be made
 - Only aggregated data can leave the research server
- **Access is gained by the researcher to the researcher computer - password, etc.**



Step 9 - Log-on to the research server via the internet



Step 10 – Data analysis and output

- Folders
 - Rawdata
 - Work data
 - Mail
- Programs
 - SAS, SPSS, STATA, GAUSS, R. etc.

Data security - Summary

- Only access to data according to a "need to know" principle
- Statistics Denmark may reduce data (samples and grouping) if data security is at stake
- All data will be de-identified
- Micro data stays on special dedicated servers placed and managed by Statistic Denmark
- The research server are separated from the rest of the servers at Statistics Denmark

Data Security _summary

Users

- The head of the research institute signs Authorization: Responsibility & supervision
- The researcher signs declaration to follow rules for access to micro data
- Possible to close an authorization if rules for access to micro data is not followed

Data

- All data will be de-identified
- Micro data stays on special dedicated servers placed and managed by Statistic Denmark
- The research server are separated from the rest of the servers at Statistics Denmark

New and on-going initiatives:

- **New data warehouse**
 - more efficient extract of data for research projects
 - data and documentation match
 - key variables consistent (easier to combine data for researcher)
- **Updated internet page – one access for researchers**
 - more user-friendly
 - information on research service – how to get started etc.
 - updated manuals etc.
 - access to updated documentation
 - information of where to get help!

Future project:

- **Online order of project data**