

Assessing and improving the quality, analytic potential and accessibility of data by linking administrative, survey and open data

European Conference on Quality in Official Statistics (Q2014)
Vienna, 5 June 2014

preliminary version

Manfred Antoni
Alexandra Schmucker

Motivation

■ Starting position:

- Increasing demand of comprehensive data (particularly longitudinal data) in social sciences
- Cumulative problems with survey data (declining reachability and cooperation of respondents, rising costs)
- More and more new data sources are explored (administrative data, big data, etc.)
- Different shortcomings of these different data sources

■ Remedy:

- Balancing the disadvantages of these data sources combining the advantages of these data

■ Implementation:

- Creating a dataset using data linkage

Advantages and disadvantages: Survey data

- Advantages:
 - Specially conducted for certain research questions
 - Additional information on behaviours, attitudes etc.
- Disadvantages:
 - Missing data (unit-nonresponse, item-nonresponse, panel attrition)
 - Misreporting (e.g. recall errors in retrospective interviews)
 - Time restrictions
 - High costs

Advantages and disadvantages: Administrative data

■ Advantages:

- Long time period
- Precise and reliable information
- Complete census

■ Disadvantages:

- Data are collected for administrative use (research is secondary use)
- Changes in the data collection method and the recorded information
- Time lag

Remedy: Data linkage

- **Several project to link different data at the Research Data Centre**
- **Establishment of the German Record Linkage Center**
- **Challenges:**
 - Matching key for data linkage
 - Legal restrictions for linkage and data access
- **Consequences:**
 - Understanding and improving data quality
 - Higher analytic potential
 - Higher cost efficiency
- **Data:**
 - Several survey data on individuals, household or establishments
 - Open data
 - Administrative data of the Institute for Employment Research and the German Federal Employment Agency

Data sources of the Research Data Centre

Administrative Data

**Social
Security
Notifications**

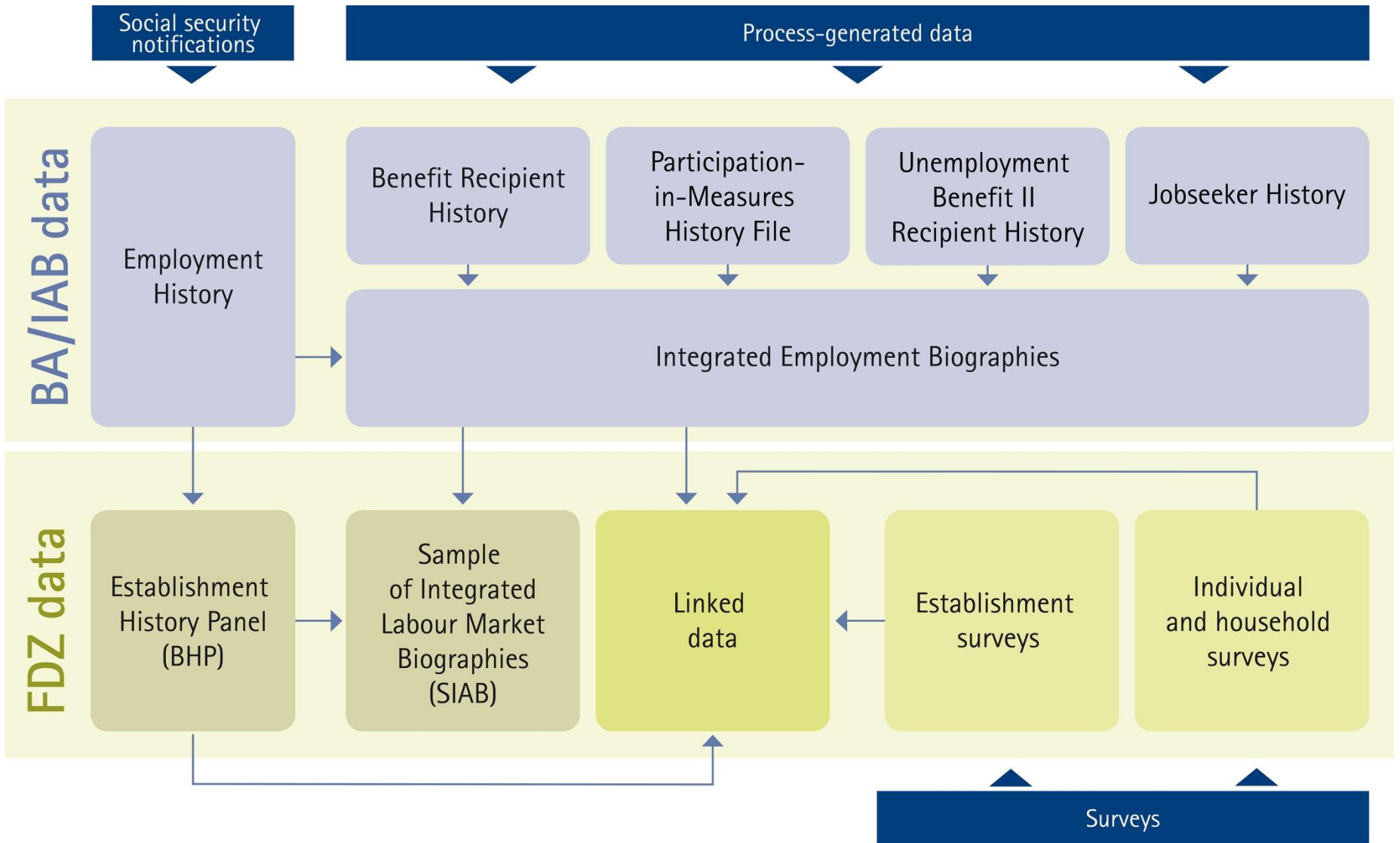
**Process-
generated
data of the
BA**

Linked data

**individuals
households
establishments**

Survey Data

Data sources of the Research Data Centre



German Record Linkage Center (GRLC): Objectives

- Sustained increase in the number and quality of Record Linkage applications in specific sciences
- Development of new data sources for research
- Performing service tasks (FDZ) and research on technical solutions (Duisburg-Essen)

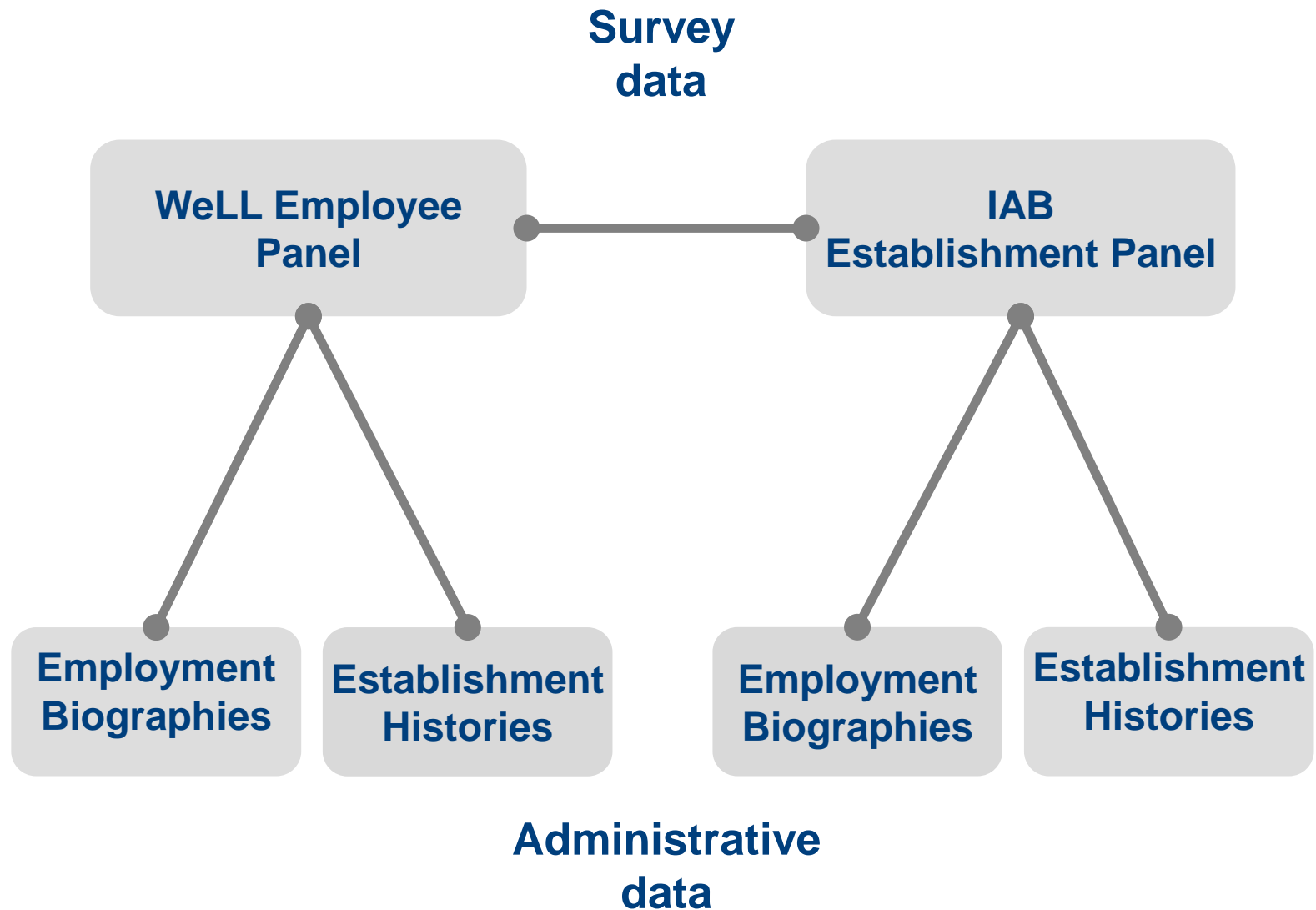
German Record Linkage Center (GRLC): Services

- Individual advice during the planning and realization stages of data linkage projects
- Conducting data linkages as commissioned work
- Compilation and provision of reference databases, which are useful for data linkages, i.e. municipality keys, geo-coordinates and corresponding zip codes
- Updating and maintaining the record linkage software MTB
- Overview of projects at record-linkage.de
- Logo des GRLC

Exemplary project I: WeLL-ADIAB

- Data sources:
 - Employee survey (project 'Further Training as Part of Lifelong Learning')
 - IAB Establishment Panel
 - Employment biographies from administrative sources
 - Establishment histories from administrative sources
- Data linkage:
 - Using the social security number and the establishment number
 - Allowance for linkage
- Analytic potential:
 - Innovative linked employer-employee dataset to analyse determinants and consequences of further training in Germany
 - Research on data quality and selectivity (unit-nonresponse or refusal of allowance)
- Data access:
 - On-site use at the Research Data Centre

Exemplary project I: WeLL-ADIAB



Exemplary project II: MPI-IC-IAB-Inventor Data (MIID)

- Data sources:
 - Patent and inventor data (German Patent and Trademark Office and PATSTAT - EPO Worldwide Patent Statistical Database)
 - Employment biographies from administrative sources
 - Establishment histories from administrative sources
- Data Linkage:
 - Record linkage using names and addresses of the inventors and individuals in the administrative data
- Analytic potential:
 - Research at the intersection of labour market processes and patenting activities of individuals
 - Topics: Socio-demographic profiles of inventors, team composition, employment careers of inventors and their co-workers (e.g. mobility)
- Data access:
 - Currently only for project members of IAB and MPI-IC
 - Access via the Research Data Centre planned in the future

Exemplary project III: Geocoding of German Administrative Data



- Data sources:
 - Geocoded addresses from the Federal Agency for Cartography and Geodesy
 - Addresses of individuals and establishments from administrative employment biographies of the IAB
- Linkage:
 - Record linkage using addresses of establishments and individuals
 - Aggregation to 2,280,864 small-area regions and grid cells (1,000 meter edge length)
- Analytic potential:
 - Analyses below the municipality level
 - Neighbourhood effects
- Data access:
 - Currently only for employees of the Research Data Centre
 - Feasibility study on data access via the Research Data Centre

References:

- Bender, Stefan; Dorner, Matthias; Harhoff, Dietmar; Hoisl, Karin; Scioch, Patrycja (2014): The MPI-IC-IAB-Inventor Data (MIID): record-Linkage of Patent Register Data with Labor Market Data of the IAB. FDZ-Methodenreport, xx/2014 (forthcoming), Nürnberg.
- Bender, Stefan; Fertig, Michael; Görlitz, Katja; Huber, Martina; Schmucker, Alexandra (2009): WeLL - unique linked employer-employee data on further training in Germany. In: Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften, Jg. 129, H. 4, S. 637-643.
- Scholz, Theresa; Rauscher, Cerstin; Reiher, Jörg; Bachteler, Tobias (2012): Geocoding of German Administrative Data. The Case of the Institute for Employment Research. FDZ-Methodenreport, 09/2012 (en), Nürnberg.

Contact:

Manfred.Antoni@iab.de

Alexandra.Schmucker@iab.de

Information on the Research Data Centre:

<http://fdz.iab.de>

Information on the German Record Linkage Center:

<http://www.record-linkage.de/>