

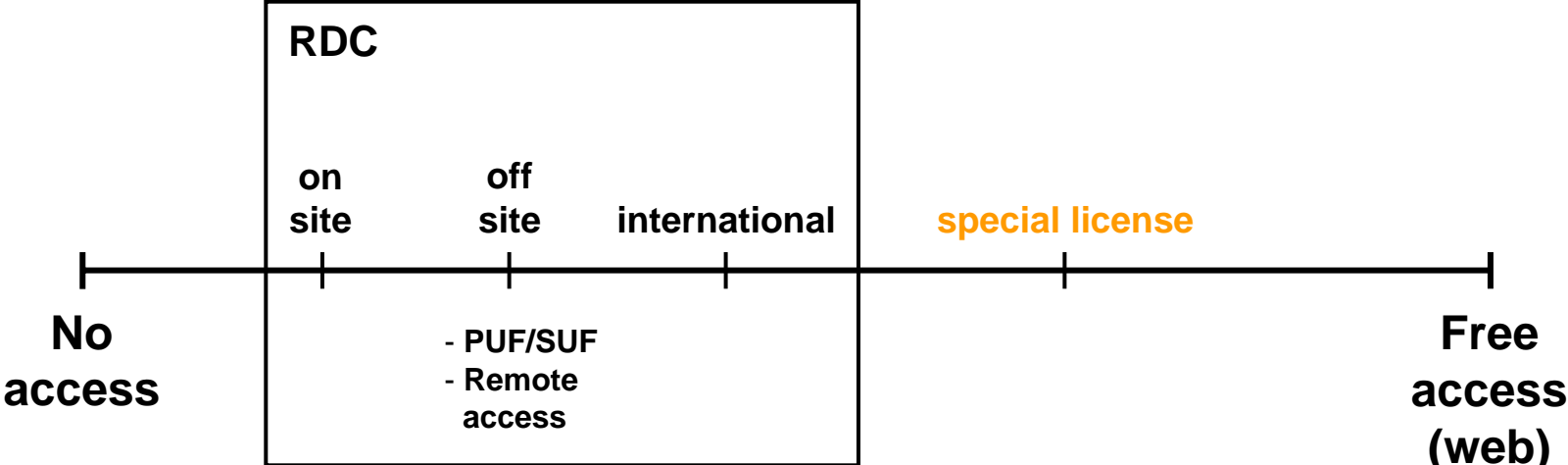


## **Providing Access to Microdata: The German Experience and the RDC of the BA in the IAB**

**Stefan Bender**  
*(Institute for Employment  
Research, Germany)*



# Modes of access



Based on: Felix Ritchie



## Topics of the talk

- Data Access in Germany: small introduction
- The RDC of the Federal Employment Agencies (BA) in the Institute for Employment Research (IAB)
- New Data Infrastructure Projects in the RDC (WeLL, KombiFiD)
- Data Access
- Conclusion



## Data Access in Germany: The 80's

- Law for the census (1987): privilege for research
- Federal Statistics Law: from absolute to factual anonymity
- First scientific use files were published



## Data Access in Germany: The 90's

- Constant pressure of the scientific community „Set the data free“
- The „Commission to Improve the Statistical Infrastructure in Cooperation with the Scientific Community and Official Statistics“ (*KVI 1999, report 2001*) established by the Federal Ministry of Education and Research (*BMBF*).
- German Council for Social and Economic Data (RatSWD).  
The Council's main purpose is to advise in the development of the German data infrastructure for empirical research in the social and economic sciences.
- Establishment of Research Data Centers (RDC) by data producers and Data Service Centers (DSC). At the beginning all co-financed by the BMBF.

## **RDCs provide researchers access to micro data for non-commercial empirical research in the fields of social security and employment**



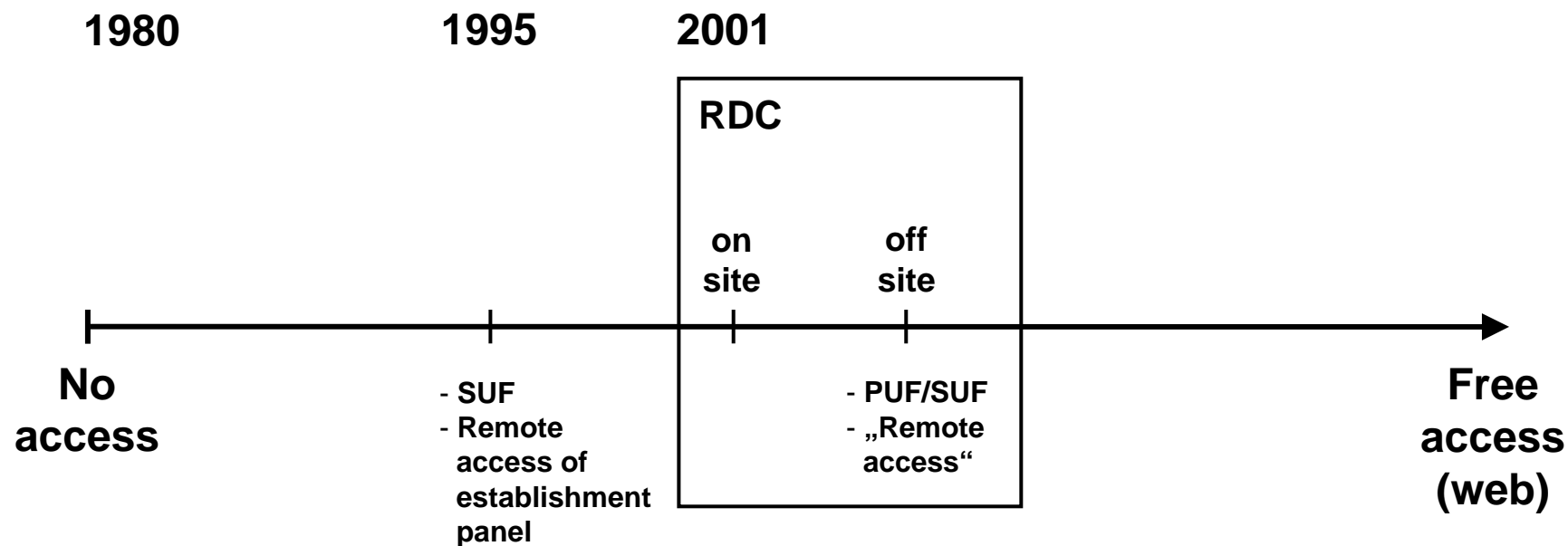
- Advisory service on data selection and data access
- Handling of remote data access
- Assistance and support for visiting researchers
- Online data documentation and documentation of methodological aspects of data
- Clarification of questions on data protection
- Updates of scientific use files and other research datasets
- Organization of workshops and user conferences



## RDC: Advantages for research and politics to get free access

- „democratic“ access to micro data (RatSWD guidelines for RDCs)
- more high quality German research
  - output
  - publications
  - competitiveness
- better base for political decision process

# Development in Germany







## Networking with IDSC

- Translations
- Remote Execution (JoSuA)
- Meta Data (DDI Alliance)



# RDCs in Germany: Summary



## Overall

- RDC/DSC is a story of success
- research access is nearly free
- network of RDC/DSC
- “quality control” (RatSWD, guidelines)



## RDC of the BA in the IAB – general idea

- Better international data access (remote data access)
- Development of data infrastructure
  - National access to IAB/BA microdata
  - International access to IAB/BA microdata
  - Access to combined German micro data (other RDC and commercial micro data)
  - Access to combined and harmonized international micro data.
- Information:  
ZAF-article of Jörg Heining (5 years of RDC) and annual reports.



## RDC of the BA in the IAB – basic information



- Started in April 2004; positive evaluation in April 2006, since December 2006 100% financed by the Federal Employment Agency
- 6 researchers, 3 non-researchers, office in Nuremberg
- Data based on:
  - the notification process of the social security system,
  - the internal procedures of the Federal Employment Agency
  - data from IAB-surveys (e.g. IAB Establishment Panel).
- Available data:
  - Household or individual surveys (like PASS or WeLL)
  - IAB Establishment Panel, Establishment History Panel
  - IAB Employment Sample, BA Employment Panel, Integrated Employment Biographies Sample of the IAB
  - Linked Employer Employee of the IAB



## Tasks of RDCs

- Advice on data selection and data access, support guest researchers, data documentation, documentation on methodological aspects of data, clarify questions on data protection, update data, organize workshops, user conferences

## Further tasks of our RDC

- Publications (FDZ-Datenreporte, FDZ-Methodenreporte)
- English
- Research

## Data Sources

- Administrative data
  - Definition: “Administrative data is information that is routinely collected by organisation, institutions, companies and other agencies in order that the organisation can carry out, monitor, archive or evaluate the function or service it provides” (Calderwood and Lessof 2006, p2)
  - notification procedure of the Social Security System (health-, pension-, unemployment- and long-term care insurance)
  - administrative data of the Federal Employment Agency
- Surveys

# Notification procedure of the Social Security System



- Intention of data collection
  - calculation of social security contribution and entitlement of benefit
  - statistics
- Procedures
  - every employee and marginal worker covered by social security system (notification requirement)
  - identification: social security number and establishment number
  - annual, begin or end of employment, employment interruption, change of health insurance
- What is notified?
  - no information about: civil servants, freelancer





# Administrative data of the Federal Employment Agency



- Receipt of benefit
- Job search
- Participation in active labour market programs  
→ computer-aided processes
- Federal Employment Agency is not allowed to store data for a long time  
→ data is transmitted to the Institute for Employment Research



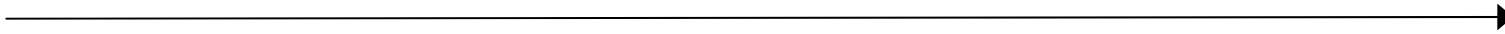
# Administrative Data

- Advantages:
  - individual information
  - complete coverage of time
  - important administrative questions are reliable
  - to report statistics and longitudinal studies
  - to create research data
  
- Disadvantages:
  - depend on the administrative system's population
  - changes in administrative systems make comparison difficult
  - lower quality of less important administrative information



apprenticeship

pension

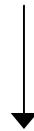


Employment history covered by social security system (1975)

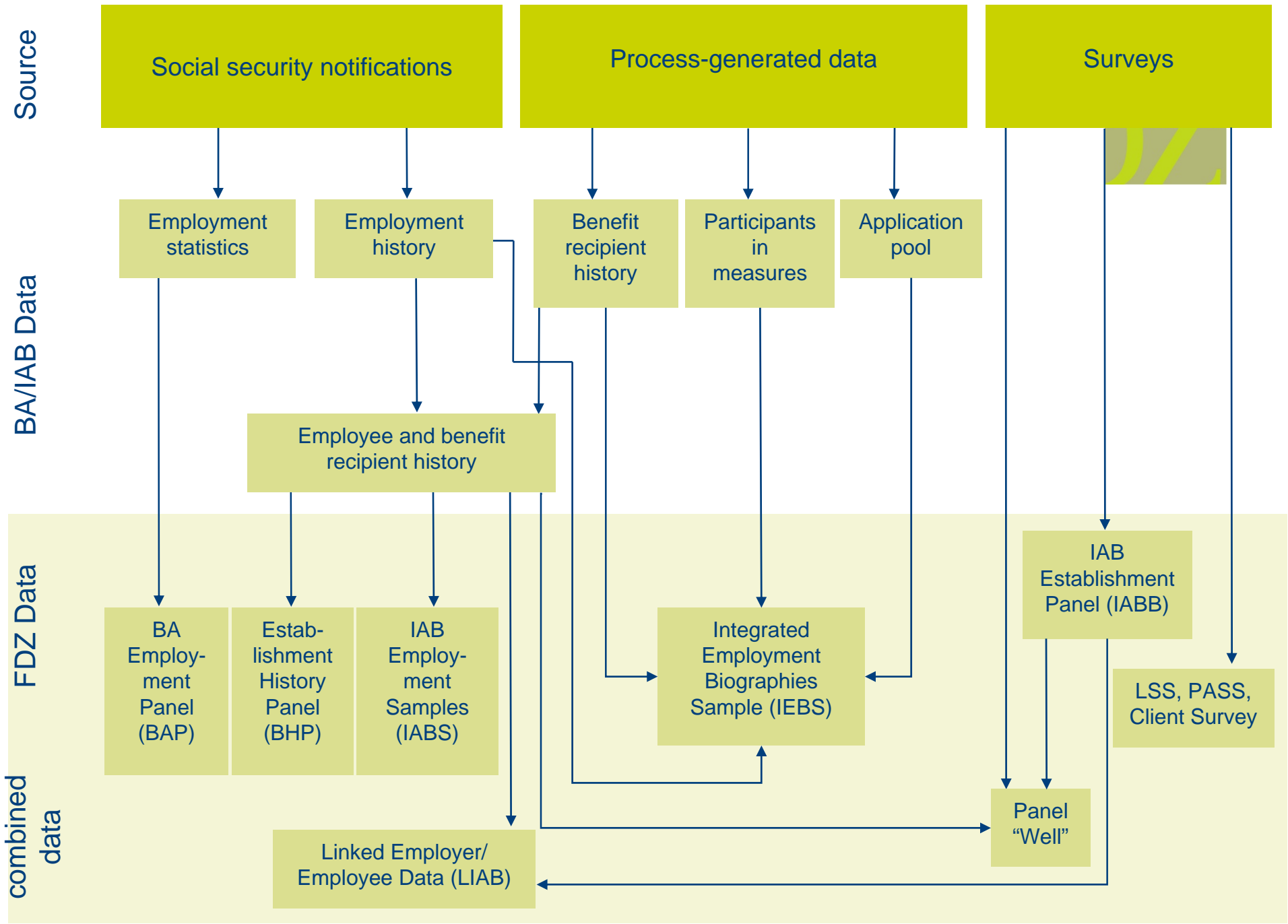
Benefit receipt (1975)

Job search (2000)

Participation in active labor market programs (2000)



IAB data is merged data



# Research data



	Individual data	Establishment data	Household Data
Administrative data	BA Employment Panel	Establishment History Panel	
	IAB Employment Samples		
	Integrated Employment Biographies Sample		
Surveys	Life Situation and Social Security 2005	IAB Establishment Panel	Labour Market and Social Security
Combined data	Linked Employer/Employee data		

## Individual data (administrative data)



	BA Employment Panel	IAB Employment Samples	Integrated Employment Biographies Sample
Period covered	1998-2007	1975-2004	1990-2008
Data collection	2% random sample	2% (or 1%) random sample	2,2% random sample
Time reference	Longitudinal data on a defined reference date (quarterly)	Longitudinal data on a daily basis	Longitudinal data on a daily basis
Number of cases	517,000 - 621,000	1,360,949	1,487,835
Topics	Socio-demographic characteristics, employment and receipt of benefit information, participation in measures, regional and <b>establishment information</b>	Socio-demographic characteristics, employment and receipt of benefit information, regional and <b>establishment information</b>	Socio-demographic characteristics, employment and receipt of benefit information, <b>information about participation in measures and jobsearch</b>

## Establishment data

	Administrative data	Survey
	Establishment History Panel	IAB Establishment Panel
Period covered	1975-2006	1993-2007
Method of data collection	50% random sample or random sample stratified by 8 establishment sizes	representative random sample, stratified according to establishment size, industry and federal state
Number of cases	between 650,000 and 1.25 million	between 4,265 and approx. 16,000 establishments per year
Time reference	30th June of each year	30th June of each year
Topics	total number: of employees, full-time employees, employees by main occupation, unpaid employees, female employees, employees by nationality, employee age structure, division of employees by educational and vocational qualifications, division of employees by occupational status, division of employees by person group, wage structure of full-time employees	Development of employment, business policy, investments, in-house innovations, government subsidies, personnel structure, vocational training and apprenticeship places, recruitments and dismissals, personnel search, wages, working hours, training programs, general company information, different main topic per year

# Combined data



	Linked Employer Employee Data	
	IAB Establishment Panel	Employee and Benefit Recipient History
Period covered	1993-2006	1993-2006
Method of data collection	Random sample	Individual employed in an establishment of the IAB Establishment Panel
Number of cases	Approx. 4,000-16,000 per year	1,8 to 2,5 mill. per year
Time reference	Annual panel	30th of June
Topics	e.g. development of employment, business policy, investments, in-house innovations, government subsidies, personnel structure, vocational training and apprenticeship places, different main topic per year	e.g. socio-demographic characteristics, employment and receipt of benefit information



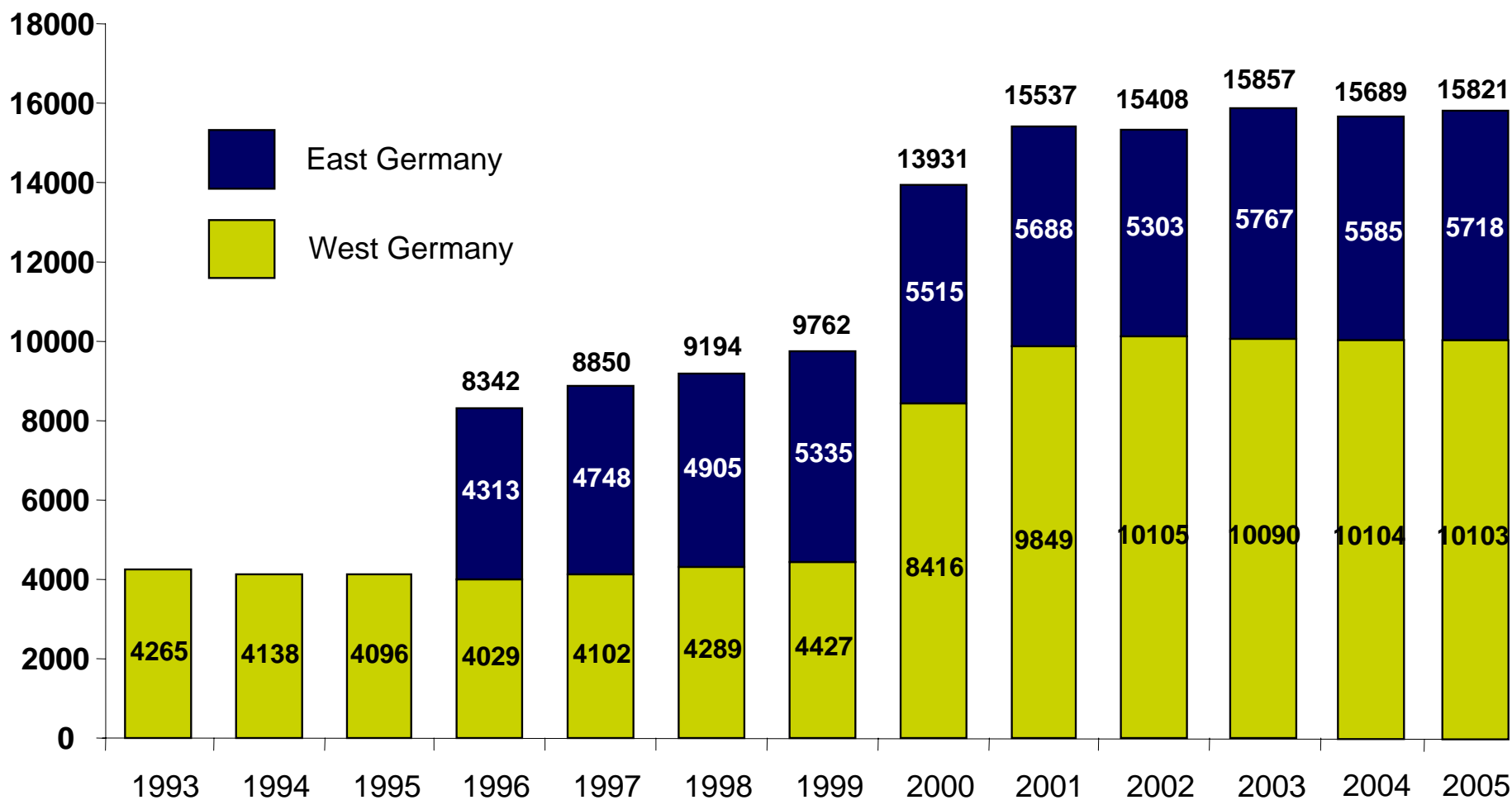


## IAB Establishment Panel

- annual representative survey on establishments (approx. 16,000 )
  - West Germany: 1993-2007
  - East Germany: 1996-2007
  - stratified sample: 17 industries and 10 size ranges
  - population: all establishment with at least one employee covered by social security on 30th June in the previous year
  - since 2001 representative on federal state level
- Contents: **development of employment, business policy and development, investments, in-house innovations, government subsidies, personnel structure, vocational training and apprenticeship places, recruitments and dismissals, personnel search, wages, working hours, training programmes, general company information**
- annual main focuses



# Number of interviewed firms





## Linked Employer/Employee Data

- matched data from the IAB Establishment Panel and the individual data from the IAB (BLH)
- simultaneous analysis of the supply and demand sides of the German labour market
- 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 (30th June).
  - 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.

# Further Training as a Part of Lifelong Learning (WeLL) - Motivation



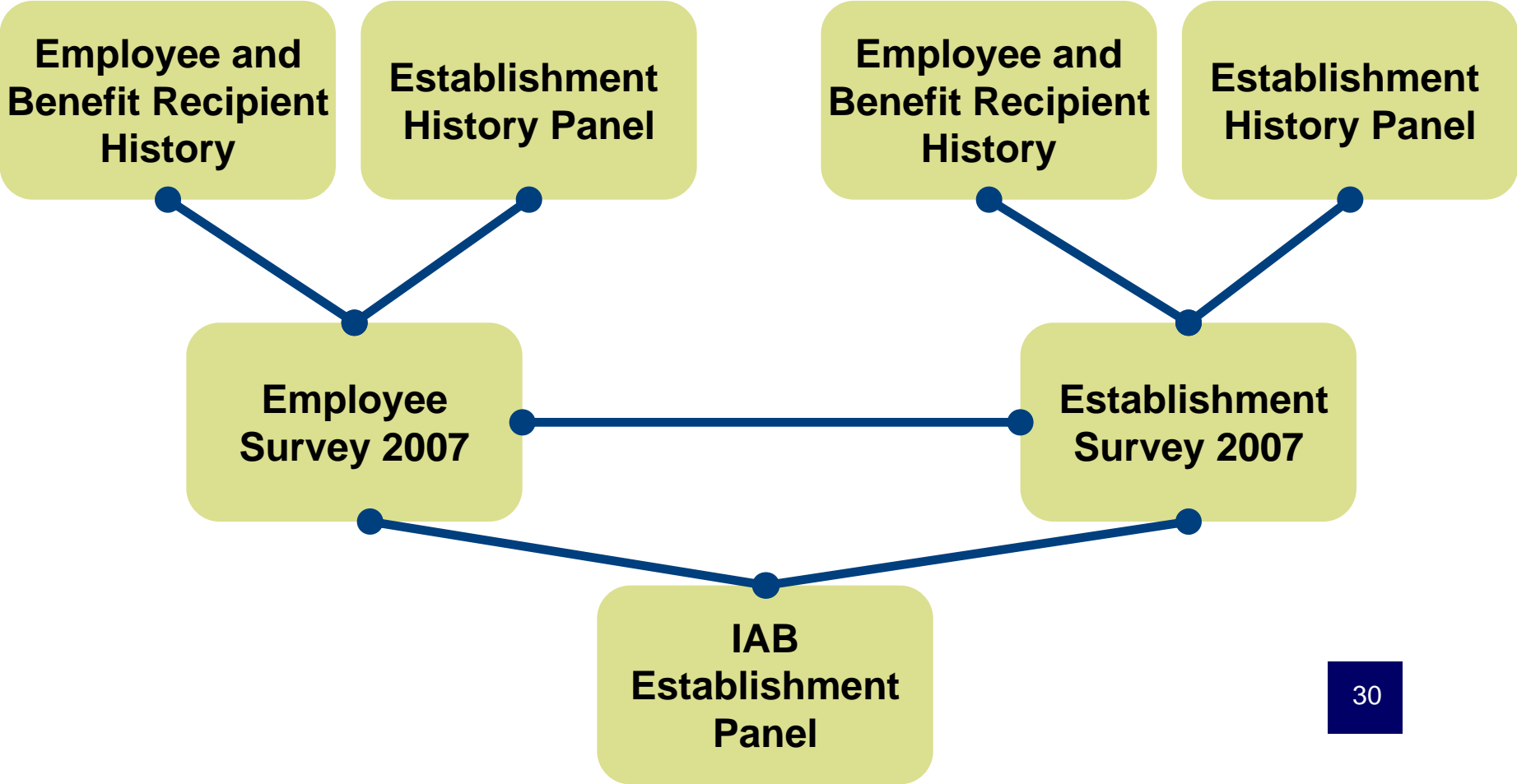
- Knowledge is an important capital in modern societies
- High qualified employees are essential for the performance of the companies
- Quick changes of knowledge
- Flexibility and continuous learning: in particular on-the-job training
  
- Are there differences in different countries?  
Comparision with nearly the same French data

## Further Training as a Part of Lifelong Learning (WeLL) - Questions



- Relationship between organisational changes and training
- Volume of the support of vocational training by the company
- (Non) Monetary costs and earnings
- Aims and reasons for participation
- Barriers for participation
- Cost absorption
- Expected and real benefit (short-, long-term)

# Further Training as a Part of Lifelong Learning (WeLL) - Linked Data



●—● Linkage based on the agreement of the employees or establishments

# KombiFiD – Combined Firm Data for Germany



**Combining firm datasets of the single data producers for the first time in Germany; time horizon 1993 - 2006**



**New dataset for researchers with:**

- a variety of variables,
- a variety of data generating processes
- a variety of economic sectors

**Reduce respondent burden for the firms**

## Selected datasets



### 1. Federal german business register (URS)

- Implemented in late 90s
- Regulated by european law (Council regulation No 2186/93)
- Master file
  
- Identifier:
  - Unique business identifier
  - All corresponding establishment identifiers
  - All corresponding tax numbers



## Selected datasets



### 2. Mix of official (sample) surveys

- Cost structure surveys  
Information about added value, turnover, subsidies,...
- Salary and wage-earning set up surveys  
Information about wages, salaries, working hours, vacation entitlement

### 3. Tax data

- Turnover tax
- Corporate income tax

## Selected datasets

Institut für Arbeitsmarkt-  
und Berufsforschung  
Die Forschungseinrichtung der  
Bundesagentur für Arbeit



### 1. Process generated data:

Personal data from the employment notification  
annual aggregation on establishment level (establishment number)

Establishment-History-Panel (BHP)

### 2. Survey data:

IAB Establishment Panel

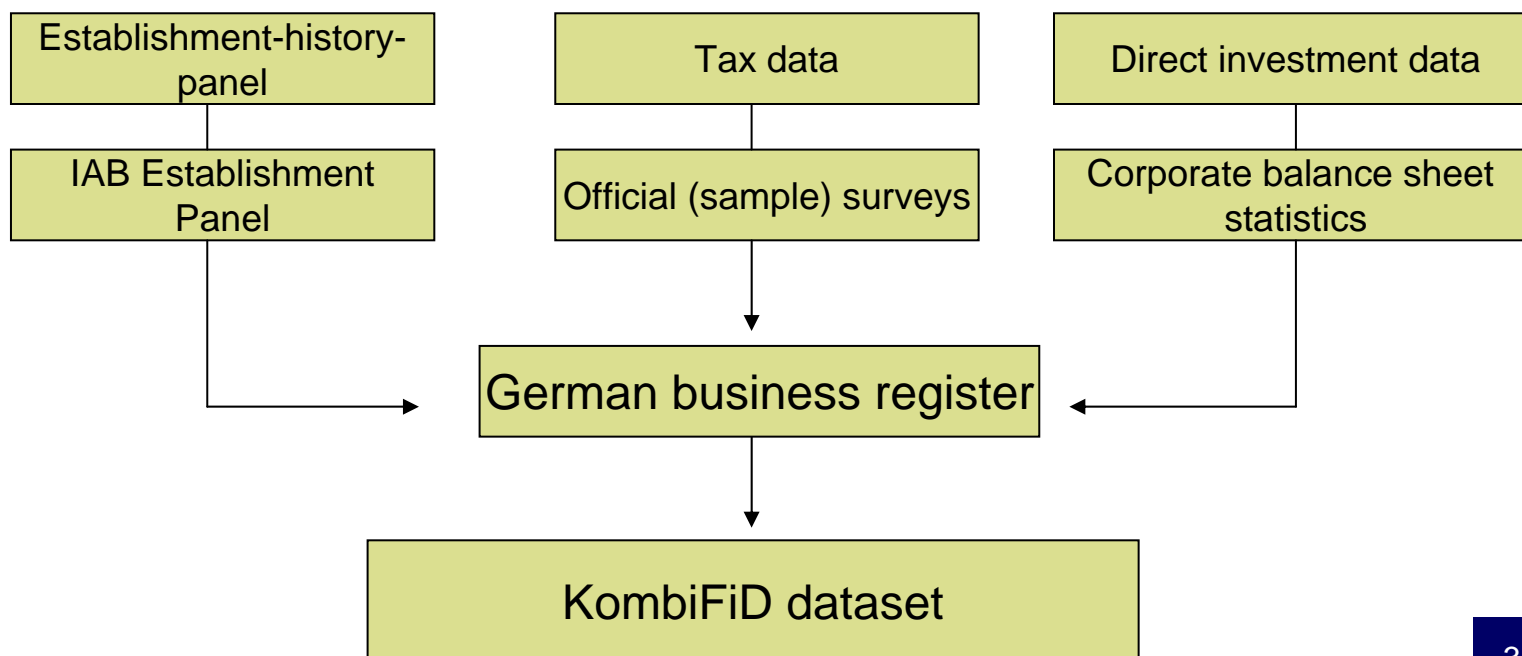
- annual panel (10.000 establishments)
- Conducted by the IAB since 1993
- Information about various and changing topics (e.g. labour demand)

## Selected datasets



### Process generated data:

- **Microdatabase Direct Investment (MiDi)**  
annual firm data on foreign direct investment stocks  
due to german trade and payment regulation
- **Corporate Balance Sheet Statistics (Ustan)**  
annual balance sheet statistics of non-financial enterprises  
(main economic sectors: manufacturing, construction, wholesale  
and retail trade)





## **Combined Patent and Administrative Data**

- **German and European patent office data for 2002.**
- **Accurate data, because payments are done with the data.**
- **Record linkage of names and addresses**
- **Most inventors are employees in a firm**
- **Result: Linked Employer Employee Data for Germany (over time)**



## Data Access at the RDC

Datasets	Data Access		
	On-site Use	Remote Data Access	Scientific Use File
IAB Employment Samples	✓	✓	✓
BA Employment Panel	✓	✓	✓
Integrated Employment Biographies Sample	✓	✓	✓
IAB Establishment Panel	✓	✓	Project
Linked-Employer/Employee Data	✓	✓	
Establishment History Panel	✓	✓	
Individual and Household Surveys (LSS 2005, PASS, WeLL, §6c)			✓

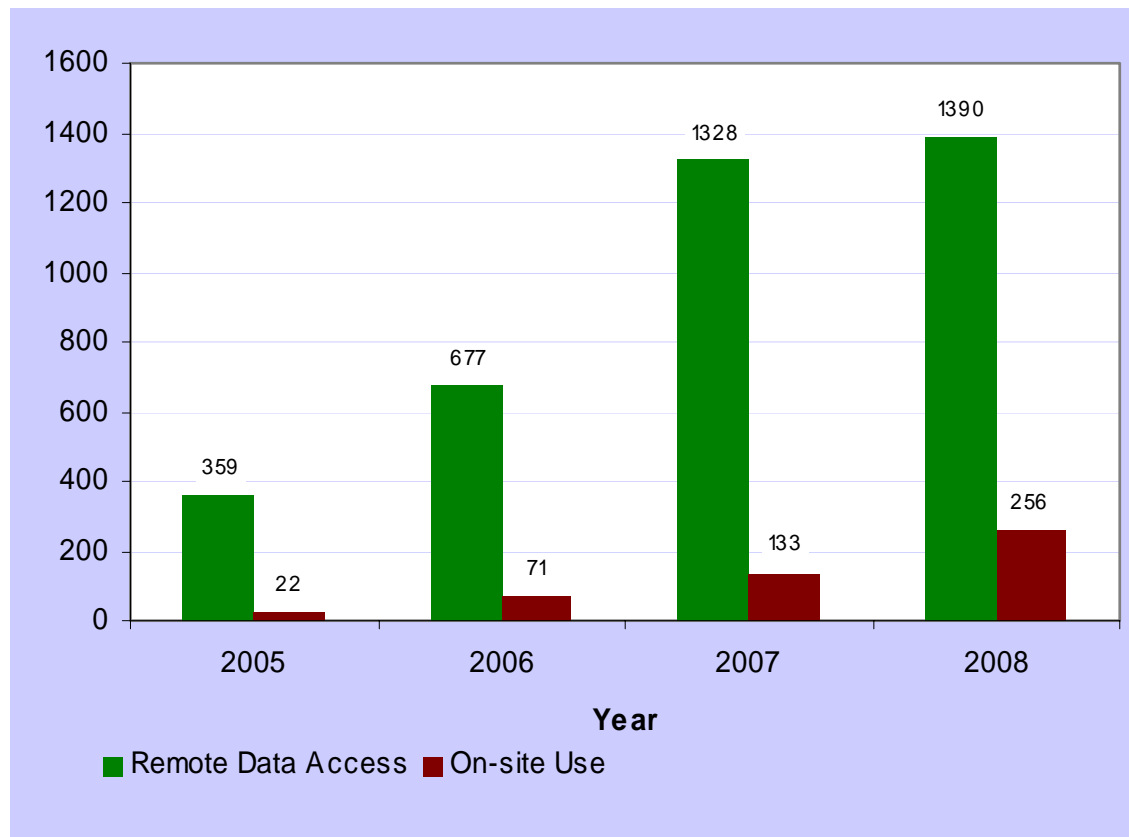


## Data Access

- Restriction: access for non-commercial empirical research in the fields of social security and employment
  - On-Site Use
  - Remote Data Access
  - Scientific Use Files
  
- **No costs**
- **Financial support for guest researchers from abroad**

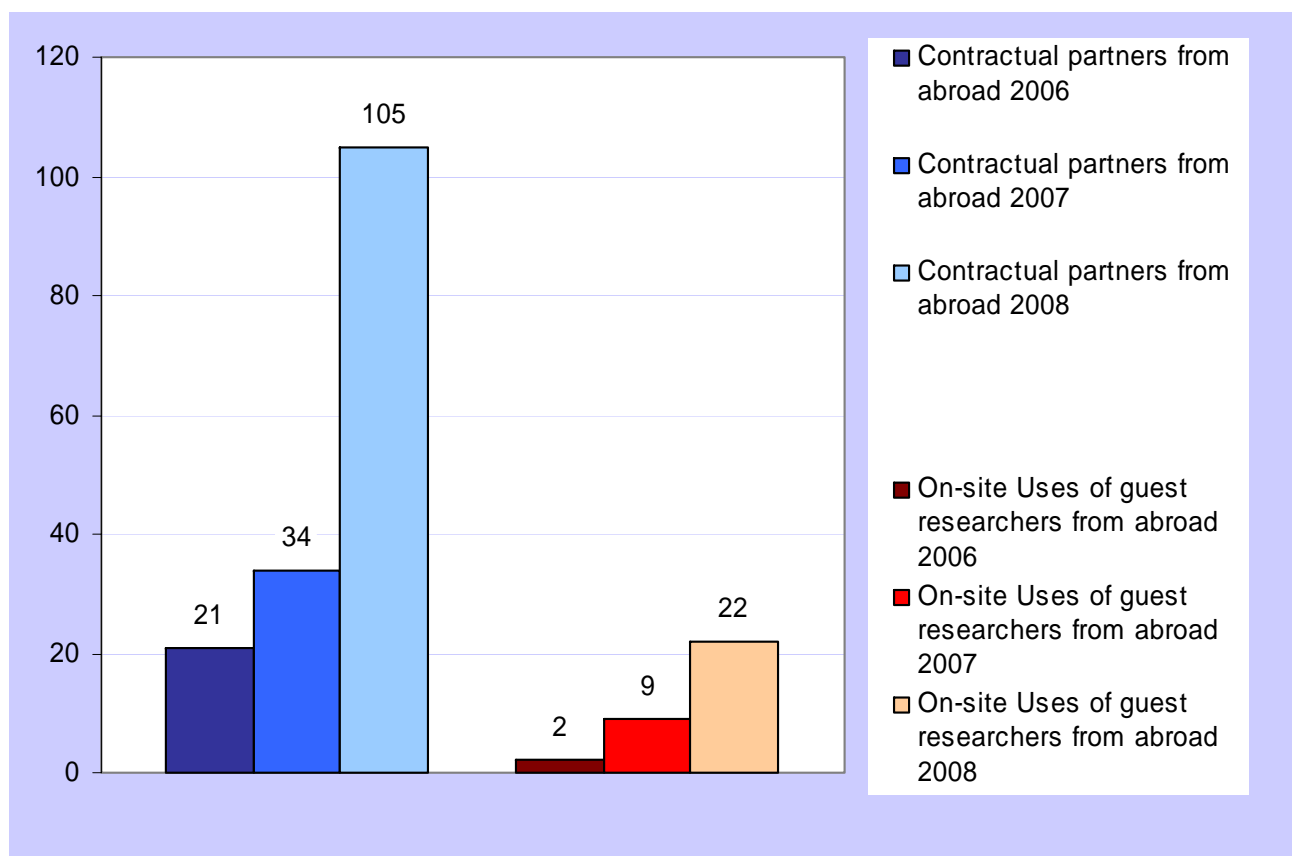


# # of remote data accesses and on-site uses





# # of users outside Germany (contractual partners, guests)





## Advantages for the IAB / RDC

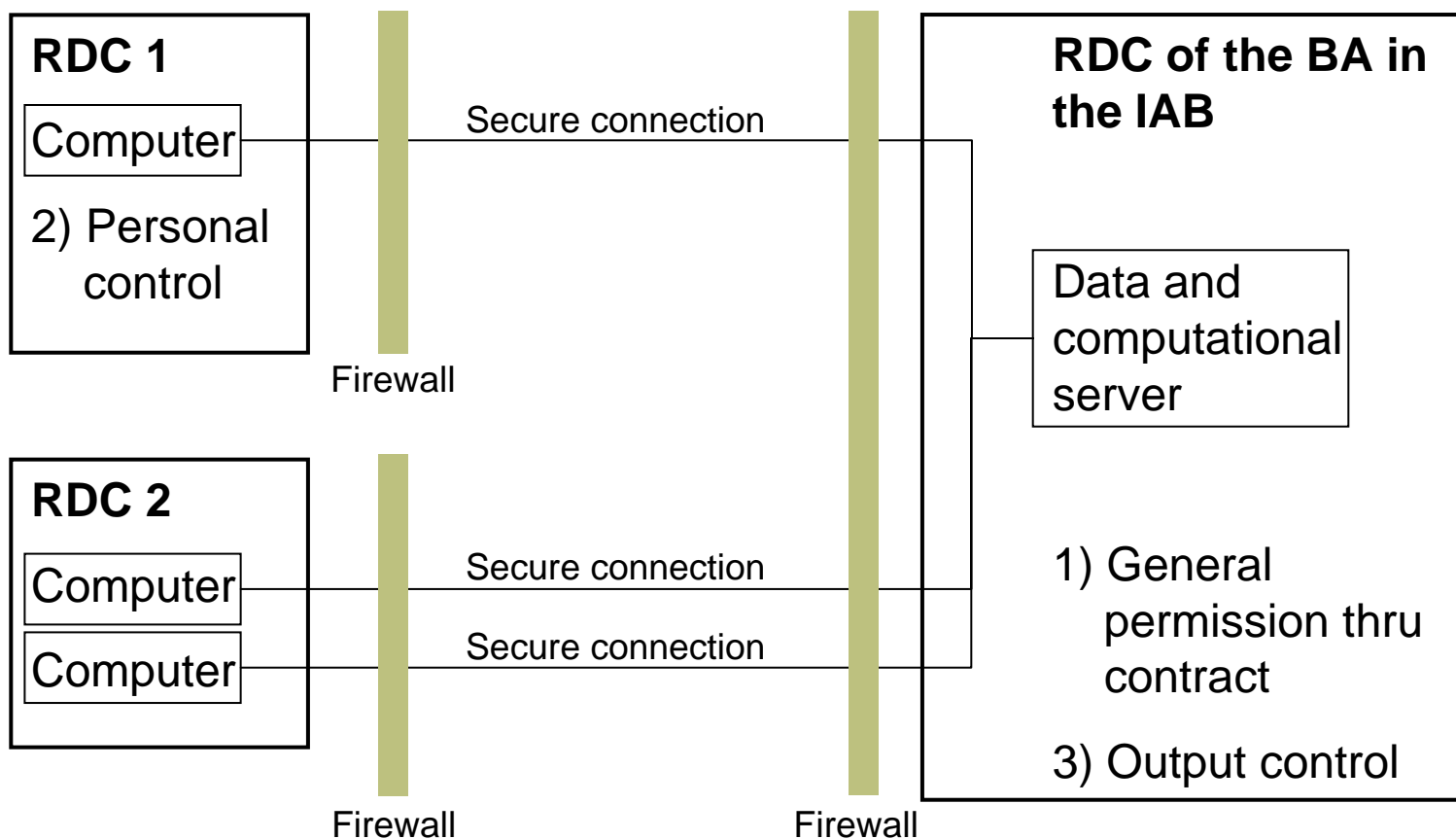
- More research on our own data
- More data quality
- Greater network on researchers
- More competence on data and research knowledge
- International acceptance of IAB-/RDC researcher
- More reputation of German research



## “RDC in RDC”

- Remote Access in Germany is far behind solutions like in Denmark or in the Netherlands.
- We have to fulfil requirements of data protection and data security.
- Problem: not every country has the same data protection standards/regulations/laws.
- Nearly the same standards in RDCs all over the world (or the standards can be established).
- Main problem for German data protection: how to control who is sitting at the PC. That is possible in other RDCs.

# Model of the RDC in RDC approach



## Advantages for you

- Interesting (unique) data sets
- Easy access
- Financial support
- High publication possible (American Economic Review, Quarterly Economic Review, Journal of Labor Economics, Review of Economics and Statistics, ...)
- Nice German medieval city
- and Bavarian Food





Contact:  
Stefan.Bender@iab.de  
<http://fdz.iab.de/en.aspx>



## BA Employment Panel (BAP)

- panel data of employees covered by social security on reference date:
  - cross sectional dataset: 40 waves 1998-2007
  - 2% sample drawn from the quarterly employment statistics of the BA
  - only one employment per person
  - data extension since 1999: marginal part-time employees
  - information on unemployment and participation in measures
  - many characteristics of establishments: size, industry, region, structure of employees



## Establishment History Panel (BHP)

- cross sectional data on establishments
  - all establishments in Germany (covered by the Employee and Benefit Recipient History - BLH)
  - with at least one employee covered by social security at reference date
  - period: 1975-2006
  - reference date: 30th June
  - contents:
    - industry and location of the establishment
    - number of employees liable to social security (since 1999 marginal part-time employees)
    - broken down by gender, age, occupational status, qualification and nationality
    - quartiles of ages and wages
- long-term strategy: Linkage with BAP, IABS or IEBS





## Data Access I

- Definition of social data:
  - individual data collected by administrations which contain individual information on the personal or material circumstances of a determined or determinable natural person
  - The IAB Establishment data are also treated as social data.
  - Aggregated tables are also social data if
    - one establishment could be identified or
    - the number of cases is too small.
  - Absolutely anonymous data aren't social data.

## Data Access II

- Social data are particularly sensitive and therefore subject to the confidentiality regulations of the German Social Code.
- legal permission for scientific use of the social data (“Wissenschaftsprivileg”)
  - § 75 SGB X
    - research project in the field of social security and assistance benefits
    - public research interest
    - authorization by the Federal Ministry of Labour and Social Affairs(BMAS)
  - § 282 Abs. 7 SGB III
    - factually anonymous data
    - employment and occupational research



## Data Access III

- factually anonymous scientific use files (SUF)
- remote data access
- data transmission to third party (Social Code Book X, Section 75)
  - costly procedure
  - contact: Data Centre of the Statistical Office of the BA
- on-site use
  - Social Code Book X, Section 75
  - standardised procedure

# RDCs and Data Service Centers in Germany



- Research Data Centers (which fulfill the guidelines):
  - the Federal Statistical Office (*Statistisches Bundesamt*) and the Statistical Offices of the German Länder (*Statistische Landesämter*)
  - the Federal German Pension Insurance (*Deutsche Rentenversicherung Bund*)
  - the Institute for Employment Research (*IAB: Institut für Arbeitsmarkt- und Berufsforschung*) within the Federal Employment Agency (*BA: Bundesagentur für Arbeit*)
- and Data Service Centers:
  - the Institute for the Study of Labor (*IZA: Institut zur Zukunft der Arbeit*).
  - the Center for Survey Research and Methodology (*ZUMA: Zentrum für Umfragen, Methoden und Analysen*)



## DSC at GESIS German Microdata Lab

- Established in 1987. The DSC - Project started in 2003 (cofinanced by the Federal Ministry of Education and Research until 2009); positive evaluation in May 2006
- Special focus on Mikrozensus
- Location: Mannheim
- Staff: 6 researchers, 1 non-researcher
- Main purposes:
  - research based service for empirical researchers
  - brokering service for data providers
  - research on methodological and substantial topics using official microdata
- Instruments
  - comprehensive metadata on Mikrozensus (<http://www.gesis.org/MISSY>)
  - workshops and user conferences
  - access support to German official microdata sets

## DSC at the Institute for the Study of Labor (IZA)



- DSC - Project started in 2005 and is cofinanced by the Federal Ministry of Education and Research until 2009; positive evaluation in February 2007
- Special focus on labor economics
- Location: Bonn
- Staff: 2 researchers, 3 non-researchers
- Main purposes:
  - support service for empirical researchers
  - brokering service for data providers
  - international promotion of German microdata sets
- Instruments
  - standardized meta information on data sets (<http://metadata.iza.org>)
  - access support to German official microdata sets (also as a prototype remotely from abroad via JoSuA)

# RDC of the Federal German Pension Insurance



- RDC - Project started in 2004 and is co-financed by the Federal Ministry of Education and Research until 2008; positive evaluation in July 2005
- Locations: Berlin, Würzburg
- Staff
  - permanent: 4 researchers, 3 non-researchers
  - temporary: 4 Ph.D. students, 1 non-researcher
- Data based on:
  - the notification process of the social security system
  - the internal procedures of the Pension Insurance
- Data available especially as Scientific Use Files for off site use:
  - retiring statistics (cross sectional and longitudinal)
  - insured statistics (cross sectional and longitudinal)
- All FDZ-RV-data are available for free



## RDC of the Federal Statistical Office

- RDC - Project started in 2001 and is financed by the Federal Ministry of Education and Research
- Locations: Wiesbaden, Berlin, Bonn
- Staff: 8 researchers, 6 non-researchers
- Well balanced range of services (number of users in 2007)
  - Scientific Use Files / Public Use File (112)
  - Remote execution and data laboratories (40)
  - free access on the web to CAMPUS Files
- Available data: microdata from almost all official statistical surveys
- Research projects:
  - Anonymized Business Panel Data
  - KombiFiD (Combination of Enterprise Data in Germany )



## RDC of the Statistical Offices of the Länder



- The RDC was set up in 2002 and is supported by the Federal Ministry of Education and Research.
- It has 16 regional locations - one in every federal state.
- Four different forms of data access are provided: Public Use Files, Scientific Use Files, scientific workstations and controlled remote data processing.
- About 70 statistics from all parts of official statistics are available (see [www.forschungsdatenzentrum.de](http://www.forschungsdatenzentrum.de)).
- Since 2004 data sets for more than 300 scientific projects have been ordered via the RDC.
- Main research project: Merging of economic and environmental statistics and developing of panel-data (AFiD = Amtliche Firmendaten für Deutschland).

## IAB Employment Samples (IABS)

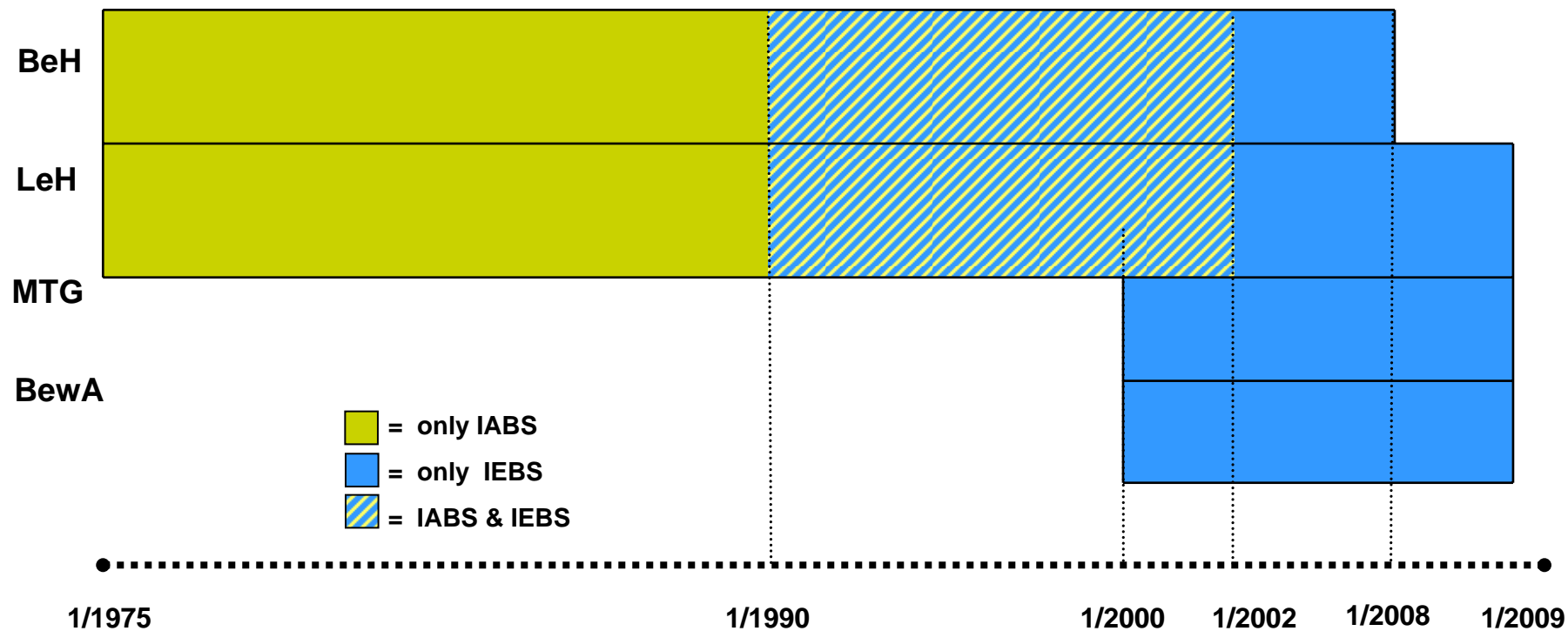
- 2% (or 1%) sample drawn from the employee and benefit recipient history (BLH)
  - information on employment from the social security notifications
  - information on benefit receipt from the process-generated data of the BA
- different versions of the IABS:
  - SUF: basic file 1975-1995 (West/East Germany, detailed information on occupations, industries and employers)
  - SUF: regional file 1975-2004 (343 regions based on districts, less detailed information on occupations, industries and employers)
  - weakly anonymous version 1975-2004

# Integrated Employment Biographies Sample (IEBS)



- Differences between IEBS and IABS
  - additional information on participation in active labour market programs and on job-seeking
  - period covered:
    - employment: 1990-2007
    - benefit receipt: 1990-2008
    - participation in active labour market programs, job-seeking: 2000-2008
  - IEBS data are edited in clearly defined cases only.
- Scientific Use File available

# Comparison of IABS and IEBS

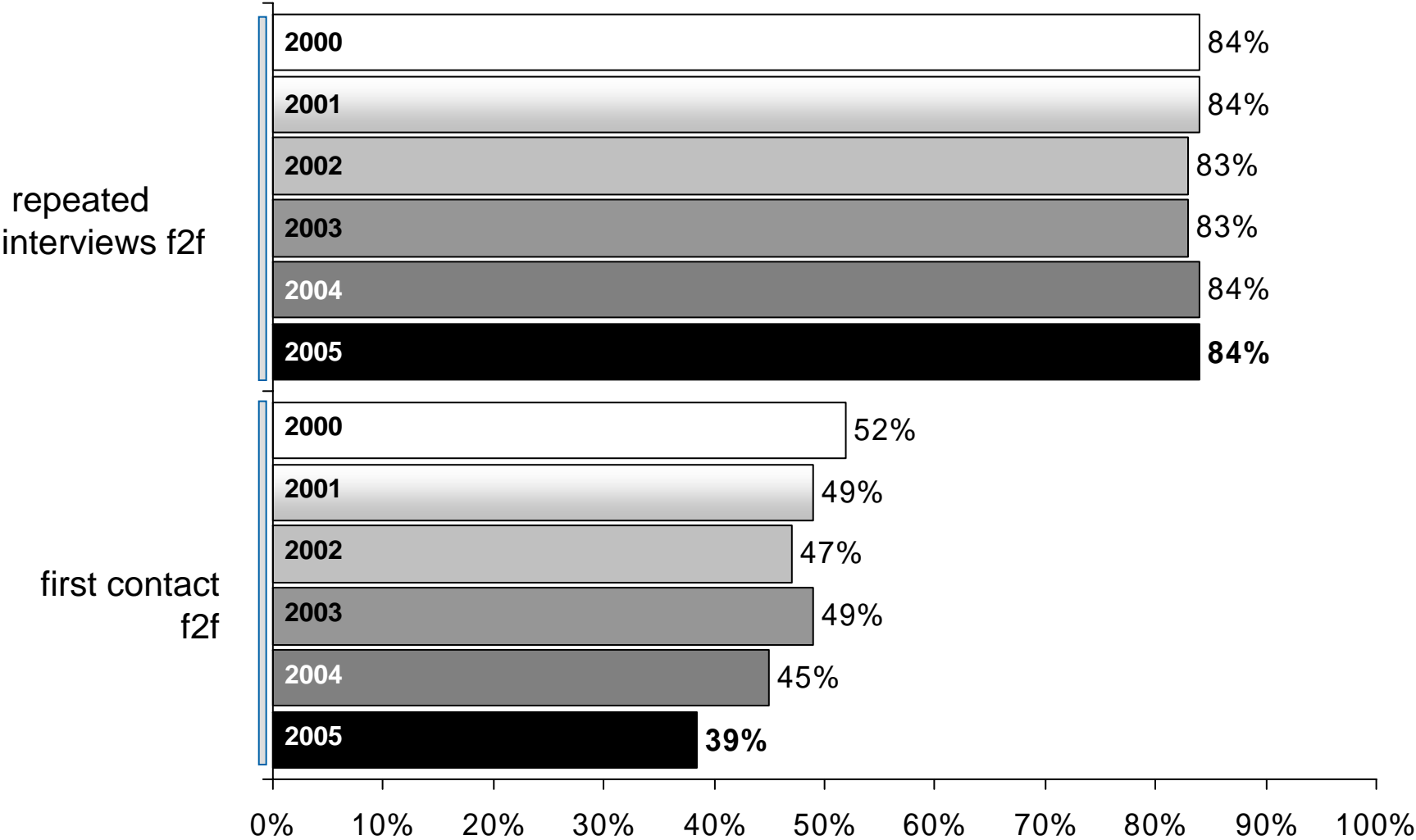


## Individual data (survey)



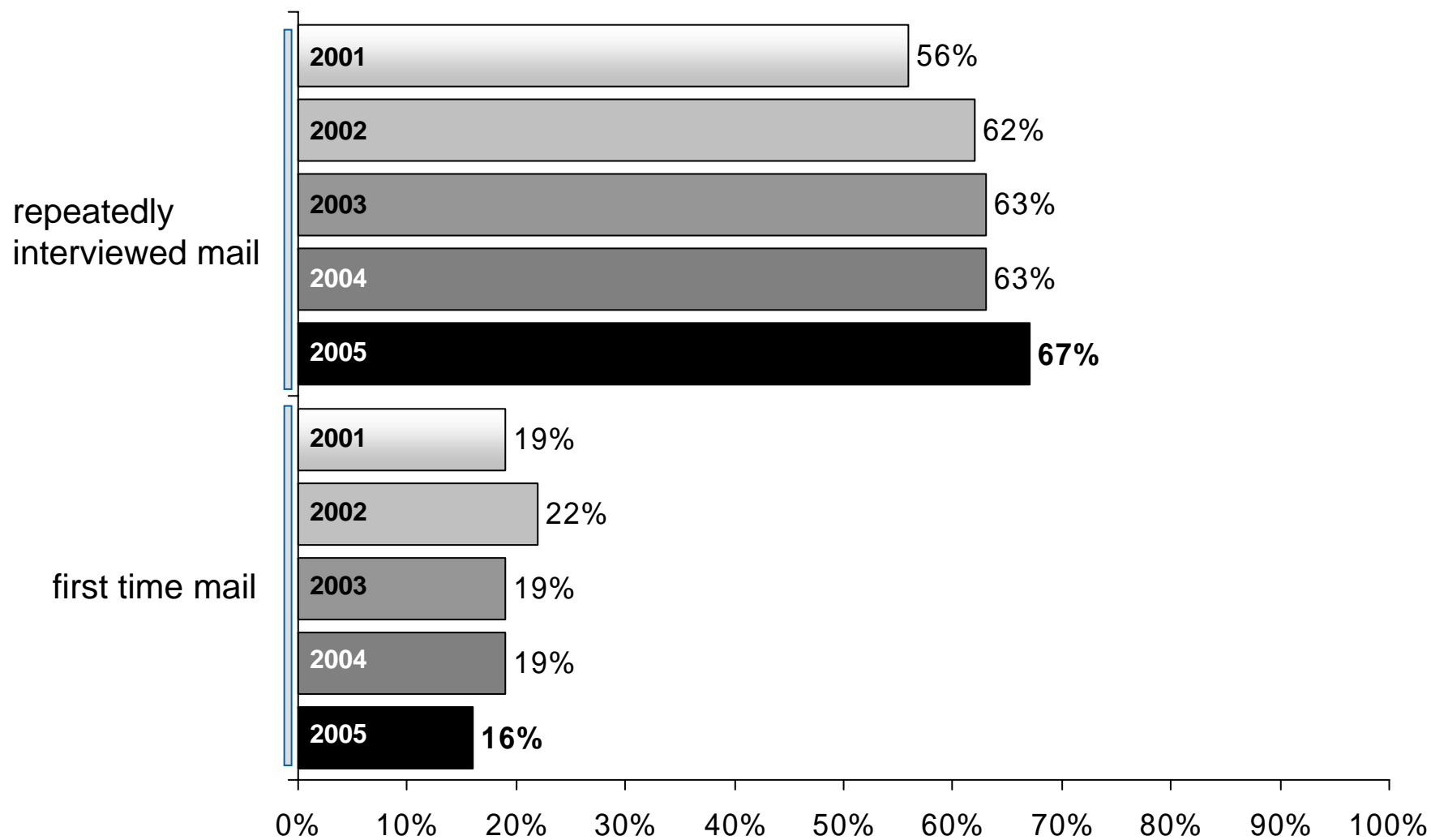
	Life Situation and Social Security 2005 (LSS 2005)	Panel Study 'Labour Market and Social Security' (PASS)
Period covered	2005-2006	2006/2007 (first wave)
Method of data collection	Clustered random sample	1 stage: Random sample 2 stage: clustered
Time reference	Cross-sectional survey	Cross-sectional survey
Study unit	Recipients of Unemployment Benefit II	Recipients of Unemployment Benefit II and low wage earners
Number of cases	20,832	12,794 households, 18,954 individuals
Topics	Household, social and financial situation, social network and social support, Employment and education cycle, participation in measures, attitudes, job search, regional context	<i>Nearly the same as Life Situation and Social Security 2005.</i>

# Response Rate 2005: f2f





## Response Rate 2005: mail





## Annual Questions

- development of employment
- business strategies
- vocational training
- structure of the workforce
- recruitment of personnel/reduction of workforce
- remuneration/wage agreements
- investments
- structural information





## Modular Questions

- Every two years
  - public subsidy
  - working time and flexibility
  - advanced vocational training
  
- Every three years
  - organisational change
  - innovation



## Special Topics

- 2000: skill needs
- 2001: innovation und modern technology
- 2002: older workers
- 2003: flexibility
- 2004: financing of innovations, investments and vocational training
- 2005: recruitment of employees, environmental protection
- 2006: factors of location, employment guarantee foreign investments (before: 1998)