



FORSCHUNGSDATENZENTRUM  
der Bundesagentur für Arbeit im Institut für  
Arbeitsmarkt- und Berufsforschung

# ANONYMISIERUNG UND DATENZUGANG AM FORSCHUNGSDATENZENTRUM DES INSTITUTS FÜR ARBEITSMARKT- UND BERUFSFORSCHUNG (IAB)

13.10.2022  
ELSA-Workshop  
Berlin

Sandra Dummert



# DAS INSTITUT FÜR ARBEITSMARKT- UND BERUFSFORSCHUNG (IAB)

---

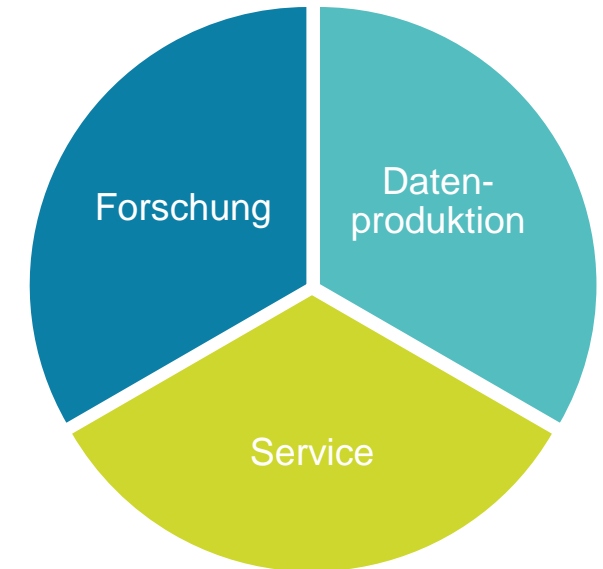
- Forschungseinrichtung der Bundesagentur für Arbeit (BA)
- Wissenschaftliche Unabhängigkeit
- Arbeitsmarkt- und Berufsforschung
- Politikberatung
- Forschungsdaten für wissenschaftliche Zwecke



# DAS FORSCHUNGSDATENZENTRUM (FDZ) DER BA IM IAB

---

- Generierung von standardisierten Datenprodukten
- Verlinkung von administrativen Arbeitsmarktdaten mit (eigenen) Befragungen
- Erstellung von Datendokumentationen
- Kostenloser Zugang zu Mikrodaten für die nicht-kommerzielle Arbeitsmarkt- und Berufsforschung
  - Gesetzlicher Auftrag (SGB III und X)
- Datenzugang im In- und Ausland
- Beratung der Datennutzenden hinsichtlich der Datenauswahl und des Beantragungsprozesses
- Forschung mit eigenen Daten



# DATENANGEBOT

<https://fdz.iab.de/unsere-datenprodukte/dateneuebersicht/>

## Administrative Daten

Meldungen zur  
Sozialver-  
sicherung

Daten aus dem  
Geschäftsprozess  
der BA

am FDZ verfügbare Daten

Befragungen

Externe  
Daten

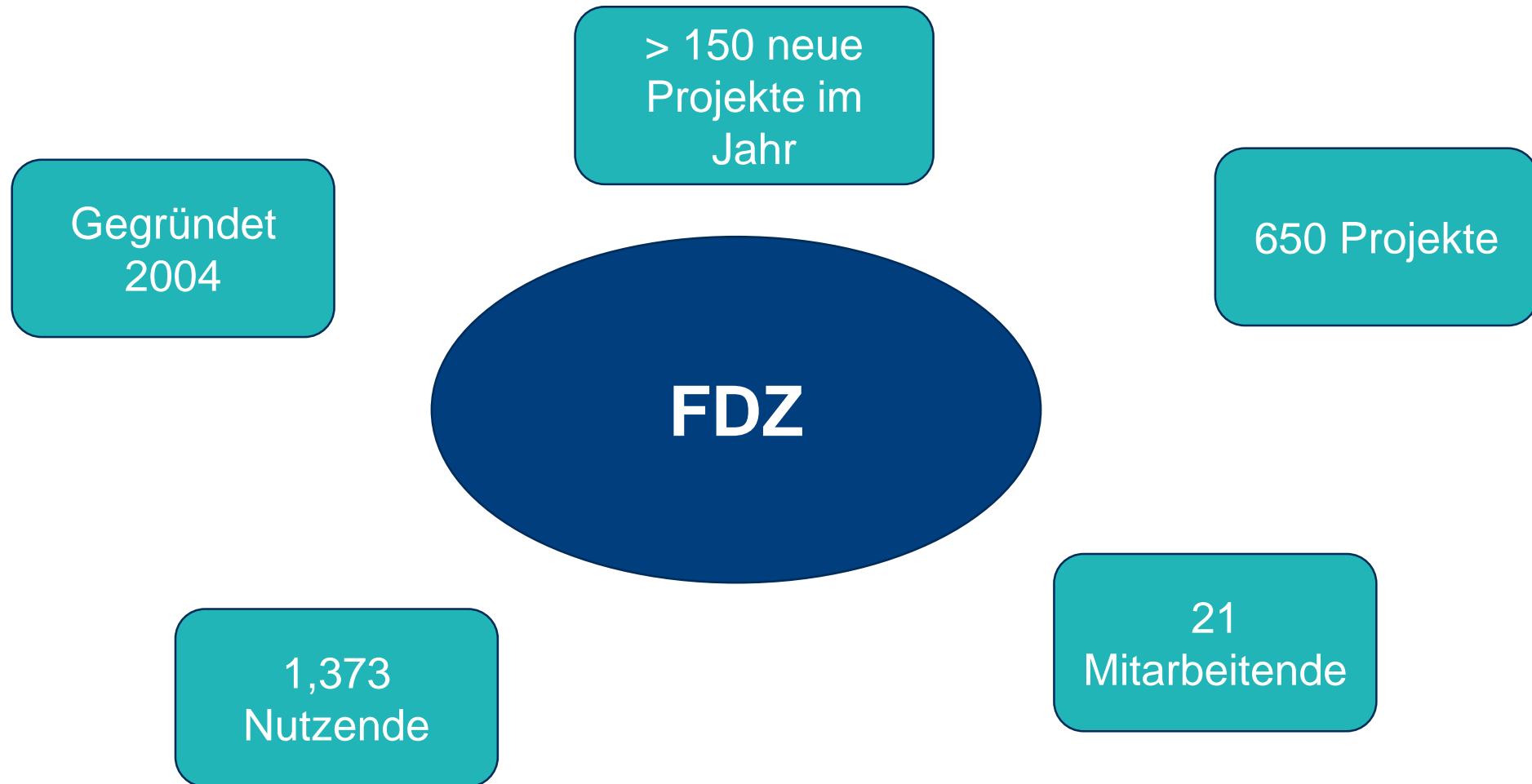
- Mikrodaten zu Personen, Haushalten und Betrieben

# BEISPIEL (TESTDATEN) – STICHPROBE DER INTEGRIERTEN ARBEITSMARKTBIOGRAFIEN (SIAB)

persnr	betnr	spell	quelle	begorig	endorig	frau	gebjahr	gebmon	nation
972180	.n Fehlt syst.	1	3 LHG Leistungshistorik Grundsicherung	19 Feb 2005	31 Dec 2005	0 Maennlich	1931	7 Juli	0 Deutschland
972180	.n Fehlt syst.	2	7 XASU Arbeitsuchendenhistorik aus XSozial	19 Feb 2005	31 Dec 2006	0 Maennlich	1931	7 Juli	0 Deutschland
972180	.n Fehlt syst.	3	3 LHG Leistungshistorik Grundsicherung	01 Jan 2006	30 Dec 2006	0 Maennlich	1931	7 Juli	0 Deutschland
972180	.n Fehlt syst.	4	7 XASU Arbeitsuchendenhistorik aus XSozial	19 Feb 2005	31 Dec 2006	0 Maennlich	1931	7 Juli	0 Deutschland
972180	.n Fehlt syst.	5	3 LHG Leistungshistorik Grundsicherung	31 Dec 2006	31 Dec 2006	0 Maennlich	1931	7 Juli	0 Deutschland
972180	.n Fehlt syst.	6	7 XASU Arbeitsuchendenhistorik aus XSozial	19 Feb 2005	31 Dec 2006	0 Maennlich	1931	7 Juli	0 Deutschland
972190	50084885	1	1 BeH Beschaeftigtenhistorik	30 Jul 1991	31 Dec 1991	1 Weiblich	1979	12 Dezember	148 Niederlande
972190	50084885	2	1 BeH Beschaeftigtenhistorik	01 Jan 1992	31 Dec 1992	1 Weiblich	1979	12 Dezember	148 Niederlande
972190	.n Fehlt syst.	3	2 LEH Leistungsempfaengerhistorik	16 Jan 1993	26 Jan 1993	1 Weiblich	1979	12 Dezember	148 Niederlande
972190	.n Fehlt syst.	4	2 LEH Leistungsempfaengerhistorik	08 Feb 1993	20 Apr 1993	1 Weiblich	1979	12 Dezember	148 Niederlande
972190	50608398	5	1 BeH Beschaeftigtenhistorik	08 Apr 1993	24 Apr 1993	1 Weiblich	1979	12 Dezember	148 Niederlande
972190	.n Fehlt syst.	6	2 LEH Leistungsempfaengerhistorik	08 Feb 1993	20 Apr 1993	1 Weiblich	1979	12 Dezember	148 Niederlande
972190	50608398	7	1 BeH Beschaeftigtenhistorik	08 Apr 1993	24 Apr 1993	1 Weiblich	1979	12 Dezember	148 Niederlande
972190	50047300	8	1 BeH Beschaeftigtenhistorik	18 Jul 1993	20 Sep 1993	1 Weiblich	1979	12 Dezember	148 Niederlande
972191	.n Fehlt syst.	1	6 ASU Arbeitsuchendenhistorik	07 Mar 2001	20 Nov 2001	0 Maennlich	1989	6 Juni	0 Deutschland
972191	.n Fehlt syst.	2	4 MTH Massnahmeteilnahmehistorik	13 Oct 2001	14 Dec 2001	0 Maennlich	1989	6 Juni	0 Deutschland
972191	.n Fehlt syst.	3	6 ASU Arbeitsuchendenhistorik	07 Mar 2001	20 Nov 2001	0 Maennlich	1989	6 Juni	0 Deutschland
972191	.n Fehlt syst.	4	4 MTH Massnahmeteilnahmehistorik	13 Oct 2001	14 Dec 2001	0 Maennlich	1989	6 Juni	0 Deutschland

# STATISTIKEN

---



# ANONYMISIERUNG

---

# ANONYMISIERUNGSSCHRITTE

---

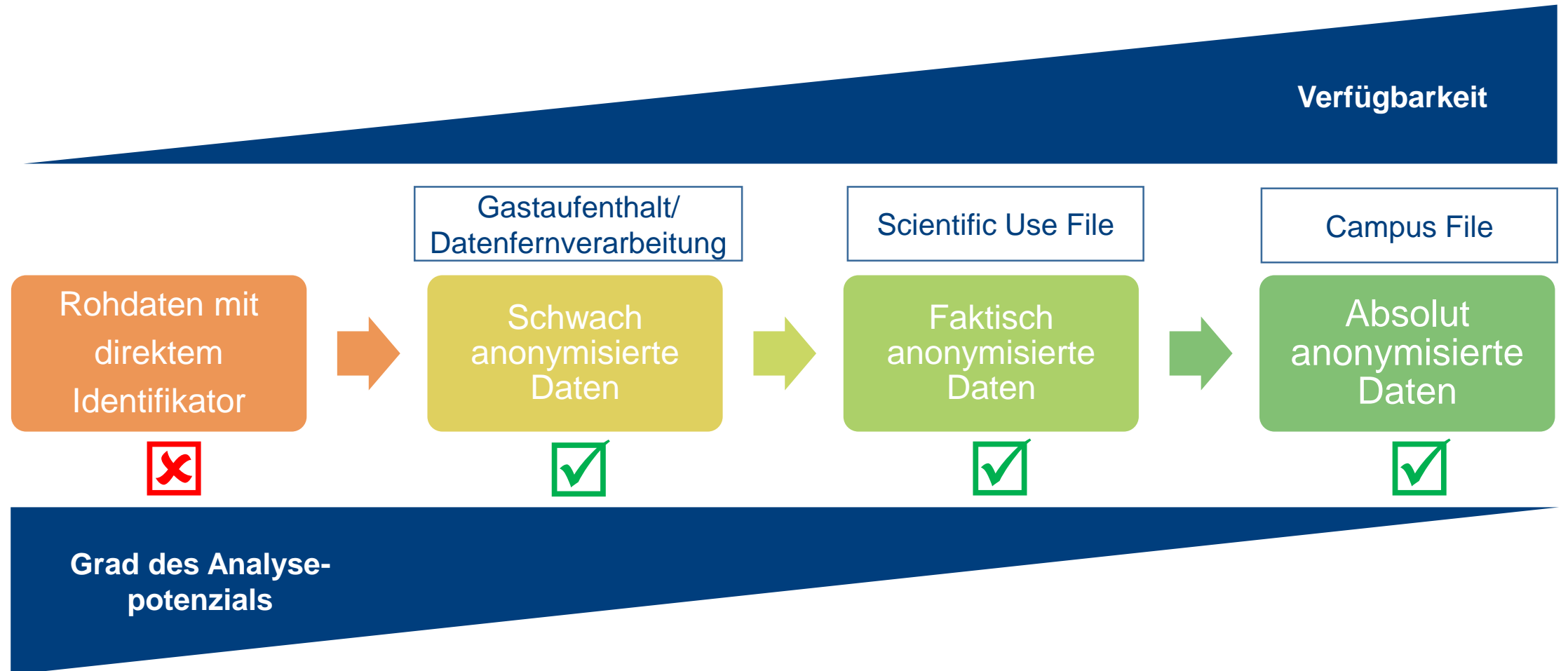
- Keine direkten Identifikatoren (Namen, Adressen, genaue Geburtsdaten,...)
- Stichprobenziehung
- Vergrößerung von Angaben
- Kleinste regionale Ebene: Kreisebene bzw. Bundeslandebene bei Befragungsdaten
- SGB-II Anonymisierung
- Sensible Merkmale (z.B. 4-Steller Berufe, Kreisangaben, 5-Steller Wirtschaftszweig) nur auf gesonderten Antrag verfügbar
- Unterschiedliche Datenzugangswege je nach Anonymisierungsgrad



# DATENZUGANG

---

# DATENZUGANGSWEGE



# GASTAUFENTHALT - WELTWEITE STANDORTE



# ZUSÄTZLICHE ANONYMISIERUNG BEI GASTAUFENTHALTEN AUßERHALB DER EU

---



- Zusätzliche Anonymisierung der Daten bei Gastaufenthalten außerhalb der EU notwendig (weitere Löschung/Vergrößerung von Merkmalen)



- Absprache der Bedingungen der Anonymisierung während des Beantragungsprozesses



- Möglichkeit von individuellen Absprachen unter Betrachtung des Forschungsvorhabens

# DATENFERNVERARBEITUNG

---

- Forschende erstellen Skripte (im Gastaufenthalt oder mit Testdatensätzen) nach transparenten Vorgaben
- Skripte werden über ein Web-Interface an das FDZ übermittelt
- Forschende erhalten nach Datenschutzprüfung die Ergebnisse
  
- **JoSuA**: Job Submission Application (entwickelt durch das IZA)



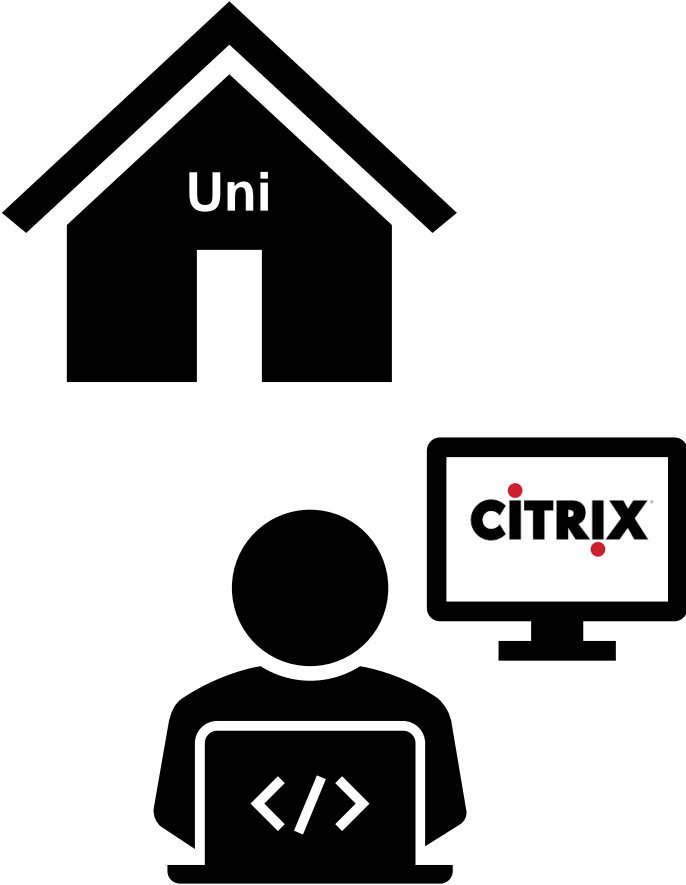
## ZWEI UNTERSCHIEDLICHE MODI

---

	<b>Internal Use</b>	<b>Presentation / Publication</b>
Geheimhaltung	Forschende verpflichtet durch Datennutzungsvertrag	Publizierbar
Herunterladbare Ergebnisse	Nein	Ja
Transfer von Output	Vorschau im JoSuA Web-Interface	Text Files
Datenschutzprüfung	Automatisch (Skript-basiert)	Manuell
Anteil an Jobs	~ 85 %	~ 15 %

# IN PLANUNG: REMOTE DESKTOP ZUGANG

---



# IN PLANUNG: REMOTE DESKTOP ZUGANG

---

- Testballon mit ausgewählten Nutzenden
  - Nutzung eines zusätzlich anonymisierten Datensatzes im Remote Desktop Verfahren
  - Weiterhin Zugang zu schwach anonymisierter Version der Daten in der Datenfernverarbeitung
  - Weiterhin Prüfung des herunterladbaren Outputs durch FDZ



# IMPORT UND EXPORT VON AGGREGIERTEN DATEN

---

- **Import von aggregierten Daten**

- Information über die Anzahl an Beobachtungen auf denen die Variable basiert ( $\geq 3$ )
- Detaillierte Beschreibung der Variablen im Datensatz
- Datenquelle
- Datenschutzkontrolle durch Mitarbeitende des FDZ

- **Export von aggregierten Daten**

- Einmal pro Projekt
- Upload des datengenerierenden Programms in JoSuA
- Datenschutzkontrolle durch Mitarbeitende des FDZ (Anzahl an Beobachtungen pro Variable:  $\geq 20$ )

# REPRODUZIERBARKEIT UND ARCHIVIERUNG

---

- FAIR-Prinzipien (**F**indable, **A**ccessible, **I**nteroperable, **R**eusable)
- Standardisierte und keine individuell aufbereiteten Datensätze
- Transparente Vorgaben zum Gastaufenthalt und der Datenfernverarbeitung
- Vorgeschriebene Struktur der Skripte (Templates)
- Speicherung der hochgeladenen Skripte und Ergebnisdateien in JoSuA
- Archivierung der Skripte bis 10 Jahre nach Ablauf des Nutzungsvertrages nach den Grundsätzen der guten wissenschaftlichen Praxis
- Archivierung der Daten nach dem Archivierungskonzept des IAB

# VIELEN DANK FÜR IHRE AUFMERKSAMKEIT!

---

Sandra Dummert

[Sandra.dummert@iab.de](mailto:Sandra.dummert@iab.de)

<https://fdz.iab.de/>