M.Sc. in ‚Transportation Systems'

Applied Statistics in Transport Travel Surveys, MID2008

Prof. Regine Gerike

Technische Universität München, mobil.TUM

regine.gerike@tum.de

Munich, 08/11/2011, 15/11/2011
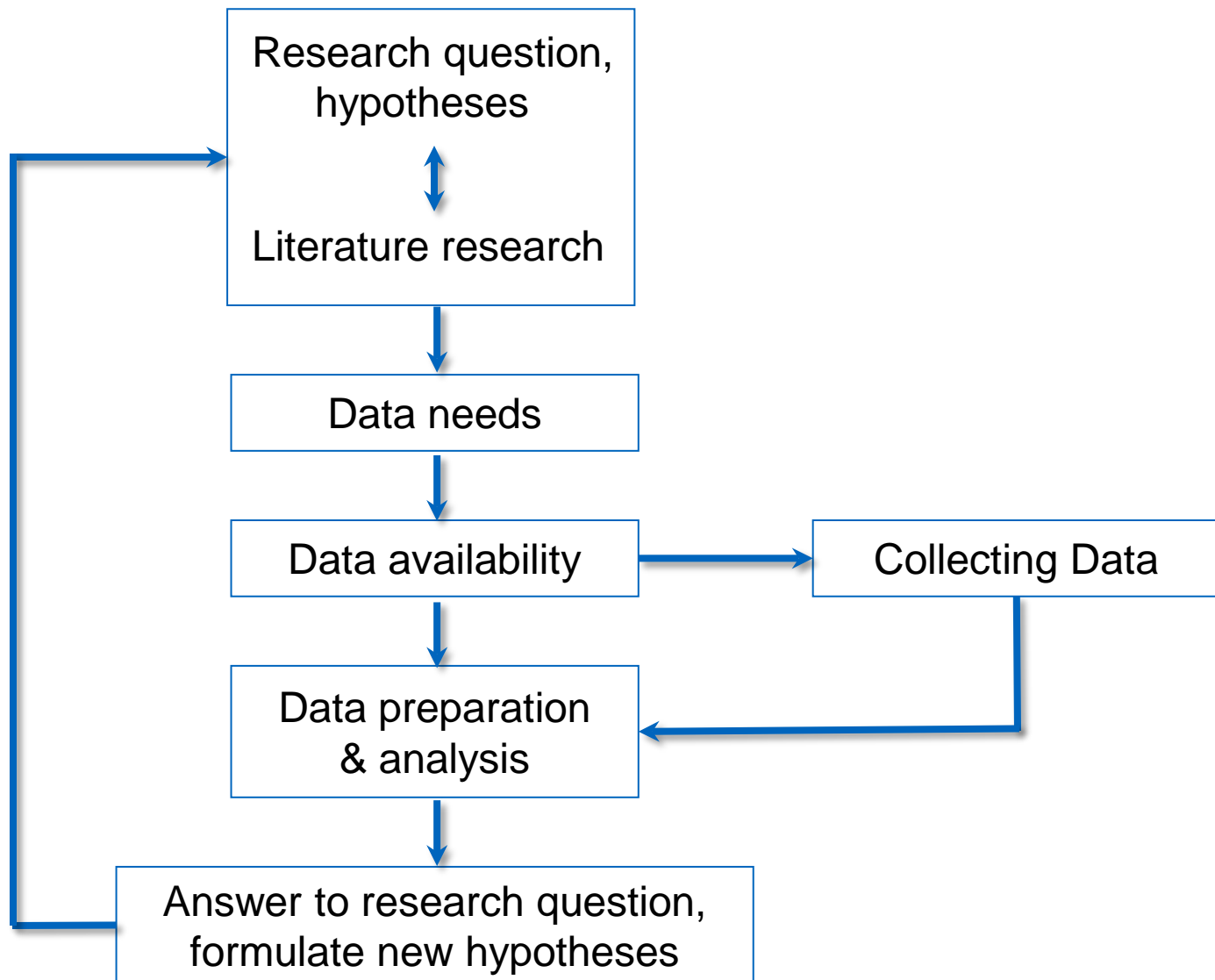
Time Table Updated, 02/11/2011

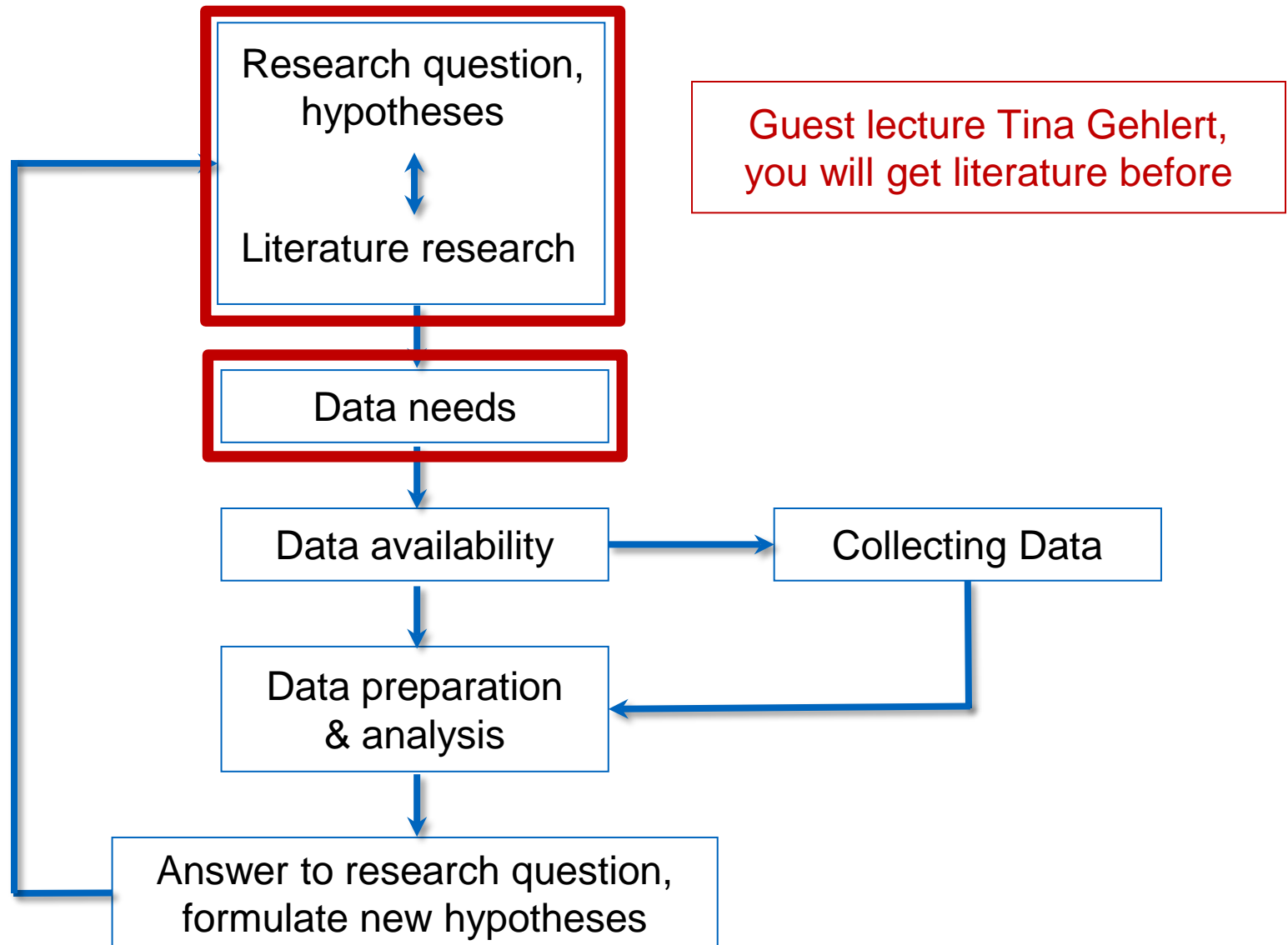| 1	18.10.2011	9:45-11:15 L	Welcome and Introduction
2	25.10.2011	8:00-9:30 LE	Introduction R, Tinn-R, Math
3	25.10.2011	9:45-11:15 L	Theory of probability
	01.11.2011	All Saints Day	
4	08.11.2011	8:00-9:30 LE	Descriptive analysis with R
5	08.11.2011	9:45-11:15 L	Descriptive statistics
6	15.11.2011	8:00-9:30 LE	Real world data input and preparation
	15.11.2011	9:45-11:15 L	Studentische Vollversammlung
7	22.11.2011	8:00-9:30 LE	Descriptive analysis with R, real world data
8	22.11.2011	9:45-11:15 L	Distributions
9	29.11.2011	8:00-9:30 LE	Reserve
10	29.11.2011	9:45-11:15 L	Distributions
11	06.12.2011	8:00-9:30 L	Inferential statistics
12	13.12.2011	8:00-9:30 L	Hypotheses testing
13	20.12.2011	8:00-9:30 L	Tests, statistical modelling
14	10.01.2012	8:00-9:30 L	ANOVA
15	17.01.2012	8:00-9:30 L	Regression
16	17.01.2012	9:45-11:15 L	Reserve
17	24.01.2012	8:00-9:30 L	Repetition
	02.02.2012	Tina Gehlert	Hypothesis-driven data analysis in transport
	03.02.2012	Tina Gehlert	Hypothesis-driven data analysis in transport
	07.02.2012	10:00-11:00	Exam

Real Data Example MID2008

Steps for project:

- Research cycle
- Research questions
- Data Preparation, validation
- Descriptive statistics, visualization
- Tests
- Models, e.g. ANOVA, regression analysis

Main Goals:

- Formulate research questions, hypotheses
- Work with real data (read and prepare the data)
- Get familiar with exploratory data analysis (descriptive measures, graphics)
- Get familiar with methods for analyzing your data beyond descriptive statistics

Empirical Research Cycle

Research question, hypotheses

Literature research

Data needs

Guest lecture Tina Gehlert, you will get literature before

Data availability → Collecting Data

Data preparation & analysis

Answer to research question, formulate new hypotheses

Step 1: Literature for hypotheses, research questions

- … Guest lecture Tina Gehlert

- Our next three lab sessions: reading and exploring data
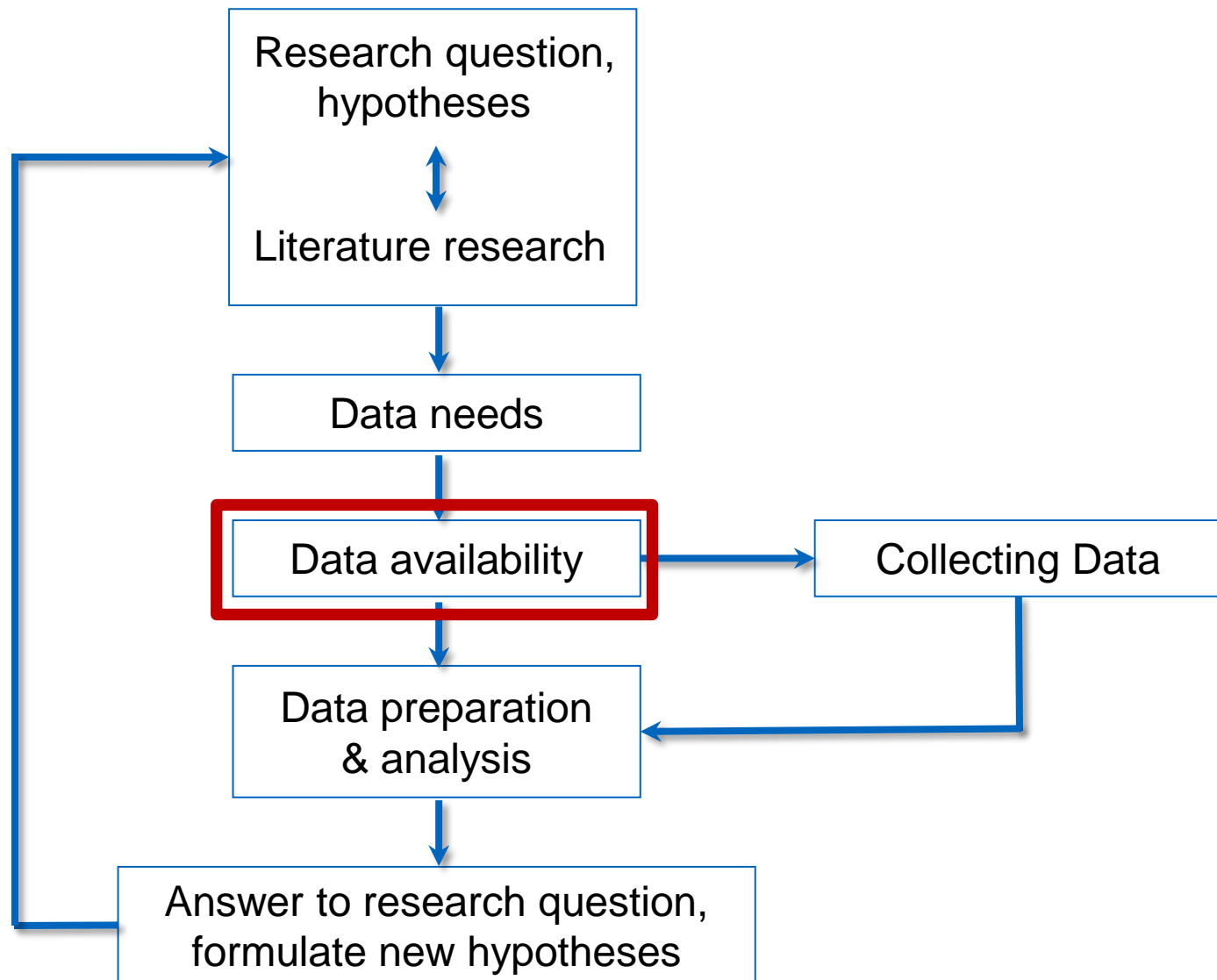
German Travel Surveys: Overview

Two diary day surveys:

- Mobilität in Deutschland: http://www.mobilitaet-in-deutschland.de/
- System repräsentativer Verkehrsbefragungen: http://www.tu-dresden.de/srv/SrV_Web/

One panel survey:

- Deutsches Mobilitätspanel: http://mobilitaetspanel.ifv.uni-karlsruhe.de/

Data as basis for transport planning and for research

## See Kunert-paper for overview of international surveys

German Travel Surveys: Overview
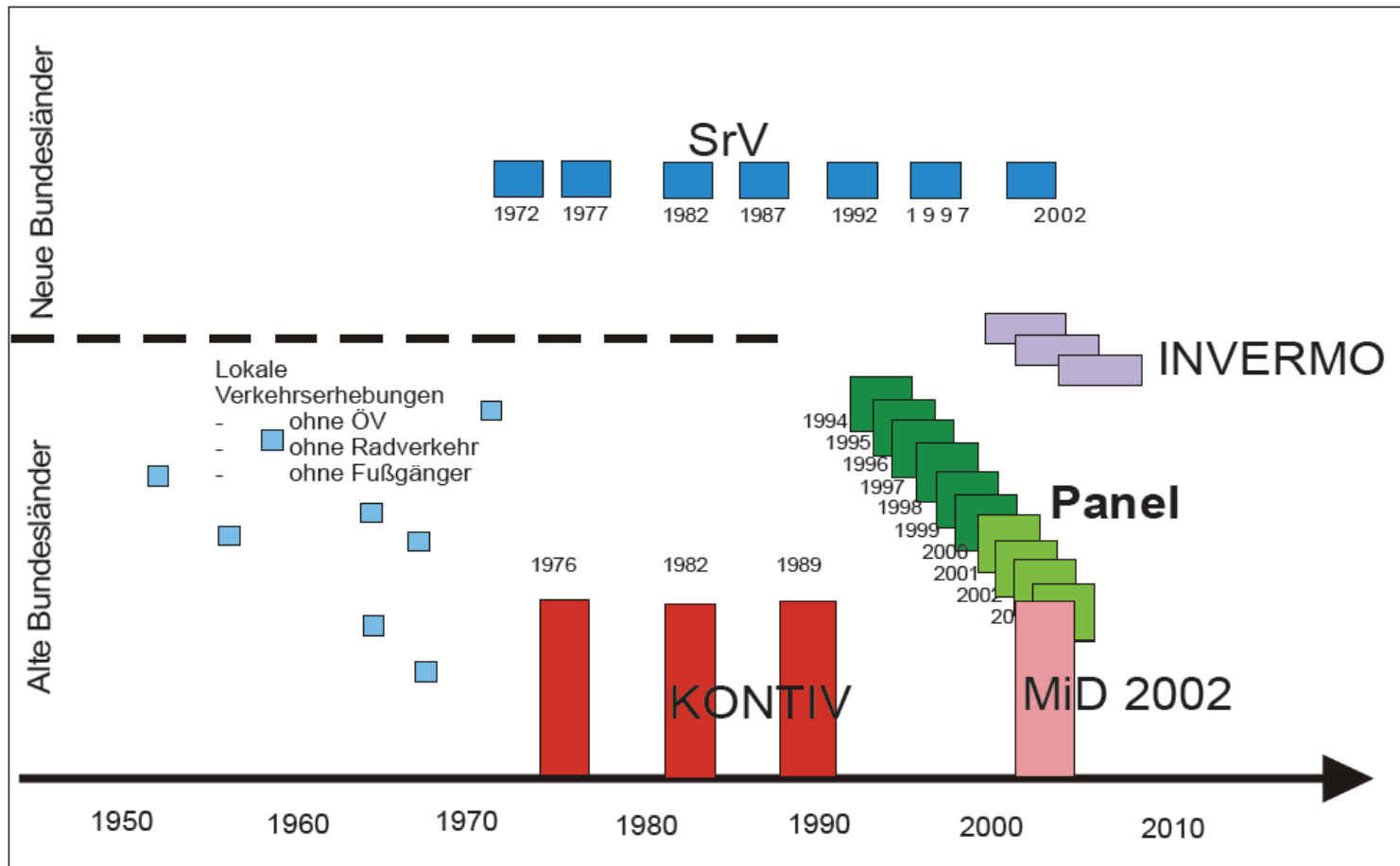
German Travel Surveys: Mobilität in Deutschland

- Size of random sample net 25,000 households based on registries of residents (covering the whole Germany)
- States or regions add-on by app. another 25,000 households, 100,000 individuals, 300,000 trips
- Collection of information of the whole household
- Survey guided by fixed diary-dates and lasting 12 months (weekdays and weekend-days)
- Non-response-study
- Written questionaires, telephone interviews, online questionaires
- Years 2002, 2008
- Similar surveys: 1976, 1982, 1989 "KONTIV" (Kontinuierliche Erhebung zum Verkehrsverhalten)

System repräsentativer Verkehrsbefragungen
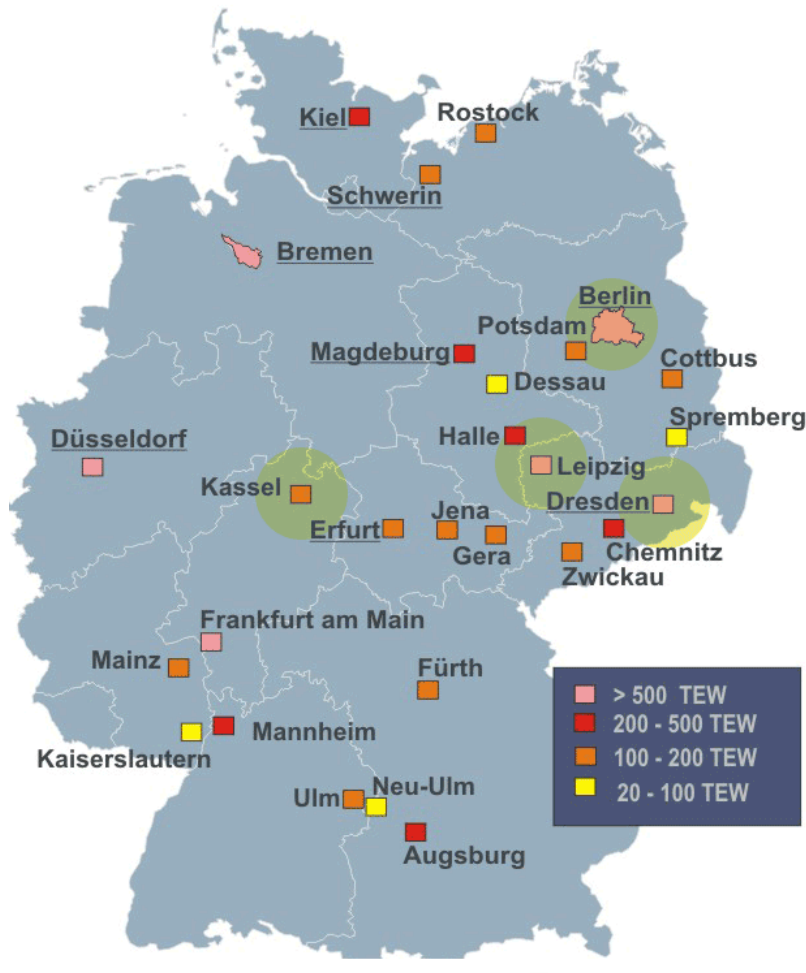

Participating cities in 2008,
http://www.tu-dresden.de/srv/SrV_Web/

German Mobility Panel, Deutsches Mobilitätspanel

- Mobility behaviour of one complete week (longitudinal continuous approach) - diary for one week
- Yearly repetition (one week in autumn)
- Mobility of complete households
- Rotating sample (3 years of participation)
- No geocoding of destinations

German Mobility Panel
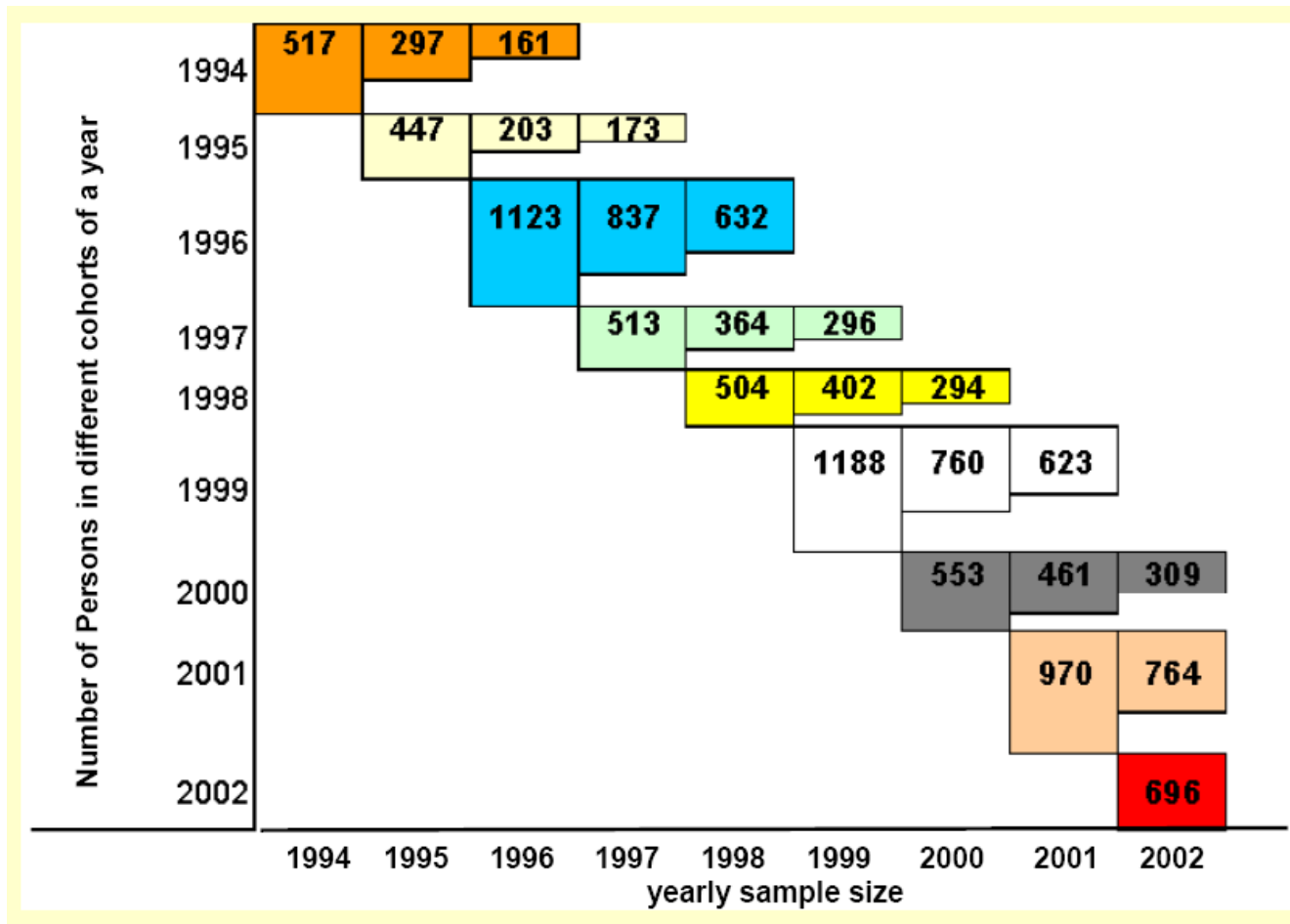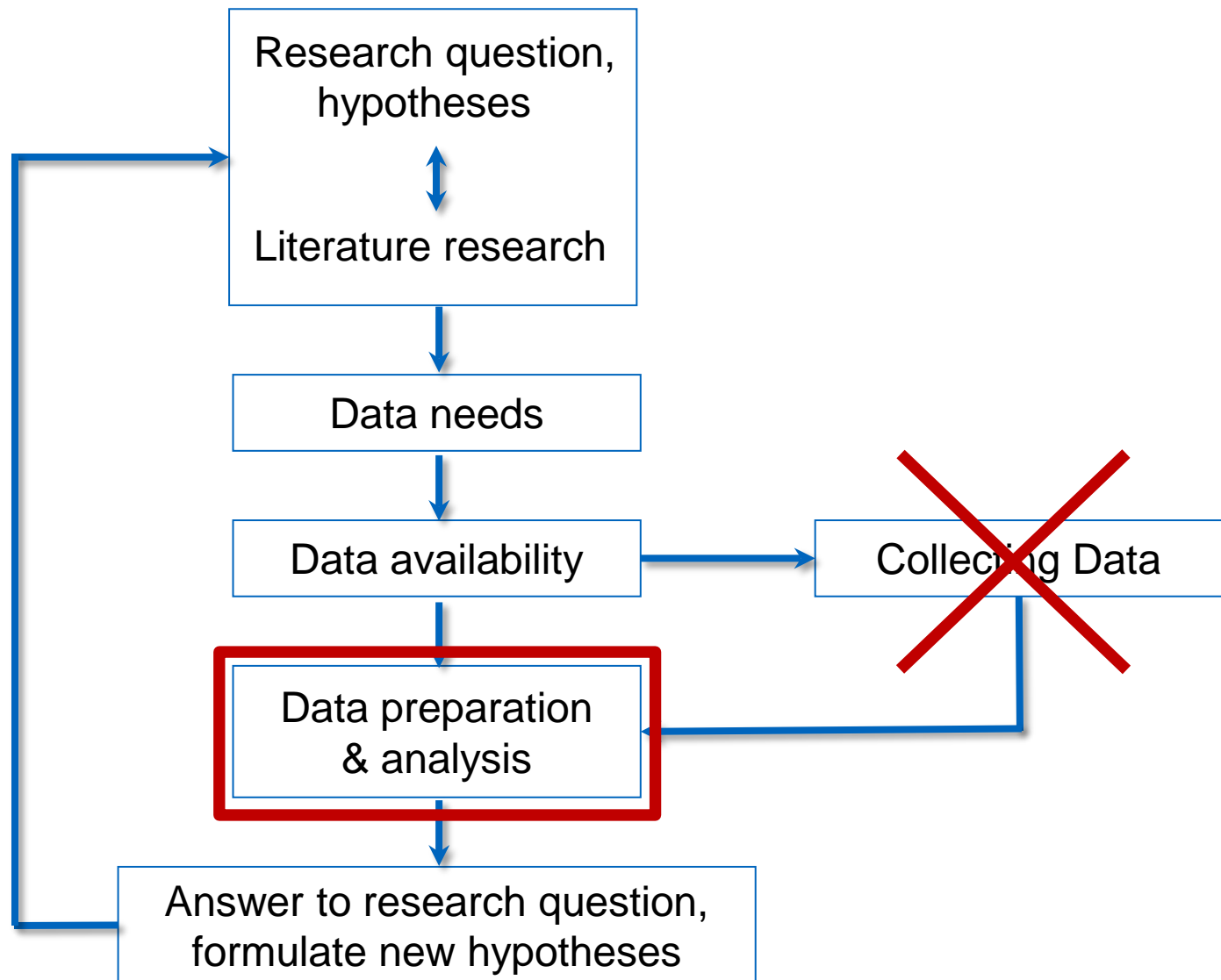
Decision on Data Set: Mobilität in Deutschland 2008

- One diary day
- Weekdays and weekend days during a whole year
- All region types
- Covering the whole household (including children, complete households in 81% of the cases)

- http://www.mobilitaet-in-deutschland.de/engl%202008/

MiD2008: Data sets, Structure

households	persons	cars	trips

in general
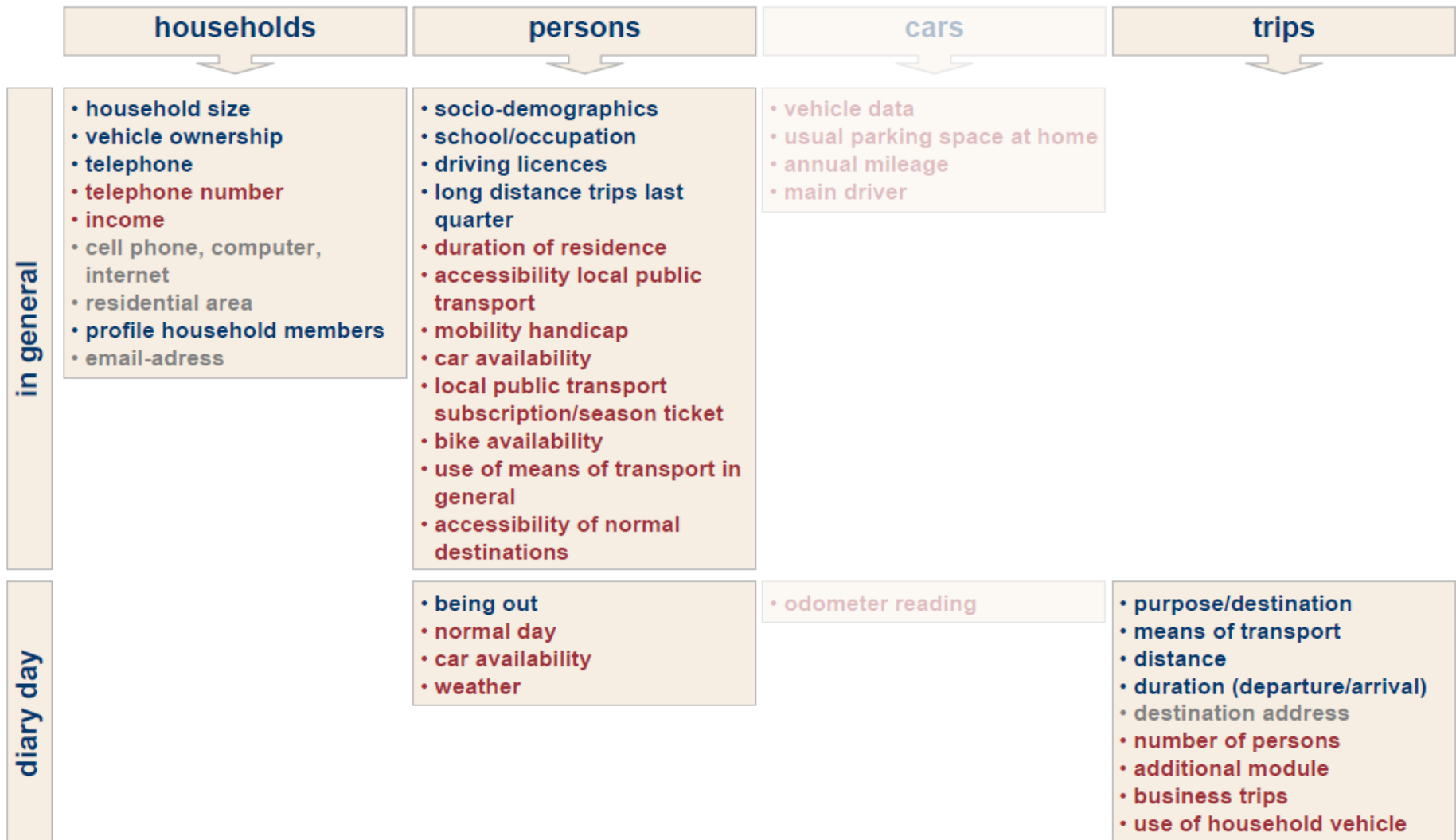
households
- household size
- vehicle ownership
- telephone
- telephone number
- income
- cell phone, computer, internet
- residential area
- profile household members
- email-adress

persons
- socio-demographics
- school/occupation
- driving licences
- long distance trips last quarter
- duration of residence
- accessibility local public transport
- mobility handicap
- car availability
- local public transport subscription/season ticket
- bike availability
- use of means of transport in general
- accessibility of normal destinations

cars
- vehicle data
- usual parking space at home
- annual mileage
- main driver

diary day

persons
- being out
- normal day
- car availability
- weather

cars
- odometer reading

trips
- purpose/destination
- means of transport
- distance
- duration (departure/arrival)
- destination address
- number of persons
- additional module
- business trips
- use of household vehicle

red: expansion compared to previous surveys of the KONTIV type (since 2002)

grey: abbreviation 2008

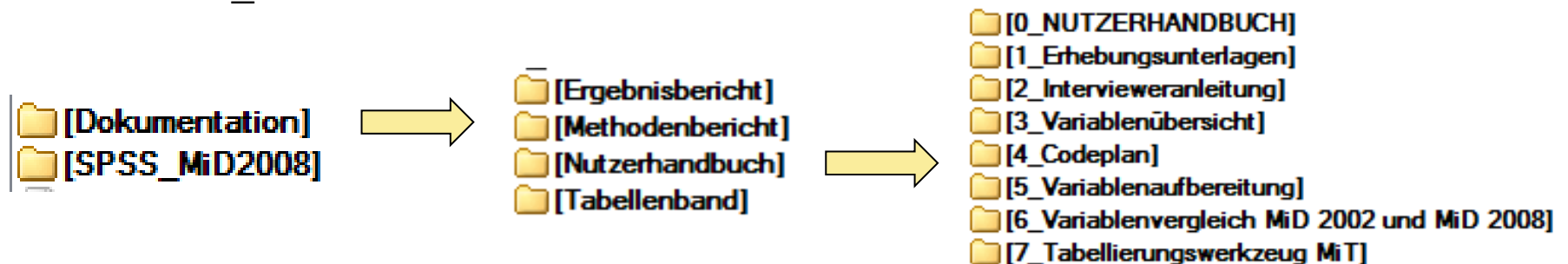# MiD2008: Material for your work, you get:

Textfiles:

- MiD2008_PUF_Wege.dat, MiD2008_PUF_Personen.dat, MiD2008_PUF_Haushalte.dat
- Each of the data sets contains information from other data sets

Codeplan:

- MiD2008 English Codeplan_20100521.xlsx

Additional material:

- Introduction of survey design, implementation and analysis: http://unstats.un.org/unsd/hhsurveys/, Household_surveys.pdf
- "Mobility in Germany 2008_Projects Presentation.pdf"
- MiD2008_Information.docx

[Dokumentation]
[SPSS_MiD2008]

→

[Ergebnisbericht]
[Methodenbericht]
[Nutzerhandbuch]
[Tabellenband]

→

[0_NUTZERHANDBUCH]
[1_Erhebungsunterlagen]
[2_Intervieweranleitung]
[3_Variablenübersicht]
[4_Codeplan]
[5_Variablenaufbereitung]
[6_Variablenvergleich MiD 2002 und MiD 2008]
[7_Tabellierungswerkzeug MiT]

Example MID2008, steps

3 sessions in the computer lab:

1. Reading the data, recoding, plausibility checks, filter checks
2. Descriptive statistics
3. Open

Guest lecture Tina Gehlert

Exploring the data – steps for data preparation

- Check individual variables for extreme and implausible values
- Check cross variable consistency (e.g. compute speed from trip distance and trip time, check with transport mode)
- Compute new variables that are needed for the analysis

- Most important:
- **Never work with original data! Use copies for your analyses!**

Thank you for your attention.

Regine Gerike

Technische Universität München

mobil.TUM

Office: 1753

Tel +49.89.289.28575

regine.gerike@tum.de

# #

- #

MiD2008

- Relevant variables:
- Number of trips on the diary day: wege1
- Travel time diary day: anzmin
- Travel distances diary day: anzkm
- Yearly distances travelled by car [km/year]: fahrlj_h
- CO_2-Emissions: co2tag_h (Household, diary day), co2tag_p (Person, diary day), co2weg (per trip)
- Household income: hheink

- Household variables: household type and household size, number of employed household members, number of driving licences, number of cars/bikes/motor cycles, reasons for not owning a car, type of region
- Person variables: car/bike availability, frequency usage car/bike/PT/airplane, gender, age, distances to and accessibility of facilities, employment status, education, type of PT-tickets, driving licence, availability handies/computers, mobility restrictions, type of region