

# Researchers Access to Micro-data

## Process 4: Data delivery

Ivan Thaulow, Jørn K. Petersen

Research Services, Statistics Denmark



# Data delivery: The background

- Data delivery embraces the largest process in FSE
- Includes:
  - Counselling (also an implicit part of process 3 – project approval)
  - Data description
  - Calculation of price (subprocess)
  - Contract
  - Extraction of data – from data warehouse
  - Check and delivery of data (automated)
  - Access of researcher (s)
- Assume process 1 and 3 is approved

# Process 4: Data delivery

## Defining the research project

1. Defining the research project (using a standard template in FSE – available at the webpage)
  1. Project description (counseling)
    1. Purpose of the project
    2. Population
    3. Registers and variables included in the project
    4. Period data is needed for
    5. Whom will have access
    6. External data (permission needed from the Danish Data Protection Agency)

# Process 4: Data delivery

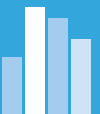
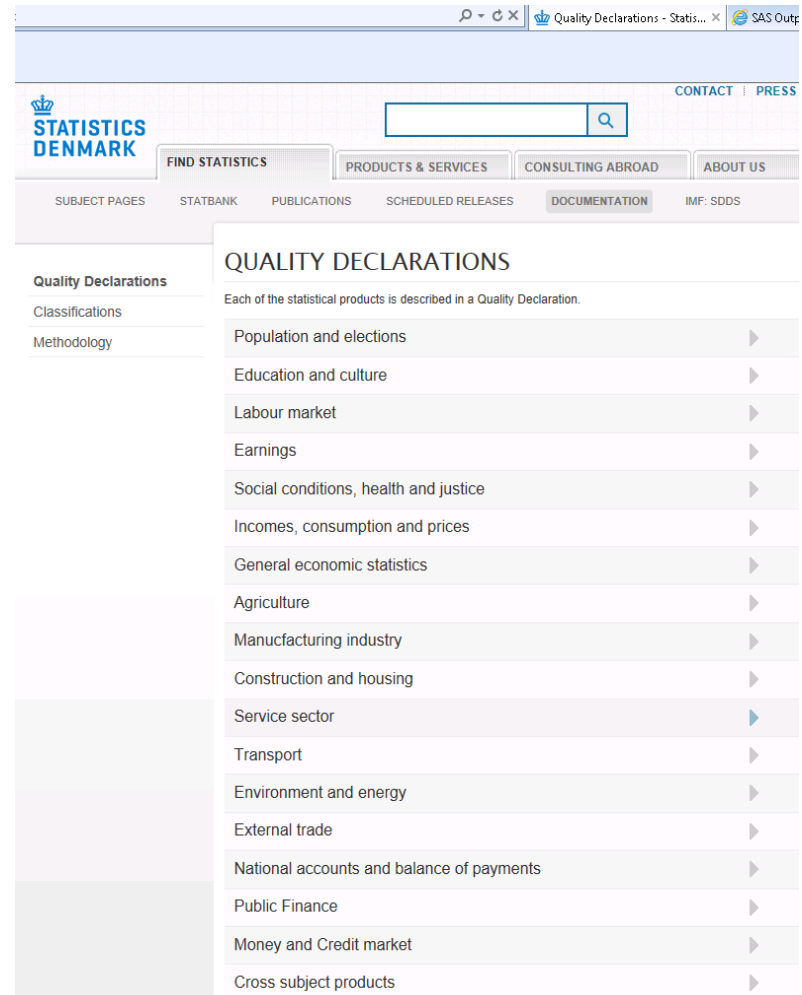
## Defining the research project

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

# Process 4: Data delivery

## Defining the research project

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



# Process 4: Data delivery

## Defining the research project

		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	Folkekirkemarkering	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

# Process 4: Data delivery

## Defining the research project

- 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/Forskningservice/Dokumentation/hojekvalitetsvariable.aspx>



# Process 4: Data delivery

## Defining the research project

<i>Education</i>	<b>EDUCATION</b>		
	<u>From UDDA</u>	<u>Description</u>	<u>Year</u>
	PNR	Personnel identification number	1981-2012
	<a href="#"><u>HFAUDD</u></a>	Highest education completed	1981-2012
	<a href="#"><u>HF_VFRA</u></a>	Time of completed	1981-2012
	<a href="#"><u>UDD</u></a>	Education in progress	1981-2012
	<a href="#"><u>HF_KILDE</u></a>	Data source	1981-2012

Further information on variable documentation:

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



# Process 4: Data delivery

## Calculation of prize

- A prize is calculated from each project
- Use a standard calculation program to estimate a prize that reflects the time used extracting the data
- The following components is included in the model
  - Time used on counseling
  - Time used on defining the research project
  - Time used on creating population
  - Time used on extracting data from data warehouse
  - Time used on handling external data

# Process 4: Data delivery

## Preparation of contract

- When the price is agreed upon a contract is prepared
- A standard template is used. Differentiate between public and private authorized institutions.
  - The contract states a price and expected time of delivery.

# Process 4: Data delivery

## Extraction of data /delivery of data

- Data are extracted for FSE's data ware house
  - Use to a wide extent standard programs (SAS)
  - Makes it easier to hand over the project to another FSE-employee
- External data are included – or integrated (e.g. if it is the population)
- Data are de identified and transferred to the research computers / hosted computers (owned by the researchers)

# Process 4: Data delivery

## FSE-data warehouse

- Data are extracted for FSE's data warehouse
  - Use to a wide extent standard programs (SAS)
  - Makes it easier to hand over the project to another FSE-employee
- External data are included – or integrated (e.g. if it is the population)
- Data are de identified and transferred to the research computers / hosted computers (owned by the researchers)
  - An automated process ensures that data are de identified when transferred

# Process 4: Data delivery

## – Discussion in groups

1. How should the process of data delivery be organized at ICBS
2. What would be the most efficient process
3. Alternative ways data delivery?
4. Your recommendations with focus on
  1. Organization?
  2. Education plan needed?