WHAT DO GEOLOCATION SMARTPHONE DATA ADD TO A SURVEY PANEL?

Available indicators from the IAB-SMART Project

BERD Academy
July 5th 2023
Florian Zimmermann, Institute for Employment Research (IAB)

MOTIVATION

- Smartphones are getting more and more present in our daily life and recently researchers started using them as a survey instrument
- Gives researchers unique access to passively collected data, e.g., geolocation data
- IAB-SMART-Study in 2018 (Kreuter et al. 2018)

IAB-SMART-MOBILITY

- IAB-SMART-Mobility module contains mobility indicators based on geodata, e.g., traveled distance
- Collaboration of

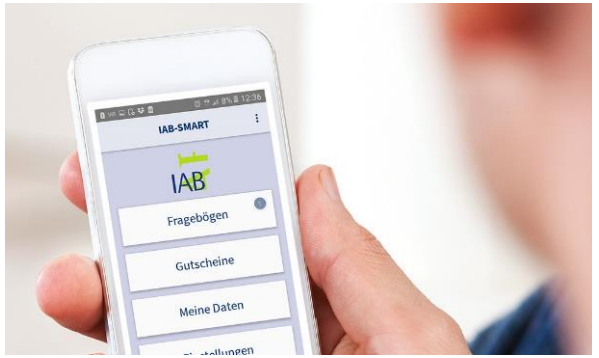
- Funded by 

OUTLINE

1. Introduction
2. IAB-SMART study
3. Data preprocessing
4. Mobility indicators
5. Link to PASS(-ADIAB)

STUDY DESIGN AND SAMPLE OF IAB SMART APP

- Usage of a smartphone app (Kreuter et al. 2020, Haas et al. 2020, Bähr et al. 2020)
  - to collect passive smartphone data and
  - to deliver short surveys for labor market research



- Research Objectives:
  - Analyses of the effects of long-term unemployment on social integration and social activity
  - Possibility to test the smartphone as a combined survey instrument

STUDY DESIGN AND SAMPLE

- Participants of the Panel Study “Labour Market and Social Security” (PASS) (Trappmann et al. 2019), restriction to respondents in 2017, who own Android Smartphone

- Random selection and invitation of **4,293** PASS participants via a postal letter

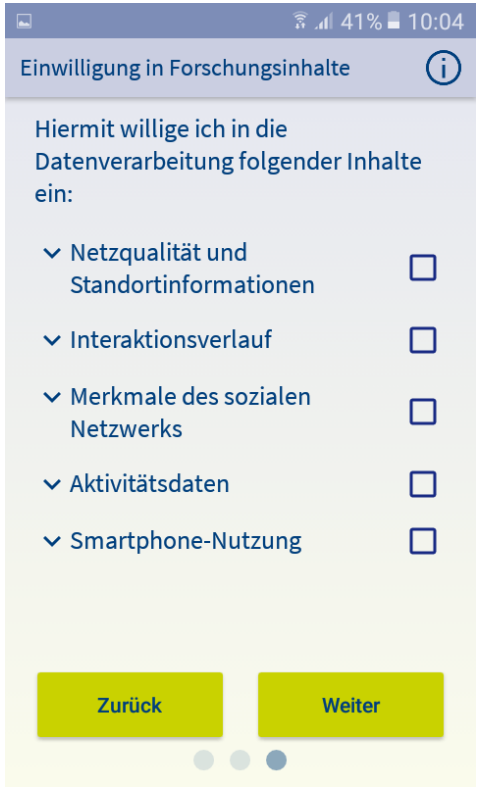
  → **623 PASS participants installed the app and answered short survey** (Keusch et al. 2020b)

- Voluntary participation and revocation was possible at any time for the whole survey or parts, such as collection of geolocation data
- Data collection (DSGVO compliant): January 2018 to August 2018

MEASUREMENTS

- Five sets of passive data collection packages using sensors on the smartphone
  - **(a) Mobile phone network quality and location information**
  - (b) Interaction history
  - (c) Characteristics of the social network
  - (d) Activity data
  - (e) Smartphone usage

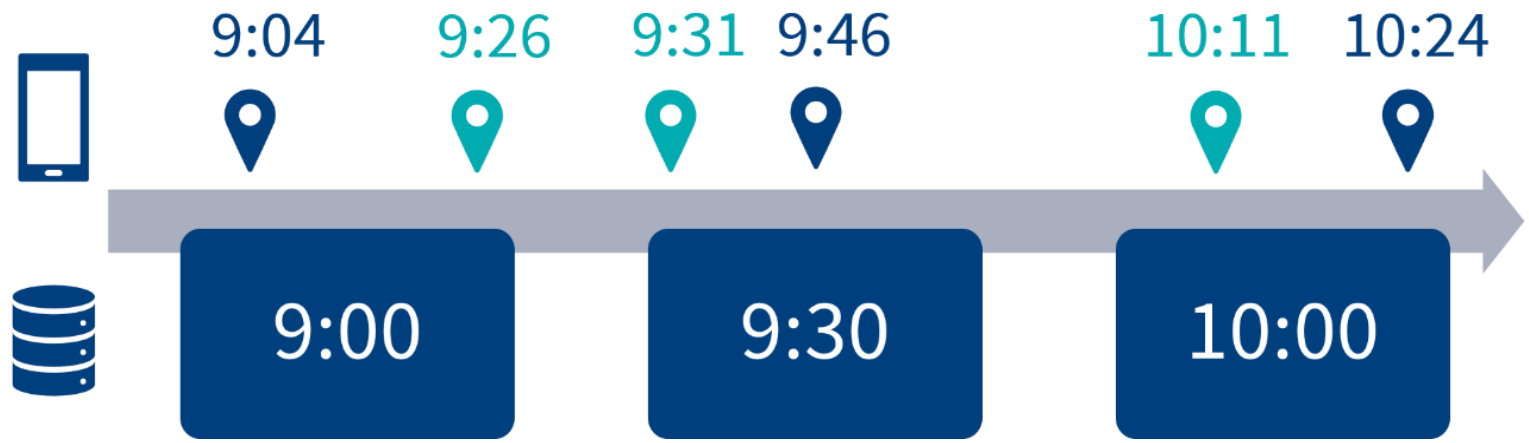
- Decision about passively collected data during installation process



OUTLINE

1. Introduction
2. IAB-SMART study
3. **Data preprocessing**
4. Mobility indicators
5. Link to PASS(-ADIAB)

ALLOCATING DATA INTO 30 MINUTE INTERVALS



- Allocation of data into 30 minute intervals
- 80% of intervals are after first installing and before uninstalling the app

	Number intervals	%
Before installing	907,938	12.96
Last information	567,323	8.10
Potential intervals	5,530,735	78.94

PREPROCESSING GEODATA

Restriction	N Individuals	% Individuals	N Geodata	% Geodata
Installed and verified	618	100%		
Geodata consent	567	92%		
Valid geolocation measurements	543	88%	1,465,541	100%
Accurate geolocation measurements (< 350m)	543	88%	1,288,050	88%
At least seven valid days	398	64%	1,218,715	83%

- These are typical problems for smartphone data and highlight the importance of documenting the preprocessing steps

OUTLINE

1. Introduction
2. IAB-SMART study
3. Data preprocessing
4. **Mobility indicators**
5. Link to PASS(-ADIAB)

CREATION AND ANONYMIZATION OF GEODATA INDICATORS

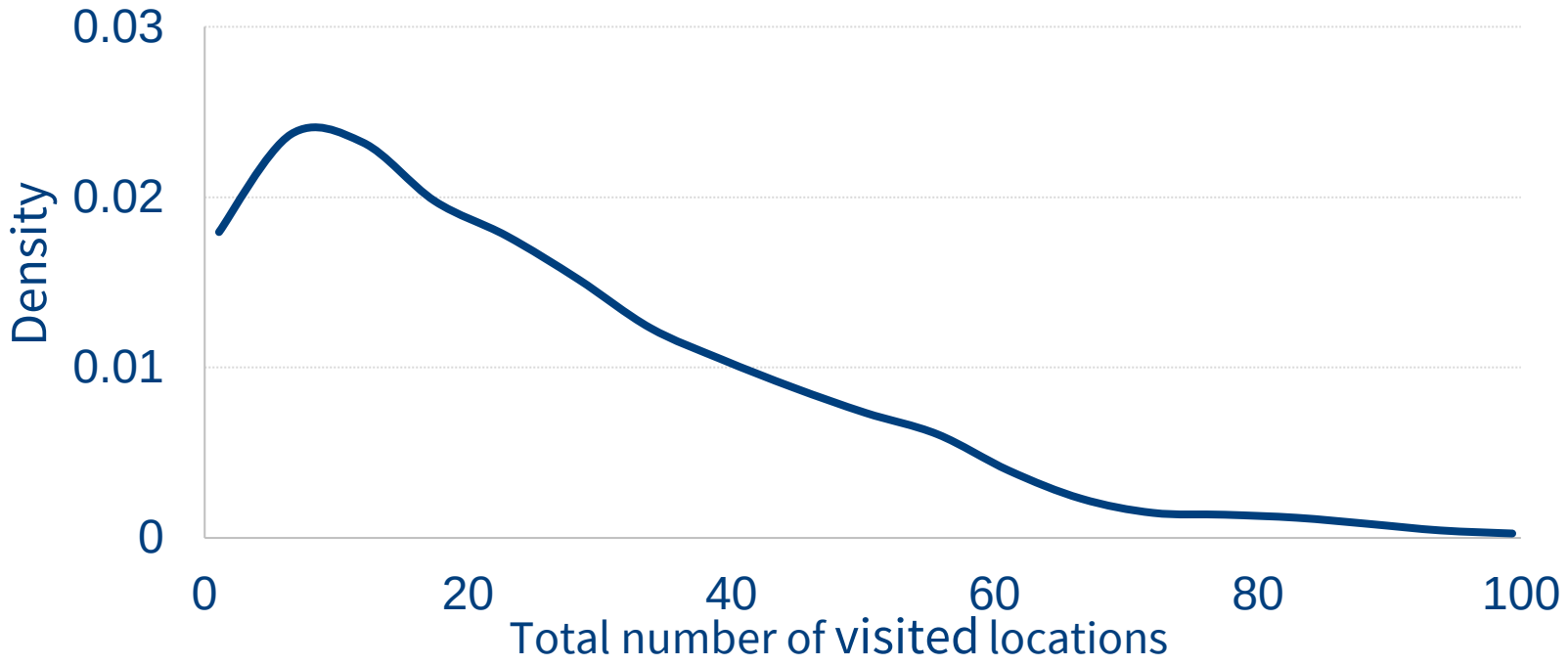
- Representing an individual's average mobility vs. risk of re-identifying individual participants
  - Geolocation trace data: particularly sensitive in itself
  - Combining multiple data sources could increase the risk of re-identification

→Aggregating data at weekday and weekend level to identify differing average mobility at weekdays and weekend

CLUSTERING UNIQUE LOCATIONS

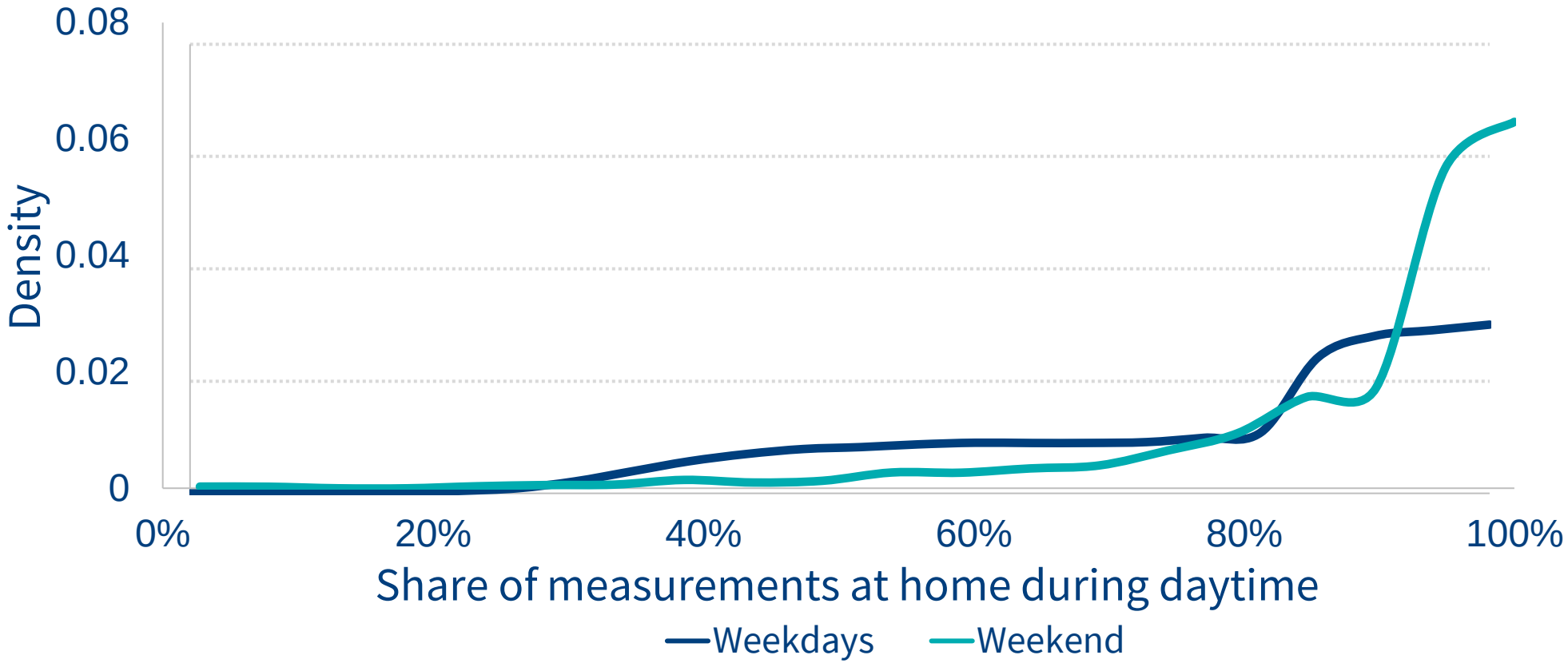
- Location is defined as a place which you visit and where you stay
- DBSCAN as clustering method (vgl. Mueller et al. 2022) with a radius of 600m (Jongs et al. 2020)

→At least 2 geodata points per location



TIME SPENT AT HOME INDICATOR

- Home is the location with most geodata points at night (midnight-6am)
- Share geolocations at home during daytime (6am-midnight)



SUMMARY MOBILITY INDICATORS

Weekday	Weekend	Total
Median visited locations	Median visited locations	Total visited locations
Median distance	Median distance	Total distance
Median variance geolocations	Median variance geolocations	Total variance geolocations
Share geolocations home cluster during daytime (6am to midnight)	Share geolocations home cluster during daytime (6am to midnight)	

SUMMARY QUALITY INDICATORS

	Total			Weekday			Weekend		
Variable	Mean	SD	N	Mean	SD	N	Mean	SD	N
Number of geolocation points	3,070.0	2,351.1	398	2,209.6	1,699.1	398	860.4	661.3	398
Number of valid days	106.5	69.5	398	76.3	50.0	398	30.2	19.8	398
Share of valid geodata	0.6	0.2	398	0.6	0.2	398	0.6	0.2	398

→Researchers can restrict the sample to ideally fit their research question

OUTLINE

1. Introduction
2. IAB-SMART study
3. Data preprocessing
4. Mobility indicators
5. **Link to PASS(-ADIAB)**

THE PASS PANEL SURVEY

- Panel study ‘Labour Market and Social Security‘ (PASS)
  - Household panel survey by IAB
  - Major data source for research into unemployment & poverty
  - Main topics include: Labor market participation, job search, benefit receipt, active labor market programs, social inclusion, health, income, deprivation

- Link of PASS to administrative employment histories (PASS-ADIAB)
  - Daily information on employment for employees liable to social security
  - Information on employment status, wages, education, establishments, etc.

→Unique combination of mobility indicators, survey data, and administrative data

CODE TEMPLATE FOR MERGING

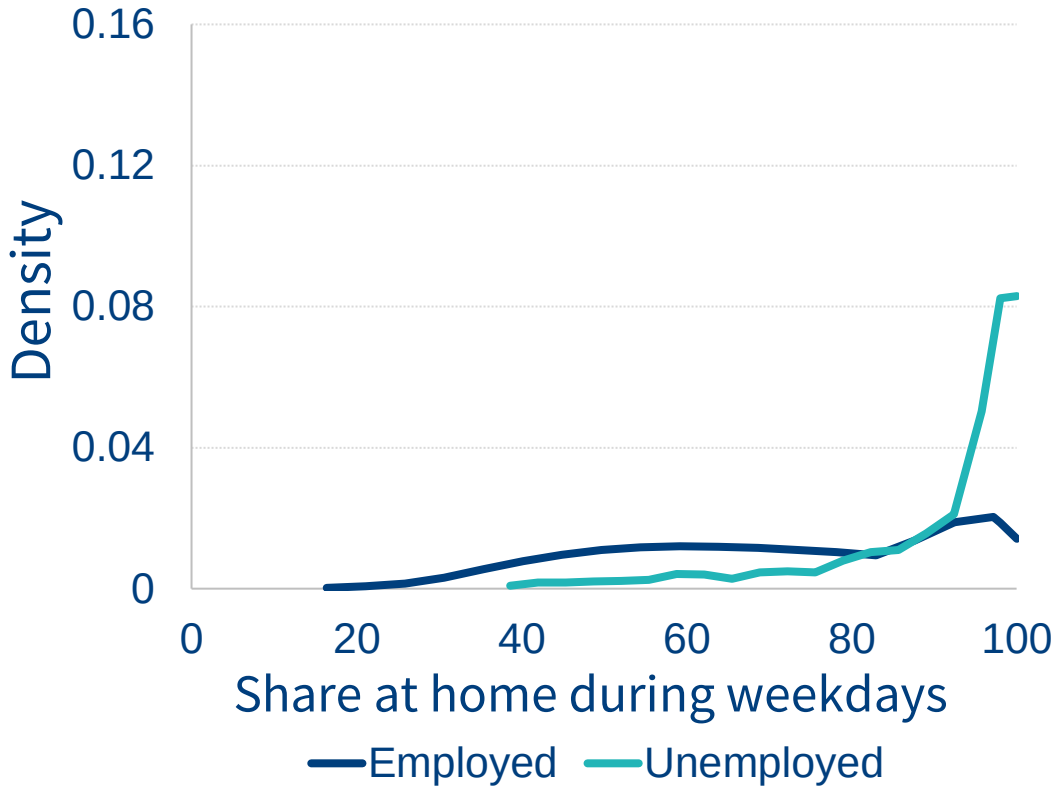
Code example Stata 17: Merging IAB-SMART-Mobility and PASS SUF

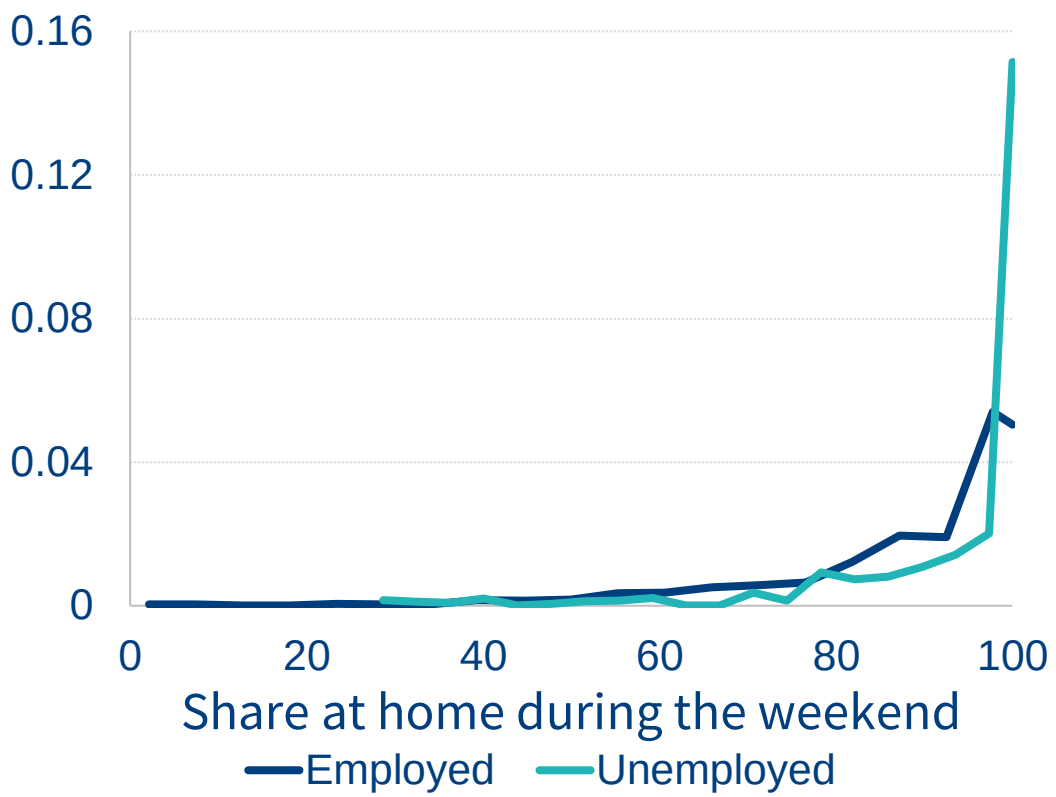
```
//Defines path for the datasets
global orig_pass   "P:\Datensaetze\_Endprodukte\PASS\PASS_0621_v2\"
global orig_smart "\\iab\dfs\017\Ablagen\D01700-Projekte\D01700-MoDeM-Daten\BERT\data\"

//Merge the data and keep only wave 11 from PASS
use "${orig_smart}/IAB-SMART-Mobilityindicators.dta", clear
merge 1:m pnr using "${orig_pass}/PENDDAT.dta", keep(match master)
keep if welle == 11 //Keep only wave 11 (from 2017)
```

TIME SPENT AT HOME BY EMPLOYMENT STATUS

- Unemployed individuals spend more time at home on average during weekdays and the weekend





SUMMARY

- Dataset with indicators based on geodata for 2018
- Addon for PASS and PASS-ADIAB

→ Combination of rich survey data, administrative employment histories, and mobility indicators based on geodata

- Method report to document data creation and anonymization for data protection:
  Zimmermann, Filser, Bähr, Haas: The IAB-SMART-Mobility module: An innovative research dataset with mobility indicators based on raw geodata. Soon in Journal of Economics and Statistics

OUTLOOK AND DATA ACCESS

- Planned release: This quarter

→ Subscribe to FDZ newsletter for updates on the release

- Free to use for labor market research
- Planned ways to access data at the FDZ
  - On-site use for PASS-ADIAB
  - SUF for PASS at the end of this year

THANK YOU!

Florian.Zimmermann@iab.de