

Crossing Obstacles for a European Research Project: From a Business Case to an Ideal World

Julien Grenet

Paris School of Economics

March 27, 2012

DwB First European Data Access Forum, Luxembourg

Introduction

- The development of longitudinal data sets linked to employment, earnings and employer information has created unique and **growing research opportunities**

Introduction

- The development of longitudinal data sets linked to employment, earnings and employer information has created unique and **growing research opportunities**
- However, the proliferation of these longitudinal datasets has simultaneously produced a complex set of challenges: **tradeoff** between the needs of researchers and confidentiality requirements

Introduction

- The development of longitudinal data sets linked to employment, earnings and employer information has created unique and **growing research opportunities**
- However, the proliferation of these longitudinal datasets has simultaneously produced a complex set of challenges: **tradeoff** between the needs of researchers and confidentiality requirements
- The emergence of **secure remote access systems** offers a promising solution to enable researchers to access, analyze, and protect confidential microdata

Introduction

- The development of longitudinal data sets linked to employment, earnings and employer information has created unique and **growing research opportunities**
- However, the proliferation of these longitudinal datasets has simultaneously produced a complex set of challenges: **tradeoff** between the needs of researchers and confidentiality requirements
- The emergence of **secure remote access systems** offers a promising solution to enable researchers to access, analyze, and protect confidential microdata
- Despite the progress made in recent years, **important obstacles** remain to transnational access to official microdata

Introduction

- **Ongoing research project** on the economic impact of Social Security contributions offers a good illustration of the constraints faced by European researchers when attempting to perform micro-based cross-country analysis

Introduction

- **Ongoing research project** on the economic impact of Social Security contributions offers a good illustration of the constraints faced by European researchers when attempting to perform micro-based cross-country analysis
- This project associates research teams from **four European countries** (France, Germany, Netherlands, UK) and uses confidential social security data on wages and employment in each country

Introduction

- **Ongoing research project** on the economic impact of Social Security contributions offers a good illustration of the constraints faced by European researchers when attempting to perform micro-based cross-country analysis
- This project associates research teams from **four European countries** (France, Germany, Netherlands, UK) and uses confidential social security data on wages and employment in each country
- A potentially interesting **case study** to highlight:
 - the research potential of newly available administrative data to improve our knowledge of policy relevant issues, in particular from a cross-country perspective
 - the practical constraints that researchers face to find, access and use these microdata

Outline of Presentation

1. Research Project: Goals, Methodology and Data
2. Accessing Restricted Data: the Current State of Affairs
3. Transnational Data Access in an Ideal World

Outline of Presentation

1. Research Project: Goals, Methodology and Data
2. Accessing Restricted Data: the Current State of Affairs
3. Transnational Data Access in an Ideal World

Research Project: Description

- **Project's title:** “The Impact of Social Security Contributions on Earnings: Evidence from administrative data in France, Germany, the Netherlands and the United Kingdom”

Research Project: Description

- **Project's title:** "The Impact of Social Security Contributions on Earnings: Evidence from administrative data in France, Germany, the Netherlands and the United Kingdom"
- **Four Partner research institutions:** *Institut des Politiques Publiques* (France), *Deutsches Institut für Wirtschaftsforschung* (Germany), Amsterdam School of Economics (Netherlands), Institute for Fiscal Studies (UK)

Research Project: Description

- **Project's title:** "The Impact of Social Security Contributions on Earnings: Evidence from administrative data in France, Germany, the Netherlands and the United Kingdom"
- **Four Partner research institutions:** *Institut des Politiques Publiques* (France), *Deutsches Institut für Wirtschaftsforschung* (Germany), Amsterdam School of Economics (Netherlands), Institute for Fiscal Studies (UK)
- Research proposal submitted in September 2011 to the **Open Research Area (ORA) call** jointly launched by the ANR (France), DFG (Germany), NWO (Netherlands) and ESRC (UK) to fund integrated projects in the social sciences by researchers coming from more than one of the four participating countries

Research Project: Description

- **Project's title:** "The Impact of Social Security Contributions on Earnings: Evidence from administrative data in France, Germany, the Netherlands and the United Kingdom"
- **Four Partner research institutions:** *Institut des Politiques Publiques* (France), *Deutsches Institut für Wirtschaftsforschung* (Germany), Amsterdam School of Economics (Netherlands), Institute for Fiscal Studies (UK)
- Research proposal submitted in September 2011 to the **Open Research Area (ORA) call** jointly launched by the ANR (France), DFG (Germany), NWO (Netherlands) and ESRC (UK) to fund integrated projects in the social sciences by researchers coming from more than one of the four participating countries
- Funding decision will be communicated by the end of June 2012

Table – Teams Involved in the SSCs Project

Country	Research Institution	Principal Investigator	Co-Investigators
France	<i>Institut des politiques publiques</i> (IPP), Paris School of Economics and CREST	Antoine Bozio	Julien Grenet Thomas Piketty
Germany	<i>Deutsches Institut für Wirtschaftsforschung</i> (DIW)	Peter Haan	Luke Haywood Kai-Uwe Muller
Netherlands	Amsterdam School of Economics, University of Amsterdam (UvA)	Casper van Ewijk	Leon Bettendorf Nicole Bosch
United Kingdom	Institute for Fiscal Studies (IFS)	Stuart Adam	Richard Blundell David Phillips

Research Project: Motivation

- **Social security contributions (SSCs)** represent a very large share of total taxation in OECD countries (25% in 2008), and an even larger share in Continental Europe (around 37% in France, Germany and the Netherlands in 2008, 19% in the UK)

Research Project: Motivation

- **Social security contributions** (SSCs) represent a very large share of total taxation in OECD countries (25% in 2008), and an even larger share in Continental Europe (around 37% in France, Germany and the Netherlands in 2008, 19% in the UK)
- These taxes were originally introduced to finance various aspects of the **welfare state** (health, pensions, unemployment, disability, etc.)

Research Project: Motivation

- **Social security contributions** (SSCs) represent a very large share of total taxation in OECD countries (25% in 2008), and an even larger share in Continental Europe (around 37% in France, Germany and the Netherlands in 2008, 19% in the UK)
- These taxes were originally introduced to finance various aspects of the **welfare state** (health, pensions, unemployment, disability, etc.)
- Faced with large budget deficits and **increasing costs** of welfare spending, many countries have increased, or are considering increasing, SSCs

Research Project: Motivation

- **Social security contributions** (SSCs) represent a very large share of total taxation in OECD countries (25% in 2008), and an even larger share in Continental Europe (around 37% in France, Germany and the Netherlands in 2008, 19% in the UK)
- These taxes were originally introduced to finance various aspects of the **welfare state** (health, pensions, unemployment, disability, etc.)
- Faced with large budget deficits and **increasing costs** of welfare spending, many countries have increased, or are considering increasing, SSCs
- However, SSCs have been blamed for being **detrimental to employment**

Research Project: Goals

- As a means of reducing these negative employment effects, a number of European countries (including France, Germany and the Netherlands) have implemented **large reductions in employers' SSCs**, funded through general taxation

Research Project: Goals

- As a means of reducing these negative employment effects, a number of European countries (including France, Germany and the Netherlands) have implemented **large reductions in employers' SSCs**, funded through general taxation
- **Several policy options** have been explored: reduce employers' SSCs for low wages (France), transfer SSCs to general income taxation (Netherlands) or VAT (Germany)

Research Project: Goals

- As a means of reducing these negative employment effects, a number of European countries (including France, Germany and the Netherlands) have implemented **large reductions in employers' SSCs**, funded through general taxation
- **Several policy options** have been explored: reduce employers' SSCs for low wages (France), transfer SSCs to general income taxation (Netherlands) or VAT (Germany)
- The **impact** of such policies on labour market outcomes (earnings, employment) ultimately depends on:
 - the economic incidence of SSCs, beyond the notional distinction between employer/employee SSCs
 - the impact of labour costs on labour supply and demand

Research Project: Goals

- This project aims to provide **new evidence on the effects of SSCs** using large administrative panel datasets in France, Germany, the Netherlands and the UK

Research Project: Goals

- This project aims to provide **new evidence on the effects of SSCs** using large administrative panel datasets in France, Germany, the Netherlands and the UK
- The project addresses **two interrelated questions**:
 - What is the economic incidence of SSCs in the short and medium run?
 - How do changes in SSCs affect labour supply and demand?

Research Project: Methodology

- Careful assessment of the structure and rates of SSCs as well as the frequent and complex changes to these over time:
micro-simulation of the four countries' tax systems
→ TAXIPP (FR), STSM (DE), MIMOSI (NL), TAXBEN (UK)

Research Project: Methodology

- Careful assessment of the structure and rates of SSCs as well as the frequent and complex changes to these over time:
micro-simulation of the four countries' tax systems

→ TAXIPP (FR), STSM (DE), MIMOSI (NL), TAXBEN (UK)
- Systematically use **variations the structure and rates of SSCs** to identify short-run effects of SSCs on earnings and employment

→ Use discontinuous changes in the rates of employer/employee SSCs, changes in the ceilings of SSCs, transfers of SSCs to general income taxation

Research Project: Methodology

- Careful assessment of the structure and rates of SSCs as well as the frequent and complex changes to these over time:
micro-simulation of the four countries' tax systems
→ TAXIPP (FR), STSM (DE), MIMOSI (NL), TAXBEN (UK)
- Systematically use **variations the structure and rates of SSCs** to identify short-run effects of SSCs on earnings and employment
→ Use discontinuous changes in the rates of employer/employee SSCs, changes in the ceilings of SSCs, transfers of SSCs to general income taxation
- Undertake a **micro-based cross-country analysis** to evaluate the medium-run effects of SSCs on earnings and employment
→ Take advantage of country-level SSCs reforms which affected different parts of the earnings distribution at different points in time

Research Project: Data

- **Limitations of survey data** for analyzing the economic impact of SSCs (e.g. French *Enquête Emploi*):
 - Moderate sample size → insufficient statistical precision, especially for subgroups of individuals
 - Rotating panel → individuals cannot be followed beyond a short period of time
 - Time span is limited → SSCs reforms in the 1970s and 1980s cannot be analyzed
 - Self-reported wages and hours → measurement error

Research Project: Data

- **Limitations of survey data** for analyzing the economic impact of SSCs (e.g. French *Enquête Emploi*):
 - Moderate sample size → insufficient statistical precision, especially for subgroups of individuals
 - Rotating panel → individuals cannot be followed beyond a short period of time
 - Time span is limited → SSCs reforms in the 1970s and 1980s cannot be analyzed
 - Self-reported wages and hours → measurement error
- Key aspect of our research project: take advantage of the large **social security data** on employment and wages recently made available to researchers France, Germany, the Netherlands and the UK

Research Project: Data

- Social security data is not currently available at the European level: have to rely on **national-level data**

Research Project: Data

- Social security data is not currently available at the European level: have to rely on **national-level data**
- **Administrative longitudinal data** on wages and hours worked at the individual and firm level represent an impressive source of information for empirical tax analysis:
 - Large sample size: 1% to almost 100% of wage earners
 - Panel dimension
 - Typical time span is 35 years
 - Accurate information on wages and hours worked

Research Project: Data

- Social security data is not currently available at the European level: have to rely on **national-level data**
- **Administrative longitudinal data** on wages and hours worked at the individual and firm level represent an impressive source of information for empirical tax analysis:
 - Large sample size: 1% to almost 100% of wage earners
 - Panel dimension
 - Typical time span is 35 years
 - Accurate information on wages and hours worked
- Because of **confidentiality and security concerns**, these administrative microdata cannot be made publicly available. Access to these data for research purposes is based on **country-specific** accreditation procedures and secure data access technologies

Table – Microdata on Wages and Hours Used in the SSCs Project

Country	Data	Time Span	Description	Coverage
France	DADS panel	1976-2009	Wages and hours	4%
	DADS fichier Postes	1993-2009	Wages and hours	100%
Germany	SIAB	1975-2008	Wages and hours	100%
	FAST	[1992-2004]	Income tax data	10%
	VSE	[1990-2006]	Wages and hours	1%
Netherlands	SSB Banen	1999-2009	Income tax data	100%
			Wages and hours	
	LSO	[1979-2002]	Wages and hours	0.1%
AMP	1999-2005	Wages and hours	1%	
United Kingdom	NES	1975-2003	Wages and hours	1%
	ASHE	2004-2009	Wages and hours	1%

[1992-2004]: not all years are available

Research Project: Data

- The SSCs project aims at carrying out an **integrated cross-country analysis** by building on the areas of expertise of each research team (structural modeling, microsimulation, reduced form analysis)

Research Project: Data

- The SSCs project aims at carrying out an **integrated cross-country analysis** by building on the areas of expertise of each research team (structural modeling, microsimulation, reduced form analysis)
- **Two needs:**
 - At a minimum, research teams from each country should be able to access other countries' data
 - Ideally, the different datasets would be stored in a single location to allow for a combined micro-based analysis to be performed

Outline of Presentation

1. Research Project: Goals, Methodology and Data
2. Accessing Restricted Data: the Current State of Affairs
3. Transnational Data Access in an Ideal World

Access to Restricted Data: Current State of Affairs

- **Considerable progress** has been made in recent years to provide researchers with secure access to confidential micro data

Access to Restricted Data: Current State of Affairs

- **Considerable progress** has been made in recent years to provide researchers with secure access to confidential micro data
- Example: the French **Centre d'Accès Sécurisé aux Données**
 - Secure Remote Access Centre
 - Distributed architecture based on a secure device terminal installed in the researcher's laboratory (SD-Box) and a remote secure servers farm hosting confidential microdata and a set of scientific software
 - Pilot version experimented in 2009
 - In 2010, became the official secure remote access solution for accessing official microdata

Access to Restricted Data: Current State of Affairs

- Despite this undeniable progress, **challenges** remain to be tackled to enhance researchers' access to confidential microdata:
 1. Information
 2. Accreditation Process
 3. Data Access

1. Information

- **Serious obstacles** in obtaining information on restricted data and on accreditation procedures

1. Information

- **Serious obstacles** in obtaining information on restricted data and on accreditation procedures
- **Information on data:**
 - Listing of available confidential microdata is often non-existent or incomplete
 - Information on content of data (coverage, dictionary of variables, classifications) is hard to find and often available only for the anonymized data extracts (scientific use files)
 - Language barrier: translated versions are unavailable or partial

1. Information

- **Serious obstacles** in obtaining information on restricted data and on accreditation procedures
- **Information on data:**
 - Listing of available confidential microdata is often non-existent or incomplete
 - Information on content of data (coverage, dictionary of variables, classifications) is hard to find and often available only for the anonymized data extracts (scientific use files)
 - Language barrier: translated versions are unavailable or partial
- Consequence: **prior knowledge of data** is required (previous research projects, personal contacts, word of mouth...)

1. Information

- **Serious obstacles** in obtaining information on restricted data and on accreditation procedures
- **Information on data:**
 - Listing of available confidential microdata is often non-existent or incomplete
 - Information on content of data (coverage, dictionary of variables, classifications) is hard to find and often available only for the anonymized data extracts (scientific use files)
 - Language barrier: translated versions are unavailable or partial
- Consequence: **prior knowledge of data** is required (previous research projects, personal contacts, word of mouth...)
 - in the SSCs, project, this information obstacle was overcome because each team had sufficient knowledge of its country's data

1. Information

- **Information on accreditation procedures:**

- Country-specific accreditation procedures (eligibility rules, timing, costs) are difficult to find
- Institutional complexity: different accreditation procedures exist side-by-side within the same country (SDS vs. ONS VML in the UK, multiple research data centres in Germany)
- Language barrier (e.g. *Comité du Secret Statistique* accreditation procedure is only available in French)

1. Information

- **Information on accreditation procedures:**
 - Country-specific accreditation procedures (eligibility rules, timing, costs) are difficult to find
 - Institutional complexity: different accreditation procedures exist side-by-side within the same country (SDS vs. ONS VML in the UK, multiple research data centres in Germany)
 - Language barrier (e.g. *Comité du Secret Statistique* accreditation procedure is only available in French)
- Lack of information on accreditation procedures constitutes a **significant obstacle to transnational access** to confidential microdata

1. Information

- **Information on accreditation procedures:**
 - Country-specific accreditation procedures (eligibility rules, timing, costs) are difficult to find
 - Institutional complexity: different accreditation procedures exist side-by-side within the same country (SDS vs. ONS VML in the UK, multiple research data centres in Germany)
 - Language barrier (e.g. *Comité du Secret Statistique* accreditation procedure is only available in French)
- Lack of information on accreditation procedures constitutes a **significant obstacle to transnational access** to confidential microdata
 - At this stage, teams involved in the SSCs project have been granted permission to access their **own country's data**. Next step: ask for transnational access.

2. Accreditation Process

- Accreditation procedures are country-specific and often **complex and time-consuming**

2. Accreditation Process

- Accreditation procedures are country-specific and often **complex and time-consuming**
- **Example:** accreditation procedure to access French DADS dataset

2. Accreditation Process

- Accreditation procedures are country-specific and often **complex and time-consuming**
- **Example:** accreditation procedure to access French DADS dataset

Three main stages:

1. Being granted data access by the *Comité du Secret Statistique* (Committee on Statistical Confidentiality)
2. Undertake training session for the the *Centre d'Accès Sécurisé aux Données* (Secure Remote Access Centre)
3. Declare statistical processing of data to the *Commission Nationale Informatique et Liberté* (Commission for Data Protection and Liberties)

⇒ Typical length of accreditation procedure is 9-12 months

Table – Example: the French Accreditation Procedure (previous project)

Stage	Main Steps	Step completed
<i>Comité du Secret Statistique (CSS)</i>	- Submit application form	February 11, 2010
	- Attend CSS meeting (4 per year)	March 18, 2010
	- Sign user agreement	April 4, 2010
<i>Centre d'Accès Sécurisé aux Données (CASD)</i>	- Attend training session	September 19, 2010
	- Obtain biometric identification card	
<i>Commission Nationale Informatique et Liberté (CNIL)</i>	- Submit statistical declaration	October 10, 2010
	- Receipt of approval	October 19, 2010
First Remote Access to Data	- Connection through SD-Box	November 5, 2010

9 months

2. Accreditation Process

- Main constraint: researchers involved in the project have to apply for **multiple accreditations** → at least 16 for SSCs project!

2. Accreditation Process

- Main constraint: researchers involved in the project have to apply for **multiple accreditations** → at least 16 for SSCs project!
- Accreditation procedures are different in each country and often involve **more than one institution on the providers' side**. Ex: in France, *Insee, Comité du Secret Statistique, Commission Nationale Informatique et Liberté, Centre d'Accès Sécurisé aux Données, Archives Nationales*

2. Accreditation Process

- Main constraint: researchers involved in the project have to apply for **multiple accreditations** → at least 16 for SSCs project!
- Accreditation procedures are different in each country and often involve **more than one institution on the providers' side**. Ex: in France, *Insee, Comité du Secret Statistique, Commission Nationale Informatique et Liberté, Centre d'Accès Sécurisé aux Données, Archives Nationales*
- Accreditations often require **multiple signatures on the applicant's side**. Ex: in France, approval has to be signed by the researcher and by the director of his/her research institution

2. Accreditation Process

- Main constraint: researchers involved in the project have to apply for **multiple accreditations** → at least 16 for SSCs project!
- Accreditation procedures are different in each country and often involve **more than one institution on the providers' side**. Ex: in France, *Insee, Comité du Secret Statistique, Commission Nationale Informatique et Liberté, Centre d'Accès Sécurisé aux Données, Archives Nationales*
- Accreditations often require **multiple signatures on the applicant's side**. Ex: in France, approval has to be signed by the researcher and by the director of his/her research institution
- Because the accreditation process is time-consuming and lengthy, it has to be initiated **before the funding decision is made** → potential waste of resources on both sides

Table – Accreditation Procedures in Countries Covered by SSCs Project

Country	Data	Application	Approval	Contract
France	DADS	CSS website	CSS	User agreement between researcher and CSS + statistical declaration to CNIL
Germany	SIAB	FDZ-IAB website	BMAS	User agreement between researcher and FDZ-IAB
	FAST/VSE	Destatis website	Destatis	User agreement between researcher and Destatis
Netherlands	SSB/LSO/AMP	Contact CBS	CCS	User agreement between researcher and CBS
UK	NES/ASHE	SDS website	SDS	User agreement between researcher and SDS

CSS: *Comité du Secret Statistique* (France)

CNIL: *Commission Nationale Informatique et Liberté* (France)

FDZ-IAB: Research Data Centre of the German Federal Employment Agency (Germany)

BMAS: Federal Ministry of Labour and Social Affairs (Germany)

Destatis: Federal Statistical Office of Germany

CBS: Statistics Netherlands

CCS: Central Commission for Statistics (Netherlands)

SDS: Secure Data Service (United Kingdom)

3. Data Access

- **Data access systems** differ from one country to another

3. Data Access

- **Data access systems** differ from one country to another
- Somewhat unclear distinction between **“on site”** vs. **“remote”** **access** in terms of its practical consequences:
 - Can the data can be accessed from the researchers’ home institution? (e.g. VML: NO vs. SDS: YES)
 - Can the data can be accessed from outside the country? (e.g. SDS: NO vs. CASD: YES)
 - Does the remote connection mean that the researcher can see the data (e.g. CASD) or only remotely execute programs (e.g. FDZ-IAB)?
- Country specific standards for **output checking**

3. Data Access

- **Two major constraints** for multi-country research projects:

3. Data Access

- **Two major constraints** for multi-country research projects:
 1. **Transnational remote access** to data is not always possible
 - ⇒ Accessing another country's data can require physical travel

3. Data Access

- **Two major constraints** for multi-country research projects:
 1. **Transnational remote access** to data is not always possible
⇒ Accessing another country's data can require physical travel
 2. **Combining datasets from different countries** to perform micro-based cross-country statistical analysis is not currently possible (would require the different datasets to be stored in a single location and accessed within a common user interface)
⇒ Cross-country analysis can only be performed using cell aggregated values from the original datasets (e.g. by gender, year of birth, year of observation, etc.). The level of aggregation depends on each country's disclosure rules

3. Data Access

- **Other** (less problematic) **differences** across data access systems:
 - Varying fees for preparation of data (per month of use vs. per project), output checking, etc. → problem when preparing research budgets
 - Duration of access: hard to predict and likely to be longer for multi-country projects
 - Available software for data analysis: general-purpose statistical software often provided (e.g. SAS, STATA, SPSS) but more rarely specialized software (e.g. GIS analysis)

Outline of Presentation

1. Research Project: Goals, Methodology and Data
2. Accessing Restricted Data: the Current State of Affairs
3. Transnational Data Access in an Ideal World

Transnational Data Access in an Ideal World

- Areas of improvement to facilitate transnational access to confidential data for research purposes
 1. Information
 2. Accreditation
 3. Data Access

1. Information

- Single point of entry instead of country-specific solutions
- Complete list of available datasets and variables
 - Data wikis (data information, syntax sharing)
 - Harmonized and multilingual metadata
- Information on country-specific accreditation processes (main steps, typical duration, costs, type of data access, etc.)

2. Accreditation

- Single accreditation process and single decision maker
- In the absence of a single decision making authority, have a common application form
- Reduce the overall length of the accreditation procedure

3. Data Access

- Secure Remote Access from the researcher's preferred location
- Common standards for available software, costs and output checking
- Ideally, possibility of combining different national datasets within a single working environment

Conclusion

- Despite recent improvements in access to official microdata for research purposes, significant obstacles remain in terms of transnational access for collaborative European research projects
- In the case of the SSCs project, these obstacles can be partly surmounted because the team includes researchers from each of countries covered by the study
- These barriers would probably be more difficult to overcome for a single team project aiming at accessing other countries' data
- Need for a harmonization of the rules that govern access to confidential data (cross-border accreditation) and – ideally – a common secure remote access system

Thank you for your attention