The Linkage of Administrative Data: Building Data for the 21th Century

Linking survey and social security data

Berlin, November 22-23, 2012

Manfred Antoni

Stefan Bender

1. Some principles of the IAB / FDZ

- Data sources should be linked if
    - Technical possible and
    - Do not harm data protection
- Reasons
    - Take the advantages of all different data sources
    - More research on quality of admin data is needed
    - Lower costs

FDZ
Research Data Centre (FDZ)
of the German Federal
Employment Agency (BA)
at the Institute for
Employment Research (IAB)

1. Available Data at the FDZ

Administrative Data

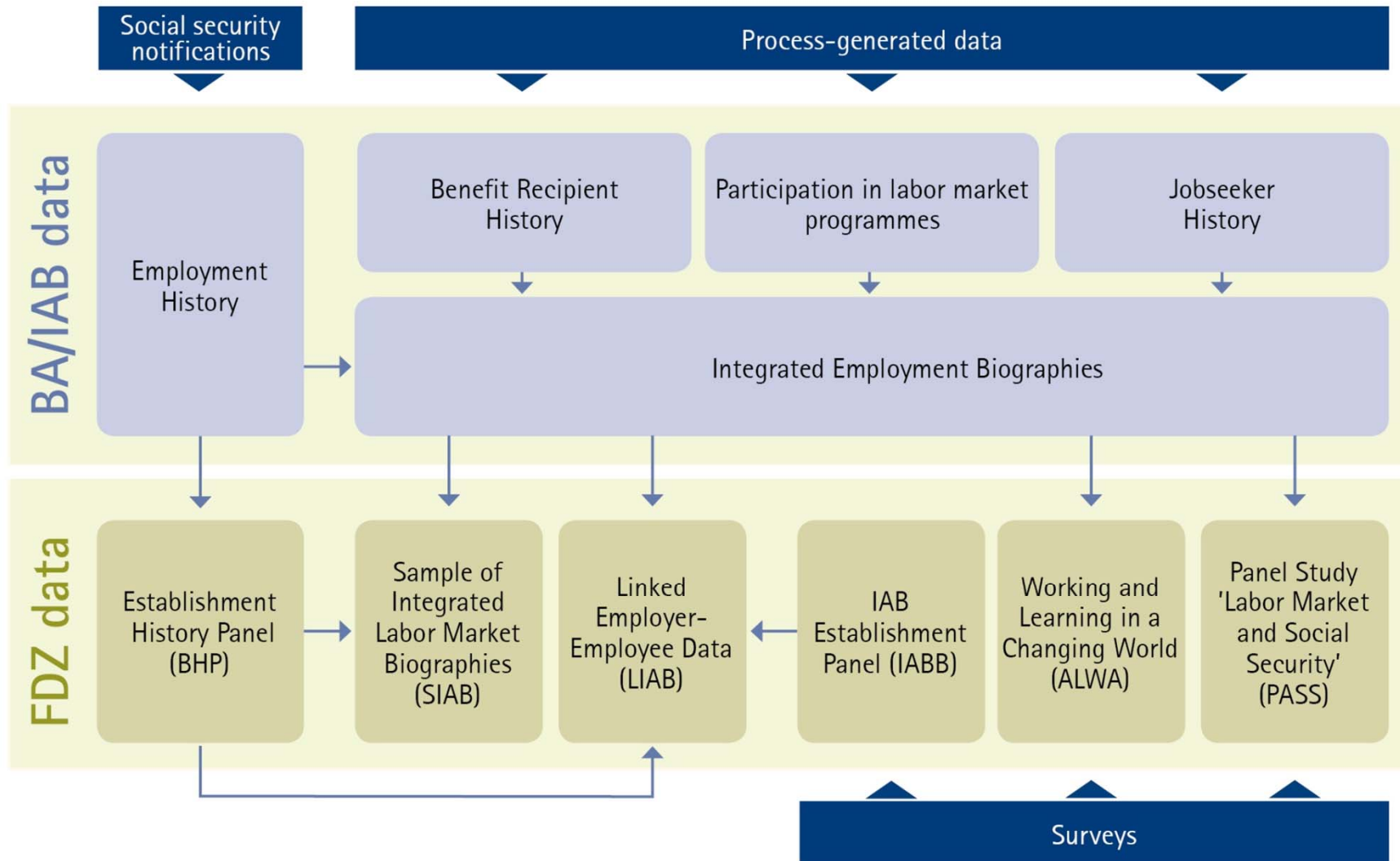Social Security
Notifications

Process-
generated data
of the BA

Data available at the FDZ

Surveys

2. Individual Admin Level Data of the IAB

- Day-to-day employment and wage spells from 1975-present (over 80% of the German labor force)
  - ⇨ This includes unemployment insurance spells, education, occupation information, and demographics
- Detailed information on training programs and job seeking since 2000
- Detailed information on benefit receipt (Hartz 4) since 2005 (household)

- (All) IAB Surveys can be linked

Social security notifications

Process-generated data

BA/IAB data

Employment History

Benefit Recipient History

Participation in labor market programmes

Jobseeker History

Integrated Employment Biographies

FDZ data

Establishment History Panel (BHP)

Sample of Integrated Labor Market Biographies (SIAB)

Linked Employer-Employee Data (LIAB)

IAB Establishment Panel (IABB)

Working and Learning in a Changing World (ALWA)

Panel Study 'Labor Market and Social Security' (PASS)

Surveys

Research Data Centre (FDZ)
of the German Federal
Employment Agency (BA)
at the Institute for
Employment Research (IAB)

FDZ

3. List of Linked Surveys with Admin Data

- ALWA – Working and Learning in a Changing World
- IAB Establishment Panel
- lidA – leben in der Arbeit. German Cohort Study on Work, Age and Health
- Minimum Wage (Construction Sector)
- PASS – „ Panel Study Labour Market and Social Security
- Panel "WeLL" – Employee Survey for the Project "Further Training as a Part of Lifelong Learning"
- Quality of Work and Economic Success: A longitudinal Study on the Development in German Plants
- ….

4. Record Linkage: ALWA

- During the ALWA survey, consent was given by 9,531 (92%) of the respondents (⟹ Joseph)

- Most of the IAB surveys can be linked thru the social security number, but NOT ALWA

- ALWA lacks unique identifiers for a direct link to admin. records

Record Linkage was done by Manfred

4. ALWA: Identifiers of respondents

- Identifiers for matching:
  - ⇨ first and last name
  - ⇨ gender
  - ⇨ day, month and year of birth
  - ⇨ postal code, place name, street name and house number
- Sources of identifiers:
  - ⇨ Field information (infas Institute for Applied Social Sciences)
  - ⇨ Personal register data (IAB)
- Reference year for both sources: 2007
- Extensive standardization of identifiers

4. ALWA: Subsequent steps of the linkage process I

- Exact matching

  - ⇨ Records from both data sources are directly matched based on all available fields.

- Probabilistic record linkage

  - ⇨ Comparison was done with Merge ToolBox (MTB, v0.7; Schnell, Bachteler, and Reiher 2005)

  - ⇨ Parameters for Jaro-Winkler metric according to prior experience with IAB data (Bachteler 2008)

  - ⇨ Blocking on postal codes (about 1,100 blocks)

  - ⇨ Classification of record pairs into links, possible links and non-links (cf. Fellegi and Sunter 1969)

4. ALWA: Subsequent steps of the linkage process II

- Manual matching
  - ⇨ Independent manual review and classification of possible links by two staff members
  - ⇨ Supervision and final decision in case of contradicting classifications by third person

Number of observations over all stages of linkage

	N	$\frac{N}{N_c}$	$\frac{N}{N_r}$
CATI respondents (N_r)	10404		100%
Consenting CATI respondents (N_c)	9531	100%	91.61%
Exact matches	5035	52.83%	48.39%
Exact and probabilistic matches (Jaro-Winkler)	7919	83.09%	76.11%
Exact, probabilistic and manual matches	8243	86.49%	79.23%

FDZ
Research Data Centre (FDZ)
of the German Federal
Employment Agency (BA)
at the Institute for
Employment Research (IAB)

5. Bringing Admin Data to the 21st Century I

- Linkages of administrative data available at the FDZ to
  - ⇨ Commercial business data (like Bureau van Dijke)
  - ⇨ Internet data (like ImmoScout 24)
  - ⇨ Patent and geo-coded Data

- Increase of record linkage (German Record Linkage Center) German RLC

- Data access!
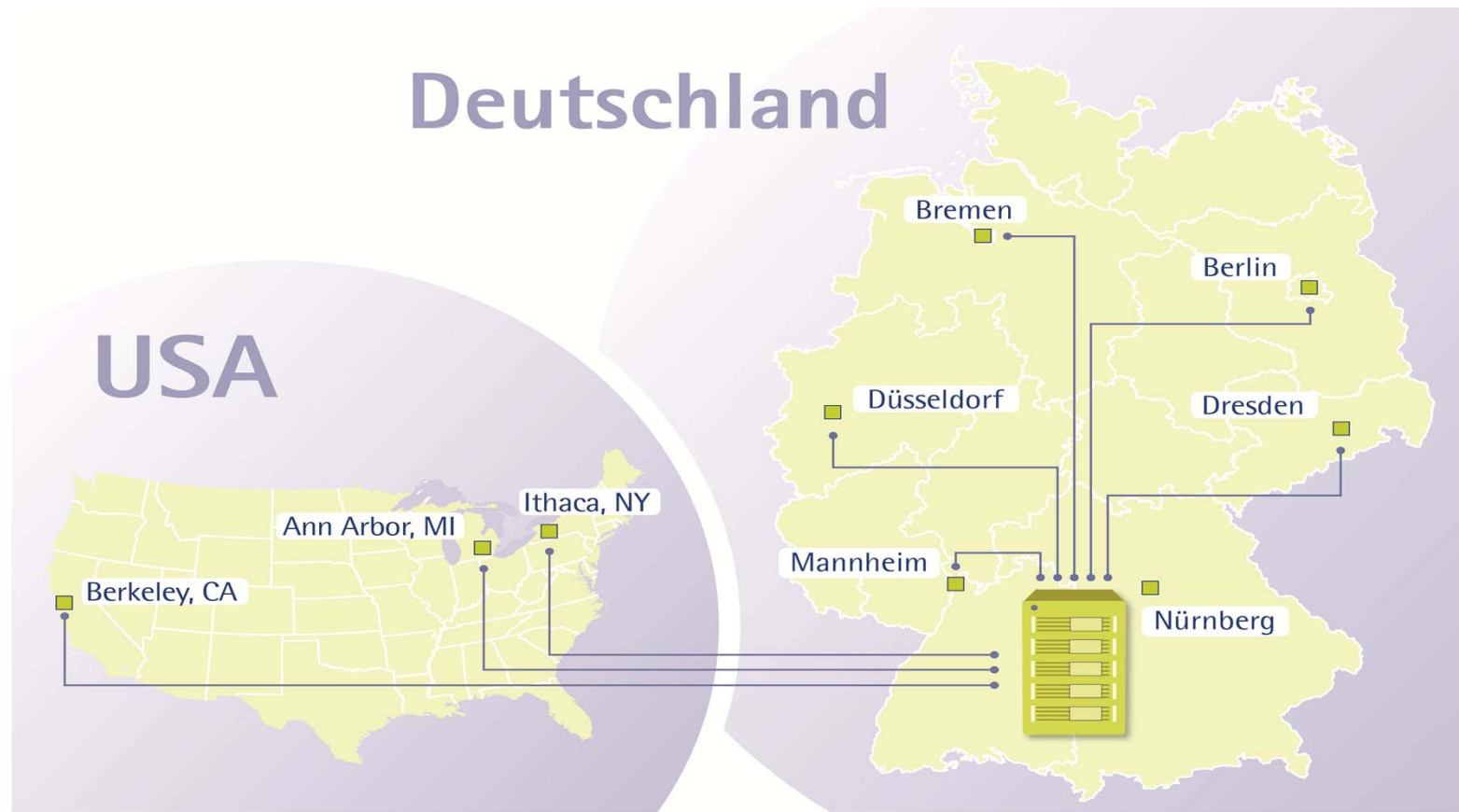
5. Bringing Admin Data to the 21st Century II

- **Geo-coding of administrative data at the FDZ (total population)**

- Place of work and place of residence

- Two cross-sections: 2008 and 2009 on June, 30th

- Geo-codes are meter-perfect

- Near future:
  - ⇨ go back in time and add external geo-coded/address information to admin data
  - ⇨ Give access to anonymized geo-coded data (first steps with multiple imputation)

5. Bringing Admin Data to the 21st Century III

■ **Patent data**

⇨ Data of the German Patent and Trade Mark Office (DPMA) linked to administrative data of IAB

⇨ In cooperation with Dietmar Harrhof (Munich)

⇨ Names and addresses of inventors for all registered patents in 2002

⇨ More years will be added in the future

The FDZ Network on Data Access (2012/13)

Goal:European network of decentralized data access

Research Data Centre (FDZ)
of the German Federal
Employment Agency (BA)
at the Institute for
Employment Research (IAB)

FDZ

http://fdz.iab.de/en.aspx

Stefan Bender

FDZ der BA im IAB

Regensburger Str. 104

90478 Nuernberg

stefan.bender@iab.de

The FDZ Network on Data Access (2012/03)

Technical Implementation: Citrix thin client solution

Ann Arbor, Mi

MICDA Enclave at ISR

Bremen

Berlin

Düsseldorf

Dresden

FDZ der Statistischen Ämter

Management Server

Nuremberg

Terminal Server with RDC Data

Thin Client

ICA connection using a Citrix Access Gateway

Connection to Management Server