Official Firm Data for Germany

The AFiD-products of the RDC: Opportunities and constraints

Kristin Nowak, regional RDC-Office Stuttgart
Diane Zabel, regional RDC-Office Bremen
Vienna, September 6th 2017

Table of contents

1. Introduction

1.1. What are research data centres (RDCs)?

- Centres for official **microdata**

- Part of the **„informational infrastructure"** for scientists in Germany

- **Service institutions that mediate** between opposing interests

- Main tasks:

  – Generate supply of microdata:

    Centralised **preparation** (anonymization, supply of metadata etc.), centralised **management** and **supply** of microdata via various forms of access

  – Accompany the demand for microdata:

    RDCs enable scientific **analysis** of microdata in compliance with confidentiality rules; they advise the scientists on how to use the data

The RDC of the German regional statistical offices:
a working group with locations in every region

STATISTISCHE ÄMTER
DER LÄNDER
FORSCHUNGSDATENZENTRUM

Kiel | Hamburg

Hannover

Bremen

Schwerin

Düsseldorf

Berlin |Potsdam

Halle

Bad Ems

Kamenz | Dresden

Saarbrücken

Erfurt

Wiesbaden | Frankfurt

München|Fürth

Stuttgart

© Statistische Ämter der Länder, Forschungsdatenzentrum
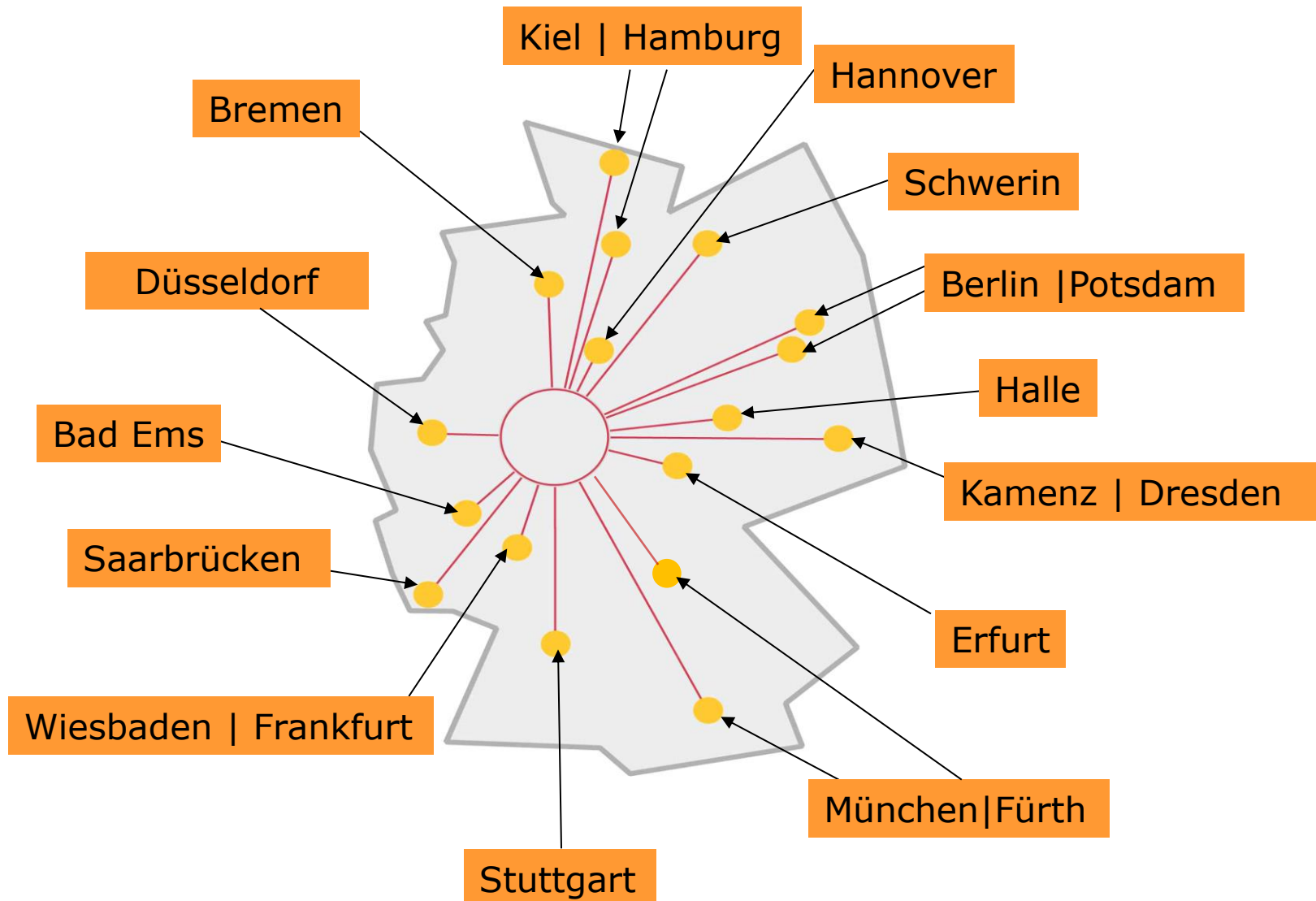
1.2.1. Microdata: advantages and drawbacks

- Higher informational content then aggregate data

- Better and more diverse options for **in-depth analysis.** Therefore answers to data questions are **more valid**, i.e. better in touch with reality and more to the point.

- Example: interrelationships between variables can be better examined, e.g. through regression analysis

- Microdata are especially well suited to (see Joachim Wagner (2015)):
    - a) discover new **stylised facts**,
    - b) test **model assumptions**,
    - c) test **hypotheses** and
    - d) to evaluate (economic) **policy** measures.

- But: Microdata are subject to **confidentiality legislation** for statistical data production, as they contain information for individuals or individual firms

- **Legislators** allow the use of official microdata under certain conditions (§ 16 BStatG)

1.2.2. Microdata in „our" RDCs

- **RDCs contain more than 100 statistics in the areas of:**
    - Economy (manufacturing, services, etc.)
    - Social themes (households, population, health, education)
    - Agriculture, energy and environment
    - Finance and taxation
    - Law and others

- => statistics most in demand by scientists enter the RDCs

- About 40% of the demand for statistics in our RDC in 2016 was about **AFiD-products**, compared to about 20% for the German „Mikrozensus"

- **Meta data** mostly on the internet: www.forschungsdatenzentrum.de

- A **fee** is levied from users (for most forms of data access)

2. The AFiD-Project: „<u>A</u>mtliche <u>Fi</u>rmendaten für <u>D</u>eutschland" (AFiD= „Official firm data for Germany")

2.1. Background, the approach and current products

AFiD: Background

- **Problems of isolated official statistics that <u>AFiD tries to solve</u>:**

    a) Limited number of <u>variables</u> (due to legal requirements)
    b) Access to well prepared microdata <u>time-series</u>

- **Requirements to solve these problems through <u>data integration</u>:**

    a) **technical:**
    - availability of common numeric identifiers for firms in the respective statistic and year
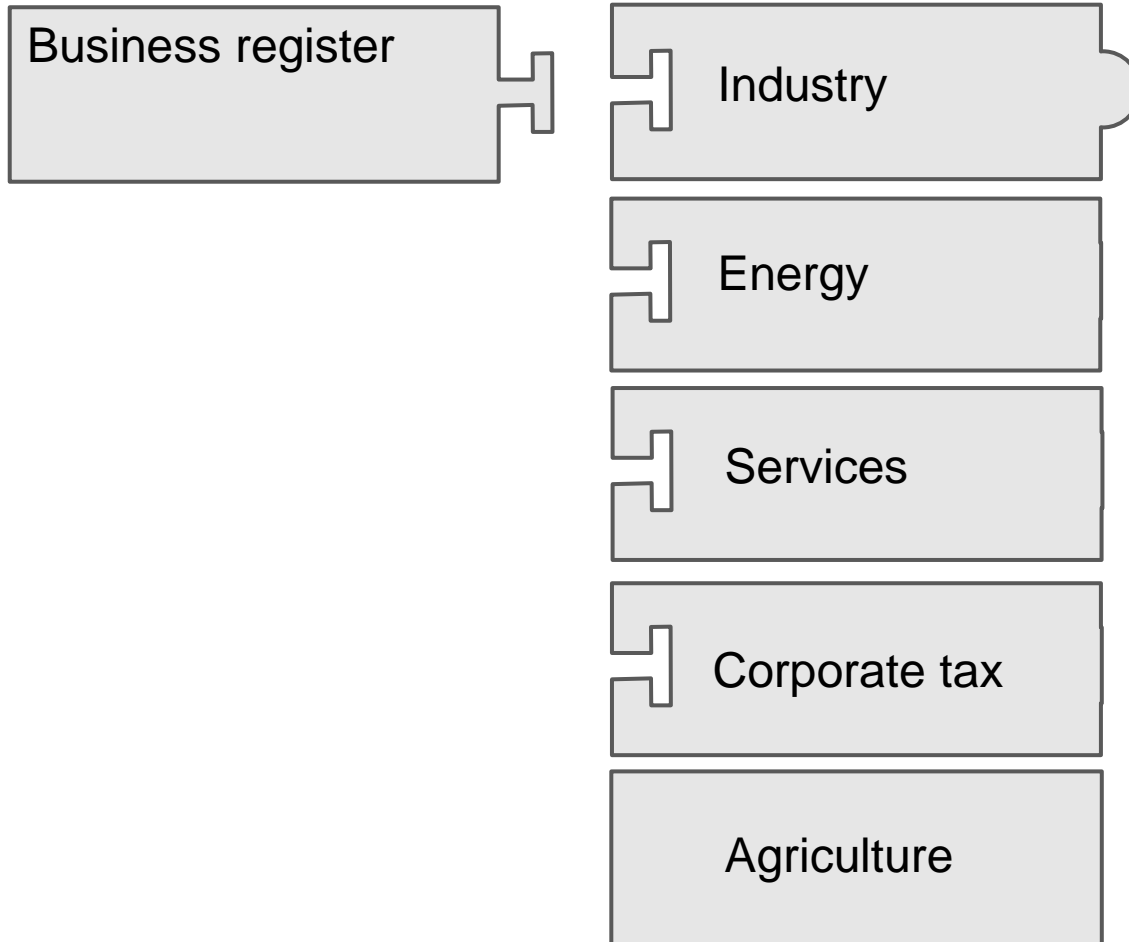
    .

    b) **legal:**
     - legitimacy of **data integrations** for firms (but not individuals): §13a BStatG;
     - legitimacy to **store** common numeric identifiers for 30 years
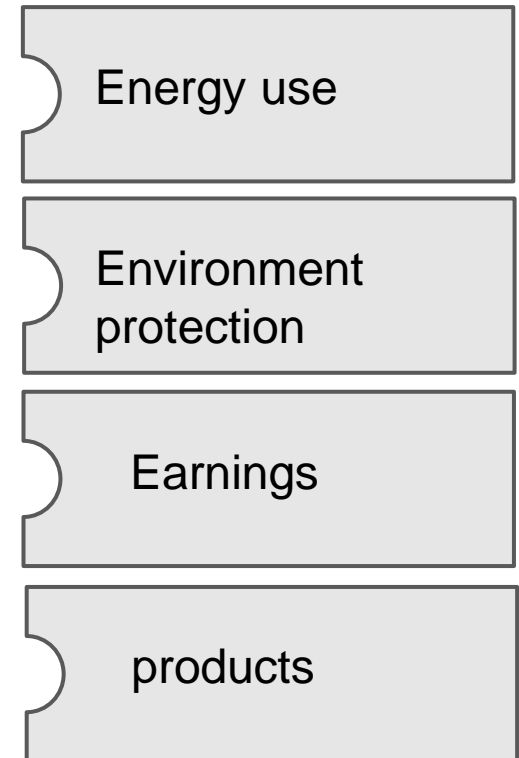
The AFiD-approach

- **Two step data integration** of firm-related microdata available in our RDC:

  1. Time-series or cross-sectional integration of official statistics:
     → Result: **Standardised AFiD-products** (panels und modules)
     → Example: AFiD-panel (local) Industrial Units
        = cross-section fusion of four official sample surveys in manufacturing

  2. Prompt and flexible add-on data matching tailored to the specific **needs of each research project**: e. g. matching of various standardised AFiD-products, like AFiD-modules and AFiD-panels.

- ➢ **Currently available outcome:**
    - ➢ eight AFiD-panels,
    - ➢ six AFiD-modules,
    - ➢ many project-specific microdata sets and
    - ➢ comprehensive meta data

- ➢ A wealth of integrated microdata for empirical research on firms in Germany

Currently available AFiD–products

AFiD-panels

AFiD-modules

Business register

Industry	Energy use

Energy	Environment protection

Services	Earnings

Corporate tax	products

Agriculture

2.2. Selected <u>AFiD-panels</u>: Content and research potential

AFiD-panels Industrial Units and Industrial Enterprises

- Availability: Industrial Units: 1995 – 2014, Industrial Enterprises: 2001 – 2014

- Panel <u>Industrial Units</u>: Variables:
  - number of employees
  - turnover
  - investments
  - wages and salaries
  - branch, official regional identifier (AGS → e.g. LAU-2-level), and other variables

- Panel <u>Industrial Enterprises</u>:
  - Additionally: information on **cost** structures including taxes and subsidies

- Research potential:
  - Differentiated analysis of enterprise success over branches, regions and time
  - Example of publication: Differences of technical efficiency across industrial branches

AFiD-panel official Statistical Business Register (URS-panel)

- Currently available for the years: 2000 – 2012.

- Variables:
    - number of employees subject to social insurance contribution (**SVB**)
    - branch (**WZ**), official regional identifier (**AGS**), **legal form** of enterprise
    - local units pertaining to an enterprise; enterprise **groups** (organizational and tax-wise)
    - various numeric **identifiers** for data matching
      (e. g.. URS-IDs, tax identifier, membership numbers (e. g. IHK), identifier for local units of the federal employment agency, identifier of the German commercial register, etc.)
    - **turnover** of enterprises – not available for general analysis – might, however, be used as control variables in complex investigations, e.g. regression

- Potential:
    - „**complete data collection**" → enables differentiated analysis across e. g. branches and regions
    - important for all **data matching** projects (URS is the ‚heart' of the AFID-project)
    - **concentration processes** can be analysed
    - However: limited themes for analysis due to lack of thematic, numerical variables (only **SVB**)

AFiD-panel Services

- Currently available years: 2003 - 2007 and 2008 - 2010, sample

- **Variables** for enterprises:
    - **staff** (number, costs)
    - revenue
    - **turnover** according to type of service (branches)
    - **input**
    - **taxes, subsidies**
    - **general** and other variables**.**

- Potential:
    - Analysis of enterprises in service

2.2. Selected <u>AFiD-modules</u>: Content and research potential

AFiD-module Earnings

- Linked. employer-employee-data, sample.
- Currently available years: 2001, 2006, 2010, 2014.
- **Variables** (for local units in manufacturing & services):

Employer-data set for local units:
- **employees** (type, gender)
- code of **collective labour agreement**
- weekly **working hours**, holidays
- general data (official regional identifier, branch)
- and other variables

Employee-data set:
- **pay scale classification** etc.
- **income**, income-tax group, tax exemption for dependent children, taxes, social-insurance contributions, etc.
- **occupation**, training
- **working hours**
- general data (branch, gender, year + month of birth, seniority) …

→ Potential: e. g. analysis of qualification, age and income structure

AFiD-module Environmental Investment

- Currently available years: 2003 - 2005 and 2006 - 2014, Sample

- Variables:
  - **Investments to protect the environment** (including rent for plant and equipment), split on subjects as:
    - Waste
    - Water
    - Air
    - Noise
    - Climate
    - Nature

- Potential:
  - Analysis of demand side of environmental goods + x
  - Exemplary publication: The effects of electricity taxation in manufacturing

2.3. Opportunities and constraints of the AFiD-project

Opportunities for matching additional data on AFiD-products

1. **Match microdata of various AFiD-products** (via URS-identifiers)
   E.g.: AFiD-module Energy Use → AFiD-panel Industrial Enterprises

2. **Match microdata of additional official statistics of the German statistical offices**, e. g. special topics surveys, process-related data (like the ‚sales tax statistic‘ and ‚foreign trade statistic‘)

3. **Match ‚generally available‘ microdata from non-official sources**, e. g. credit-worthiness data on Enterprises by Creditreform

4. **Match microdata of additional official data producers,** like the ‚German Central Bank‘ and the ‚German labour agency‘ („KomBiFiD pilot project“; enhanced options due to current amendment of the basic German Statistical Law (BStatG))

5. **Join external aggregated (grouped) data on the AFiD data via grouping variables** like official regions (AGS), branch-identifiers (WZ) and legal form

Outline on AFiD-projects

© Statistische Ämter der Länder, Forschungsdatenzentrum
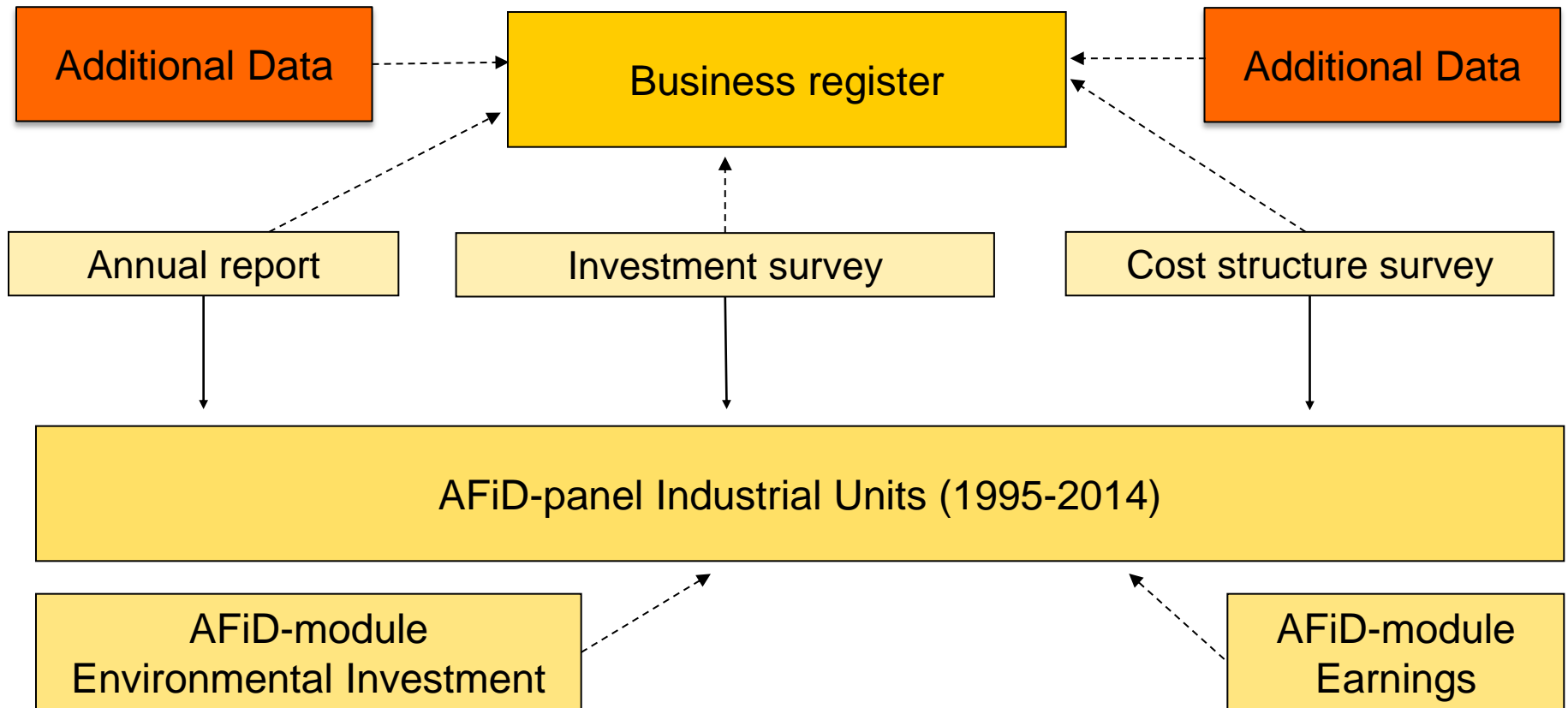
Constraints of the AFiD-project

Limitations of the data:

- **Discontinuities in certain variables** over time (e.g. official regions, branches)
  → may produce problems in data processing and analysis

- **Complete counts** are **often not** available
  → limits to e.g. the regional or branch-related differentiation of analyses

- Discontinuities in **sample selection**
  → participation numbers for balanced Panels for longer time-series may reduced

- The **number of available variables** is largely fixed by law
  → not all questions may be answered with the data

Organisational, legal and personal constraints:

- **Preparatory times** for complex AFiD-projects may be long.
  → current efforts go in the direction of delegating final work of data preparation to scientists

- **RDC fees** may be rather high – especially for complex projects with many datasets and fusions

- Current draft of the amendment of the **German official statistical law** (BStatG): limited storage time (10 years) for the URS-identifiers might impinge on the AFiD-project

- Data users need **programming know-how** in either STATA, SAS, SPSS (or R)

Thank you for your attention!

Contact:

Statistische Ämter der Länder, Forschungsdatenzentrum (RDC)

www.forschungsdatenzentrum.de und www.forschungsdatenzentrum.de/afid.asp

Regional RDC-office Stuttgart
in: Statistisches Landesamt Baden-Württemberg
Kristin Nowak: (49) 711 / 641-2864
E-Mail: forschungsdatenzentrum@stala.bwl.de

Regional RDC-office Bremen
In: Statistisches Landesamt Bremen
Diane Zabel: (49) 421 / 361-18456
E-Mail: forschungsdatenzentrum@statistik.bremen.de