



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

# Administrative Data in the IAB Metadata Management System

North American DDI Conference  
University of Kansas

2013/04/02, Lawrence, KA (USA)

David Schiller (IAB)

Ingo Barkow (DIPF)



# Outline

1. Introduction
2. Data source provided by IAB
3. Need for a useful data documentation
4. Example of an exemplary software implementation
5. Administrative data; only a step into the future of research data
6. Conclusion

Administrative Data in the IAB Metadata Management System

# INTRODUCTION

# FDZ of BA at IAB

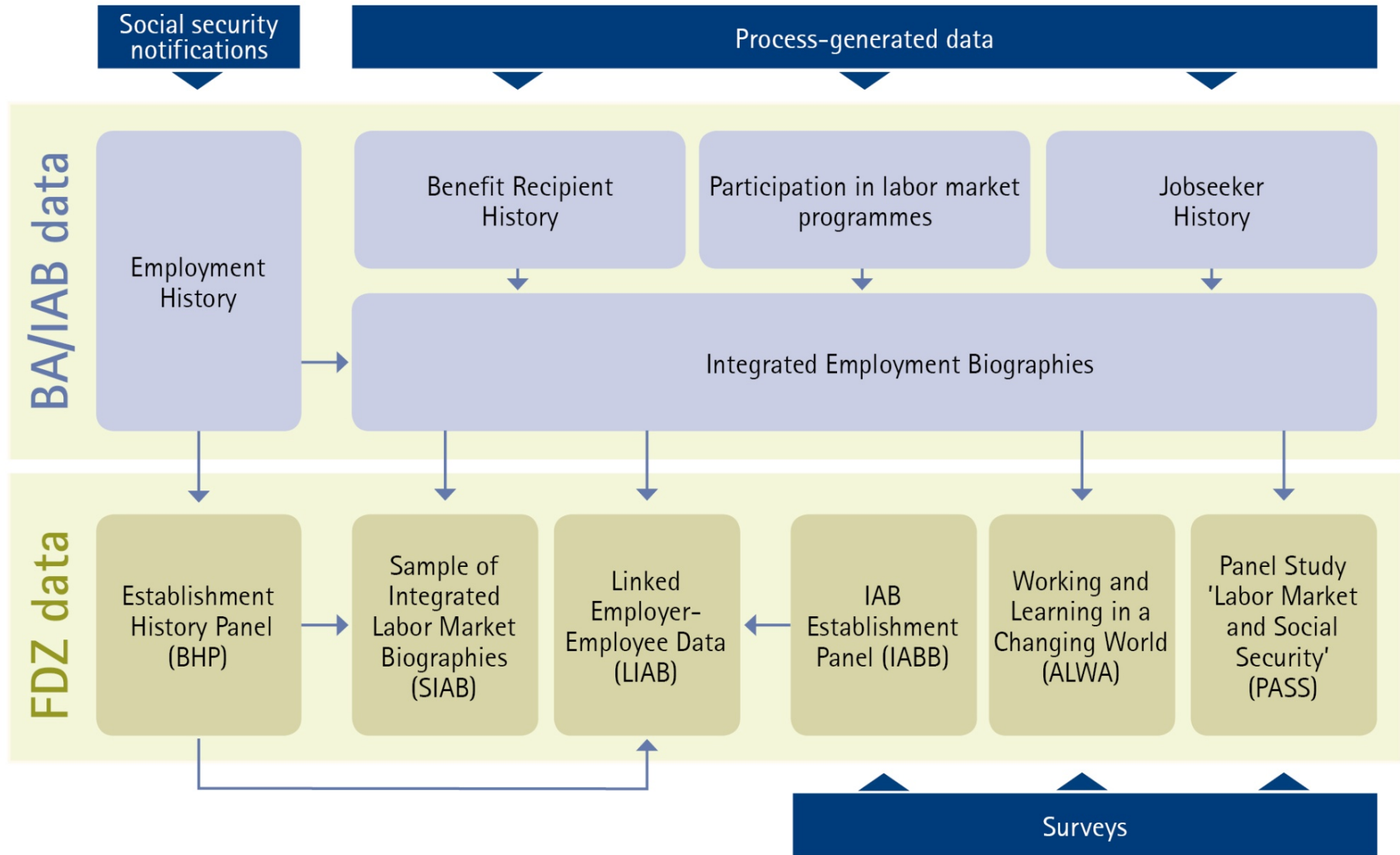
- Clarification of the Acronyms:
  - FDZ (Research Data Centre) of the
  - BA (German Federal Employment Agency) at the
  - IAB (Institute for Employment Research)
- But why?
  - It is a legal thing...
  - BA is only allowed to store data for the administrative process
  - IAB, as research institution, does not have this limitation

# FDZ of BA at IAB

- Aim of IAB
  - Research institution for the BA (est. 1967)
  - Independent scientific controlling of BA activities
  - Two statutory mandates justify the work of IAB
- Aim of FDZ
  - Research data centre for the BA (est. 2004)
  - Provide IAB survey data and BA administrative data as research data to the scientific community
  - Statutory mandate justifies the work of FDZ

Administrative Data in the IAB Metadata Management System

# **DATA SOURCE PROVIDED BY IAB**

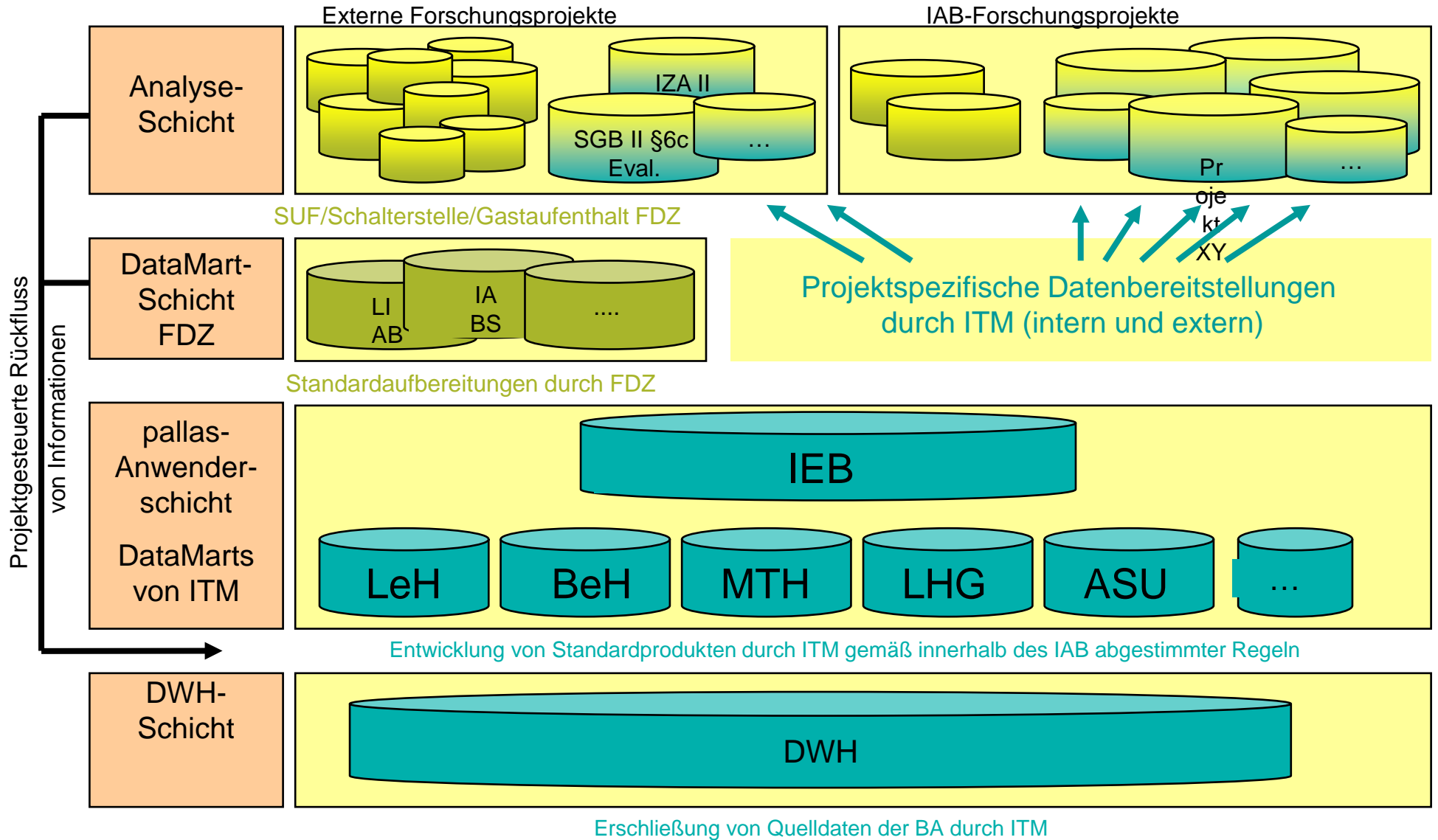


# Collection process for administrative data

- German Federal Employment Agency
  - 700 job centre in every larger German city
  - Staff fills in forms (supported by special software tools)
    - Benefit Recipient
    - Participation in labor market programs
    - Jobseeker
- Social security notification
  - Yearly notification done by the employer for every employee
  - Relevant for pension payment
- Data flows into the BA Data Warehouse (DWH)



# Schichtenmodell prozessproduzierter Individualdaten



# Collection process for administrative data

## 1. BA

1. Forms supported by different software tools
2. DWH of the BA
3. Specialized Data Marts

## 2. ITM

1. Extracts in SAS (due to size of dataset)
2. Subsample of SAS files in Stata (95% use this software)

## 3. FDZ

1. Data editing for scientific use in Stata
2. In parallel: data documentation

# Short summary

1. Data not collected for research purposes
2. No influence on and little knowledge about data selection process
3. Different storage formats
4. Goal of data editing procedure: build survey-like datasets
5. Data products of IAB/BA:
  1. Administrative data
  2. Survey data
  3. Integrated datasets
  1. Establishment data
  2. Individual / Household data
  3. Integrated establishment and individual

Administrative Data in the IAB Metadata Management System

# **NEED FOR A USEFUL DATA DOCUMENTATION**

# Data documentation requirements

## 1. Researcher view

1. Standardized documentation
2. Easy to understand documentation (do not cover every possibility)
3. Easy access able documentation (centralized web portal)
4. Software tools for search functionalities

## 2. Data provider view

1. Standardized documentation
2. Easy to understand documentation (do not cover every possibility)
3. Uncomplicated preparation of documentation
4. Software tools for preparation
5. Software tools for exchange of documentation

# Special needs for administrative data

1. Concentrate on data collection process
2. Different data quality topics, e.g.:
  1. Why was data collected
  2. How was data collected
  3. How was data modified
3. Need for interfaces to upstream data editing processes  
(you can only document what you know)
4. What special disclosure issues arise with administrative data

# Short summary: documentation needs

## 1. In general:

1. Standardized standards are needed
  1. No major changes
  2. Manageable size
  3. Not everything can be covered
2. Machine readable, supported by software, interoperable with common storage formats

## 2. Specifically:

1. Coverage of different data collection modes
2. At the same time – stay as close as possible to the standardized standard

Administrative Data in the IAB Metadata Management System

# **EXAMPLE OF AN EXEMPLARY SOFTWARE IMPLEMENTATION**



# IAB-Metadata Project

- Done by a consortium (tba21, DIFP, Colectica, OPIT, Alerk Amin)
- Work in process (runtime 24 months)
- Implementing into the BA IT infrastructure
- The steps are:
  - Requirements
  - DDI implementation
  - Software building and implementation
- Currently only a documentation of collected data

# IAB-Metadata Project

- Aim of the project
  - Update of the IT infrastructure of the FDZ
  - But money into the development
  - Enable interoperability between institutions
  - Build and use it
  - Merge research data and recarding data documentation
  - Be aware of future data sources

# DDI Concept for IAB Tools Project

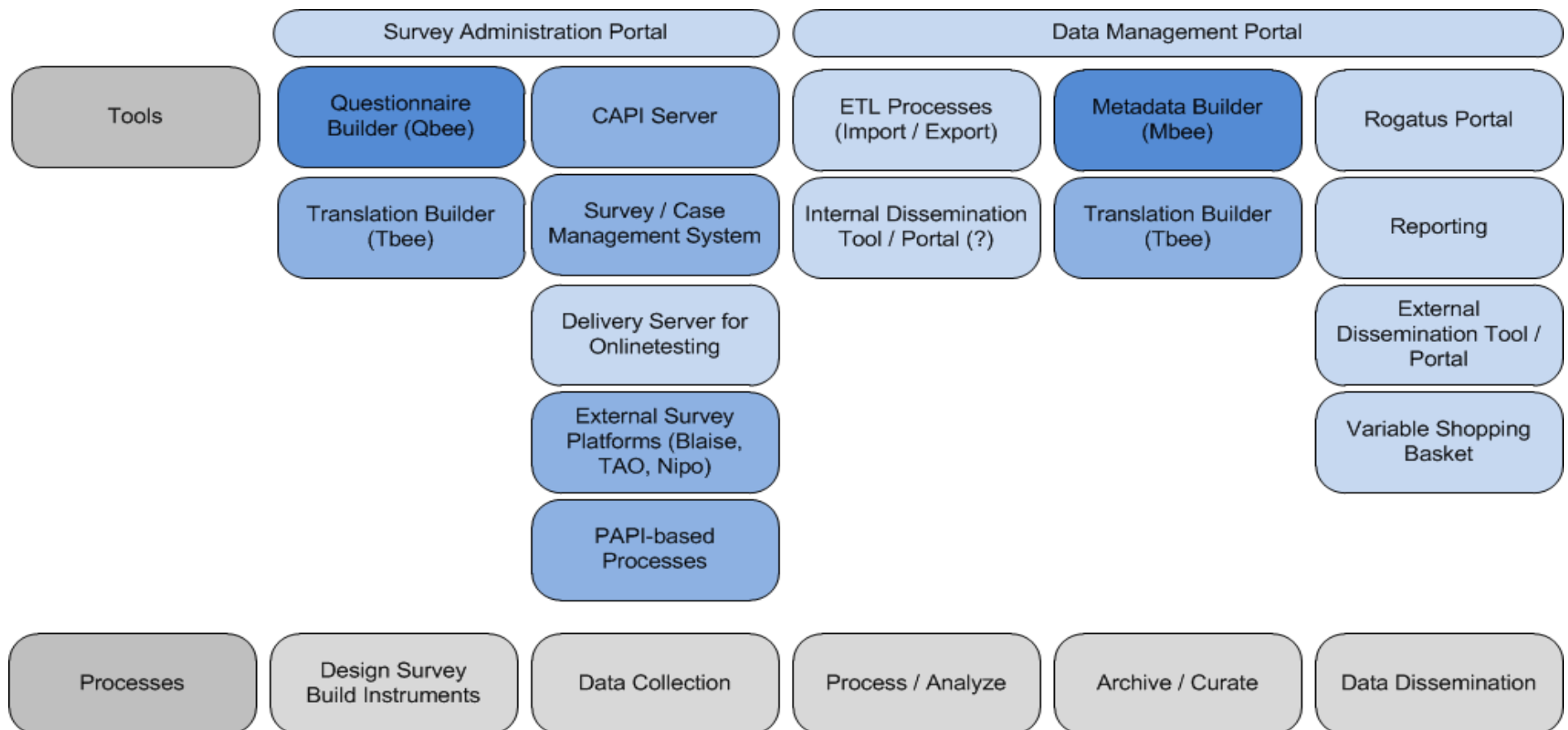
Datei Bearbeiten Ansicht Einfügen Format Daten Tools Hilfe Lesezugriff

fx | DDI Element

☰ ↶ ↷ 📄 \$ % 123 ▾ Arial ▾ | 10 ▾ | **B** *I* A ▾ 🗑️ 📏 📐 ☰ ⬇️ ⬆️ 🗨️ 📊 🔍 Σ ▾

	A	B	C	D	E	F
1	DDI Element	IAB Term	Qbee 0.2	Colectica	Questasy	IAB
2	@time			y	y (*1)	
3	@captureInstrument			y	y (*1)	
4	@panel			y	y (*1)	
5	@geography			y	y (*1)	
6	@dataProduct			y	y (*1)	
7	@languageRelationship			y	y (*1)	
8	Citation			y	y	
9	Citation.Title	Name	Name	y		
10	Abstract	Abstract	Abstract	y	y	
11	AuthorizationSource		-	n	n	
12	UniverseReference	Data unit	-	y	y	
13	SeriesStatement			y	n	
14	ExPostEvaluation		-	n	n	
15	FundingInformation		Funding, Funding From, Funding To	y	y	
16	ProjectBudget		-	n	n	
17	Purpose		-	y	y	
18	KindOfData	Data type	-	y	n	
19	Coverage		-	y	y	
20	Coverage.TopicalCoverage		-	y		
21	Coverage.TopicalCoverage.Subjects	Topics	-	y		
22	Coverage.TemporalCoverage.ReferenceDate.StartDate	Period Covered from	-	y		
23	Coverage.TemporalCoverage.ReferenceDate.EndDate	Period Covered to	-	y		
24	OtherMaterial		Bibliography	y	y	
25	Embargo		-	y	n	
26	StudyUnitReference		Survey waves	y	n	
27	SubGroupReference		-	n	n	

# Rogatus tool overview



# Short summary

- Standard documentation is essential
- Standard can only survive if it is used
- Build, communicate, enlarge...
- Administrative data can be the back-bone for merged data
  - Survey data
  - Statistical data
  - Future data

Administrative Data in the IAB Metadata Management System

# **ADMINISTRATIVE DATA; ONLY A STEP INTO THE FUTURE OF RESEARCH DATA**

*Examples of electronic recordings which can provide data for official statistics:*

- 1. Credit card transactions**
- 2. Commodity (RFID) tracking**
- 3. Toll road (RFID) recording**
- 4. Electronic tickets for travelling**
- 5. Public services offered electronically**
- 6. Immigration control.**
- 7. Mobile phone use**
- 8. Internet and social media use**
- 9. GPS tracking of traffic and transport**
- 10. Mixed active/passive recording**

Slide from EDDI 2012, Key note from Svein Nordbotten

# OECD report: Categories of “Future Data”

- Category A: Data stemming from the transactions of government, for example, tax and social security systems.
- Category B: Data describing official registration or licensing requirements.
- Category C: Commercial transactions made by individuals and organisations.
- Category D: Internet data, deriving from search and social networking activities.
- Category E: Tracking data, monitoring the movement of individuals or physical objects
  - subject to movement by humans.
- Category F: Image data, particularly aerial and satellite images but including land-based
  - video images.



# DDIs way into the future...

- Survey data
- Administrative data
- Future data
  
- No extensions but looking for a generalized approach
- Need for tools that are able to cover the whole data production process

Administrative Data in the IAB Metadata Management System

# CONCLUSION

# Conclusion

- Not administrative data is important – data collection processes are
- DDI-paper: “Documenting a wider Variety of data using the Data Documentation Initiative 3.1”
- Build solutions together
- Important is the data documentation standard
- But also:
  - Support data life-cycle
  - Standard for supporting tools is important as well

# Thanks for listening

David Schiller, [david.schiller@iab.de](mailto:david.schiller@iab.de)

Ingo Barkow, [barkow@dipf.de](mailto:barkow@dipf.de)

<http://fdz.iab.de>