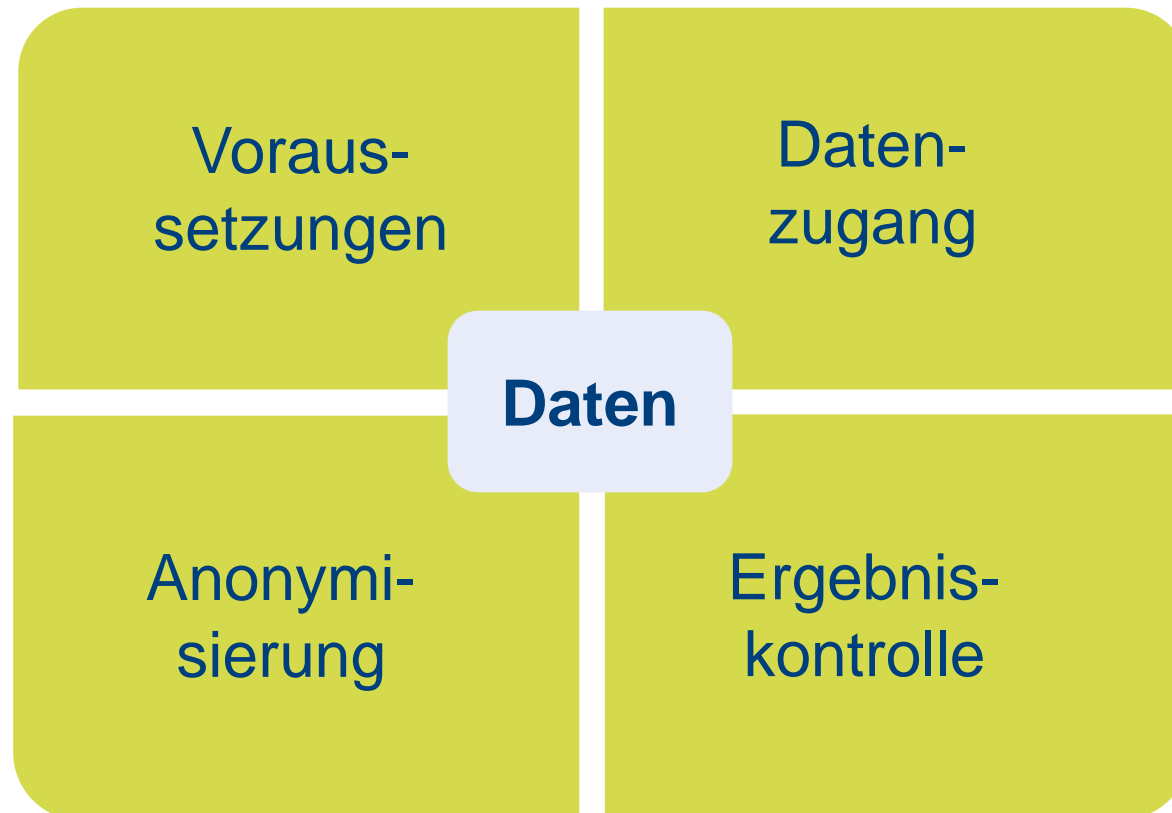


Spannungsfeld Sozialdatenschutz und Nutzung von Sozialdaten durch die Wissenschaft - Handhabung bei der BA (Teil 2)

Workshop:
Scientific Use Files aus verknüpften Sozialdaten
17.11.2015, Berlin

Manfred Antoni
Alexandra
Schmucker

Portfolio des Datenschutzes



Voraussetzungen für den Datenzugang

- wissenschaftliche Forschung
- Bezug zum sozialen Sicherungssystem oder Arbeitsmarkt- und Berufsforschung
- wiss. Forschungseinrichtung
- Zweckbindung
- Erforderlichkeit der Daten

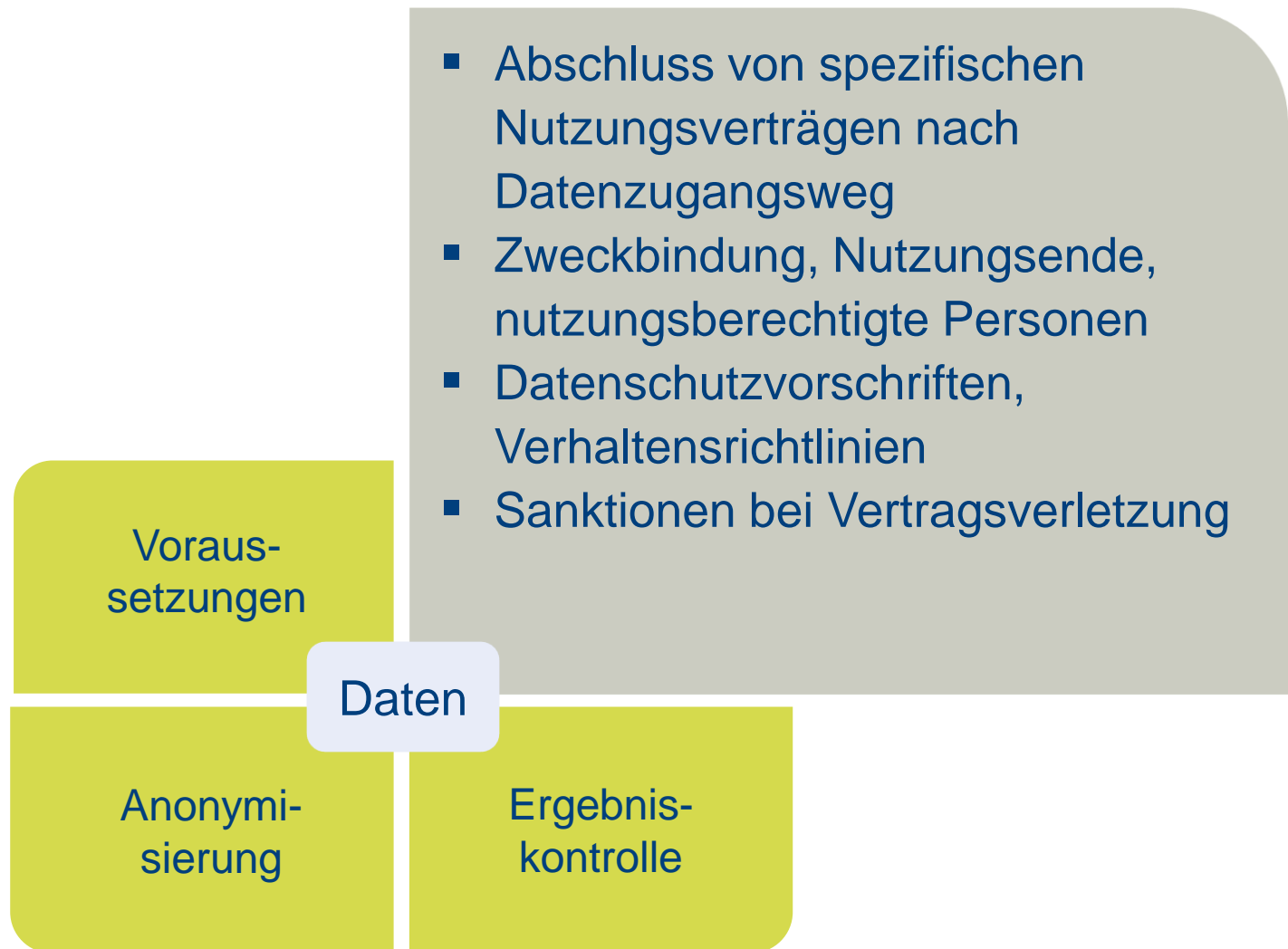
Daten-
zugang

Daten

Anonymi-
sierung

Ergebnis-
kontrolle

Regelungen zum Datenzugang

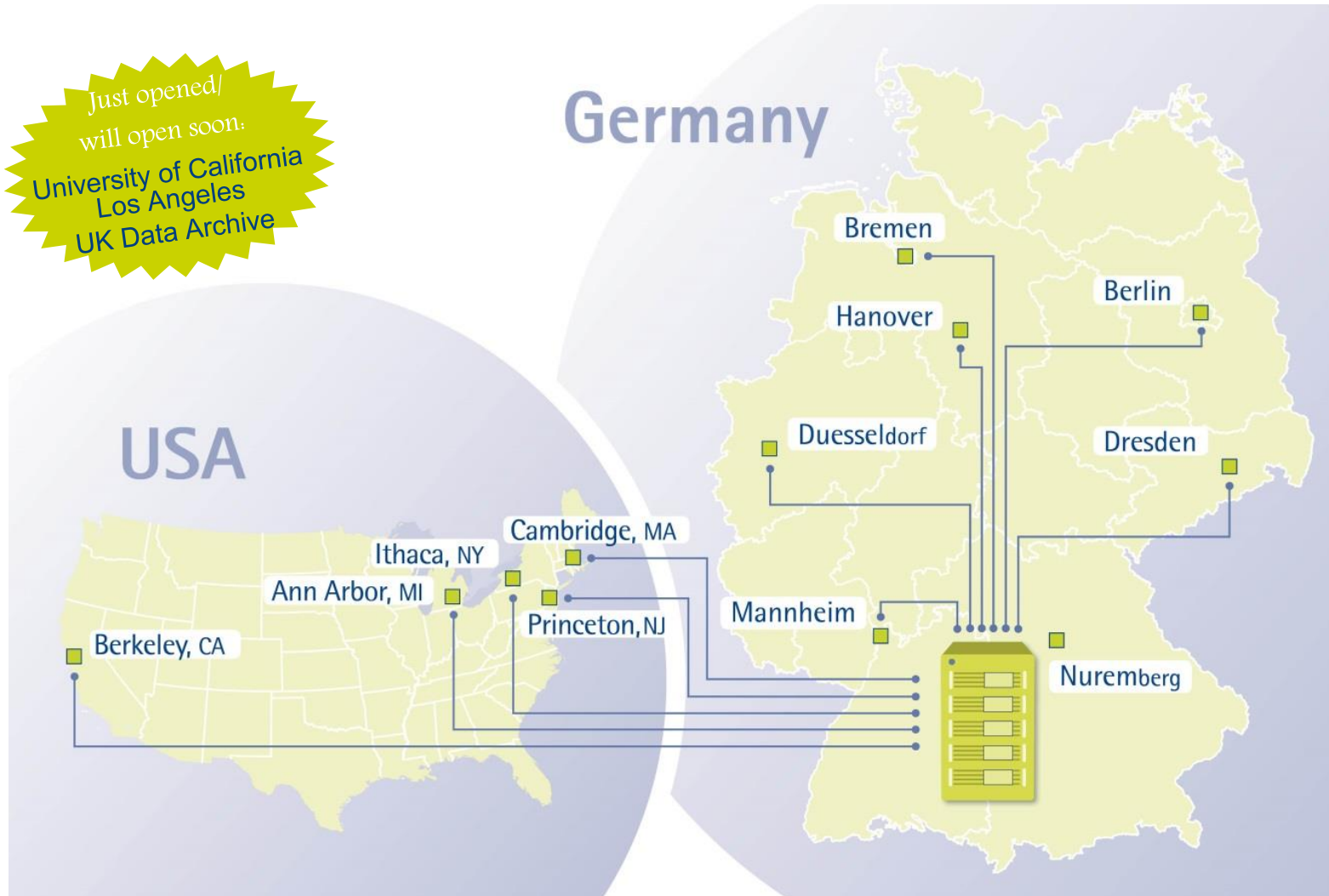


Datenzugangswege des FDZ

Off-Site Use		On-Site Use	
Campus Files	Scientific Use File	Datenfernverarbeitung mittels JoSuA	
Absolut anonym	Faktisch anonymisiert	Schwach anonymisiert	
Lehre	Datenverarbeitung in der wiss. Einrichtung	Übermittlung der Ergebnisse nach der Datenschutzprüfung	

Datenzugangswege (I): Gastaufenthalt

Just opened/
will open soon:
University of California
Los Angeles
UK Data Archive



Datenzugangswege (II): Datenfernverarbeitung

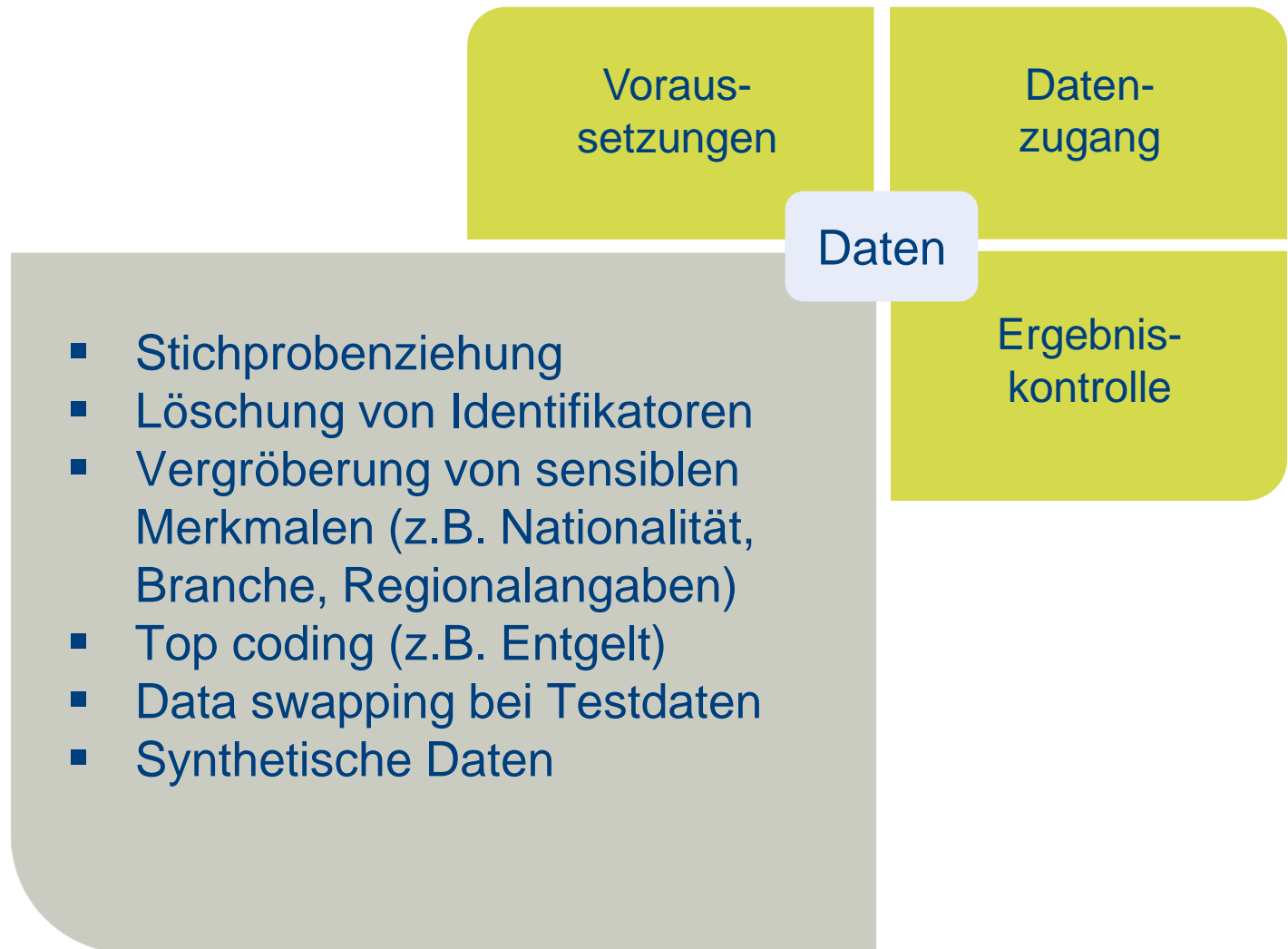


Job Submission Application

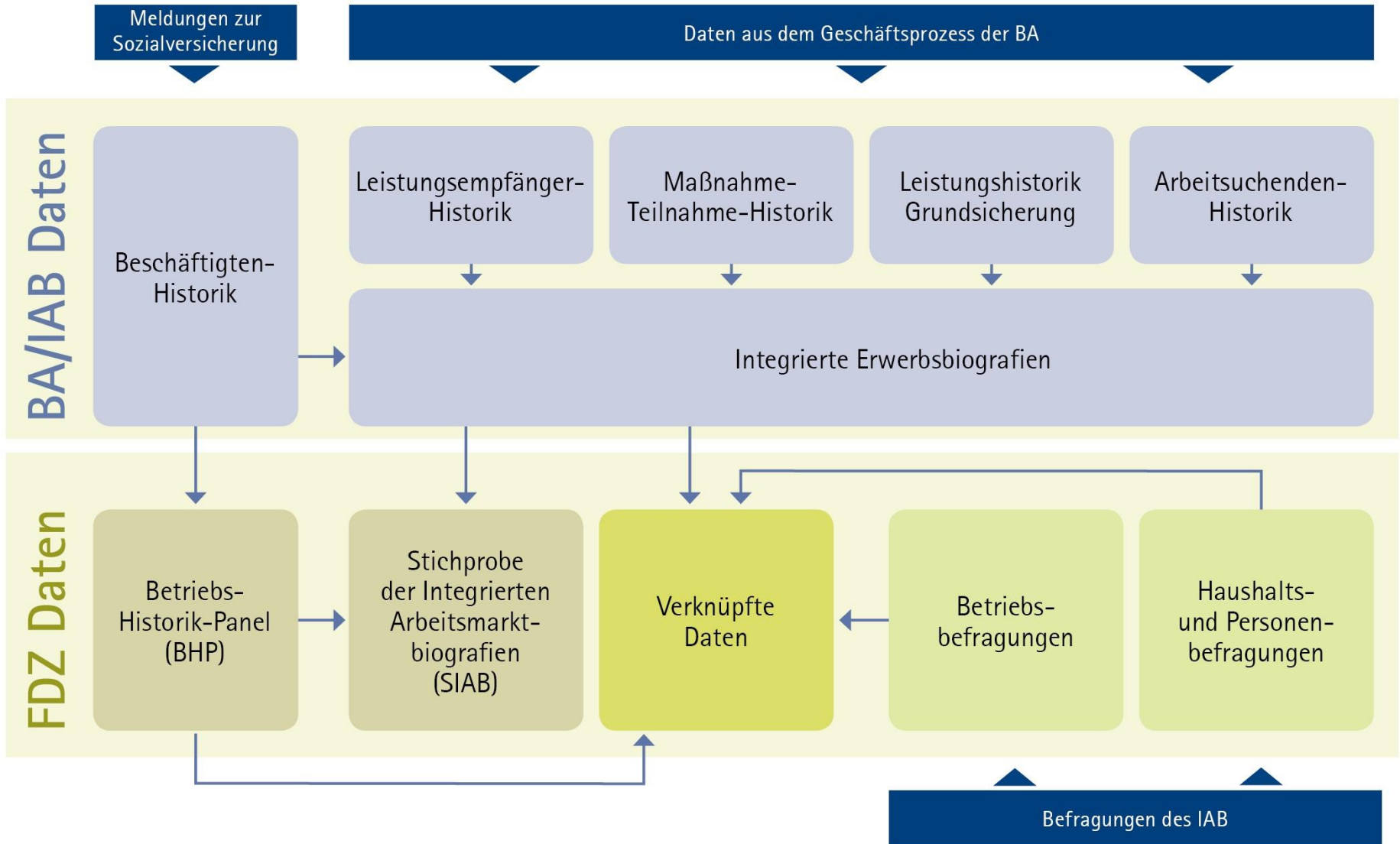
Web-Interface: <https://josua.iab.de/gui/index.html>

Manual: <https://josua.iab.de/gui/manual.pdf>

Anonymisierung von Mikrodaten



Datenüberblick



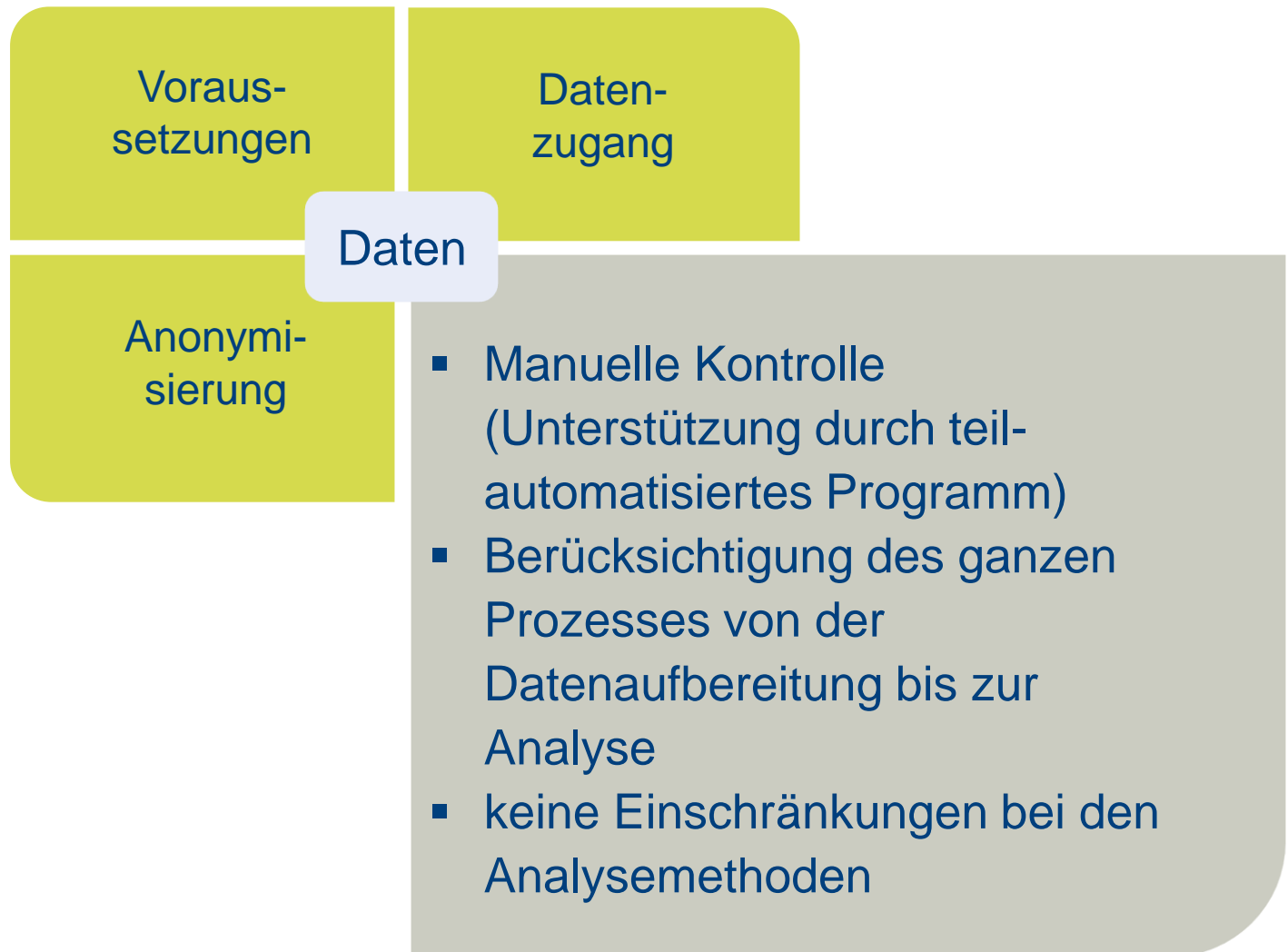
Vom FDZ aktuell angebotene verknüpfte Daten

- Verknüpfte Personen- oder Haushaltsdaten:
 - ALWA-Befragungsdaten verknüpft mit administrativen Daten des IAB (ALWA-ADIAB)
 - PASS-Befragungsdaten verknüpft mit administrativen Daten des IAB (PASS-ADIAB)

- Verknüpfte Betriebs- und Personendaten:
 - Linked-Employer-Employee-Daten des IAB (LIAB)
 - Linked Personnel Panel (LPP)
 - WeLL-Befragungsdaten verknüpft mit administrativen Daten des IAB (WeLL-ADIAB)

- Alle verknüpften Datenprodukte ausschließlich per Gastaufenthalt mit anschließender Datenfernverarbeitung zugänglich

Richtlinien der Ergebniskontrolle



Kontakt

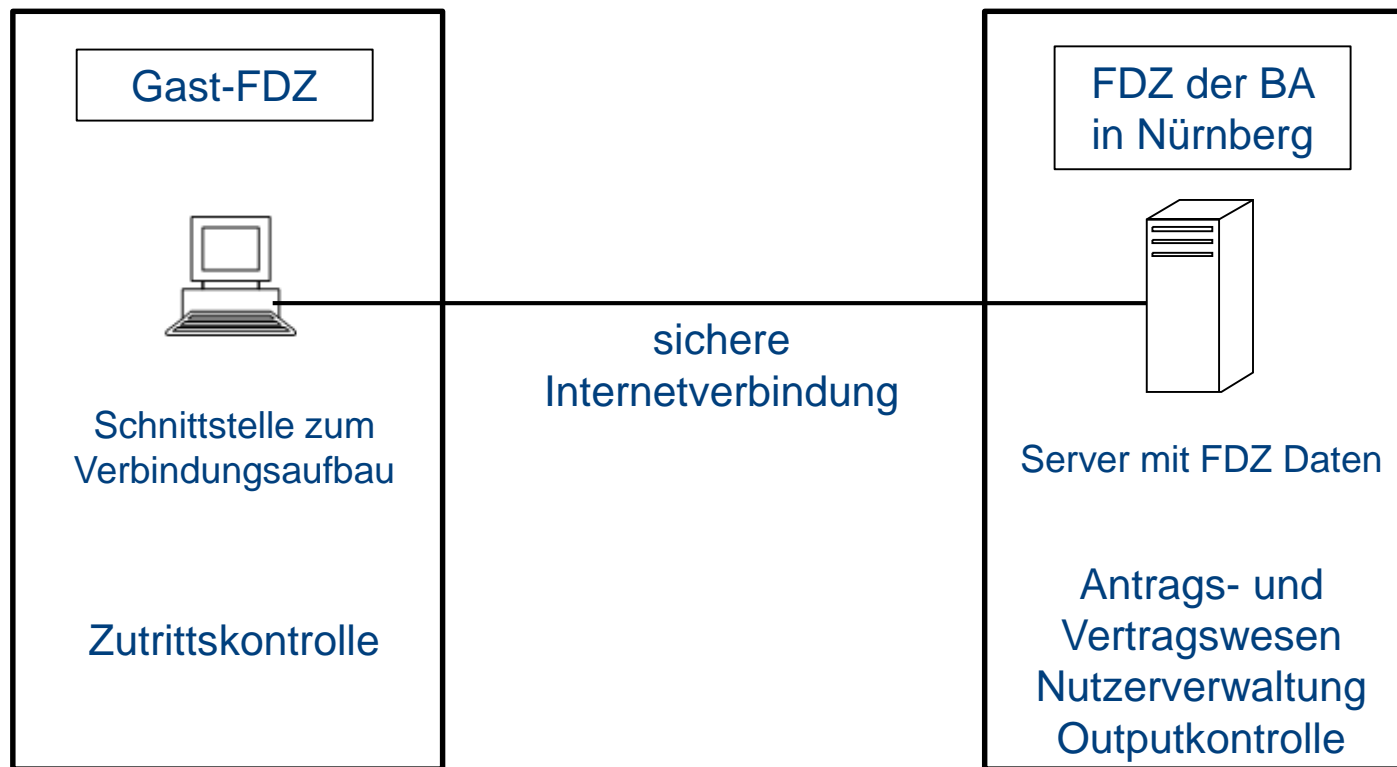
Manfred Antoni
(manfred.antoni@iab.de)

Alexandra Schmucker
(alexandra.schmucker@iab.de)

Homepage des Forschungsdatenzentrums:
<http://fdz.iab.de/>

FDZ-in-FDZ: Grundidee

- Zugang zu den Daten des FDZ in den Räumen eines anderen FDZ (Gast-FDZ)



FDZ-in-FDZ: Technische Umsetzung

- Citrix-Thin-Client-Lösung:



- „Reduzierter“ Rechner
- Benutzerschnittstelle
- Nur Software, die zum Aufbau der Verbindung benötigt wird
- Citrix: Firmenname
- Sichere Kommunikationsverbindung
- Kombination aus Software (Access Gateway) und Server
- Internationaler Standard (Statistics Netherlands, Banken, BA: client2015)

Datenschutzprüfungen am FDZ: Voraussetzungen

Nutzerinnen und Nutzer müssen folgende Regeln bei der Analyse von schwach anonymisierten Daten beachten:

- Syntax-gestützte Programmierung
- Ausführlicher Dokumentation der Analyseschritte zur Nachvollziehbarkeit für das FDZ
- Protokollierung der Programmcodes und der dazugehörigen Resultate in Ergebnisdateien
- Anlegen einer Master-Datei zum Starten aller relevanten Syntaxdateien in der richtigen Reihenfolge

Richtlinien: http://doku.iab.de/fdz/access/Vorgaben_DAFE.PDF

Beispiel für Master-Datei

```
capture cd U:\fdzXXX\  
  
* Master-Datei des Projektes "Projektname"  
* Nutzer: T. Huber (th) und W. Müller (wm)  
  
version 13  
clear  
set more off // damit das Programm durchläuft  
set linesize 255 // damit Tabellen nicht unnötig umbrechen  
adopath ++ $prog // Speicherort der ado-Dateien (falls benötigt)  
  
*do "$prog/th01_PersDat.do" // Aufbereitung Personendaten  
*do "$prog/th02_BetrDat.do" // Aufbereitung Betriebsdaten  
*do "$prog/th03_DatXY.do" // Zusammenspielen  
*do "$prog/th04_AnalyseA.do" // Untersuchung Hypothese 1 (...)  
*do "$prog/th05_AnalyseB.do" // Untersuchung Hypothese 2 (...)  
  
*do "$prog/wm01_Datagen.do" // Aufbereitung Personendaten  
do "$prog/wm02_AnalyseX.do" // Untersuchung Hypothese 5 (...)  
do "$prog/wm03_AnalyseY.do" // Untersuchung Hypothese 6 (...)  
  
* Ausgabe einer Dateiliste in eine log-Datei  
capture log close  
log using "$log/Dateiliste.log", replace  
dir $prog\  
dir $log\  
dir $data\  
dir $orig\  
log close
```


Datenschutzprüfungen am FDZ: Praxis

- Primärsperrung: Löschen von Ergebnissen, die auf weniger als 20 Beobachtungen basieren
- Sekundärsperrung: Löschen von weiteren Werten, um „Rückrechnen“ von gelöschten Werten anhand von Randsummen zu verhindern
- Dominanzsperrung: Löschung von Ergebnissen, um die Identifikation von Dominanzbetrieben zu verhindern
- Keine Einbeziehung von „alten“ Ergebnissen in die aktuelle Prüfung
- Vertragliche Verpflichtung der Nutzerinnen und Nutzer, keine Rückrechnen aus den einzelnen übermittelten Ergebnissen vorzunehmen

Beispiel 1: Anonymisierung von Kreuztabellen I

Tabelle 1: vorher

Ostdeutschland

Anzahl SVB	Betriebsrat		
	Ja	Nein	Total
1-4 SVB	43	1.380	1.423
5-9 SVB	39	547	586
10-19 SVB	89	487	576
20-49 SVB	250	590	840
50-99 SVB	255	245	500
100-199 SVB	290	110	400
200-499 SVB	283	65	348
500 + SVB	142	16	158
Total	1.391	3.440	4.831

Tabelle 2: nachher

Ostdeutschland

Anzahl SVB	Betriebsrat		
	Ja	Nein	Total
1-4 SVB	43	1.380	1.423
5-9 SVB	3 *	54 *	586
10-19 SVB	89	487	576
20-49 SVB	250	590	840
50-99 SVB	255	245	500
100-199 SVB	290	110	400
200-499 SVB	283	65	348
500 + SVB	14 *	/	158
Total	1.391	3.440	4.831

Beispiel 1: Anzahl der Betriebe mit und ohne Betriebsrat für unterschiedliche Betriebsgrößenklassen in Ostdeutschland. Die Ergebnisse basieren auf Berechnungen mit den Testdaten des IAB-Betriebspanels.

Beispiel 2: Anonymisierung von Kreuztabellen II

Tabelle 1: vorher

Anzahl SVB	Ostdeutschland			Westdeutschland			Gesamtdeutschland		
	Betriebsrat			Betriebsrat			Betriebsrat		
	Ja	Nein	Total	Ja	Nein	Total	Ja	Nein	Total
1-4 SVB	43	1.380	1.423	64	2.461	2.525	107	3.841	3.948
5-9 SVB	39	547	586	54	847	901	93	1.394	1.487
10-19 SVB	89	487	576	130	762	892	219	1.249	1.468
20-49 SVB	250	590	840	364	853	1.217	614	1.443	2.057
50-99 SVB	255	245	500	365	370	735	620	615	1.235
100-199 SVB	290	110	400	391	165	556	681	275	956
200-499 SVB	283	65	348	402	90	492	685	155	840
500 + SVB	142	16	158	198	22	220	340	38	378
Total	1.391	3.440	4.831	1.968	5.570	7.538	3.359	9.010	12.369

Beispiel 2: Anzahl der Betriebe mit und ohne Betriebsrat für unterschiedliche Betriebsgrößenklassen in Ost-, West- und Gesamtdeutschland. Die Ergebnisse basieren auf Berechnungen mit den Testdaten des IAB-Betriebspanels.

Beispiel 2: Anonymisierung von Kreuztabellen II

Tabelle 2: nachher

Anzahl SVB	Ostdeutschland			Westdeutschland			Gesamtdeutschland		
	Betriebsrat			Betriebsrat			Betriebsrat		
	Ja	Nein	Total	Ja	Nein	Total	Ja	Nein	Total
1-4 SVB	43	1.380	1.423	64	2.461	2.525	107	3.841	3.948
5-9 SVB	3*	54*	586	5*	84*	901	93	1.394	1.487
10-19 SVB	89	487	576	130	762	892	219	1.249	1.468
20-49 SVB	250	590	840	364	853	1.217	614	1.443	2.057
50-99 SVB	255	245	500	365	370	735	620	615	1.235
100-199 SVB	290	110	400	391	165	556	681	275	956
200-499 SVB	283	65	348	402	90	492	685	155	840
500 + SVB	14*	/	158	19*	2*	220	340	38	378
Total	1.391	3.440	4.831	1.968	5.570	7.538	3.359	9.010	12.369

Beispiel 2: Anzahl der Betriebe mit und ohne Betriebsrat für unterschiedliche Betriebsgrößenklassen in Ost-, West- und Gesamtdeutschland. Die Ergebnisse basieren auf Berechnungen mit den Testdaten des IAB-Betriebspanels.

Beispiel 3: Anonymisierung statistischer Kennzahlen

Tabelle 1: vorher

Branche	Anzahl Azubis	Anzahl Betriebe
Land-, Forstwirtschaft, ...	22	27
Bergbau	39	5
Textilien, Lederwaren, ...	11	17
Reparatur, ...	4	12
Information, ...	28	20
Architekturbüros, ...	2	20
Forschung, Entwicklung	22	12
Erziehung, Unterricht	134	69
Total	262	182

Tabelle 1: nachher

Branche	Anzahl Azubis	Anzahl Betriebe
Land-, Forstwirtschaft, ...	22	27
Bergbau	/	/
Textilien, Lederwaren, ...	/	/
Reparatur, ...	/	/
Information, ...	28	20
Architekturbüros, ...	/	20
Forschung, Entwicklung	/	/
Erziehung, Unterricht	134	69
Total	262	182

Beispiel 3: Anzahl der übernommenen Azubis in ausgewählten Branchen.
Die Ergebnisse basieren auf Berechnungen mit den Testdaten des IAB-Betriebspanels.

Beispiel 4: Anonymisierung von Mittelwerten

Tabelle 1: vorher

Variable	Obs	Mean	Std. Dev.	Min	Max
Ausbildungsvertrag	201	2.373134	.9192794	1	3
Ausbildungsplätze	140	.0857143	.2809469	0	1
Anz. Ausbildungsplätze	73	2.219178	2.340742	1	15

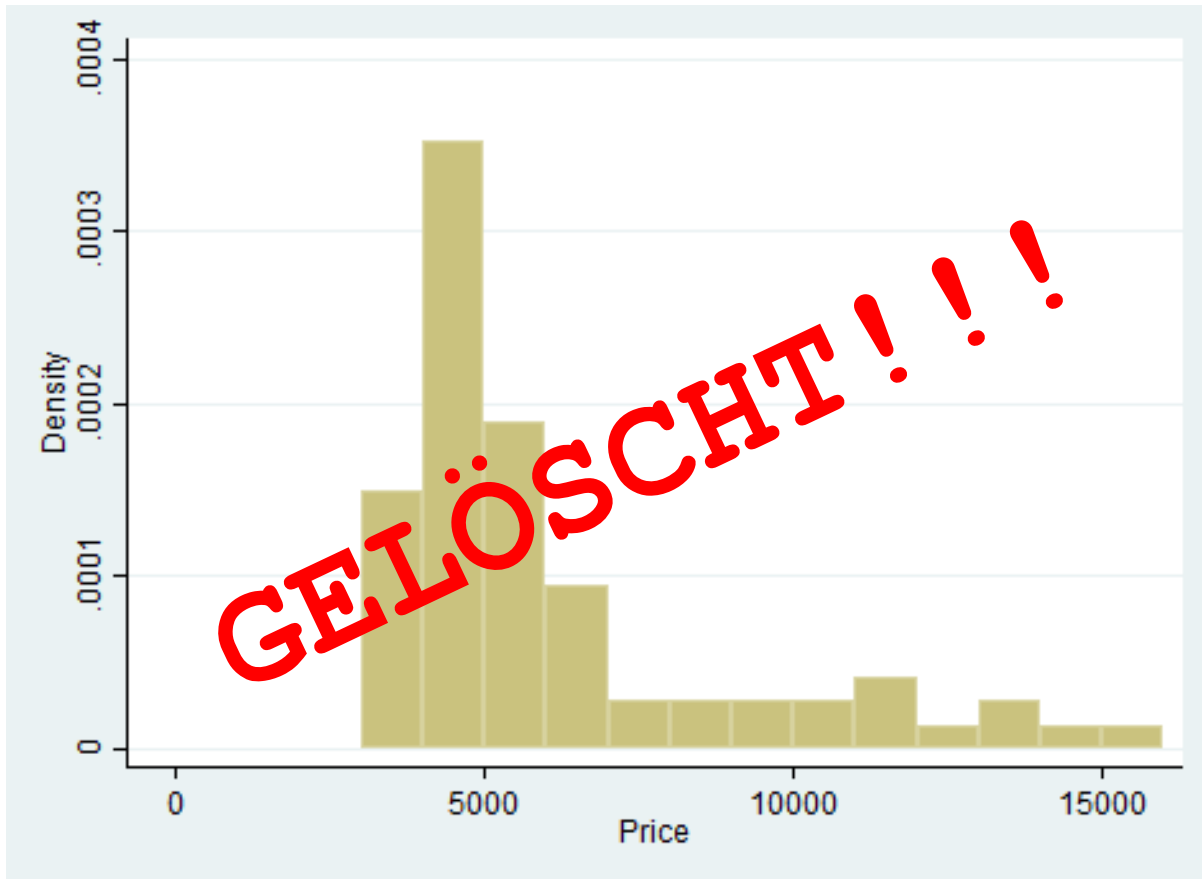
$140 \times 0.0857143 = 12$ ➔ 12 Betriebe bieten einen Ausbildungsplatz an

Tabelle 1: nachher

Variable	Obs	Mean	Std. Dev.	Min	Max
Ausbildungsvertrag	201	2.373134	.9192794	1	3
Ausbildungsplätze	140	/	/	/	/
Anz. Ausbildungsplätze	73	2.219178	2.340742	1	15

Beispiel 4: Mittelwerte der Variablen „Ausbildungsvertrag abgeschlossen“, „Ausbildungsplatz angeboten“ und Anzahl der angebotenen Ausbildungsplätze. Die Ergebnisse basieren auf Berechnungen mit den Testdaten des IAB-Betriebspanels.

Beispiel 5: Anonymisierung von Histogrammen

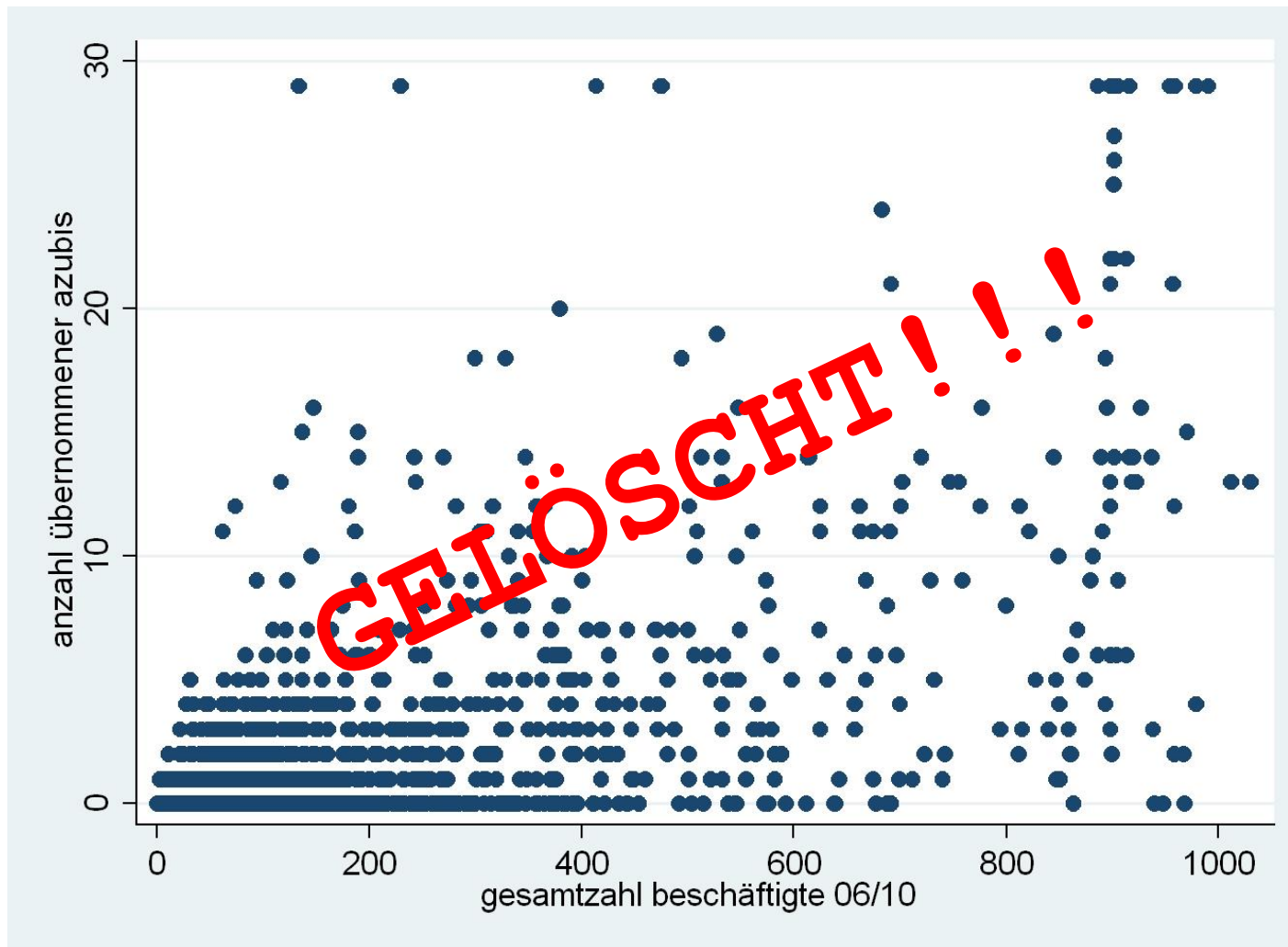


nachher

dens	freq	x
/	/	/
.0003514	26	4500
/	/	/
/	/	/
/	/	/
/	/	/
/	/	/
/	/	/
/	/	/
/	/	/
/	/	/
/	/	/
/	/	/
.	.	0

Beispiel 5: Preisverteilung der Automobile. Die Ergebnisse basieren auf Berechnungen mit dem Stata-Beispieldatensatz „auto.dta“.

Beispiel 6: Anonymisierung von Scatterplots



Beispiel 6: Scatterplot zur Anzahl der übernommenen Azubis nach Anzahl der Beschäftigten insgesamt. Die Ergebnisse basieren auf Berechnungen mit den Testdaten des IAB-Betriebspanels.

Datenschutzprüfungen am FDZ: weitere Aspekte

- **Perzentile:**
 - Mindestens 20 Beobachtungen pro Perzentil
- **Gewichtungen und Hochrechnungen:**
 - Prüfung anhand der ungewichteten Fallzahlen
 - analoge Löschung in den gewichteten/hochgerechneten Ergebnissen
- **aggregierte Datensätze:**
 - Erstellung einer Variable mit Anzahl der zugrundeliegenden Beobachtungen für jedes aggregierte Merkmal
 - Projektspezifische Absprachen zur Implementation der Datenschutzprüfung
 - Einmalige Herausgabe pro Projekt
- **Ausgabeformate:**
 - Ausschließlich Text-Dateien
 - keine gesonderten Dateien neben den Syntax-generierten Ergebnisdateien