Accessing German Social Security Records Through the Research Data Center (FDZ) of the BA at IAB

Berkeley, December 2, 2013
Use of Administrative Data in Publications in Leading Journals, 1980-2010 (Raj Chetty)
1. The FDZ-BA

- FDZ-BA: Research Data Center of the German Federal Employment Agency (BA)

- Located at the Institute for Employment Research (IAB) in Nuremberg, Germany

- Established in 2004

- Facilitates access to survey and administrative labor market data for non-commercial empirical research

- Since 2011: External sites
  - Research Data Centers of the Statistical Offices in Berlin, Bremen, Düsseldorf and Dresden
  - Institute for Social Research (ISR), University of Michigan in Ann Arbor, MI, USA
  - New sites: Berkeley, Cornell and Mannheim
1. The FDZ-BA

- Micro labor market data on individuals/households and establishments

Administrative Data

- Social Security Notifications
- Process-generated data of the BA

Data available at the FDZ

Surveys
1. The FDZ-BA

- **Availability**
  - On-site: complete microdata
  - Off-site: confidential microdata, de-facto anonymised microdata, fully anonymised microdata

- **Degree of analysis potential**
  - Stronger anonymisation method
  - Fully anonymised microdata

Available at FDZ-BA
Agenda

1. The FDZ-BA
2. Administrative Data
3. Data available at the FDZ-BA (selection)
4. Data Access
5. Other Services provided by the FDZ
6. Future developments in the EU (Data without Boundaries)
2. Administrative Data

Social Security Notifications

- Procedure:
  - Employers submit notifications to the social security system
  - For every employee and marginal worker covered by the social security system (notification requirement)
  - Annually. Or: Begin or end of employment, employment interruption, change of health insurance
  - Identification: social security number and establishment number

- Purpose of data collection:
  - Calculation of social security contributions and (unemployment) benefits
  - Statistics
**Meldung zur Sozialversicherung**

**Notification to the Social Security System:**

- Social Security Number
- Establishment Number
- Last Name
- First Name
- Address
- Reason for Notification
- Times of Employment (on a daily basis)
- Nationality
- School Education
- Vocational Training
- Type of Employment
- Wages
- Occupational Status
2. Administrative Data

- No information about civil servants, freelancers or self-employed

- **Internal processes of the Federal Employment Agency**
  - Payment/Receipt of unemployment benefits
  - Participation in labour market programs
  - Registered job search

- Exact start and end dates
- Computer-aided processes

- **Since 2011 new information**: new occupation classification, working hours
2. Administrative Data

- Federal Employment Agency transmits data to IAB
- IAB merges social security notifications and BA data
  ⇒ (complete) individual employment biographies

Employment history covered by social security system (since 1975)
  - Unemployment benefit receipt (since 1975)
  - Registered job search (since 2000)
  - Participation in labor market programs (since 2000)
3. Data available at the FDZ - SIAB

Sample of Integrated Labour Market Biographies (SIAB)

- Approximately 2% random sample drawn from the Integrated Employment Biographies of the IAB

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Time reference</td>
<td>Employment biographies on a day-to-day basis</td>
</tr>
<tr>
<td>Number of cases</td>
<td>1,639,325 individuals</td>
</tr>
<tr>
<td>Topics</td>
<td><strong>Socio-demographic characteristics (examples):</strong></td>
</tr>
<tr>
<td></td>
<td>Gender, year of birth, nationality, marital status, number of children, school leaving qualification, etc.</td>
</tr>
</tbody>
</table>
3. Data available at the FDZ - SIAB

<table>
<thead>
<tr>
<th>Topics (continued)</th>
<th>Information on employment, benefit receipt, job search (examples): Begin and end of employment, benefit receipt and job search, daily wage, daily benefit rate, occupation, termination of last job, etc.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Information on location (examples): Place of residence, place of work (federal state, county, regional directorate, employment agency)</td>
</tr>
<tr>
<td></td>
<td>Information on establishments (examples): Branch of industry, total number of employees, median wage of the establishment, etc. (\Rightarrow) Establishment History Panel</td>
</tr>
</tbody>
</table>
Establishment History Panel (BHP)

- Cross sectional dataset on all establishments in Germany with
  - at least one employee liable to social security (until 1998)
  - since 1999 also with at least one marginal worker

- Aggregations of the individual administrative data by establishment ids
- May be combined with SIAB
- 50% random sample or random sample stratified by 8 establishment sizes
### 3. Data available at the FDZ - BHP

| Period covered          | West-Germany: 1975 - 2010  
|                        | East-Germany: 1992 – 2010 |
| Time reference         | June 30                   |
| Number of cases        | Between approx. 650,000 and 1.3 million establishments p.a. (in the samples) |
| Topics                 | **Establishment information (examples):**  
                        | • Total number of employees  
                        | • Number of employees by main occupation  
                        | • Employee age structure  
                        | • Division of employees by educational and vocational qualifications  
                        | • Wage structure of full-time employees  
                        | • Branch of industry  
                        | • Number of inflows and outflows  
                        | • Classification of establishment entry and exit  
                        | • Etc. |
IAB Establishment Panel

- Annual establishment survey
  - West-Germany since 1993, East-Germany 1996
  - Representative on federal state level since 2001

- Tries to follow establishments continuously over time

- Sample, stratified according to establishment size, industry and federal state

- Annually repeated topics and a special focus every year
### 3. Data available at the FDZ – Establishment Panel

| Period covered | West Germany: 1993 to 2011  
East Germany: 1996 to 2011 |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Time reference</td>
<td>June 30</td>
</tr>
<tr>
<td>Number of cases</td>
<td>Between 4,265 and approx. 16,000 establishments per year</td>
</tr>
<tr>
<td>Topics</td>
<td><strong>Annually repeated topics (examples):</strong></td>
</tr>
<tr>
<td></td>
<td>• General company information</td>
</tr>
<tr>
<td></td>
<td>• Development of employment,</td>
</tr>
<tr>
<td></td>
<td>• Business policy and development,</td>
</tr>
<tr>
<td></td>
<td>• Investments, in-house innovations,</td>
</tr>
<tr>
<td></td>
<td>• Government subsidies,</td>
</tr>
<tr>
<td></td>
<td>• Vocational training and apprenticeship places,</td>
</tr>
<tr>
<td></td>
<td>• Recruitments and dismissals,</td>
</tr>
<tr>
<td></td>
<td>• Wages,</td>
</tr>
<tr>
<td></td>
<td>• Etc.</td>
</tr>
</tbody>
</table>
3. Data available at the FDZ – LIAB

Linked-Employer-Employee Data (LIAB)

- **IAB Establishment Panel** matched with individual data from the **Integrated Employment Biographies**

- Simultaneous analysis of the supply and demand sides of the German labour market

- May be combined with **Establishment History Panel**
3. Data available at the FDZ – LIAB

- Two different models:
  
  - LIAB cross-sectional model:
    The cross-sectional LIAB model contains both information on individuals and data from the IAB Establishment Panel matched on a specific reference date (June 30).
  
  - LIAB longitudinal model:
    The data in the longitudinal LIAB model are not referenced to a specific date, but contains personal details related to a certain period of time.
Biographical data of selected insurance agencies in Germany (BASiD)

- 1% sample of the population of the German Pension insurance related to the sample of insured persons and their insurance accounts 2007 (Versichertenkontenstichprobe - VSKT)
  - Employment liable to social security (in the data since 1951)
  - Marginal employment (in the data since 1999)
  - Self-employment with voluntary pension insurance
  - Receipt of unemployment benefits according to the German Social Code Book III (in the data since 1975) or German Social Code Book II (in the data since 2005)

- Information from the Integrated Employment Biographies (IEB) is added:
  - times of job-seeking registered at BA (in the data since 2000)
  - (planned) participation in labor market programs
Panel ‘Labour Market and Social Security’ (PASS)

- Annual household survey started in 2006/2007: Labor market and poverty situation in Germany

- Situation of recipients of benefits in accordance with the German Social Code Book II (recipients of Unemployment Benefit II (“Hartz IV”))

- Two random samples from two partial populations:
  - Persons and households in receipt of Unemployment Benefit II
  - All persons and households registered as residents of Germany

- First wave of the panel study was continued.
- Refreshment samples of households that newly claimed Unemployment Benefit II were drawn.
3. Data available at the FDZ – ALWA

Working and Learning in a Changing World (ALWA)

- Survey which contains information about more than 10,400 life histories

- Allows longitudinal analysis of
  - Schooling and training decisions
  - Labor market re-entries and labor market behavior
  - Processes of family formation
  - Regional mobility

- Combined data: ALWA – administrative data (ALWA-ADIAB)
3. Data available at the FDZ – the future (selection)

- Patent data:
  - Data from the German Patent and Trade Mark Office (DPMA) linked to administrative data of IAB
  - Name and address of inventors for all registered patents in 2002 (on the way to have additional years)

- Data on short-time allowance/work for Nuremberg

- Geocoded data for 2008/2009

- Commercial business data (Bureau van Dijk)
3. Patent data: linked to admin data

- German inventors from 2002
  - 30,000 inventors are linked to admin data
  - First paper is on the way:
    Dorner/Bender/Harhoff/Hoisl: Patterns and Determinants of Inventor Mobility – Evidence from the Employment Biographies of Inventors in Germany

- German inventors from 1999-2012
  - Adress preprocessing completed. Record linkage has started.

- Patents from the former East Germany
  - Around 75,000 inventors from the former GDR can be found afterwards in admin Data.
Anzahl Beschäftigte am 30.06.2009 pro 1km Gitterzelle

Quelle: georeferenzierte Daten des IAB
Geocoded data: social benefit

Anzahl Empfänger von Leistungen nach SGB II am 30.06.2009 pro 1km Gitterzelle

Legende
- >20%
- 16,01-20%
- 12,01-16%
- 8,01-12%
- 4,01-8%
- 0,01-4%
- 0%
- keine Personen

Quelle: georeferenzierte Daten des TAB
Bureau van Dijk

- Commercial data provider
- Firm data and business intelligence
- Compiles data from different sources
- Over 20 databases
  - orbis – global firm data
  - bankscope – worldwide banks
  - zephyr – M&A data
  - dafne – German firm data
Over 1 million firms in Germany, 1999-current

Main source: Creditreform Rating AG

Firms subject to compulsory financial statements are
- public limited companies (Aktiengesellschaft),
- limited liability companies (Kapitalgesellschaft and GmbH) and firms with a mercantile trader/partner liable to unlimited extent (Kaufmann/persönlich haftender Gesellschafter)

Only firms with certain legal statuses and reaching specific employee number or business volume thresholds.

Underrepresentation of small and medium-sized firms
Content
- profit/loss statements
- balance sheet
- (in)solvency
- ownership, crossholdings and subsidiary information

2 year time-lag of data

Linkage of firm data with administrative data to get „innovative“ data

Time table: will get addresses in December and will start record linkage
3. Data available at the FDZ - publications

- American Economic Review
- The Quarterly Journal of Economics
- The Review of Economics and Statistics
- European Economic Review
- Journal of Business and Economic Statistics
- The Scandinavian Journal of Economics
- Journal of Labor Economics
- Journal of Urban Economics
- Labor Economics
- Journal of Population Economics
- and many more
## 4. Data Access

<table>
<thead>
<tr>
<th>Off-Site Access</th>
<th>On-Site Access</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Download" /></td>
<td><img src="image" alt="Remote Execution" /></td>
</tr>
<tr>
<td>Download (Scientific Use File)</td>
<td>Remote Execution</td>
</tr>
<tr>
<td>Factually anonymous</td>
<td>Weakly anonymous (= confidential)</td>
</tr>
</tbody>
</table>
Projects On Data Access

PFiF

DwB

German RLC

Josua

OPENING DATA TO RESEARCH
4. Data Access – On-site Access

- **Guest-RDC (external site)**
  - Interface to establish a communication link
  - Admission control/Identity check

- **FDZ in Nuremberg**
  - Server with FDZ data
  - Applications, contracts, disclosure review, IT
4. Data Access – New Sites

- UKDA, Essex (UK)
4. Data Access

Easy, Quick and Cheap

- **Easy**
  - Non-technical project proposal
  - Approval by RDC (off-site use) or Federal Ministry (on-site use)
  - Use agreement with the institution of the researcher

- **Quick**
  - (Estimated) Time until user/institution receives contract:
    - 2 weeks for off-site access (scientific use file)
    - 6 weeks for on-site access

- **Cheap**
  - Data access is free of charge
  - No lab fees
  - No restrictions on hours/visits of on-site use facilities or runs of remote executions
4. Berkeley Site

1. Make an application
2. Use Agreement
3. Appointments
4. Preparation for the On-site Use
5. After the On-site Use
6. Publishing the results

- Further details:
  

- Thembi Jackson will support the Berkeley site (THX!!!)
- Jörg (Heining) is coming Jan-April 2014
4. Costum Shaped Data

- Standardized data are free of charge, but not always shaped for user needs

- Custom Shaped Administrative Data for the Analysis of Labor Markets (CADAL)
  - Will (hopefully) established mid of 2014
  - Possible to get the IAB data you need (if possible combined with other sources, too)
  - Costs for shaping the data
5. Other Services Provided by the FDZ

- Software available: Stata, R, (SAS)
- Additional software at the Nuremberg site: TSP, MatLab, Gauss
- Disclosure review service
- Bilingual documentation (German and English)
- Advisory service and assistance
6. A German Roadmap for Administrative Data

- Access to administrative Data: establishing RDCs
  - Need for documentation standards
  - Need for a decentralized data access: Remote Data Access or RDC in RDC approach
  - Need for research in RDCs

- Making admin Data fit for the Big Data Ages: German Record Linkage Center

- Need for coordinating, stimulating and roadmapping: German Data Forum (RatSWD)
6. Expanding the German Roadmap for Administrative Data into an EU Roadmap (Data without Boundaries)

- Developments of a German RDC for international access
  - Data documentation
  - International decentralized access to admin data
- Data without Boundaries
  - Need for standards (from meta data to accreditation to access)
- Next steps after DwB for a European Roadmap
Data without Boundaries (DwB)

- A four-year EU-funded FP7-13 project (2011-2015) with 29 partners
- Aims:
  - Linking the capacity of the research community with the important resources of the official micro data in Europe
  - Enhancing researchers access to official micro data in Europe
    - Surveys and administrative datasets, combined files
    - Focus on confidential (highly detailed) data
    - Focus on crossing national boundaries
- Mechanism = Coordination of existing infrastructures
  - CESSDA Data Archives, and the ESS (NSIs coordinated by Eurostat, ECB)
DwB Project Main Issues

• Building a central point of access: what are the available data? How can they be accessed?
• Metadata standards and interoperability
• Legal issues and accreditation: towards a European accreditation
• Servicing the use of OS data: provide tools (format, routines for harmonization), train the researchers for using European micro data
• Technical, standardization and methodological issues in developing a European distributed remote access both for national and for European micro data, flexible to national institutional arrangements (NSI or data archives as provider): propose and implement a test case
Next Major Steps after DwB (German view)

- European Remote Access Network (Eu-RAN)
  - Secured connections
  - Supports **different access points**:
    - Safe Centre (e.g. by checking hardware certificate)
    - Given institution (e.g. by checking IP-number)
    - Any device (e.g. by checking username/password)
  - Connects **researchers** from all over Europe with **research data** from all over Europe
  - Uses a central **Service Hub** to harmonize accreditation, ensures security standards and offer services to researchers
Next Major Steps after DwB (German view)

- **European Service Center**
  - Information portal for researchers, NSIs and Archives

- **Service Hub** to support European Research
  - Secured **Virtual Research Environment** reachable through the European Remote Access Network
  - Offers the following services to researchers and research groups:
    - Statistical packages, editors, wiki, forum, calendar, access to research data via remote desktop or remote execution, output in secure environment, contract management, etc.
Conclusion and Outlook

- Have fun with the German admin data in Berkeley

- Many THX for the support in Berkeley

- Jörg is coming Jan-April 2004
http://fdz.iab.de

iab.fdz@iab.de
fon: ++49-(0)911-179-1752
fax: ++49-(0)911-179-1728
Dafne in FDZ

- Unconsolidated financial data according to IFRS for stock market companies on a quarterly basis and unconsolidated financial data to German GAAP standards on a yearly basis for all others.
- Extensive productivity and profitability measures (sales, profits, return on equity, return on assets, EBITA, etc.).
- Information on the solvency of firms, capital, assets and invested capital, export quotas, industry sectors, information on top and second level management, as well as information on the affiliation of a firm into a larger corporate group.
- GAAP stands for „generally accepted accounting principles“ and refers to German accounting and reporting requirements. German GAAP standards are similar to but not identical with IFRS.