Remote Access zu Daten der amtlichen Statistik und der Sozialversicherungsträger

Veröffentlichungsversion / Published Version
Abschlussbericht / final report

Empfohlene Zitierung / Suggested Citation:
Rat für Sozial- und Wirtschaftsdaten (RatSWD). (2019). *Remote Access zu Daten der amtlichen Statistik und der Sozialversicherungsträger.* (Output Series, 5 (6)). Berlin. https://doi.org/10.17620/02671.42

5 | Output
6. Berufungsperiode

RatSWD.
Rat für Sozial- und WirtschaftsDaten

Remote Access

zu Daten der amtlichen Statistik und der Sozialversicherungsträger

GEFÖRDERT VOM

Bundesministerium
für Bildung
und Forschung

Rat für Sozial- und Wirtschaftsdaten (RatSWD)

Remote Access

zu Daten der amtlichen Statistik und der Sozialversicherungsträger

Inhaltsverzeichnis

Abstract

Daten der amtlichen Statistik des Bundes und der Länder und der Sozialversicherungsträger stellen eine wichtige Quelle für die empirische Sozial- und Wirtschaftsforschung dar. Ein großer Teil des Mikrodatenbestandes kann bislang jedoch nur ortsgebunden an den Gastwissenschaftsarbeitsplätzen (GWAP) der Forschungsdatenzentren (FDZ) der Datenproduzenten oder über die kontrollierte Datenfernverarbeitung (KDFV) genutzt werden.

Wenngleich die Tatsache, dass die Daten in einen transparenten und strukturierten Verfahren für die Forschung zugänglich sind, im Vergleich zu der Situation vor 16 Jahren einen erheblichen Fortschritt darstellt, sind diese Arten des Datenzugangs für die Forschung mit hohem Aufwand verbunden und schränken den Forschungsprozesses zeitlich und organisatorisch deutlich ein. Vor diesem Hintergrund ist es wichtig, über eine Modernisierung des Datenzugangs in Deutschland nachzudenken, mit dem Ziel, den Datenzugang sowohl aus der Perspektive der Forschung als auch der Datenproduzenten effizienter zu gestalten. Vorbilder können die statistischen Ämter der nordischen Länder oder auch der Niederlande sein, die schon seit geraumer Zeit die rechtlichen und technischen Voraussetzungen geschaffen haben, um ihre Daten für die Forschenden im eigenen Land an deren Arbeitsplatz via Remote Access in Form eines Remote Desktop Verfahren zugänglich zu machen.

Mit dem vorliegenden Papier werden verschiedene Vorgehensweisen dargestellt und diesbezügliche Empfehlungen formuliert: In Deutschland sollte ein Remote Desktop Verfahren zu Daten der amtlichen Statistik und der Sozialversicherungsträger über Pilotprojekte etabliert werden. Die statistischen Ämter des Bundes und der Länder, die Sozialversicherungsträger, Wissenschaft und Datenschutz sollten in diesen Pilotprojekten eng kooperieren. In Hinblick auf den Remote Access zu Daten der amtlichen Statistik empfiehlt der Rat eine Änderung des §16 Abs. 6 im Bundesstatistikgesetz dahingehend, dass ein Remote Desktop Zugang zu formal anonymisierten Daten der amtlichen Statistik möglich wird. Für die rechtssichere und hochverfügbare Implementierung der Pilotprojekte, die damit wachsende Attraktivität des Forschungsstandortes Deutschland wünscht sich der RatSWD Unterstützung seitens der forschungsfördernden Institutionen.

1 Einleitung

Amtliche Mikrodaten, sei es nun von den statistischen Ämtern oder den Sozialversicherungsträgern[1], sind eine wichtige Datenquelle für die empirische Sozial- und Wirtschaftsforschung. Seit der Etablierung des RatSWD im Jahr 2004 wurde die Erschließung dieser Datenquellen für die wissenschaftliche Forschung sukzessive über die sogenannten Forschungsdatenzentren (FDZ) vorangetrieben.

Wenngleich die Tatsache, dass die Daten in einen transparenten und strukturierten Verfahren für die Forschung zugänglich sind, im Vergleich zu der Situation vor 16 Jahren einen erheblichen Fortschritt darstellt, sind diese Arten des Datenzugangs für die Forschung mit hohem Aufwand verbunden und schränken den Forschungsprozesses zeitlich und organisatorisch deutlich ein. Vice versa verursachen die derzeitigen Zugangswege auch für die Forschungsdatenzentren einen hohen Aufwand, während zugleich die Nachfrage der Wissenschaft steigt. Vor diesem Hintergrund sowie aufgrund zwischenzeitlich eingetretener Gesetzesänderungen und technischer Entwicklung sollten die Zugangsoptionen geprüft werden.

Das Datenangebot der FDZ der Statistischen Ämter des Bundes und der Länder deckt inzwischen nahezu vollständig die Erhebungsbereiche der amtlichen Statistik zu den Themen der Sozial-, Wirtschafts-, Finanz-, Steuer-, Rechtspflege- sowie der Agrar-, Energie- und Umweltstatistiken ab. Das Mikrodatenangebot umfasst mittlerweile über 150 Statistiken, die in circa 1.650 Datenprodukten verfügbar sind. Damit ist eine Dateninfrastruktur im nationalen Umfeld gegeben, die gut von der empirisch arbeitenden Wissenschaft angenommen wird.[2]

Das Angebot des FDZ der Rentenversicherung (FDZ-RV) und des FDZ der Bundesagentur für Arbeit (BA) im Institut für Arbeitsmarkt und Berufsforschung (IAB) umfasst Datenprodukte, die aus den Meldungen zur Sozialversicherung von Seiten der Arbeitgeber sowie aus den prozessproduzierten Daten der gesetzlichen Rentenversicherung bzw. der Bundesagentur für Arbeit[3] stammen. Beide Einrichtungen bieten zudem weitere Datenprodukte für die Forschungsgemeinschaft an: Zum einen stellt das FDZ der BA im IAB (FDZ BA im IAB) die eigenen Befragungsdaten des IAB zur Verfügung. Zum anderen bieten beide FDZ mit administrativen Daten verknüpfte Befragungsdaten an.

Auf Basis all dieser Daten sind inzwischen eine Vielzahl von Forschungsarbeiten durchgeführt worden und nationale wie auch internationale Publikationen entstanden, die das Potenzial der Daten verdeutlichen.[4] Allerdings wäre es kurzsichtig, sich mit dem Erreichten zufrieden zu geben, denn

1 Noch stellen nicht alle Sozialversicherungsträger ihre Daten für Forschungsvorhaben zur Verfügung. Wenn wir hier und im Folgenden von Sozialversicherungsträgern sprechen, beziehen wir uns daher ausschließlich auf die FDZ der Bundesagentur für Arbeit (BA) und der Deutschen Rentenversicherung Bund (RV).

2 Die FDZ der Statistischen Ämter des Bundes und der Länder haben seit ihrer Einrichtung 2001/2002 insgesamt rund 3.000 Forschungsprojekte betreut. Im Jahr 2004 lag die Zahl der betreuten Projekte noch bei 66, im Jahr 2017 wurden rund 250 Projekte betreut.

3 Seit der Einrichtung des FDZ BA im IAB sind die Nutzendenzahlen jährlich gestiegen. Während es 2005 noch knapp über 200 Nutzende waren, stieg die Zahl innerhalb von 11 Jahren auf über 1000 Nutzende an (Müller und Möller 2019). Im Dezember 2018 verzeichnete das FDZ BA im IAB über 1.200 aktive Nutzende in über 600 Projekten vgl. https://fdz.iab.de/de/figures.aspx (Zugriff am 04.02.2019).

4 Seit Implementierung der Literaturdatenbank des FDZ der Statistischen Ämter des Bundes- und der Länder http://www.forschungsdatenzentrum.de/literaturdatenbank.asp (Zugriff am 26.10.2018) im Jahre 2013 werden dort durchschnittlich 90 Veröffentlichungen pro Jahr verzeichnet. Für das Erscheinungsjahr 2017 wurden insgesamt 151 Publikationen gemeldet, die zum Teil in einschlägigen nationalen und internationalen Zeitschriften im Peer-Review Verfahren veröffentlicht wurden.
Literaturverzeichnis der Rentenversicherung http://forschung.deutsche-rentenversicherung.de/FdzPortalWeb/FdzLiteraturSearchStart.do?chmenu=ispvwNavEntriesByHierarchy181 (Zugriff am 27.10.2018).
Publikationen auf Grundlage der Daten des FDZ BA im IAB finden sich hier: https://fdz.iab.de/de/FDZ_Publications.aspx (Zugriff am 18.11.2018)

ein großer Teil dieses Mikrodatenbestands kann bislang nicht vollumfänglich am Arbeitsplatz der Forschenden, sondern nur in den Einrichtungen der FDZ genutzt werden. Diese Zugangswege binden jedoch bei der Forschung wie auch in den FDZ beträchtliche Ressourcen, die an anderer Stelle besser genutzt werden könnten. Deshalb ist es wichtig, über eine Modernisierung des Datenzugangs in Deutschland nachzudenken, mit dem Ziel, den Datenzugang sowohl aus der Perspektive der Forschung als auch der Datenproduzenten effizienter zu gestalten. Vorbilder können die statistischen Ämter der nordischen Länder (hier repräsentiert durch Schweden, Dänemark, Finnland und Norwegen) oder auch der Niederlande sein, die schon seit geraumer Zeit die rechtlichen und technischen Voraussetzungen geschaffen haben, um ihre Daten für die Forschenden im eigenen Land an deren Arbeitsplatz via ***Remote Access***[5] in Form eines ***Remote Desktop Verfahrens*** zugänglich zu machen.

Ähnliche Modelle des Datenzugangs am eigenen Arbeitsplatz für Forschende im Sinne von Remote Desktop bieten bspw. auch die statistischen Ämter in Estland, Slowenien, Irland, Frankreich, Kroatien, Lettland und Litauen. In Deutschland gibt es derartige Remote Desktop Lösungen bislang nur für im Wissenschaftskontext generierte Daten. So bieten z. B. das Leibniz-Institut für Bildungsverläufe (LIfBi) und das Deutsche Zentrum für Hochschul- und Wissenschaftsforschung (DZHW) diese Option für ihre Daten an (vgl. ausführlich dazu Kapitel 3).

Zwischenzeitlich arbeiten die nordischen Länder an einer weiteren Modernisierung ihres Datenzugangs (Statistics Finland et al. 2014). Auf der Basis der Europäischen Datenschutz-Grundverordnung (DS-GVO) wird angestrebt, den Datenzugang der Forschung zukünftig auch länderübergreifend zu harmonisieren, mit dem Ziel einen transnationalen Zugang innerhalb dieses Länderverbundes zu ermöglichen. Beispielsweise können dann Forschende an einer dänischen Universität an ihrem Arbeitsplatz nicht nur mit dänischen, sondern auch mit norwegischen, schwedischen und finnischen amtlichen Daten arbeiten. Bemerkenswert ist die Begründung für dieses Vorhaben, die nationalen, im Wesentlichen aus Verwaltungsregistern entstammenden, amtlichen Daten über Remote Desktop für länderübergreifende Forschungsvorhaben zugänglich zu machen: **„The registers in the Nordic countries are a goldmine for research, and give the nordic countries an opportunity to play a pivotal role in international scientific research"** (Kvalheim und Myhren 2017, S. 2). De facto gelten die nordischen Länder aus der Perspektive der Forschung auch schon jetzt als Goldstandard, nicht nur weil viele (administrative) Mikrodatenquellen die jeweilige Grundgesamtheit meist vollständig abdecken, sondern weil die unterschiedlichen Datenquellen für Forschungsvorhaben auch miteinander verknüpft werden können (Drechsel-Grau et al. 2015; Loeffler et al. 2015; Isungset et al. 2019). Ein Beispiel jüngeren Datums ist eine Studie, die sich mit dem Ausmaß von Steuerhinterziehung beschäftigt (Alstadsaeter et al., 2018)[6], und zu diesem Zweck administrative Einkommensdaten aus Schweden, Dänemark und Norwegen mit anderen Datenquellen zusammenführt.

Ähnlich lässt sich – wie eingangs dargestellt – auch für Deutschland die Relevanz von amtlichen Daten als eine wichtige Datenquelle für die Forschung hervorheben, allerdings mit der Einschränkung, dass die Wege des Datenzugangs deutlich hinter den – jetzt schon existierenden – Möglichkeiten zurückbleiben.

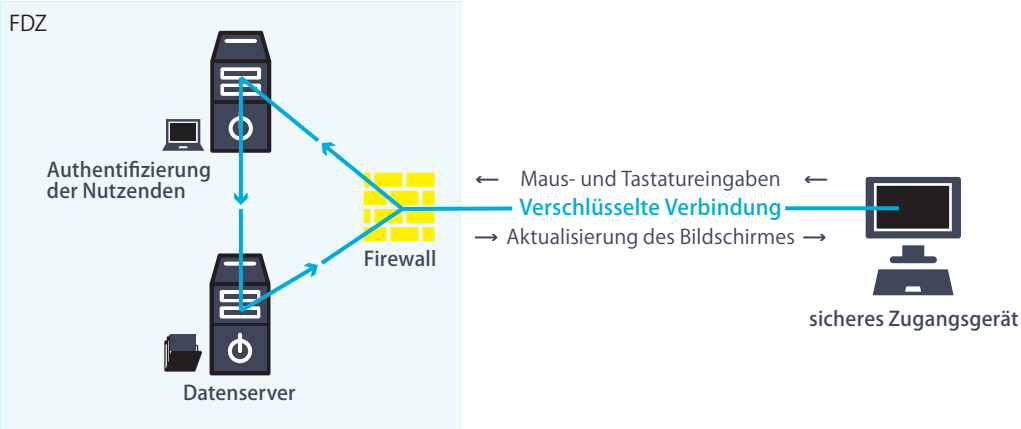
5 In der Literatur wird der Begriff 'Remote Access' häufig synonym für das Remote Desktop Verfahren verwendet. D.h. es wird nur zwischen Remote Execution und Remote Access unterschieden (vgl. z. B. Schiller/Welpton 2013; Bujnowska 2017)

6 Für eine differenzierte Beschreibung der Daten und der Vorgehensweise siehe: Alstadsaeter et al., 2018, http://gabriel-zucman.eu/files/AJZ2017.pdf (Zugriff am 04.04.2019).

Infokasten 1: Remote Access Verfahren

Remote Access: Mit dem Begriff Remote Access wird allgemein der Zugriff aus der Ferne auf eine IT-Infrastruktur beschrieben, wobei die Ausgestaltung bzw. Umsetzung dieses Zugriffs variieren kann.

Abb. 1: Remote Access als Remote Desktop Verfahren



Quelle: In Anlehnung an Schiller und Welpton, 2014, S. 7

Remote Desktop und **Remote Execution** stellen unterschiedliche Formen des Remote Access dar (vgl. hierzu Kapitel 2 und 3, ebenso Schiller et al. 2017).

Beim **Remote Desktop Verfahren** erfolgt die Datenspeicherung und -verarbeitung ausschließlich auf den Servern der datenhaltenden Organisation (z. B. einem FDZ). Die Benutzeroberfläche wird über eine gesicherte Verbindung auf den Bildschirm der Forschenden übertragen (virtueller Desktop). Das Zugangsgerät der Forschenden dient lediglich dazu, mit dem Server zu kommunizieren. Die Applikationen und Daten sind physikalisch ausschließlich auf dem FDZ-Server, wobei das Sichten und Browsen der Ergebnisse und der Daten innerhalb einer gewohnten Desktopumgebung möglich ist. D.h. es kann iterativ gearbeitet werden. Wünschen die Forschenden die Mitnahme der Ergebnisse so werden diese nach einer Datenschutzprüfung durch das FDZ-Personal an die Forschenden übermittelt.

Im Unterschied zu Remote Desktop wird bei **Remote Execution** gewissermaßen blind gearbeitet. Das Sichten und Browsen von Daten oder Ergebnissen auf dem Bildschirm ist nicht möglich. Die Forschenden schreiben ein Analyseskript, wobei für die Überprüfung der Syntax zumeist ein Strukturdatensatz zur Verfügung steht. Das Skript wird an den FDZ-Server übermittelt (oder per E-Mail an das FDZ-Personal geschickt) und dort auf die Daten angewandt. Die (Zwischen-)Ergebnisse werden nach einer Datenschutzüberprüfung durch das FDZ an die Forschenden übermittelt. Erst dann lässt sich feststellen, ob die Analysen inhaltlich sinnvoll sind, bzw. wo diese verändert werden müssen, d.h. der Kreislauf beginnt erneut. Eine effiziente Nutzung von Remote Execution setzt daher hervorragende Datenkenntnisse voraus. Explorationen des Datensatzes sind hingegen auf diesem Wege im Allgemeinen sehr zeitaufwendig (vgl. z. B. Schiller und Welpton 2013).

2 Zugangswege

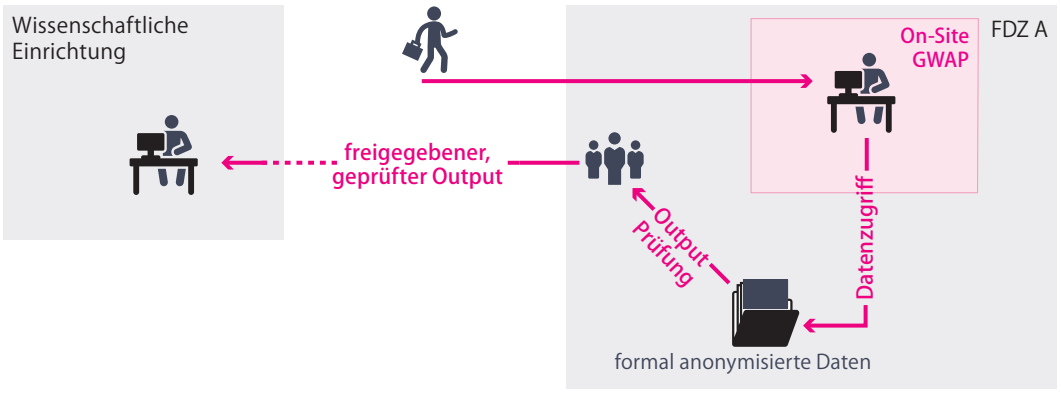
Die Forschung hat in Deutschland gegenwärtig vier Möglichkeiten des Zugangs zu Mikrodaten der amtlichen Statistik und der Sozialversicherungsträger (vgl. Abbildungen 2a - d).

Zugangsweg 1: Scientific Use Files (SUF): Bei SUF handelt es sich um faktisch anonymisierte Daten[7], die ausschließlich für wissenschaftliche Vorhaben auf der Basis des fachgesetzlich verankerten Wissenschaftsprivilegs und mit vertraglicher Bindung zur Verfügung gestellt werden. SUF können von der Forschung direkt am eigenen Arbeitsplatz in der beantragenden wissenschaftlichen Einrichtung gespeichert und genutzt werden, haben jedoch aufgrund der Anonymisierung ein geringeres Analysepotential als formal anonymisierte Daten[8] (siehe Zugangsweg 3).

Abb. 2a: Zugangsweg 1: Scientific Use File (SUF)

Zugangsweg 2: On-site Gastwissenschaftsarbeitsplatz (GWAP)[9]: Dieser bietet den Forschenden die Möglichkeit, vor Ort in den Räumen der Datenproduzenten mit formal anonymisierten Mikrodaten zu arbeiten. Vor einer physikalischen Freigabe der Ergebnisse an die Forschenden zur weiteren Verwendung außerhalb des GWAP werden diese durch die Datenproduzenten auf Einhaltung der statistischen Geheimhaltung geprüft (Outputkontrolle).

Abb. 2b: Zugangsweg 2: On-site Gastwissenschaftsarbeitsplatz (GWAP)



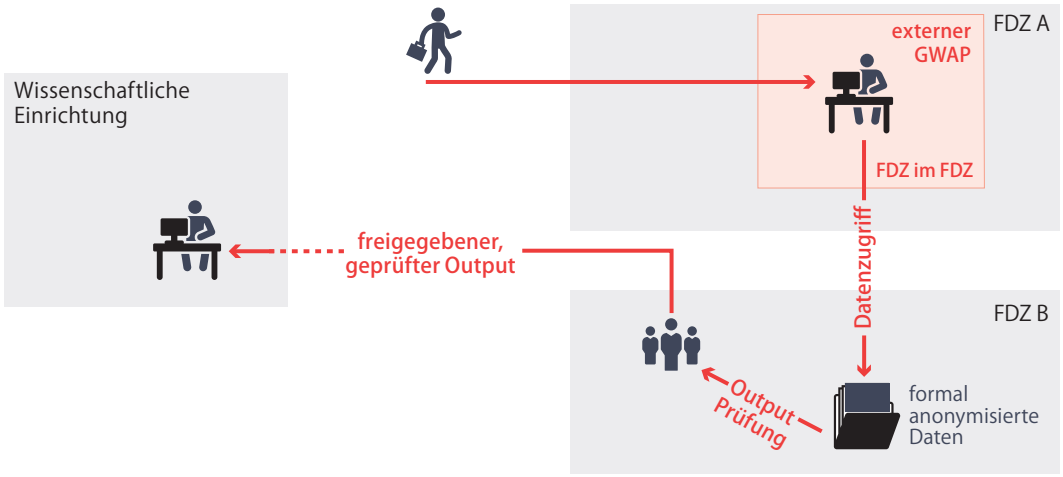
7 Als faktisch anonym werden Daten bezeichnet, bei welchen eine Reidentifikation von Einzeldatensätzen – sofern überhaupt – nur mit einem unverhältnismäßig hohen Aufwand an Zeit, Kosten und Arbeitskraft möglich ist (Müller et al., 1991; Wirth 2016).

8 In der Begründung zum BStatG 20.10.2016 Drucksache 557/12 wird von formal anonymisierten Einzeldaten gesprochen als Einzelangaben ohne Hilfsmerkmale (S. 84), ohne dass innerhalb des Gesetzestextes eine Definition erfolgt. Nach §12 BStatG sind die Hilfsmerkmale von den Erhebungsmerkmalen frühestmöglich zu trennen und ggf. zu löschen. Ausnahmen sind §12 Absatz 21 BStatG, §10 Absatz 22 und §133 BStatG sowie besondere Regelungen in einzelstatistischen Gesetzen. Darüber hinaus definieren sich die formal anonymisierten Einzelangaben nach der Legaldefinition nach §5a Absatz 3 Satz 1 BStatG ohne Name und Anschrift. D.h. die zur Verfügung stehenden formal anonymisierten Einzelangaben für wissenschaftliche Zwecke sind die Einzelangaben ohne Namen und Anschrift sowie ohne weitere Hilfsmerkmale.

9 Eine Übersicht zu den GWAP Standorten der amtlichen Statistik und der Sozialversicherungsträger findet sich im Anhang 2.

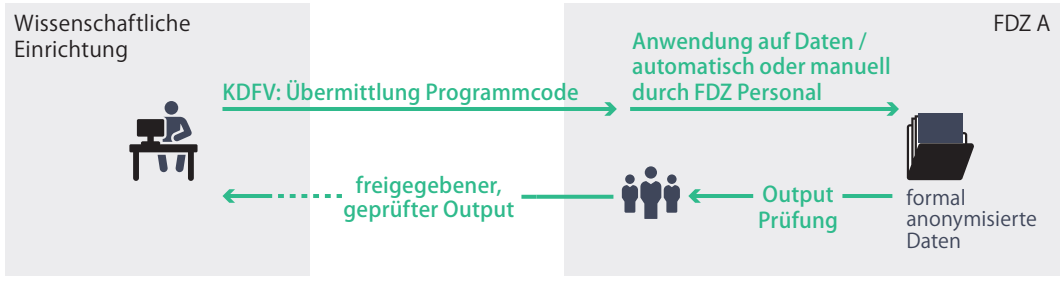
Zugangsweg 3: Externer GWAP (FDZ im FDZ): Die Forschenden haben die Möglichkeit in einem Safe-Room eines Datenproduzenten (z. B. FDZ A) über ein Remote Desktop Verfahren mit formal anonymisierten Mikrodaten eines anderen Datenproduzenten (z. B. FDZ B)[10], zu arbeiten. Das ist z. B. dann sinnvoll, wenn sich die/der Forschende an einem Ort befindet, wo kein GWAP des Datenproduzenten B lokal verfügbar ist. Auch hier erfolgt vor Freigabe der Ergebnisse eine Prüfung auf statistische Geheimhaltung (Outputkontrolle) – in diesem Fall von Seiten des Datenproduzenten B.

Abb. 2c: Zugangsweg 3: Externer GWAP (FDZ im FDZ)



Zugangsweg 4: Kontrollierte Datenfernverarbeitung (KDFV) bzw. Remote Execution: Bei der kontrollierten Datenfernverarbeitung wird gewissermaßen blind gearbeitet. Die Forschenden erstellen ein Analyseskript, wobei für die Überprüfung zumeist ein Strukturdatensatz zur Verfügung steht. Dieses Analyseskript wird per Email oder über eine andere Schnittstelle an das FDZ übermittelt, dass das Skript dann entweder manuell oder automatisiert auf die Daten anwendet. Vor einer Übermittlung der (Zwischen-)Ergebnisse an die Forschenden werden diese von Seiten des FDZ-Personals auf statistische Geheimhaltung überprüft.[11]

Abb. 2d: Zugangsweg 4: Kontrollierte Datenfernverarbeitung (KDFV) bzw. Remote Execution



10 So bietet das FDZ BA im IAB die Möglichkeit eines Remote Access (FDZ im FDZ) in den sicheren Räumen von fünf Einrichtungen in Deutschland. Dazu gehören die regionalen Standorte Berlin, Hannover und Dresden des FDZ der Statistischen Ämter der Länder, die GESIS in Köln und Mannheim und die Bibliothek der Hochschule der BA. Darüber hinaus ist der Datenzugang in den sicheren Räumen ausgewählter FDZ, Archiven und Universitäten in Frankreich, UK, den USA und Kanada möglich. Im Unterschied hierzu kann die amtliche Statistik einen Remote Access nur in klar definierten Räumen der amtlichen Statistik zulassen (Standorte siehe Anhang 2). Auf die Unterschiede hierbei, und inwieweit dies für die Option des Remote Access relevant ist, gehen wir in Kapitel 4 ein.

11 Kontrollierte Datenfernverarbeitung bzw. Remote Execution findet z. B. beim FDZ-IAB Anwendung (https://fdz.iab.de/de/FDZ_Data_Access/FDZ_Remote_Data_Access.aspx) aber auch im FDZ der Rentenversicherung http://forschung.deutsche-rentenversicherung.de/FdzPortalWeb/dispcontent.do?id=main_fdz_fernrechnen&chmenu=ispvwNavEntriesByHierarchy118 (Zugriff am 07.10.2019)

Wenngleich Scientific Use Files (SUF) (Zugangsweg 1) im Vergleich zu formal anonymisierten Daten zumeist ein geringeres Analysepotenzial aufweisen, wird diese Form des Datenzugangs von der Forschung in Deutschland sehr geschätzt. Die unmittelbare Verfügbarkeit der Daten und Ergebnisse am eigenen Arbeitsplatz und die damit verbundene zeitliche Flexibilität bei der Analyse kommt dem in der Regel nicht geradlinig verlaufenden und immer wieder auch stark explorativen Forschungsprozess entgegen (Smith 1991; Wirth 2016). Oder wie Zühlke und Christians (2005:9) formulieren: „Für Wissenschaftler/-innen ist die Auswertung von Einzelangaben am eigenen Arbeitsplatz die bequemste Lösung". Darüber hinaus ist die Nutzung der bereits faktisch anonymisierten SUF für Forschende und Datenproduzenten natürlich auch zeiteffizient, da z. B. die Übermittlung der Analyseskripte, Terminvereinbarungen, Reise- und Bearbeitungszeiten, Betreuung der Forschenden vor Ort wie auch die Überprüfung der Ergebnisse auf statistische Geheimhaltung durch die FDZ entfallen.

Abhängig vom Datenproduzenten wird jedoch gegenwärtig nur ein Bruchteil der verfügbaren amtlichen Daten als SUF (Zugangsweg 1) angeboten. Dies hat im Wesentlichen zwei Ursachen:

- Erstens ist der zeitliche und personelle Aufwand für die erstmalige Generierung und die laufende Aktualisierung eines SUF hoch, da für unterschiedliche Berichtszeiträume zumeist Anpassungen erforderlich sind. Der Aufwand für die Generierung eines SUF lohnt sich nur dann, wenn es sich um Daten handelt, die von der Forschung sehr stark nachgefragt werden (wie z. B. der Mikrozensus, die Einkommens- und Verbrauchsstichprobe, die Lohn- und Einkommensstatistik, die Stichprobe der Integrierten Arbeitsmarktbiografien oder die Versicherungskontenstichprobe).
- Zweitens gibt es spezifische Daten, wie z. B. Wirtschafts- und Betriebsdaten, die von Seiten der Forschung zwar stark nachgefragt werden, aber vor einer Weitergabe als SUF so stark mit datenverändernden Verfahren anonymisiert werden müssten (z. B. das IAB-Betriebspanel), dass sie ihr Analysepotential verlieren und somit für Forschungsvorhaben nicht mehr nutzbar sein würden. Gleiches gilt ebenfalls für verstärkt nachgefragte kleinräumige Informationen oder verknüpfte Daten, die aufgrund bestimmter Merkmalskombinationen ein hohes Reidentifikationsrisiko beinhalten und deshalb nicht als SUF angeboten werden.

Ist kein SUF verfügbar oder mit hinreichendem Analysepotential generierbar, wie häufig bei Betriebs- und Wirtschaftsdaten, besteht die Möglichkeit, diese Daten am GWAP (Zugangsweg 2 und Zugangsweg 3) und/oder mittels der KDFV/Remote Execution (Zugangsweg 4) zu nutzen. Diese Zugangswege haben aus Perspektive der Forschung jedoch den Nachteil, den Forschungsprozess gewissermaßen zu ‚bürokratisieren' und zu verlangsamen. Sie sind mit einem zusätzlichen Zeit- und Kostenaufwand verbunden, wie sie auch bei den FDZ selbst einen hohen Aufwand verursachen. So ist die Nutzung des GWAP Zugangs in den Räumen der Datenproduzenten (Zugangsweg 2) bzw. sonstigen Standorten (FDZ in FDZ: Zugangsweg 3) an die dortigen Öffnungszeiten gebunden, die Zahl der Arbeitsplätze begrenzt und geht in der Regel mit Reisekosten sowie einem höheren Zeit- und Bearbeitungsaufwand einher. Entsprechend muss die Analysestrategie genau geplant und außerordentlich gut vorbereitet sein. Im Arbeitsalltag des Forschungsprozesses auftretende spontane Ideen oder Fragen können, anders als mit dem Datenzugang am eigenen Arbeitsplatz, nicht unmittelbar bearbeitet werden.

Hinzu kommt, dass die Forschungsergebnisse bzw. der Output zunächst nur am GWAP einsehbar sind und den Forschenden erst nach einer Überprüfung auf statistische Geheimhaltung Tage oder in Einzelfällen Wochen später zur weiteren Verwendung übermittelt werden. Wie oben angeführt, ist der Forschungsprozess aber selten geradlinig, sondern eher ein kontinuierliches Reflektieren von unterschiedlichen Ideen, Fragen und Methoden. Die ersten Ergebnisse werden häufig verworfen und andere Ansätze überprüft, bevor man sich für eine Analyse entscheidet, deren Ergebnisse dann auch veröffentlicht werden können. Da die in diesem Findungsprozess am GWAP erzeugten Zwischenergebnisse vor einer Übermittlung jeweils auf statistische Geheimhaltung überprüft werden müssen, ist damit ein hoher Zeitaufwand für die FDZ verbunden. Für die Forschenden wiederum ist dieser Prozess zeitineffizient, da diese Zwischenergebnisse oftmals nur für die weiteren Überlegungen benötigt werden, nicht aber für die letztendliche Publikation gedacht sind. Zudem können gegenwärtig geprüfte und herausgegebene Zwischenergebnisse, obgleich sie nicht weiterverwendet wurden, verhindern, dass später erstellte Ergebnisse freigegeben werden, da sie zusammengenommen Geheimhaltungsfälle ergeben.

Bei KDFV/Remote Execution **(Zugangsweg 4)** kann zwar vom eigenen Arbeitsplatz gearbeitet werden. Wie eingangs ausgeführt, wird hier aber gewissermaßen ‚blind' gearbeitet. Eine effiziente Nutzung von Remote Execution setzt hervorragende Datenkenntnisse voraus; explorative Analysen sind hingegen zeitaufwändig. Folgt man Schiller und Welpton (2013:k.S.) ist Remote Execution auch für die Datenanbieter ressourcenintensiv:„(...) disadvantages of Remote Execution: a large staff is required to execute the statistical programming syntax submitted by the researcher; provision of assistance to researchers in the event that statistical programming syntax does not execute; and the costly preparation of ‚synthetic or ‚fake' data to enable the researchers to accurately programme their syntax."

Die nordischen Länder aber auch die Niederlande und Frankreich lösen diese Probleme im Spannungsfeld zwischen Geheimhaltung und flexiblen Forschungsbedingungen, indem sie der Forschung einen Remote Desktop Zugang zu formal anonymisierten amtlichen Mikrodaten am eigenen Arbeitsplatz an der Universität oder einer externen Forschungseinrichtung[12] ermöglichen (vgl. Tab 1, S. 23; vgl. Statistics Finland et al. 2014; siehe Kap. 3.3). Unter bestimmten Restriktionen können die Forschenden im eigenen Büro am Bildschirm mit diesen Daten arbeiten. Physikalisch verbleiben die Daten auf dem Server des Datenproduzenten, dort findet auch die Datenverarbeitung statt. Daten und Zwischenergebnisse können am Bildschirm eingesehen aber nicht ausgedruckt oder auf den eigenen Computer heruntergeladen werden. Liegen die Endergebnisse vor, erfolgt eine Prüfung auf statistische Geheimhaltung, bevor der Output den Forschenden zur weiteren Verwendung physikalisch übermittelt wird. Da die Prüfung von Endergebnissen in der Summe deutlich weniger umfangreich ist als eine etwaige Prüfung aller Zwischenergebnisse, ergibt sich durch dieses Vorgehen eine erhebliche Zeitersparnis für alle Beteiligten.

Bedingt durch die kontinuierliche Veränderung und die steigende Nachfrage der Wissenschaft, stehen die FDZ in Deutschland vor der Herausforderung, ihre Angebote zu verbessern und die Prozesse für die Forschenden wie auch die FDZ effizienter zu machen. Es stellt sich daher die Frage, ob ein Remote Desktop Zugang am eigenen Arbeitsplatz, wie er von der amtlichen Statistik der nordischen Länder und vom Leibniz-Institut für Bildungsverläufe (LIfBi) sowie dem Deutschen Zentrum für Hochschul- und Wissenschaftsforschung (DZHW) angeboten wird, nicht auch für die amtliche Statistik und die Sozialversicherungsträger in Deutschland ein Weg ist, um den Datenzugang für die Forschung zu modernisieren. Insbesondere vor dem Hintergrund, dass die Daten der amtlichen Statistik und der Sozialversicherungsträger nicht nur immer stärker nachfragt werden, sondern auch die Komplexität der nachgefragten Daten (z. B. Verknüpfung unterschiedlicher Datenquellen) zunimmt, würden nicht nur die Forschenden sondern auch die FDZ stark profitieren.

Im Folgenden wird dieser Gedanke aufgegriffen und versucht, die Bedingungen eines Remote Desktop Zugangs am eigenen Arbeitsplatz herauszuarbeiten. In Kapitel 3 werden Varianten erläutert und ein Überblick über bereits etablierte Lösungen in Deutschland und europäischen Ländern gegeben. Kapitel 4 behandelt die rechtlichen Bedingungen, unter welchen gegenwärtig Daten der Statistischen Ämter der Länder und des Bundes sowie der Sozialversicherungsträger für Forschungsvorhaben bereitgestellt werden. Kapitel 5 stellt die technischen, organisatorischen und rechtlichen Aspekte auf dem Wege zur Umsetzung von Remote Access dar. In Kapitel 6 werden Empfehlungen für die Etablierung von Remote Access durch die amtliche Statistik und die Sozialversicherungsträger ausgesprochen.

12 Eine Ausnahme hiervon ist Norwegen. In Norwegen ist der Remote Access gegenwärtig im Aufbau. Bislang werden die Daten (nach erfolgreichen Antragsverfahren) physikalisch an die Forschenden übermittelt.

3 Remote Access in Deutschland und ausgewählten europäischen Ländern

3.1 Beschreibung des Remote Access-Verfahrens

Remote Access bzw. Remote Desktop Verfahren kommen aus unterschiedlichen Gründen zum Einsatz und haben daher unterschiedliche technische Konfigurationen: Zum Beispiel können die erlaubten Zugangsgeräte hardwareseitig vorgegeben sein. In anderen Fällen ist der Remote Access ausschließlich an festgelegten Orten möglich (z. B. vom Büro der Forschenden mit eindeutiger IP-Adresse, die eine Zuordnung zu einer Institution der unabhängigen Wissenschaft zulässt).[13] So erlaubt das niederländische Modell (CBS Niederlande) zum Beispiel zugelassenen Forschenden den Datenzugang mit Einschränkungen ortsunabhängig im In- und Ausland.

Diese Art des sicheren und geschützten Zugriffs von außen hat inzwischen in vielen Unternehmen und in der Verwaltung Einzug gehalten, um so Arbeitsflexibilität, Home-Office zur Vereinbarkeit von Familie und Beruf etc. IT-technisch und datenschutzkonform umzusetzen. So bieten beispielsweise auch die BA (und das IAB) und die Rentenversicherung ihren Mitarbeitenden die Möglichkeit des mobilen Arbeitens z. B. im Homeoffice. Das Bundesamt für Sicherheit in der Informationstechnik (BSI) veröffentlichte hierzu bereits 2010 eine Studie zur technischen und organisatorischen Umsetzung.[14]

Je nach Ausgestaltung würde die Einführung eines Remote Desktop Zugangs zu den Mikrodaten der amtlichen Statistik und der Sozialversicherungsträger zu einer deutlichen Erhöhung der Nutzungsfreundlichkeit für die Forschenden führen. Im Vergleich zur Arbeit an einem GWAP sind hier insbesondere der zeitunabhängige Datenzugang am eigenen Arbeitsplatz, die Vermeidung von Wartezeiten sowie der Wegfall von Reisezeit und -kosten zu nennen.

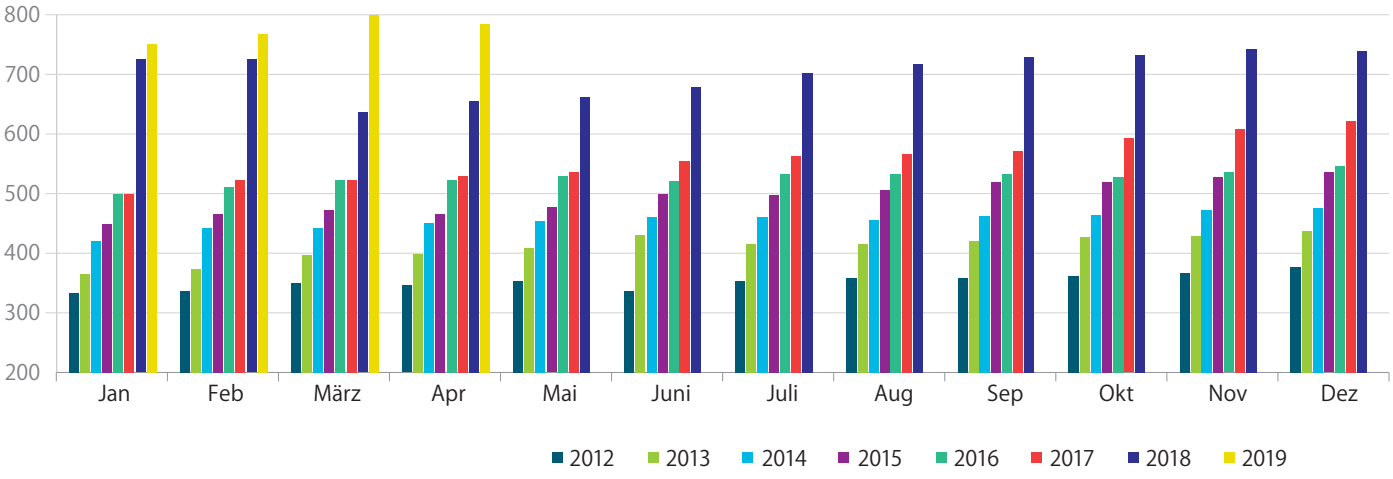
Aber auch für die FDZ ist eine Verringerung der projektspezifischen Arbeitsbelastung bezüglich der Betreuung der Forschenden vor Ort und der Outputkontrolle zu erwarten. Denn das Bedürfnis der Wissenschaftlerinnen und Wissenschaftler sich prophylaktisch alle Zwischenergebnisse vom GWAP übermitteln zu lassen, um sich am eigenen Arbeitsplatz weiter damit zu beschäftigen, wird vermutlich deutlich sinken, wenn die Befunde grundsätzlich via Remote Desktop jederzeit eingesehen und reflektiert werden können. Allerdings ist auch zu erwarten, dass bei einer Einführung die Zahl der Forschenden, die mit amtlichen Daten arbeiten werden, zunehmen wird, während die Nachfrage am GWAP zurückgehen wird, so zumindest die Erfahrung in den Niederlanden (vgl. Abb. 3a und 3b).

Abb. 3a gibt die Zahl der mittels Remote Desktop auf niederländische Daten zugreifenden Forschenden pro Monat (im Jahresvergleich) an. So haben im März 2012 knapp über 300 Forschende mindestens einmal remote gearbeitet. Im März 2019 waren es fast 800. Im Vergleich dazu ist die Anzahl der Reservierungen für GWAP (Abb. 3b) im gleichen Zeitraum von knapp über 40 (März 2012) auf weniger als fünf Nutzende (März 2019) zurückgegangen. Über die Jahre ist ein stetiger Anstieg des Remote Desktop Zugangs zu beobachten (Abb. 3a), während die Zahl der GWAP Nutzung bei starken Schwankungen nur in Ausnahmefällen über 80 pro Monat steigt (Abb. 3b). Das erklärt sich dadurch, dass die Nutzung des GWAP durch die Anzahl der verfügbaren Arbeitsplätze im Vergleich zum Remote Zugang limitiert ist. Nach Auskunft von CBS werden die GWAP inzwischen nur noch selten und hauptsächlich zu Schulungen genutzt.

13 Auch bei der gegenwärtig von den FDZ der Statistischen Ämter des Bundes und der Länder sowie den FDZ der Sozialversicherungsträger praktizierten Safe Room Lösung in speziell gesicherten Räumen an sonstigen Standorten (z. B. FDZ-in-FDZ) handelt es sich (i.d.R.) um ein Remote Desktop Verfahren. Allerdings werden hier nicht nur das Zugangsgerät und der Raum, in welchem das Zugangsgerät steht, festgelegt, sondern der Raum muss darüber hinaus spezifischen Bedingungen genügen. Wesentlich ist, dass es eine Zugangs- und Zutrittskontrolle zum GWAP gibt. Das FDZ Personal kann Einblick in die Arbeit der Nutzenden nehmen. Außerdem ist die Nutzung bzw. schon das mit sich führen von Laptops, Massenspeichergeräten, Mobilfunkgeräten und Geräten zur Bildaufnahme am GWAP im Safe Room nicht erlaubt. Die Geräte erlauben kein Speichern, Drucken etc., USB-Schnittstellen sind deaktiviert.

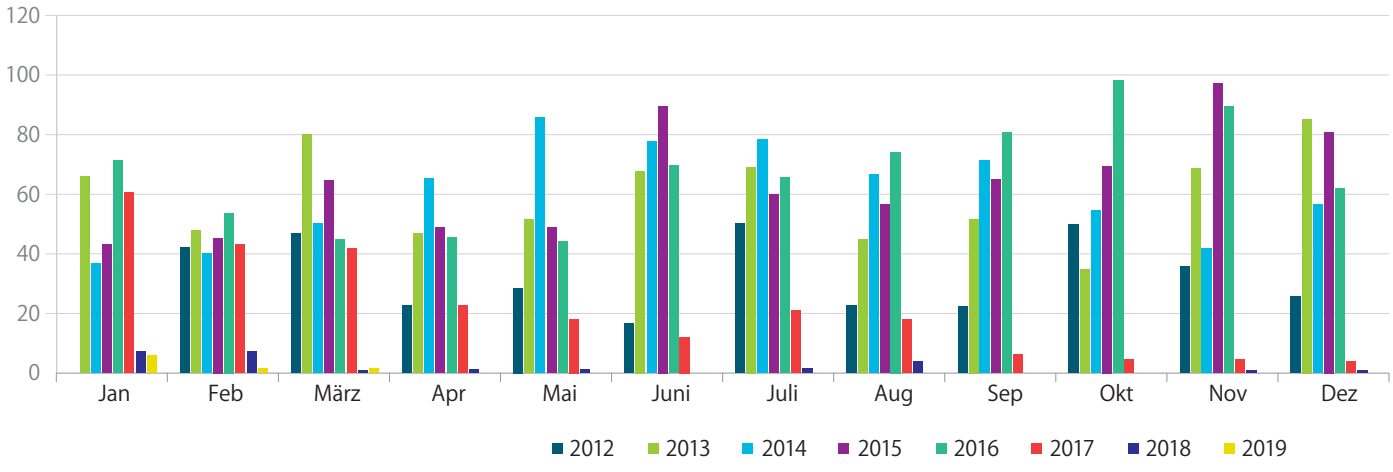
14 Sicherer Fernzugriff auf das interne Netz (ISi-Fern) https://www.bsi.bund.de/SharedDocs/Downloads/DEBSI Internetsicherheit/isi_fern_studie_pdf.pdf (Zugriff am 31.08.2019)

Abb. 3a: Anzahl der aktiven Forschenden pro Monat (2012-2019)



Quelle: E-Mail Auskunft CBS Niederlande

Abb. 3b: Anzahl der GWAP Reservierungen pro Monat (2012-2019)



Quelle: E-Mail Auskunft CBS Niederlande

3.2 Ausgewählte Remote Desktop-Lösungen bei Forschungseinrichtungen in Deutschland

Die folgenden Unterabschnitte skizzieren die Remote-Lösungen des Leibniz-Instituts für Bildungsverläufe (LIfBi) und des Deutschen Zentrums für Hochschul- und Wissenschaftsforschung (DZHW). Der von LIfBi und DZHW verfolgte Ansatz ist an den Bedürfnissen der Forschung und des Forschungsprozesses ausgerichtet, was in der Natur der Sache liegt, da die hier verfügbaren Daten von der Forschung für die Forschung erhoben werden.[15] Die Nutzung der Daten von Seiten der Forschung ist primäres Ziel und nicht – wie bei amtlichen Daten – ein nachgeordneter Aspekt.

3.2.1 Remote Desktop-Lösung des LIfBi

Seit 2009 bietet das LIfBi Forschenden die Möglichkeit, über das eigens implementierte System RemoteNEPS[16] sensible faktisch anonymisierte Daten des Nationalen Bildungspanels (NEPS) via Remote Desktop kostenfrei für Forschungsvorhaben zu nutzen.

Die technischen Voraussetzungen sind einfach gehalten: neben einem Internetzugang ist lediglich ein aktueller Browser notwendig,[17] sodass sowohl Arbeitszeit als auch Arbeitsort von den Forschenden flexibel gehandhabt werden können.

Voraussetzung für die Nutzung von RemoteNEPS ist der Abschluss von zwei Verträgen: Der NEPS-Datennutzungsvertrages[18] spezifiziert die Anbindung an eine Forschungseinrichtung sowie den Zweck und die Dauer der Datennutzung inkl. Konsequenzen bei Vertragsverstößen. Der RemoteNEPS-Ergänzungsvertrag[19] spezifiziert die Zusatzbestimmungen zum Datenschutz, der Speicherung von Analysen und der Output-Kontrolle. Forschende sind darüber hinaus verpflichtet, zuvor an einer NEPS-Schulung teilzunehmen und ihr persönliches Tippverhalten als biometrisches Merkmal zu registrieren. Die biometrische Authentifizierung für den Zugang zum Remote Server erfolgt durch diese Schriftrhythmuserkennung (Tippbiometrie, engl. keystroke biometrics), welche neben der Eingabe eines individuellen Logins mit Passwort zum mehrstufigen Anmeldungsverfahren gehört.

Der Zugang zu RemoteNEPS ist ortsunabhängig, d. h. ein Zugriff ist auch aus dem Ausland unabhängig von der Nationalität der Forschenden kostenlos möglich. Für Forschende aus dem Ausland gelten die gleichen Zugangsvoraussetzungen wie für Forschende in Deutschland und deutsche Forschende.

Ist der Zugang gewährt, können Forschende über eine gesicherte Internetverbindung in einer virtuellen Arbeitsumgebung arbeiten. In dieser Arbeitsumgebung stehen Standard-Analysesoftwarepakete wie R, Stata oder SPSS sowie Anwendungen für Text- und Tabellenbearbeitung zur Verfügung. Die Arbeitsprozesse und (Zwischen-)Ergebnisse können daher jederzeit eingesehen werden.

Über das Forschungsdatenzentrum des LIfBI können auch externe Daten in RemoteNEPS importiert werden. Die Arbeit in Kooperationsprojekten wird unterstützt, d. h. Forschende aus unterschiedlichen Institutionen können jeweils am eigenen Arbeitsplatz in gemeinsamen Verzeichnissen mit denselben Dateien arbeiten, da individuelle Zugriffsrechte in RemoteNEPS definiert und über eine Nutzerdatenbank administriert werden können.

Bevor die Endergebnisse auf Anfrage aus der virtuellen Arbeitsumgebung an die Forschenden physikalisch zur Übermittelung bereitgestellt werden, erfolgt eine datenschutzrechtliche Prüfung durch das Forschungsdatenzentrum des LIfBi.

15 Auch das im Aufbau befindliche Forschungsdatenzentrum des Deutschen Zentrums für Integration und Migration (DeZIM) plant die Etablierung eines Remote Access.

16 Skopek et al. (2016) sowie Fuß und Wenzig (2019) informieren ausführlich über RemoteNEPS.

17 Vgl. auch https://www.neps-data.de/de-de/datenzentrum/datenzugang/remoteneps.aspx (Zugriff am 31.08.2019).

18 Vgl. https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Datenzugangswege/Vertraege/NEPS_Datennutzungsvertrag_de.pdf (Zugriff am 31.08.2019).

19 Vgl. https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Datenzugangswege/Vertraege/NEPS_RemoteNEPS_Ergaenzungsvertrag_de.pdf (Zugriff am 31.08.2019)

3.2.2. Remote Desktop-Lösung des DZHW

Das Forschungsdatenzentrum des DZHW (fdz.DZHW) ermöglicht Forschenden seit seiner Eröffnung 2017 ebenfalls einen kostenlosen Remote Zugriff auf seinen Mikrodatenbestand.[20] Die Voraussetzung für den Datenzugang ist der Abschluss eines Datennutzungsvertrages. Dieser setzt (wie auch bei LIfBi) einen Datennutzungsantrag voraus, in welchem das Forschungsvorhaben beschrieben wird und die beantragten Daten, der gewünschte Datenzugang und die benötigte Analysesoftware aufgeführt werden. Nach positiver Prüfung des Antrages wird der Datennutzungsvertrag geschlossen, welcher neben der datenschutzrechtlichen Verpflichtung gemäß der DS-GVO die Nutzungsdauer und den Datenzugang festlegt.[21] Die Datennutzenden verpflichten sich, die Daten gemäß des vereinbarten Vertrages zu verwenden (keine gewerbliche Nutzung,[22] keine Weitergabe der Mikrodaten und des Zugangscodes an Dritte, keine Re-Identifikationsversuche). Die Nutzung des Remote-Zugangs des fdz.DZHW ist unter gleichen Bedingungen auch für ausländische Forschende und deutsche Forschende aus dem Ausland möglich.

Nach erfolgtem Login mit einem individuellen Passwort können Forschende die Mikrodaten über eine gesicherte Verbindung in einer virtuellen Arbeitsumgebung mit den im Datennutzungsantrag bestellten Statistikprogrammen wie R, Stata oder SPSS[23] analysieren. Wie bei jedem Remote Desktop Verfahren verbleiben die Daten physikalisch auf den Servern des fdz.DZHW und werden dort verarbeitet.

Das fdz.DZHW implementiert derzeit eine neue technische Lösung. Der Aufbau der sicheren Verbindung zum Remote Access Server ist ohne die Installation zusätzlicher Software auf dem Zugangsgerät möglich, sofern ein Standard Internetbrowser genutzt wird.[24]

Auch beim fdz.DZHW können individuelle Zugangsrechte verteilt werden, sodass in Kooperation arbeitende Forschende am jeweils eigenen Arbeitsplatz dieselben Daten nutzen können. Die Zwischenergebnisse können eingesehen werden. Vor der Freigabe der letztendlichen Analyseergebnisse erfolgt eine datenschutzrechtliche Prüfung von Seiten des fdz.DZHW.

3.3 Zugang zu Daten der amtlichen Statistik in den nordischen Ländern und den Niederlanden

Eine Reihe von statistischen Ämtern in Europa (z. B. Frankreich, Niederlande, Schweden, Dänemark, Finnland, Estland, Slowenien, Irland, Kroatien, Lettland und Litauen) bieten zwischenzeitlich, wie einführend erwähnt, Forschenden die Möglichkeit des Datenzugangs am eigenen Arbeitsplatz. Stellvertretend für andere skizzieren wir im Folgenden die Remote Access Modelle der statistischen Ämter in Finnland, Schweden, Dänemark, den Niederlanden und Norwegen.[25] Berücksichtigt wird a) der geographische Ort des Zugangs – Inland/Ausland und b) unter welchen Rahmenbedingungen jeweils ausländische – also auch deutsche – Forschende Zugang zu den Daten erhalten. Allen Ländern ist gemeinsam, dass (1) der Datenzugang nach dem ‚need-to-know' Prinzip erfolgt,[26] (2) in der Regel sogenannte de-identified[27] (formal anonymisierte) Daten genutzt werden können und (3) der Datenzugang kostenpflichtig ist. Für eine zusammenfassende Darstellung siehe Tab. 1, S. 23.

20 Vgl. hierzu https://fdz.dzhw.eu/datennutzung/zugang (Zugriff am 31.08.2019).
21 Bei Vertragsverletzung können Vertragsstrafen bis zu 10.000 Euro verhängt werden. Darüber erfolgt eine Mitteilung (Vorfall und Namen) an die anderen FDZ.
22 Es besteht die Möglichkeit, eine gewerbliche Nutzung der Daten zu beantragen. Bisher gab es noch keine diesbezüglichen Anträge.
23 Die Bereitstellung der jeweiligen Software in der virtuellen Arbeitsumgebung erfolgt je nach Verfügbarkeit vgl. Antrag zur Nutzung von Mikrodaten des fdz.DZHW unter: https://fdz.dzhw.eu/pdf/fdz_datennutzungsantrag_dt_engl.pdf (Zugriff am 09.09.2019).
24 Derzeit ist diese Lösung für Mozilla Firefox möglich, die Konfigurierung für Google Chrome und weiterer Browser wird geprüft. Darüber hinaus ist auch die lokale kostenlose Installation des Clients auf dem Zugangsgerät möglich.
25 Für eine ausführliche Beschreibung der verschiedenen Zugangsmodelle in den nordischen Ländern siehe Statistics Finland et al. (2014).
26 Das heißt, es werden nur die für das Forschungsprojekt benötigten Daten zur Verfügung gestellt.
27 De-identified bezieht sich darauf, dass alle direkten Identifikatoren aus den Daten entfernt wurden. "Data are said to be de-identified when all variables that provide an explicit link to a data-subject, such as social security numbers, name, address, insurance number, etc. are replaced by random numbers." (Statistics Finland et al. 2014: 103). Demnach entspricht der Begriff de-identified im Deutschen der formalen Anonymisierung (BStatG) bzw. der Pseudonymisierung (BDSG).

Statistics Finland

3.3.1 Die Remote-Lösung von Statistics Finland (FIONA)

Über das System FIONA bietet Statistics Finland Forschenden einen kostenpflichtigen[28] Mikrodatenzugriff via Remote Desktop auf de-indentified Datenbestände an. Eine physikalische Weitergabe der nur de-indentified Daten an die Forschenden ist nicht möglich, dafür müssten die Daten stärker anonymisiert werden. Aufgrund des damit verbundenen Arbeitsaufwands ist Statistics Finland daran interessiert, dass die Forschung die Möglichkeiten des Remote Desktop nutzt.

Die Voraussetzungen für den Datenzugang sind die Teilnahme an einer Schulung, die Verpflichtung auf Einhaltung des Statistikgeheimnisses[29] und der Abschluss eines Datennutzungsvertrags. Der Zugang ist ortsabhängig, d.h. er kann nur von dem Ort (Adresse) durchgeführt werden, der im Projektantrag angegeben wurde und ist an die dazugehörige spezifizierte IP-Adresse gebunden, Änderungen bedürfen eines Antrages.

Nutzende loggen sich mit einer persönlichen User ID und einem Passwort ein. Darüber hinaus erfolgt die Authentifizierung durch eine sogenannte ‚flash message' mit Zugangscode an das Mobiltelefon der Nutzenden. In der virtuellen Arbeitsumgebung stehen verschiedene Statistikprogramme (R, Stata, SPSS etc.) das Geoinformationssystem QGIS sowie Textverarbeitungsprogramme (z. B. Libre Office oder Latex) zur Verfügung.

Auf Antrag kann ein Remote Desktop Zugang von zuhause oder aus dem Ausland gewährt werden; dies gilt auch für deutsche Nutzende, sofern sie mit einer finnischen Institution kooperieren und die obigen Zugangsvoraussetzungen erfüllen.[30]

Die endgültigen Analyseergebnisse werden vor der Übermittlung an Forschende durch Statistics Finland auf die Einhaltung der datenschutzrechtlichen Bestimmungen geprüft. Statistics Finland gibt hierbei in seinen ‚Rules and Instructions of Research Services' spezifische Regeln vor.[31] Die Forschenden sind vertraglich verpflichtet, ihre Analyseergebnisse (unabhängig ob in numerischer, tabellarischer oder grafischer Form) vorab selbst auf Einhaltung dieser Geheimhaltungsregeln zu prüfen. Bei Vertragsverstößen kann den Forschenden oder der ganzen Forschungsinstitution der Datenzugang entzogen werden und/oder eine Geld- oder Gefängnisstrafe verhängt werden.

28 Die Kosten für den Datenzugang durch Remote Access reichen von 200 Euro pro Monat/2400 Euro pro Jahr für einen Nutzenden bis hin zu 600 Euro pro Monat/7200 Euro pro Jahr für neun Nutzende. Vgl. https://www.stat.fi/tup/hinnat/tutkimuspalvelut_en.html (Zugriff am 25.03.2019)

29 Vgl. https://tilastokeskus.fi/sivusto/lomakkeet/salassapitositoumus_en.pdf (Zugriff am 10.04.2019).

30 Vgl. „Rules and instructions of Research Services" https://www.stat.fi/static/media/uploads/tup_en/mikroaineistot/rules_researcherservices.pdf (Zugriff am 12.01.2018)

31 https://www.stat.fi/static/media/uploads/tup_en/mikroaineistot/rules_researcherservices.pdf (Zugriff am 12.01.2018). Auch bei den FDZ der amtlichen Statistik in Deutschland gibt es entsprechende Vorgaben https://www.forschungsdatenzentrum.de/sites/default/files/fdz_nutzungsbedingungen_regelungen.zip (Zugriff am 21.01.2019). Ähnliche Regeln finden sich auch bei den FDZ der Sozialversicherungsträger http://doku.iab.de/fdz/access/Vorgaben_DAFE.PDF (Zugriff am 09.09.2019)

3.3.2 Die Remote Desktop-Lösung von Statistics Sweden

Statistics Sweden ermöglicht den Zugang zu formal anonymisierten Mikrodaten mit einem Remote Desktop Verfahren über das System MONA (Microdata Online Access).[32] Die Datennutzung auf diesem Weg ist für Forschende, die mit öffentlichen Mitteln aus Schweden gefördert werden, kostenfrei. Liegt keine öffentliche Förderung vor, fallen Kosten an.[33]

Forschende müssen für den Zugang über eine sichere Internetverbindung verfügen (kein offenes WLAN) und die Zugangsvoraussetzungen erfüllen (ggfls. kann ein Ethikvotum notwendig sein).

Die Authentifizierung erfolgt im ersten Schritt durch die Eingabe eines Einmal-Passwortes (OTP – one-time password) – generiert über eine Sicherheitskarte oder eine Smartphone App – , im zweiten Schritt durch die Eingabe von Username und Passwort. Die Forschungsdaten können innerhalb der Remote Desktop Umgebung mit der dort bereitgestellten Software (u. a. R, Stata, SPSS, QGIS oder FreeMat)[34] analysiert werden.

Statistics Sweden ist nach eigenen Aussagen relativ restriktiv, ausländischen Forschenden an ausländischen Universitäten über MONA einen Zugang zu schwedischen Mikrodaten zu erlauben. Prinzipiell kann er beantragt (und genehmigt) werden, unter der Voraussetzung, dass eine Kooperation mit einer schwedischen Universität vorliegt und die ausländische Forschungseinrichtung in der EU, einem EWG-Land oder einem anderen Land angesiedelt ist, das von der EU-Kommission als sicher für den Datentransfer eingestuft wurde.[35] Die Daten müssen über eine schwedische Universität bei Statistics Sweden beantragt werden und die beantragende Einrichtung ist für den sicheren Umgang mit den Daten verantwortlich.

Das Hinzufügen von eigenen Daten ist in MONA möglich, der Import wird von Statistics Sweden vorgenommen und kontrolliert. Den Forschenden werden die Ergebnisse ihrer Analysen automatisch per Mail zugesandt, dieser Vorgang wird derzeit in einen direkten Download geändert. Eine Outputkontrolle erfolgt nur stichprobenartig nach dem Zufallsprinzip; erfolgt der Zugang über MONA aus dem Ausland, werden alle Outputs kontrolliert.

Durch die in den Nutzungsbedingungen festgelegten Geheimhaltungspflichten nimmt Statistics Sweden die beantragende Institution in die Pflicht, die Einhaltung der Datenschutzvorgaben und der vertraglichen Bedingungen durch die Forschenden zu gewährleisten.

Nutzenden, die gegen die Geheimhaltungspflicht verstoßen, kann der Zugriff auf MONA entzogen werden. Langfristig kann sich ein Verstoß auch auf die Entscheidung bei zukünftigen Anfragen der beantragenden Institution auswirken.

32 Vgl. https://www.scb.se/mona-en (Zugriff am 01.11.2018)

33 Die Höhe der Kosten ist abhängig von der Komplexität der Datenbereitstellung. Eine relativ standardisierte Datenlieferung mit Daten aus wenigen Registern kostet ca. 3000-6000 Euro (30.000-60.000 SEK) (Beispielrechnung). SCB ist rechtlich verpflichtet, kostendeckend zu arbeiten. Vgl. https://www.scb.se/vara-tjanster/prissattning-pa-scb (Zugriff am 25.03.2019)

34 Vgl. https://www.scb.se/en/services/guidance-for-researchers-and-universities/mona--a-system-for-delivering-microdata/software-available-to-users-in-mona (Zugriff am 23.03.2019)

35 Vgl. https://ec.europa.eu/info/law/law-topic/data-protection/data-transfers-outside-eu/adequacy-protection-personal-data-non-eu-countries_en (Zugriff am 18.02.2019)

STATISTICS DENMARK

3.3.3 Die Remote Desktop-Lösung von Statistics Denmark

Statistics Denmark bereitet die beantragten formal anonymisierten Mikrodaten kostenpflichtig[36] auf und stellt sie über einen gesicherten Zugang in einer virtuellen Arbeitsumgebung zur Verfügung.[37] Voraussetzung dafür ist, dass Forschende mit einer dänischen Institution zusammenarbeiten, die von Statistics Denmark für den Remote Desktop Zugang autorisiert ist.[38]

Öffentliche und private gemeinnützige Forschungseinrichtungen sowie staatliche Behörden können autorisiert werden, unter besonderen Umständen ist dies auch für Unternehmen möglich.[39]

Die Authentisierung der Forschenden erfolgt über ein Passwort sowie einen von Statistics Denmark bereitgestellten Token (RSA SecurID card). Der Zugriff auf die virtuelle Arbeitsumgebung erfolgt über eine gesicherte Verbindung. Dort steht den Forschenden Standard-Analysesoftware (z. B. R, Stata, SPSS, SAS oder GAUSS) zur Verfügung.

Der Zugriff für Forschende aus dem Ausland ist möglich, wenn Forschende mit einer für den Remote Desktop autorisierten dänischen Institution zusammenarbeiten. In diesem Fall ist die dänische Institution dafür verantwortlich, dass sich die ausländischen Forschenden an die Geheimhaltungsregeln halten. Diese Regelung gilt auch für dänische Forschende, die aus dem Ausland auf die Daten zugreifen.[40]

Ein Datendownload oder Screenshot ist, wie in den anderen Ländern, nicht möglich bzw. verboten. Die Analyseergebnisse werden den Forschenden per Mail zugesandt und stichprobenartig von Statistics Denmark auf Einhaltung der datenschutzrechtlichen Bestimmungen kontrolliert.

Forschenden, die gegen die Nutzungsbedingungen verstoßen, wird der Zugang zu Daten von Statistics Denmark für mindestens drei Jahre entzogen.

36 155 Euro/Stunde für Beratung und Datenbereitstellung vgl. https://www.dst.dk/da/TilSalg/Forskningsservice/Meddelelser (Zugriff am 26.03.2019).

37 Bedingungen: https://www.dst.dk/en/TilSalg/Forskningsservice (Zugriff am 19.11.2018). Für anekdotische Evidenz hinsichtlich des bürokratischen und finanziellen Aufwands für die Datenbereitstellung siehe z. B. Kvalheim und Høgetveit Myhren (2017: 5).

38 Statistics Denmark will not grant authorization to single persons. https://www.dst.dk/ext/645846915/0/forskning/Access-to-micro-data-at-Statistics-Denmark_2014.pdf

39 ebd.

40 Vgl. Access to micro data at Statistics Denmark_2014 https://www.dst.dk/en/TilSalg/Forskningsservice (Zugriff am 06.11.2018).

Statistics Netherlands

3.3.4 Die Remote Desktop-Lösung von Statistics Netherlands (CBS)

CBS stellt Forschenden die für das Forschungsprojekt benötigten formal anonymisierten Daten kostenpflichtig[41] in einer virtuellen Arbeitsumgebung bereit.[42]

Forschende erhalten auf Leihbasis einen personalisierten Token, um den Remote Access zu ermöglichen. Neben den üblichen vertraglichen Verpflichtungen und einer persönlichen Einführung, ist dem eigentlichen Datenzugang jedes Mal aufs Neue ein zufallsgenerierter Testfragebogen vorgeschaltet, der der Sensibilisierung im Umgang mit Mikrodaten dient. Nur nach bestandenem Test wird der Datenzugang freigeschaltet. Für die Datenanalyse stehen u. a. R, Stata oder SPSS zur Verfügung.[43]

Forschende aus dem Ausland sollten in einer Arbeitsbeziehung mit einer Institution stehen, die von CBS für den Zugriff auf die Mikrodaten autorisiert ist. Für den Remote-Zugang zu Mikrodaten sind folgende Einrichtungen zugelassen: Universitäten, öffentliche Forschungsinstitutionen, Organisationen für Politikberatung und -analyse, Statistikbehörden anderer EU Länder sowie sonstige Forschungsinstitutionen, die für die Arbeit mit den Mikrodaten von CBS[44] autorisiert sind (über einen Autorisierungsantrag).[45]

Beim Import von eigenen Daten mit dem Ziel, diese mit den bereitgestellten Mikrodaten zu verknüpfen, erfolgt vorab – wie in allen anderen dargestellten Modellen auch – eine datenschutzrechtliche Prüfung durch CBS. Wie bei Statistics Denmark sind die Forschenden dazu angehalten, ihre Ergebnisse vorab (selbstständig) auf Einhaltung der Geheimhaltungsregeln zu überprüfen und erst danach einen Export zu beantragen. Vor dem Export von Analyseergebnissen prüft CBS zudem auf Einhaltung der vorgegebenen Regeln. Werden die Vorgaben von CBS nicht eingehalten, werden dem Forschenden bzw. der Institution zusätzliche Kosten für die Outputkontrolle in Rechnung gestellt. Verstöße gegen die Nutzungsbedingungen können je nach Schweregrad des Verstoßes geahndet werden (z. B. vom Widerruf bis hin zur Aussetzung der Projektvereinbarung Zugangsberechtigung für einen Monat bei Login über ein öffentliches Netzwerk bis hin zur Aussetzung der Projektvereinbarung für alle beteiligten Forschenden falls Mikrodaten kopiert werden).[46]

41 Zur Gebührenstruktur von CBS vgl. https://www.cbs.nl/-/media/cbs%20op%20maat/zelf%20onderzoek%20doen/181201%20services%20catalogue2019.pdf (Zugriff am 22.03.2019).

42 Vgl. https://www.cbs.nl/en-gb/our-services/customised-services-microdata/microdata-conducting-your-own-research (Zugriff am 01.11.2018).

43 Außerdem sind noch folgende Programme verfügbar: MS Access, MS Word, MS Excel, MS Powerpoint, Winzip, Adobe Acrobat Reader, Blaise, WinEdt, Winbugs.

44 Im Rahmen des IDAN Projektes hat z. B. GESIS einen Remote Access Zugang zu CBS eingerichtet (Stand Mai 2019).

45 Vgl. https://www.cbs.nl/en-gb/our-services/customised-services-microdata/microdata-conducting-your-own-research (Zugriff am 18.04.2019).

46 Vgl. die Sanktionsmaßnahmen von CBS im PDF „Remote access sanctioning policy" unter https://www.cbs.nl/en-gb/our-services/customised-services-microdata/microdata-conducting-your-own-research/rules-and-sanctioning-policy (Zugriff am 06.11.2018).

Statistisk sentralbyrå
Statistics Norway

3.3.5 Der Remote Access bei Statistics Norway (SSB)

In Norwegen erfolgte der Zugang zu amtlichen Mikrodaten bislang vorwiegend über eine physikalische Übermittlung von Daten (data release bzw. ‚hand out data to researchers'). Nach erfolgreicher Beantragung von Mikrodaten werden diese den Forschenden formal anonymisiert (de identified)[47] kostenpflichtig zur Verfügung gestellt und können am eigenen Arbeitsplatz genutzt werden.

Das norwegische Statistikgesetz erlaubt keine Übermittlung von nur formal anonymisierten Daten ins Ausland. Dementsprechend müssen Daten vor einer Übermittlung an Forschungsinstitute im Ausland so anonymisiert werden, dass kein Personenbezug hergestellt werden kann. Eine Alternative hierzu ist die Kooperation mit norwegischen Forschungseinrichtungen, in diesem Fall ist auch die Nutzung von de-identified Daten möglich.[48]

In Ergänzung hierzu wird seit März 2018 darüber hinaus ein weiterer Datenzugangsweg zu amtlichen Mikrodaten getestet, das sogenannte RAIRD[49]-Projekt (Remote Access Infrastructure for Register Data). Mittels RAIRD können Forschende Mikrodaten über microdata.no direkt ohne Beantragungsprozess analysieren. Eine spezielle Vorbereitung der Daten durch Statistics Norway entfällt. Voraussetzung ist, dass die Forschenden über eine Affiliation zu einer für den Datenzugang autorisierten Institution, verfügen. Die Autorisation kann über die Webseite microdata.no beantragt werden.

In Hinblick auf die bislang dargestellten Remote Access Modelle unterscheidet sich das norwegische Modell unter anderem dahingehend, dass die Forschenden zwar auch auf (nur) formal anonymisierte Daten zugreifen, aber die Daten am Bildschirm nicht eingesehen werden können. Einsehbar sind nur die Datendokumentation und sonstige Metadaten. Wie bei einem Remote Execution Modell wird die Syntax ‚blind' geschrieben und an den Server geschickt. Der generierte Output wird automatisch anonymisiert, zumindest wenn man der Beschreibung von RAIRD folgt.[50] In welcher Form und in welchem Ausmaß anonymisierend in die Datenanalyse eingegriffen wird, ist im RAIRD User Manual[51] (nur auf Norwegisch verfügbar) beschrieben. Über die Folgen dieser Eingriffe in die Reliabilität und Validität der erzeugten Ergebnisse können an dieser Stelle keine Aussagen getroffen werden.

In der online Arbeitsumgebung von microdata.no sind keine der Standard-Statistikprogramme verfügbar, die Nutzenden können gegenwärtig nur mit den Python Modulen SciPy, NumPy, Pandas und Statsmodels arbeiten.[52] Es ist nicht möglich, eigene Daten mit den in RAIRD verfügbaren Daten zu verknüpfen. Darüber hinaus sind die Analysemöglichkeiten deutlich limitiert (z. B. sind keine Panelanalysen möglich). ***De facto soll RAIRD auch nicht das bisherige System der physikalischen Datenweitergabe ersetzen, sondern vielmehr zu mehr Forschung mit den norwegischen Registerdaten anregen.*** Als Beispiel werden etwa Master-Studierende angeführt, an die im Regelfall aufgrund der hohen Kosten keine Weitergabe von norwegischen Registerdaten erfolgt.[53]

47 https://www.ssb.no/en/omssb/personvern/behandling-av-personopplysninger/_attachment/123246?_ts=13f3ce570c8 (Zugriff am19.09.2019).
48 Vgl. https://www.ssb.no/en/omssb/tjenester-og-verktoy/data-til-forskning#Microdata_applications_from_outside_Norway (Zugriff am 27.03.2019)
49 http://raird.no (Zugriff am 08.02.19)
50 Vgl. https://microdata.no (Zugriff am 27.03.2019)
51 Vgl. RAIRD User Manual (Norwegian only). https://microdata.no/brukermanual.pdf (Zugriff am 11.04.2018)
52 Vgl. https://microdata.no/en/faq (Zugriff am 27.03.2019)
53 "Norway has many good sources of register data, and Norwegian law allows them to be used for research purposes. However, the application process for obtaining de-identified microdata from registers is long-winded, and adapting datasets for researchers is time-consuming and costly. microdata.no gives researchers instant access to register data and enables them to adapt the data without going through the application process." https://microdata.no/en/faq (Zugriff am 04.07.2019).

Tab. 1: Zugang zu Daten der amtlichen Statistik in den nordischen Ländern und den Niederlanden

	Voraussetzungen RA Inland	RA aus dem Ausland	RA für Ausländer	Kostenpflichtig	Anonymisierungsgrad	Outputkontrolle	Sanktionen bei Vertragsverstößen
Statistics Finland	IP-Adresse finnischer Institution, Nutzerschulung, Datennutzungsvertrag	Ja	Ja	Ja	De-identified/ pseudonymisiert	Ja	Ja
Statistics Sweden	Affiliation mit autorisierter Forschungsinstitution, Datennutzungsvertrag	Ja	Ja	Nein, wenn mit öffentlichen Mitteln aus Schweden geförderten Forschende Ja, nicht öffentlich geförderte Forschung	De-identified/ pseudonymisiert	Ja, nach Zufallsprinzip, bei Zugang aus dem Ausland wird alles kontrolliert	Ja
Statistics Denmark	Affiliation mit autorisierter Forschungsinstitution, Datennutzungsvertrag	Ja	Ja	Ja	De-identified/ pseudonymisiert	Ja, nach Zufallsprinzip	Ja
Statistics Netherlands	Affiliation mit autorisierter Forschungsinstitution, Datennutzungsvertrag	Ja	Ja	Ja	De-identified/ pseudonymisiert	Ja	Ja
Statistics Norway	Affiliation mit autorisierter Forschungsinstitution	Unbekannt	Unbekannt	Ja	De-identified/ pseudonymisiert	Nein, nicht relevant	Ja
LIfBi Deutschland	Affiliation mit anerkannter Forschungsinstitution, Voraussetzungen: ein NEPS-Datennutzungsvertrag, ein Remote Access Ergänzungsvertrag, Nutzerschulung tippbiometrische Registrierung	Ja	Ja	Nein	Moderater Anonymisierungsgrad	Ja	Ja
DZHW Deutschland	Voraussetzungen: ein NEPS-Datennutzungsvertrag, ein Remote Access Ergänzungsvertrag, Nutzerschulung tippbiometrische Registrierung	Ja	Ja	Nein	Moderater Anonymisierungsgrad	Ja	Ja

4 Aktuelle Rahmenbedingungen der Datenbereitstellung der amtlichen Statistik und der Sozialversicherungsträger

Im Folgenden werden zunächst die geltenden rechtlichen Rahmenbedingungen zur Bereitstellung von Mikrodaten der Statistischen Ämter des Bundes und der Länder für die Wissenschaft (Kapitel 4.1) dargelegt. Anschließend werden die entsprechenden Regelungen für die Daten der Sozialversicherungsträger skizziert

4.1 Datenbereitstellung nach dem Bundesstatistikgesetz

Die FDZ der Statistischen Ämter des Bundes und der Länder bieten Hochschulen und sonstigen Einrichtungen mit der Aufgabe unabhängiger wissenschaftlicher Forschung Zugang zu Mikrodaten aus Erhebungen der amtlichen Bundesstatistik[54] für wissenschaftliche Zwecke. Vertragspartner der FDZ (Datenproduzenten) sind hierbei nicht die individuellen Forschenden, sondern die jeweiligen Hochschulen oder Forschungseinrichtungen mit der Aufgabe der unabhängigen Forschung. Die rechtlichen Grundlagen für die Bereitstellung der Einzelangaben werden in den § 3, 3a und 16 des Gesetzes über die Statistik für Bundeszwecke[55] (BStatG) geregelt.

Demnach gehört es zu den Aufgaben des Statistischen Bundesamtes, *„[...] Einzelangaben nach Maßgabe dieses Gesetzes oder einer anderen Rechtsvorschrift für wissenschaftliche Zwecke bereitzustellen; die Zuständigkeit der Länder, diese Aufgabe ebenfalls wahrzunehmen, bleibt unberührt“* (§ 3 Abs. 1 Nr. 4 BStatG).

Die Zusammenarbeit zwischen dem Statistischen Bundesamt und den statistischen Ämtern der Länder ist in § 3a Abs. 1 und 2 BStatG geregelt. § 16 Abs. 7 BStatG sieht die explizite Verpflichtung der Forschenden auf Geheimhaltung vor, bevor die Arbeit mit den Daten begonnen werden kann.[56] Diese Verpflichtung ermöglicht eine strafrechtliche Verfolgung bei einem Verstoß gegen § 203 Abs. 2 Nr. 6 StGB. Davon abgeleitet wird nach vorherrschender Rechtsauslegung, dass die Weitergabe von SUF an im Ausland arbeitende Forschende (unabhängig von ihrer Nationalität) grundsätzlich nicht möglich ist, da eine strafrechtliche Verfolgung bei einer im Ausland begangenen Tat aus rechtlichen und tatsächlichen Gründen ins Leere laufen würde.[57]

Der § 16 Abs. 8 BStatG regelt mit der strikten Zweckbindung für das Forschungsvorhaben die Verwendungsmöglichkeiten der bereitgestellten Daten. Darüber hinaus werden die empfangenden Institutionen in die Pflicht genommen, mittels technischer und organisatorischer Maßnahmen dafür Sorge zu tragen, dass nur die nach § 16 Abs. 7 BStatG verpflichteten Personen Datenzugang haben.

§

§ 16 Abs. 8 BStatG: Die aufgrund einer besonderen Rechtsvorschrift oder der Absätze 4, 5 oder 6 übermittelten Einzelangaben dürfen nur für die Zwecke verwendet werden, für die sie übermittelt wurden. In den Fällen des Absatzes 6 Satz 1 Nummer 1 sind sie zu löschen, sobald das wissenschaftliche Vorhaben durchgeführt ist. Bei den Stellen, denen Einzelangaben übermittelt werden, muss durch organisatorische und technische Maßnahmen sichergestellt sein, dass nur Amtsträger, für den öffentlichen Dienst besonders Verpflichtete oder Verpflichtete nach Absatz 7 Satz 1 Empfänger von Einzelangaben sind.“

54 Der Zugang zu Daten aus Erhebungen, denen keine Bundesgesetzgebung zugrunde liegt, so z. B. bei der Schulstatistik, ist über diese Gesetzgebung nicht geregelt.

55 In der Fassung der Bekanntmachung vom 20. Oktober 2016 (BGBl. I.S. 2394), das zuletzt durch Art. 10 Abs. 5 des Gesetzes vom 30. Oktober 2017 (BGBl. I S. 3618) geändert worden ist.

56 § 16 Abs. 7 BStatG: Personen, die Einzelangaben nach Abs. 6 erhalten sollen, sind vor der Übermittlung zur Geheimhaltung zu verpflichten, soweit sie nicht Amtstragende oder für den öffentlichen Dienst besonders Verpflichtete sind. § 1 Abs. 2, 3 und 4 Nr. 2 des Verpflichtungsgesetzes vom 2. März 1974 (BGBl. I S. 469, Artikel 42), das durch Gesetz vom 15. August 1974 (BGBl. I S. 1942) geändert worden ist, gilt entsprechend.

57 Hier wäre zu prüfen, ob innerhalb der EU zwischenzeitlich andere Regelungen Anwendung finden.

Wie schon bei der Datenschutzrichtlinie ist auch in der DS-GVO nicht definiert, was unter dem Begriff der „Übermittlung“ verstanden werden soll. Die in § 3 Absatz 4 Nr. 3 Bundesdatenschutzgesetz alte Fassung enthaltene Festlegung, wird jedoch de facto als weiterhin gültig angesehen. Danach ist „Übermitteln das Bekanntgeben gespeicherter oder durch Datenverarbeitung gewonnener personenbezogener Daten an einen Dritten in der Weise, dass a) die Daten an den Dritten weitergegeben werden oder b) der Dritte zur Einsicht oder zum Abruf bereitgehaltene Daten einsieht oder abruft“.

Grundsätzlich gilt, dass **das rechtlich vorgesehene Schutzniveau für die Daten nicht durch die Art der Übermittlung untergraben werden darf**. Diese Vorgabe schlägt sich dann auch in § 16 Abs. 6 BStatG nieder, der einerseits ein Wissenschaftsprivileg im Datenzugang vorsieht und sich andererseits mit einem erforderlichen Anonymisierungsniveau der Daten in Abhängigkeit von der Art der Datenbereitstellung befasst.

§

*§ 16 Abs. 6 BStatG: Für die Durchführung **wissenschaftlicher Vorhaben** dürfen das Statistische Bundesamt und die statistischen Ämter der Länder Hochschulen oder sonstigen Einrichtungen mit der Aufgabe unabhängiger wissenschaftlicher Forschung.*

*1) Einzelangaben übermitteln, wenn die Einzelangaben nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft zugeordnet werden können (**faktisch anonymisierte** Einzelangaben),*

*2) innerhalb speziell abgesicherter Bereiche des Statistischen Bundesamtes und der statistischen Ämter der Länder Zugang zu **formal anonymisierten** Einzelangaben gewähren, wenn wirksame Vorkehrungen zur Wahrung der Geheimhaltung getroffen werden.*

Faktisch anonymisierte Daten dürfen nach § 16 Abs. 6 Nr. 1 BStatG in Form von SUF direkt an die Forschung übermittelt werden. Im Unterschied hierzu ist für die Nutzung von formal anonymisierten Daten in § 16 Abs. 6 Nr. 2 BstatG festgelegt, dass ein Datenzugang nur *„innerhalb speziell abgesicherter Bereiche des Statistischen Bundesamtes und der statistischen Ämter der Länder“* erfolgen kann und zudem *„wirksame Vorkehrungen zur Wahrung der Geheimhaltung“* getroffen werden.

Der entscheidende Unterschied zu den in Kapitel 3 skizzierten Remote Desktop Lösungen der amtlichen Statistik in anderen europäischen Ländern ist demnach vor allem in § 16 Abs. 6 Nr. 2 BStatG begründet, der einen Zugang zu formal anonymisierten Daten der deutschen amtlichen Statistik aktuell nur in speziell geschützten Bereichen der amtlichen Statistik erlaubt.

Die Formulierung *„innerhalb von speziell abgesicherter Bereiche des Statistischen Bundesamtes und der statistischen Ämter der Länder“* legt die Auslegung nahe, dass es sich hier um bestimmte Räumlichkeiten handeln muss, womit folglich ein Remote Desktop Zugang außerhalb abgesicherter Bereiche, so am Uni-Arbeitsplatz der Forschenden, derzeit nicht möglich ist.

Dass sich diese speziellen Bereiche nicht immer in den Dienstgebäuden der amtlichen Statistik befinden müssen, zeigen die FDZ Standorte im DIW in Berlin, an der TU Dresden, am ZEW in Mannheim, am ifo Institut in München und an der Universität Frankfurt am Main, die jedoch alle die geforderte Absicherung über das vor Ort anwesende Personal der amtlichen Statistik realisieren. Letztlich ist das jedoch nur eine mögliche Operationalisierung für das unstrittig erforderliche Schutzniveau.

4.2 Datenbereitstellung nach dem Sozialgesetzbuch

Die Übermittlung von Daten, die anlässlich der Verwaltung von Leistungen der sozialen Sicherung und weiterer Aufgaben der Sozialleistungsträger entstehen, an Dritte ist im Sozialgesetzbuch geregelt. Für einige Bereiche gibt es spezielle Einzelnormen, für alle gemeinsam gilt § 75 SGB X als grundlegende Norm.

Nach § 75 SGB X ist eine Übermittlung von Sozialdaten nach Absatz 1 zulässig:

§

soweit sie erforderlich ist für ein bestimmtes Vorhaben

1) der wissenschaftlichen Forschung im Sozialleistungsbereich oder der wissenschaftlichen Arbeitsmarkt- und Berufsforschung oder

2) der Planung im Sozialleistungsbereich durch eine öffentliche Stelle im Rahmen ihrer Aufgaben

und schutzwürdige Interessen der betroffenen Person nicht beeinträchtigt werden oder das öffentliche Interesse an der Forschung oder Planung das Geheimhaltungsinteresse der betroffenen Personen erheblich überwiegt. Eine Übermittlung ohne Einwilligung der betroffenen Person ist nicht zulässig, soweit es zumutbar ist, ihre Einwilligung einzuholen.

Die Übermittlung bzw. Bereitstellung der Daten unterliegen somit einer Zweckbindung. Ferner muss die Datenübermittlung von der zuständigen obersten Bundes- oder ggf. durch eine Landesbehörde genehmigt werden (§ 75 SGB X).

Da ein Zugang zu den formal anonymisierten Daten der FDZ BA im IAB und der Rentenversicherung ausschließlich an den jeweiligen GWAP möglich ist und somit keine **physikalische** Übermittlung der Daten, sondern nur von geprüften Ergebnissen stattfindet, erfolgt für jede Datennutzungsantrag **vorab** eine individuelle Prüfung, ob eine Genehmigung der obersten Bundes- oder ggf. einer Landesbehörde erforderlich ist.

Folgende spezielle Normen hinsichtlich der Zugangsbedingungen und der Forschungsthemen sind ebenfalls zu berücksichtigen: § 282 Abs. 7 SGB III für die Daten der Bundesagentur für Arbeit,[58] § 206 SGB VII für die Forschung zu Berufskrankheiten mit Daten der gesetzlichen Unfallversicherung und von behandelnden Ärzten.[59]

Werden personenbezogene Daten (Art. 4 Nr. 1 EU Datenschutz-Grundverordnung) ausreichend stark anonymisiert, gegebenenfalls auch nur pseudonymisiert[60], können andere standardisierte Verfahren des Datenzugangs, wie z. B. SUF, zur Anwendung kommen.

Im Falle von anonymisierten Daten ist der Vertrag für die Datennutzung nicht mehr als genehmigungspflichtig der obersten Bundesbehörde vorzulegen. Dies beschleunigt die Datenweitergabe. Es werden Datennutzungsverträge geschlossen, in denen die Bedingungen der Datennutzung (zeitliche Begrenzung der Nutzung auf die Projektlaufzeit, Datensicherheit, inhaltliche Begrenzung auf das Projekt, persönliche Begrenzung auf die genannten Personen, Datenschutzverpflichtung der Forschenden und ein Verbot der Deanonymisierung) geregelt sind.

Die Etablierung des FDZ der RV und des FDZ der Bundesagentur für Arbeit (BA) im IAB hatte zum Ziel, den Zugang der Forschung zu den Sozialdaten zu vereinfachen. Dies wird durch die Erstellung von standardisierten Datenprodukten für die Forschungsgemeinschaft im In- und Ausland umgesetzt, wobei der Datenzugang außerhalb Deutschlands restriktiver gehandhabt wird.

Eine dem BStatG (§ 16 Abs. 6 Nr. 2) vergleichbar formulierte Einschränkung, den Zugang zu formal anonymisierten Daten der Sozialversicherung nur in speziell geschützten Bereichen der Sozialversicherungsträger zu ermöglichen, findet sich im Sozialgesetzbuch nicht. Durch das ersatzlose Entfallen des § 3 Abs. 4 Nr. 3b BDSG sowie des wortgleichen § 67 Abs. 6 Nr. 3b SGB X (ab 25.05.2018), steht das Gesetz der Einführung einer Remote Desktop Lösung **nicht explizit** entgegen. Insbesondere, da das o.a. Genehmigungsverfahren im Zugang zu formal anonymisierten Daten auch unter den Bedingungen des Remote Desktop nicht außer Kraft gesetzt wird.

---

58 Sozialgesetzbuch (SGB III), Drittes Buch, Arbeitsförderung, https://www.sozialgesetzbuch-sgb.de/sgbiii/282.html (Zugriff am 19.11.2018)

59 Sozialgesetzbuch (SGB VII) Siebtes Buch, Gesetzliche Unfallversicherung, https://www.sozialgesetzbuch-sgb.de/sgbvii/206.html (Zugriff am 19.11.2018)

60 „Pseudonymisierung" die Verarbeitung personenbezogener Daten in einer Weise, dass die personenbezogenen Daten ohne Hinzuziehung zusätzlicher Informationen nicht mehr einer spezifischen betroffenen Person zugeordnet werden können, sofern diese zusätzlichen Informationen gesondert aufbewahrt werden und technischen und organisatorischen Maßnahmen unterliegen, die gewährleisten, dass die personenbezogenen Daten nicht einer identifizierten oder identifizierbaren natürlichen Person zugewiesen werden (Art. 4. Nr. 5 EU DS-GVO).

Für eine Umsetzung in die Praxis wäre z. B. auch zu prüfen, wie die Arbeitsplätze, von denen aus der Remote Zugang erfolgen soll, gestaltet sein müssen und wie eine sichere Authentisierung der Nutzenden erfolgen kann. So müssen bei einer physikalischen Übermittlung der Daten nach § 75 SGB X und § 282 Abs. 7 SGB III geeignete technische und organisatorische Maßnahmen getroffen werden (Art. 32 EU DS-GVO und § 64 BDSG), um die Datensicherheit zu gewährleisten. Dabei muss das Schutzniveau dem Risiko der Deanonymisierung und den daraus entstehenden Schäden/Folgen angepasst sein.[61] Derartige Regelungen müssen für den Remote Desktop Zugang noch definiert werden.

4.3 Zwischenfazit

Die bisherige Darstellung legt nahe, dass unter der aktuellen Gesetzeslage die Etablierung eines Remote Desktop Zugangs zu formal anonymisierten Daten im Falle der Sozialversicherungsträger einer datenschutzkonformen und praxistauglichen Operationalisierung bedarf, aber eine Gesetzesänderung nicht notwendig erscheint. Im Falle der Amtlichen Statistik hingegen ist eine klarstellende Gesetzesanpassung erforderlich.

In den nordischen Ländern und in den Niederlanden wird der Zugang via Remote Desktop als ein sehr sicherer Weg angesehen, um der Forschung einen Zugang zu formal anonymisierten amtlichen Mikrodaten zu ermöglichen. Die Daten verbleiben physikalisch bei der amtlichen Statistik und der Output unterliegt ihrer Kontrolle. Dies gibt der Statistik die Möglichkeit „to control that the rules governing data security laid down by the NSI's and the national laws are, in general, complied with" (Statistics Finland et al. 2014: 20). Darüber hinaus wird der Remote Zugang im Vergleich zu anderen Zugangsmodellen (GWAP onsite, Scientific Use File) als für die Datenproduzenten weniger ressourcenaufwändig angesehen.

In diesen Ländern funktionieren die Remote Modelle in ihren verschiedenen Varianten, da u. a. von einem grundsätzlich rechtskonformen Handeln der Forschung – also auch der deutschen Forschenden – ausgegangen wird. In Deutschland beinhaltet das BStatG zwar ein Wissenschaftsprivileg für den Zugang zu formal anonymisierten Daten, erachtet aber eine vertragliche Bindung an geltendes Recht für nicht hinreichend zur Realisierung der erforderlichen Schutzmaßnahmen. Stattdessen wird auf vorbeugende Kontrolle gesetzt, die neben der expliziten Verpflichtung auf Geheimhaltung und den vertraglichen Bindungen derzeit zusätzlich die räumliche Eingrenzung des Zugangs auf gesicherte Bereiche innerhalb der amtlichen Statistik vorsieht.

Dies führt zu der paradoxen Situation, dass deutschen Forschenden in anderen Ländern bei Kooperation mit nationalen Einrichtungen weitergehende Möglichkeiten des Datenzugangs eingeräumt werden, als in Deutschland selbst, was auch zum Standortnachteil und zur Verlagerung von Forschungskapazitäten auf ausländische Datenbestände führt und den Wissenschaftsstandort Deutschland schwächt. Zudem erfolgen die meisten Erhebungen der amtlichen Statistik in Deutschland aufgrund von europäischen Gesetzen und Regelungen, wobei die EU Qualitätsstandards und die zu erhebenden Merkmale festlegen. D.h. es existieren vom Merkmalskatalog her identische Datenbestände in jedem EU-Land (Lieferverpflichtung an Eurostat), die – wie oben geschildert – in einigen Ländern unter Einhaltung aller Datenschutzbestimmungen formal anonymisiert als tauglich für einen Remote Zugang erachtet werden, in Deutschland jedoch nicht.

Letztlich würden nicht nur die Forschung, sondern auch die FDZ der amtlichen Statistik und der Sozialversicherungsträger von einer Modernisierung des Datenzugangs profitieren. Die Etablierung eines Remote Zugangs vergleichbar den vorgestellten Desktop-Modellen in Kapitel 3 wird als ein äußerst anspruchsvolles, aber – aus Perspektive der Forschung und der Datenproduzenten – lohnendes Ziel gesehen.

61 Bundesdatenschutzgesetz (BDSG) § 64 Anforderungen an die Sicherheit der Datenverarbeitung https://www.buzer.de/64_BDSG_Bundesdatenschutzgesetz.htm (Zugriff am 06.02.2019). Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates Artikel 32 Sicherheit der Verarbeitung https://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=CELEX:02016R0679-20160504&from=DE (Zugriff am 06.02.2019)

5 Mögliche Wege zu Remote Desktop-Lösungen in Deutschland

Kapitel 5 skizziert technische und organisatorische Herausforderungen, die bei der Etablierung eines Remote Zugangs für die amtliche Statistik sowie die Sozialversicherungsträger auftreten. Der aus der Sicht der Forschenden ideale Zustand wäre selbstredend der ortsunabhängige Remote Desktop, z. B. entsprechend dem niederländischen Modell.[62] Auf dem Weg zum Idealzustand sind aber auch Zwischenlösungen in Form von ortsabhängigem Remote Zugang am eigenen Arbeitsplatz (wie in den nordischen Ländern) oder in spezifischen Räumen der wissenschaftlichen Einrichtungen denkbar, die das Arbeiten mit amtlichen Mikrodaten zeiteffizienter gestalten und zu einer deutlichen Vereinfachung des Forschungsprozesses beitragen würden.

An dieser Stelle sei darauf hingewiesen jedoch nicht weiter ausgeführt, dass die Etablierung eines Remote Zugangsweges mit erheblichen finanziellen Investitionen für die Implementierung und die laufenden Kosten verbunden ist. Durch eine Ausweitung des Angebotes wird sich vermutlich die Anzahl der gleichzeitig arbeitenden Forschenden erhöhen, da die Plätze nicht wie bei GWAP limitiert sind. Daher müssen auf der Seite der FDZ entsprechend skalierbare Server und ausreichende Software-Lizenzen bereitgestellt sowie das aktuelle Modell der Nutzungsentgelte, bei dem die Forschenden den Aufwand pauschal mitfinanzieren, angepasst werden. Damit verbunden ist auch die Frage, ob und in welcher Form sich die Forschung an den einmaligen und laufenden Kosten eines Remote Zugangs beteiligt. Ob dies über eine pauschale Finanzierung oder eine Anpassung der Nutzungsentgelte erfolgen kann, wäre an anderer Stelle zu diskutieren. Langfristig werden sich diese Investitionen jedoch sowohl für die datenbereitstellenden wie auch die forschungsfinanzierenden Institutionen u.a. durch eine Verschlankung der organisatorischen Prozesse und einer intensiveren Nutzung der verfügbaren Daten durch die Forschung rentieren. Darauf deuten zumindest die Erfahrungen der nordischen Länder und der Niederlande hin.

5.1 Grundsätzliche Überlegungen - Berechtigtenkreis und Zugangskontrolle

Für die folgende Skizzierung möglicher Remote Desktop Szenarien ist das am GWAP realisierte Schutzniveau die Referenz. Wie bei Schiller/Welpton (2013:8) ausgeführt, ist der GWAP Zugang technisch identisch zu einem Remote Desktop Verfahren. Der wesentliche Unterschied zwischen GWAP und Remote Desktop Verfahren besteht darin, dass sich der GWAP in einem abgeschlossenen Raum (Safe Room) befindet, einer persönlichen Zutrittskontrolle durch das Personal unterliegt und sich nur Geräte in dem Raum befinden, die für den GWAP benötigt werden.

Für eine Remote Desktop Lösung ist zu prüfen, ob **insgesamt** die technischen und organisatorischen Maßnahmen mit hinreichender Wahrscheinlichkeit verhindern werden, dass sich a) **Unbefugte** Zugriff auf die Daten verschaffen, und dass b) **Befugte** gegen Datenschutzbestimmungen und sonstige Auflagen verstoßen. D.h. es müssen Maßnahmen ergriffen werden, die den Datenzugang kontrollieren, das zeitnahe Aufdecken von Fehlverhalten ermöglichen und unterbinden sowie ein Fehlverhalten sanktionieren.

Eine Maßnahme, um Fehlverhalten zu verhindern und ggf. zu sanktionieren, ist auch schon jetzt die **vertragliche Bindung** der wissenschaftlichen Institution und der Forschenden sowie ihre Verpflichtung auf die jeweils einschlägige Gesetzgebung. Die internationalen Beispiele haben gezeigt, dass Institutionen und Forschende vertraglich in die Pflicht genommen werden und Verstöße ggf. nicht nur zum Ausschluss vom Datenzugang des Forschenden, sondern auch der gesamten Institution

62 An dieser Stelle sei nochmals darauf hingewiesen, dass das Kostenmodell der Niederlande für diesen forschungsfreundlichen Zugang einen Preis von 1.700 Euro für ein neues Projekt vorsieht. Pro Datentyp kommen 170 Euro hinzu. Für die Outputprüfung fallen je nach Aufwand zwischen 105 und 210 Euro an (Stand Januar 2019).

führen können. Interessant hierbei ist, dass manche Ämter die Outputkontrolle entsprechend vorgegebener Regeln auf Einhaltung der statistischen Geheimhaltung an die Forschenden delegieren und bei Nichteinhaltung der Regeln die Forschenden sanktioniert werden können. D. h., in diesen Ländern geht man davon aus, dass die vertraglichen Bindungen inklusive Sanktionsmaßnahmen als Sozialisationsmechanismen greifen. Folglich ist für Deutschland eine zweckmäßige Anpassung der existierenden Verträge für einen Remote Desktop Zugang zu prüfen. Darüber hinaus ist zu erwägen, ob die Teilnahme an einer Datenschutz- und Datensicherheitsschulung als eine verpflichtende Voraussetzung vor Gewährung eines Remote Zugangs eingeführt werden sollte. Ergänzend ist vorstellbar, dass – wie im niederländischen Modell – dem eigentlichen Datenzugang jedes Mal aufs Neue ein zufallsgenerierter Test vorgeschaltet wird. Hintergrund dieser Erwägung ist, dass sich die Forschenden in einem solchen Szenario regelmäßig aktiv mit den Voraussetzungen des Datenzugangs beschäftigen und dies die Regeln stärker im Bewusstsein verankert, als die einmalige Unterzeichnung eines Vertrags.

→ Zugangswege S. 10

Ein Teil des Sicherheitskonzepts beim GWAP beruht darauf, den Zugriff auf Daten, die ein höheres Re-Anonymisierungsrisiko haben als ein Scientific Use File, nur nach **persönlicher Identitätskontrolle vor Ort (Zugangskontrolle)** zu erlauben.[63] Für den Remote Desktop muss hier auf andere Maßnahmen umgestiegen werden, denkbar wäre etwa ein biometrisches Verfahren vergleichbar dem beim LIfBi implementierten Zugang oder eine Lösung mit einem vom betreuenden FDZ zur Verfügung gestellten Fingerabdruckscanner. Für eine solche Lösung hat sich bspw. auch das französische Secure Data Access Centre (CASD) entschieden, das via Remote Desktop Forschenden im In- und Ausland Zugang zu Mikrodaten der französischen amtlichen Statistik und der Sozialversicherungsträger ermöglicht.[64]

Durch die **Arbeitssituation** am GWAP ist es den FDZ-Beschäftigten grundsätzlich möglich, vertragswidriges Verhalten zu erkennen. Aktuell sind am GWAP z. B. keine mobilen Endgeräte und bei einigen FDZ auch keine Schreibutensilien erlaubt. Diese Maßnahmen sollen gewährleisten, dass keine ungeprüften Materialien via Foto oder Papier den GWAP verlassen.[65] Des Weiteren soll verhindert werden, dass über die mobilen Endgeräte Zusatzinformationen im GWAP genutzt werden, die für eine Reidentifikation genutzt werden könnten. Dieses Maßnahmenbündel wird als hinreichend angesehen, um das erforderliche Sicherheitsniveau zu gewährleisten. Das derzeitige Mitnahmeverbot von Materialien und Geräten an den und vom GWAP hat das Ziel, den Kontroll- und Überwachungsaufwand für die FDZ-Beschäftigten zu reduzieren. Nicht die Mitnahme an den GWAP ist problematisch, sondern die vertragswidrige Verwendung. Es ist offensichtlich, dass diese Maßnahmen bei Remote Lösungen nur mit erheblichem technischem Aufwand nachgebildet werden könnten. Eine Kameraüberwachung ohne Aufzeichnung würde der bisherigen Sichtkontrolle am GWAP nahekommen, müsste jedoch dann auch FDZ-seitig stichprobenmäßig kontrolliert werden. Diesen Aufwand erachten Länder, die bereits Remote Desktop etabliert haben (vgl. Kapitel 3), für nicht erforderlich und stützen sich auf im jeweiligen Nutzungsvertrag fixierte und i.d.R. strafbewährte Untersagungen.

Geht man davon aus, dass Forschende das Risiko eines Vertragsbruchs in Form des Abschreibens/Fotografierens nicht auf Geheimhaltung geprüfter Outputs vermutlich nur deshalb in Erwägung ziehen könnten, damit sie ihre Arbeit an anderer Stelle ohne Zeitverzug fortsetzen können, entfällt die Motivation bei einem Remote Zugriff-Szenario, da die Analysen unabhängig von Öffnungszeiten und Lokalisation der FDZ in der eigenen Einrichtung oder komplett ortsunabhängig durchgeführt werden können. Hervorzuheben ist auch, dass Remote Execution Lösungen z. B. bei dem FDZ BA im IAB bereits jetzt Forschenden eine Sichtung von automatisch, mittels Scripts, geprüften Zwischenergebnissen ermöglichen, die deutlich als solche gekennzeichnet sind. Dadurch erhalten die Forschenden die Möglichkeit, schnell erste und vorläufige Ergebnisse und Datenaufbereitungsschritte zu prüfen, um am Ende nur noch die relevanten Ergebnisse für Publikationen und Präsentationen, die manuell geprüft werden, zu erhalten. Technisch unterbunden werden zudem ausgewählte Befehle im Statistik-

63 Siehe auch BSI IT-Grundschutz M 2.17: https://www.bsi.bund.de/DE/Themen/ITGrundschutz/ITGrundschutzKataloge/Inhalt/_content/m/m02/m02017.html (Zugriff am15.02.2019)

64 Vgl. https://www.casd.eu/en/le-centre-dacces-securise-aux-donnees-casd/le-casd (Zugriff am 09.09.2019)

65 Bei dem derzeitigen Verfahren wird übersehen, dass schon jetzt bei entsprechender krimineller Energie die Absicherung unentdeckt unterlaufen werden kann, da kleinste Aufnahmegeräte und „Datenbrillen“ auf dem Markt erhältlich sind. Die Möglichkeiten der wissenschaftlichen Verwendung von Daten und Ergebnissen, die auf diese Art und Weise erschlichen wurden, ist jedoch mehr als fraglich.

paket, um die Ausgabe von Einzelinformationen zu verhindern und das Herunterladen vorläufiger nur per automatischen Skript geprüfte Ergebnisse durch die Forschenden (Eberle 2017). D.h. ähnlich wie im Beispiel der nordischen Länder, geht das FDZ BA im IAB von einem rechtskonformen Verhalten der Forschung aus. Wenn dies bei Remote Execution möglich ist, ist es analog auch für den Remote Access vorstellbar. Technisch ist das Anspielen von Daten durch die Forschenden unterbunden, die Bereitstellung erfolgt immer über die FDZ-Beschäftigten. Zudem würden Verstöße gegen die Vertragsbedingungen auch bei der FDZ-seitigen Prüfung der Ergebnisse auffallen. Folgt man insgesamt dieser Argumentation und zieht ebenfalls die inzwischen mehr als 15-jährige Erfahrung in Betracht, könnte sich eine über die Protokollierung aller Arbeitsschritte hinausgehende technische Überwachung als obsolet erweisen.

Basierend auf diesen Überlegungen werden im folgenden Kapitel mögliche Szenarien für die Operationalisierung eines Remote Access skizziert.

5.2 Zugangsszenarien

Für den Remote Desktop lassen sich räumlich drei Grundszenarien unterscheiden, die hier natürlich nur prinzipiell beschrieben und nicht bis ins Detail[66] ausgearbeitet werden können.

1) Remote Desktop in einem dedizierten Raum in einer wissenschaftlichen Einrichtung für mehrere Forschende, wobei mehrere Ausstattungsvarianten denkbar sind[67]
2) Remote Desktop innerhalb der wissenschaftlichen Einrichtung am eigenen Arbeitsplatz
3) Remote Desktop ortsunabhängig innerhalb von Deutschland

Die Implementierung des Zugriffs auf einen entfernten virtuellen Arbeitsplatz wird von der weiteren Betrachtung ausgeklammert, da auf langjährig erprobte BSI-konforme Standards zurückgegriffen werden kann (BSI 2010; 2014). Forschende, die sich über eine gesicherte und verschlüsselte Verbindung über einen Remote Access Client an der jeweiligen FDZ-Infrastruktur angemeldet haben, befinden sich an einem Arbeitsplatz, welcher ihnen alle Optionen wie ein GWAP bietet, wo weder Export noch Import – also auch kein Copy/Paste der Daten oder das Zuspielen externer Datenquellen – möglich ist. Die Daten verbleiben auf dem sicheren Server des Datenproduzenten, auf das Zugangsgerät des Forschenden werden gewissermaßen nur ‚Bilder' geschickt (Schiller und Welpton 2013). Für einen solchen virtuellen FDZ-Desktop existieren z. B. bei dem IAB bereits technische Spezifikationen und Implementierungen.

Bei **Szenario 1** – dem Remote Desktop in dedizierten Räumen (Cluster/Pool) –, die für mehrere Forschende zur Verfügung stehen, kann die Zutrittskontrolle z. B. über personalisierte ID-Karten, PIN-Eingabe und/oder biometrische Einlassmechanismen[68] realisiert werden. Falls als erforderlich erachtet, könnten als zusätzliche Identitätskontrollen randomisierte Abfragen biometrischer Merkmale während einer Arbeitssitzung eingeführt werden. Für die Datenanbieter besteht nach IT-Grundschutz die Pflicht, die Einhaltung der Zutrittsregelungen wirksam zu kontrollieren, was jedoch nicht mit einer Dauerüberwachung gleichzusetzen ist.

Eine Zutrittskontrolle bei **Szenario 2** kann z. B. über Türschlüssel oder PIN-Eingabe realisiert werden. Vertraglich wäre zu regeln, dass bei Abwesenheit die Türen verschlossen zu halten sind. Da es sich um die normalen Arbeitszimmer der Forschenden handelt, können sich natürlich auch Personen, die nicht im Rahmen des FDZ-Projekts verpflichtet, also in diesem Sinne Unbefugte sind, im Raum aufhalten. Das Risiko der Einsichtnahme von Daten und Ergebnissen auf dem Bildschirm durch Unbefugte kann durch sachgerechte Bildschirmpositionierung, Sichtwinkelumschaltungen bzw. entsprechende Sichtschutz-Folien reduziert werden. Wie oben bereits für die Pool-Lösung erörtert,

66 Ein solches Detail wäre z. B., dass alle Bildschirme mit einer Sichtschutzfolie versehen werden.

67 A) Nachbildung eines FDZ GWAP, d.h. alle benötigten Geräte und das Equipment zur Umsetzung des Remote Access Sicherheitskonzepts werden bereitgestellt und werden über ein dediziertes Netzsegment zum FDZ geroutet. B) Es werden Docking Stationen bereitgestellt und die Rechner sind von den Forschenden mitzubringen.

68 Bspw. müssen sich im französischen Modell die Forschenden beim Remote Access über ihre Fingerabdrücke authentifizieren. Als zusätzliche Kontrollmaßnahme könnten auch Zutrittszähler installiert werden, damit nachvollziehbar ist, wie viele Personen sich jeweils im Raum befinden.

reduzieren Arbeitssession begleitende biometrische Abfragen das Risiko des unbefugten Zugriffs bei unerlaubtem Zutritt bei Abwesenheit des Forschenden. Ergänzend könnte das Sicherheitsniveau über für den FDZ-Zugang-dedizierte abschließbare Ethernetdosen in diesen Räumen angehoben werden. Während es bei Szenario 1 noch denkbar wäre, die Pools mit FDZ-eigenen Zugangsgeräten auszustatten, würden Forschende am eigenen Arbeitsplatz ihre eigene Technik nutzen. Das Sicherheitsniveau kann hier durch die FDZ-seitige automatisierte Abfrage von eindeutig identifizierenden Geräteparametern sowie eine Multifaktor-Authentifizierung erhöht werden (M 4.441 IT-Grundschutz). Zudem können die Forschenden verpflichtet werden, für den Zugriff benötigte Zusatzgeräte sicher vor Unbefugten verschlossen aufzubewahren.

Bei **Szenario 3** entfallen die physischen auf Räumlichkeiten und IT-Infrastruktur der jeweiligen Einrichtung bezogenen Maßnahmen. Voraussetzung sind die Möglichkeit, das genutzte Gerät FDZ-seitig eindeutig identifizieren zu können sowie eine Zwei-Wege-Authentifizierung. Auch hier ist die randomisierte Abfrage biometrischer Parameter während der Arbeitssession ein Mittel der Wahl. Auflagen zur Arbeitsumgebung müssen vertraglich detailliert fixiert werden, z. B. könnte das Arbeiten in der Bahn und öffentlichen Räumen untersagt werden. Zu berücksichtigen ist ebenfalls, dass nicht alle über FDZ verfügbaren Daten außerhalb von Deutschland genutzt werden dürfen, so z. B. die Daten der amtlichen Statistik. In diesen Fällen muss technisch feststellbar sein, wo sich das Endgerät befindet. Das heißt aber auch, dass eine VPN-Einwahl in einer sich nicht in Deutschland befindlichen wissenschaftlichen Einrichtung und eine dann erfolgende Einwahl in das FDZ-Netz als solche erkannt und ggf. unterbunden werden muss. Darüber hinaus könnte auch die VPN-Einwahl in das Netz der wissenschaftlichen Einrichtung und das darauf erfolgende Routing zur FDZ-Infrastruktur eine weitere Maßnahme zur Erhöhung der Zugangssicherheit sein.

Zusammenfassend lassen sich die drei Szenarien auch als eine Umschichtung von Verantwortlichkeiten sehen: Sicherheitsmaßnahmen, die am GWAP aufgrund der Ortsgebundenheit und des Personals umsetzbar waren, werden durch eine Erweiterung der Maßnahmen am Zugangsgerät und der Zugangs-Infrastruktur abgelöst. Die wissenschaftlichen Einrichtungen werden bei allen Remote Szenarien stärker als bisher in die Verantwortung einbezogen, haben aber den Gewinn, dass ihre Forschenden effektivere Arbeitsbedingungen bekommen – so wird technisch ein 24/7 Betrieb FDZ-seitig grundsätzlich möglich sein. Allerdings muss klar kommuniziert und von der Forschung auch akzeptiert werden, dass ein Support und insbesondere die Outputkontrolle nur innerhalb der regulären Arbeitszeiten[69] der FDZ erfolgt. Die Forschenden zahlen für einen Remote Zugang den Preis, dass sie Kontrollmaßnahmen während ihrer Arbeit, wie etwa den Einsatz biometrischer Verfahren, und ggf. technische Anpassungen im Umfeld ihres Arbeitsplatzes akzeptieren müssen, ansonsten bleibt nur wie bisher das Arbeiten am GWAP.

69 Beispielhaft beschränkt sich die Beratung beim Remote Access System (Mona) von Statistics Sweden auf Werktage und die Zeiten 8:00-11:30 und 12:30-16:00 Uhr. Siehe https://www.scb.se/en/services/guidance-for-researchers-and-universities/mona--a-system-for-delivering-microdata (Zugriff am 21.1.2019). Den Nutzenden eines Remote Access wäre zu verdeutlichen, dass spezifische Funktionen wie z. B. erstmalige Authentifizierung und die Verpflichtung auf die Einhaltung der relevanten Gesetze nur an Werktagen innerhalb eines definierten Zeitrahmens nutzbar sind.

6 Zusammenfassung und Empfehlungen

Das Arbeiten über Remote Desktop mit Daten der amtlichen Statistik sowie der Sozialversicherungsträger ist international kein Neuland. Das in anderen Ländern und bei einigen wenigen deutschen FDZ vorliegende Erfahrungswissen kann genutzt werden, um für Deutschland Implementierungsszenarien zu entwickeln. Durch die Etablierung eines Remote Desktop Zugangs profitieren die Wissenschaft wie auch die Datenproduzenten und das hohe Analysepotential der Daten kann besser als bisher ausgeschöpft werden.

Aufgrund fachspezifischer Gesetze, denen die bereitzustellenden Mikrodaten unterliegen, können die Remote Desktop Szenarien unterschiedliche Ausprägungen haben – wünschenswert wäre natürlich eine weitgehend einheitliche Vorgehensweise. Das FDZ des IAB könnte in Kooperation mit den anderen Anbietern eine Vorreiterrolle übernehmen, da die Etablierung eines Remote Desktops für diesen Datenbestand vermutlich keiner Gesetzesänderung bedarf und somit in naher Zukunft mit einem Pilotprojekt begonnen werden könnte.

Beim Remote Desktop Zugang sollten die bereitgestellten Daten grundsätzlich das gleiche Analysepotential haben wie derzeit die Daten am GWAP, ansonsten würde neben den schon jetzt existierenden Zugangswegen nur ein weiterer geschaffen. Die langfristige Zielsetzung, dass Forschende wie auch schon in anderen Ländern Europas auf Gastwissenschaftsarbeitsplätze nur noch in Ausnahmefällen zurückgreifen müssen, würde nicht eingelöst.

Vorbehaltlich weiterer Prüfungen ist derzeit die Implementierung eines Remote Desktop Verfahrens für **formal anonymisierte** Daten der Sozialversicherungsträger und für **faktisch anonymisierte** Daten der amtlichen Statistik möglich. Während also für die Sozialversicherungsträger vermutlich kein gesetzlicher Änderungsbedarf erforderlich ist, um auch formal anonymisierte Mikrodaten Remote bereitzustellen, sehen die Vertreter der amtlichen Statistik einen gesetzlichen Regelungs- bzw. Klarstellungsbedarf. Neben den Wissenschaftlerinnen und Wissenschaftlern sollten daran letztlich auch die forschungsfördernden Institutionen sowie das Bundesministerium für Forschung ein großes Interesse haben, die Forschungsbedingungen für deutsche Wissenschaftlerinnen und Wissenschaftler zu verbessern und im europäischen Wettbewerb hinsichtlich der Datenzugänglichkeit aufzuholen.

Empfehlung 1

Pilotprojekte zur Implementierung

Es ist wünschenswert, dass sich die forschungsfördernden Institutionen für die Verbesserung des Forschungsstandortes Deutschland einsetzen und Pilotprojekte für eine rechtssichere und hochverfügbare Implementierung unterstützen.

Empfehlung 2

Schaffung einer Rechtsgrundlage für Remote Desktop Access zu den Daten der amtlichen Bundesstatistik

Das Bundesministerium des Inneren sollte eine Änderung des §16 Abs. 6 Bundesstatistikgesetzes dahingehend einleiten, dass ein Remote Desktop zu formal anonymisierten Einzeldaten der amtlichen Statistik ermöglicht wird.

Hierbei kommt der Begründung des Gesetzestextes eine besondere Bedeutung zu, da im Gesetz selber die Operationalisierung der Implementierungsstandards nicht benannt werden, aber die Interpretationsspielräume möglichst klein gehalten werden sollten. In der Begründung muss näher dargelegt werden, wie der sichere Datenzugriff zu operationalisieren und wie zu evaluieren ist. Gleiches gilt für die wirksamen Vorkehrungen zur Wahrung der Geheimhaltung. Insgesamt muss das Sicherheitsniveau des Remote Desktop Zugangs vergleichbar hoch sein wie das des GWAP. Dabei sollte aber zugleich kritisch reflektiert werden, ob die Bedingungen des GWAP verhältnismäßig und in ihrer derzeitigen Implementierung technisch zeitgemäß sind. Die derzeitige sehr restriktive Umsetzung des Zugangs zu Mikrodaten in Deutschland steht dem berechtigten Anliegen gegenüber, dass Wissenschaft betrieben wird, um Erkenntnisse mit der Scientific Community und der Öffentlichkeit zu teilen und auch die Grundlage für auf empirischer Evidenz basierende Politikberatung ist.

Die Erfahrungen über nunmehr 16 Jahre seit Gründung der FDZ zeichnen insgesamt das Bild von verantwortungsvoll handelnden Wissenschaftlerinnen und Wissenschaftlern. Letztlich sind die wissenschaftlichen Institutionen die Vertragsschließenden und damit in der zentralen Verantwortung, missbräuchlichen Nutzungen vorzubeugen, da der Imageschaden für die Einrichtung bei Fehlverhalten ihrer Forschenden kaum zu überschätzen ist. Darüber hinaus sind Sanktionen in Form einer Sperrung des Datenzugangs für das ganze Institut, wie z. B. in den nordischen Ländern vertraglich vorgesehen, sehr effiziente Mittel dem Missbrauch vorzubeugen, denn letztendlich ist der Datenzugang die Grundlage ihrer Arbeit. Das alles schließt natürlich mögliches individuelles Fehlverhalten nicht aus. Insgesamt erscheint eine erneute, transparente und dezidierte Abwägung von Maßnahmen angezeigt, die bei der Gründung der FDZ, als mit der erstmaligen standardmäßigen Bereitstellung völliges Neuland betreten wurde, (noch) nicht möglich war.

Da es bei den Sozialversicherungsträgern vermutlich keiner gesetzlichen Änderung bedarf, um via Remote Access die Daten anzubieten, die auch am GWAP verfügbar sind, kann ein Pilotprojekt zur Bestimmung möglicher Szenarien begonnen werden. Im Hinblick auf eine möglichst generische Remote Desktop-Lösung sollte abgewogen werden, ob dies nicht auch unter Beteiligung der Statistischen Ämter im Rahmen eines Pilotprojekts erfolgen könnte, auch wenn dies aktuell nur mit faktisch anonymen Mikrodaten möglich wäre.

Empfehlung 3

Zeitnahe Umsetzung von Pilotprojekten zum Remote Access

In Kooperation sollten die Statistischen Ämter des Bundes und der Länder sowie die Sozialversicherungsträger, die Wissenschaft und der Datenschutz in einem Pilotprojekt unter Berücksichtigung der internationalen Erfahrungen Remote Desktop Szenarien entwickeln und unter Verwendung existierender und anhand der jeweiligen aktuellen Rechtslage hinreichend anonymisierter Datenbestände erproben.

Literaturverzeichnis

Alstadsaeter, Anette; Niels Johannesen und Gabriel Zucman (2018): „Tax Evasion and Inequality", Working Paper, https://gabriel-zucman.eu/files/AJZ2017.pdf (Zugriff am 04.04.2019)

Alexander, J. Trent; Micheal Davern und Betsey Stevenson (2010): The polls review: inaccurate age and sex data in the census PUMS files: evidence and implications. Public Opinion Quarterly, 74 (3), 551-569, https://doi.org/10.1093/poq/nfq033 (Zugriff 07.10.2019)

Bujnowska, Aleksandra (2017): Item 5.3 New modes of access: remote execution, remote access. Presentation at 'Microdata Access Network Group Meeting'. 20 June 2017. https://ec.europa.eu/eurostat/cros/system/files/5_3_new_modes_of_access.pdf (Zugriff am 12.09.2019)

BSI (2010): BSI-Studie zur Internet-Sicherheit (ISi-S). Sicherer Fernzugriff auf das interne Netz. https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Internetsicherheit/isi_fern_studie_pdf.pdf?__blob=publicationFile&v=2 (Zugriff am 01.11.2018).

BSI (2014): IT-Grundschutz-Kataloge. 14. Ergänzungslieferung – 2014. https://download.gsb.bund.de/BSI/ITGSK/IT-Grundschutz-Kataloge_2014_EL14_DE.pdf (Zugriff am 20.09.2019).

Drechsel-Grau, Moritz; Andreas Peichl und Kai Daniel Schmid (2015): Einkommensverteilung und gesamtwirtschaftliche Entwicklung – eine Erwiderung. Wirtschaftsdienst. 95/12, 864-867

Eberle, Johanna; Dana Müller und Jörg Heining (2017): A modern job submission application to access IAB's confidential administrative and survey research data. FDZ-Methodenreport, 01/2017 (en), Nürnberg. http://doku.iab.de/fdz/reporte/2017/MR_01-17_EN.pdf (Zugriff am 28.03.2019)

Fuß, Daniel und Knut Wenzig (2019): The Research Data Center: Making National Educational Panel Study Data Available for Research. In: Blossfeld, H.-P.; Jutta von Maurice, Michael Bayer und Jan Skopek (Hrsg.): The German National Educational Panel Study (NEPS). Second Revised Version 361–378,. Wiesbaden, Germany, Springer Fachmedien Wiesbaden.

Hochfellner, Daniela; Dana Müller; Alexandra Schmucker und Elisabeth Roß (2012): Datenschutz am Forschungsdatenzentrum. FDZ-Methodenreport, 06/2012 (de), Nürnberg, http://doku.iab.de/fdz/reporte/2012/MR_06-12.pdf (Zugriff am 28.03.2019)

Kvalheim, Vigdis und Marianne Høgetveit Myhren (2017): New legislation – a unique opportunity for harmonising the legal framework for research in the Nordic countries, Norwegian Centre for Research Data (NSD). http://www.nsd.uib.no/personvernombud/dok/position-paper-new-legislation.pdf (Zugriff am 01.11.2018).

Löffler, Max; Andreas Peichl, Christian Wittneben und Carina Neisser (2015): Möglichkeiten zur Verbesserung der statistischen Datengrundlage zur Beschreibung höchster Einkommen und Vermögen, Bundesministerium für Arbeit und Soziales, https://www.zew.de/PU70495 (Zugriff am 10.05.2019).

Isungset, Martin Arstad; Mats Lillehagen und Elisabeth Ugreninov (2019): One Order Fits All? Birth Order and Education in Immigrant Families, European Sociological Review, jcz040, https://doi.org/10.1093/esr/jcz040 (Zugriff 07.10.2019)

Müller, Dana und Joachim Möller (2019): Giving the International Scientific Community Access to German Labor Market Data: A Success Story. In: Nuno Crato und Paolo Paruolo (Hrsg.) Data-Driven Policy Impact Evaluation. Cham, Schweiz, Springer Open. 101-117, https://doi.org/10.1007/978-3-319-78461-8_7 (Zugriff am 07.10.2019)

Schiller, David; Johanna Eberle, Daniel Fuß, Jan Goebel, Jörg Heining, Tatjana Mika, Dana Müller, Frank Röder, Michael Stegmann und Karsten Stephan (2017): Standards des sicheren Datenzugangs in den Sozial- und Wirtschaftswissenschaften: Überblick über verschiedene Remote Access Verfahren, RatSWD Working Paper 261/2017. Berlin, Rat für Sozial- und Wirtschaftsdaten. https://10.17620/02671.15 (Zugriff am 28.03.2019)

Schiller, David und Richard Welpton (2014): Distributing Access to Data, not Data - Providing Remote Access to European Microdata. IQ (IASSIST Quarterly) 38 (3), 6-14, https://iassistquarterly.com/pdfs/iqvol38_3_schiller.pdf (Zugriff 07.10.2019)

Schiller, David und Richard Welpton (2013): Providing Remote Access to European Microdata. https://www.researchgate.net/publication/259900882_Providing_Remote_Access_to_European_Microdata/download (Zugriff 12.07.2019).

Skopek, Jan; Tobias Koberg und Hans-Peter Blossfeld (2016): RemoteNEPS – An Innovative Research Environment. In: Blossfeld, H.-P.; Jutta von Maurice, Michael Bayer und Jan Skopek (Hrsg.): Methodological Issues of Longitudinal Surveys. Wiesbaden, VS Verlag für Sozialwissenschaften, 611-626.

Smith, James. P. (1991): Data confidentiality: a researcher's perspective. In: Proceedings of the American Statistical Association, Social Statistics Section, 117-120. Alexandria, VA: American Statistical Association. https://econwpa.ub.uni-muenchen.de/econ-wp/lab/papers/0403/0403006.pdf (Zugriff am 09.10.2019).

Statistics Finland, Statistics Denmark, Statistics Iceland, Statistics Norway, Statistics Sweden, Statistics Greenland (2014): Feasibility Study Regarding Access to Nordic Microdata. https://simsam.nu/wp-content/uploads/2016/08/Feasibility-study-regarding-research-access-to-nordic-microdata.pdf (Zugriff am 12.04.2019).

Wirth, Heike (2016): Analytical Potential versus Data Confidentiality. Finding the Optimal Balance. In: Christof Wolf, Dominique Joye, Tom W. Smith und Yang-chih Fu (Hrsg.):The SAGE Handbook of Survey Methodology, 486-499. https://dx.doi.org/10.4135/9781473957893 (Zugriff am 08.10.2019)

Zühlke, Sylvia und Helga Christians (2005): Datenangebot und Datenzugang im Forschungsdatenzentrum der Statistischen Landesämter. Amtliche Mikrodaten für die Sozial- und Wirtschaftswissenschaften. Beiträge zu den Nutzerkonferenzen des FDZ der Statistischen Landesämter 2005. https://www.forschungsdatenzentrum.de/sites/default/files/nutzerkonferenzbeitr%C3%A4ge_2005_band_I.pdf (Zugriff am 07.10.2019)

Anhang

Anhang 1: Ansätze für die Identitätskontrolle bei Remote Access-Verfahren

Im Folgenden sind unterschiedliche Authentisierungsmethoden, die für den Remote Access genutzt werden könnten, aufgelistet. Details zu den Nummern 1, 2 und 4 finden sich z. B. in BSI (2010, 17–19; BSI 2014).

1. Einmalpasswörter

- Zugangscodes, die nur für eine einmalige Authentisierung gültig sind: (Funktionsweise analog zu den z. B. im Internet-Banking genutzten Transaktionsnummern (TANs))

2. 2-Faktor- Authentisierung

- Für die Identitätskontrolle sind zwei Faktoren aus den folgenden Bereichen nötig:
  - Wissen (z. B. PIN, Passwort, TAN)
  - Besitz (z. B. Smartcard oder Hardware-Token (elektronischer Schlüssel))
  - Biometrisches Merkmal (z. B. Fingerabdruck, Keystroke-Erkennung, Gesichtserkennung oder Augenhintergrund)

- Die 2-Faktor-Authentisierung nutzt dann die folgenden Paare zur Identitätskontrolle (ggf. kann noch ein Standard-Login vorgeschaltet werden → Multi-Faktor-Authentisierung):
  - Wissen und Besitz
  - Wissen und biometrisches Merkmal
  - Besitz und biometrisches Merkmal

3. Weitere auf Schlüsseln basierende Verfahren (Details s. BSI (2010, 17–19))

- Statische Schlüssel
- Public-Key-Verfahren (Zertifikate)
- Challenge-Response
- Pre-Shared Keys (PSK, Shared Secret)

Anhang 2: Kennzahlen und Standorte der Forschungsdatenzentren der amtlichen Statistik und der Sozialversicherungsträger in 2018

Abb. 4a:
Kennzahlen der Forschungsdatenzentren der amtlichen Statistik und der Sozialversicherungsträger

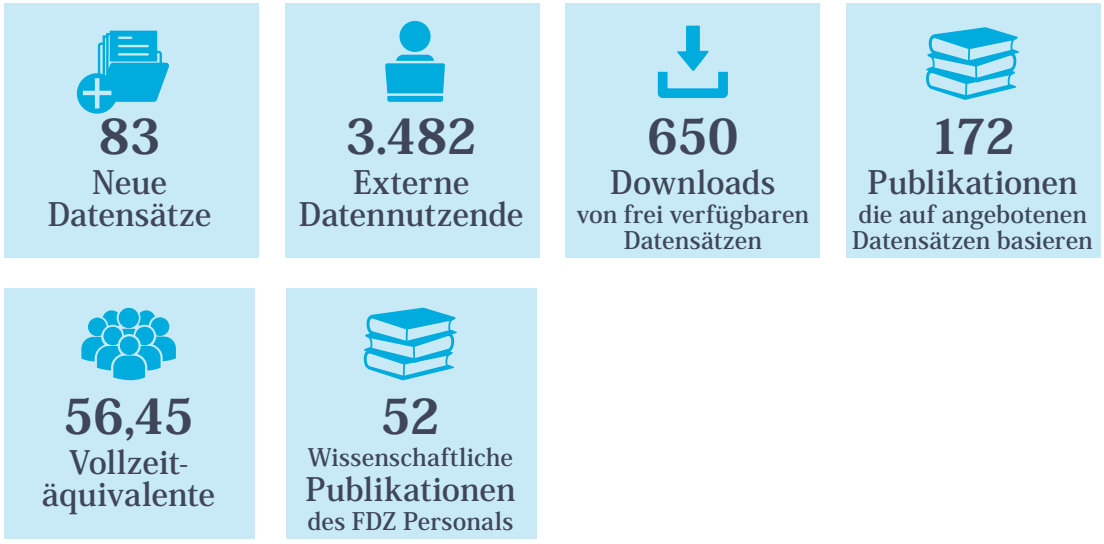
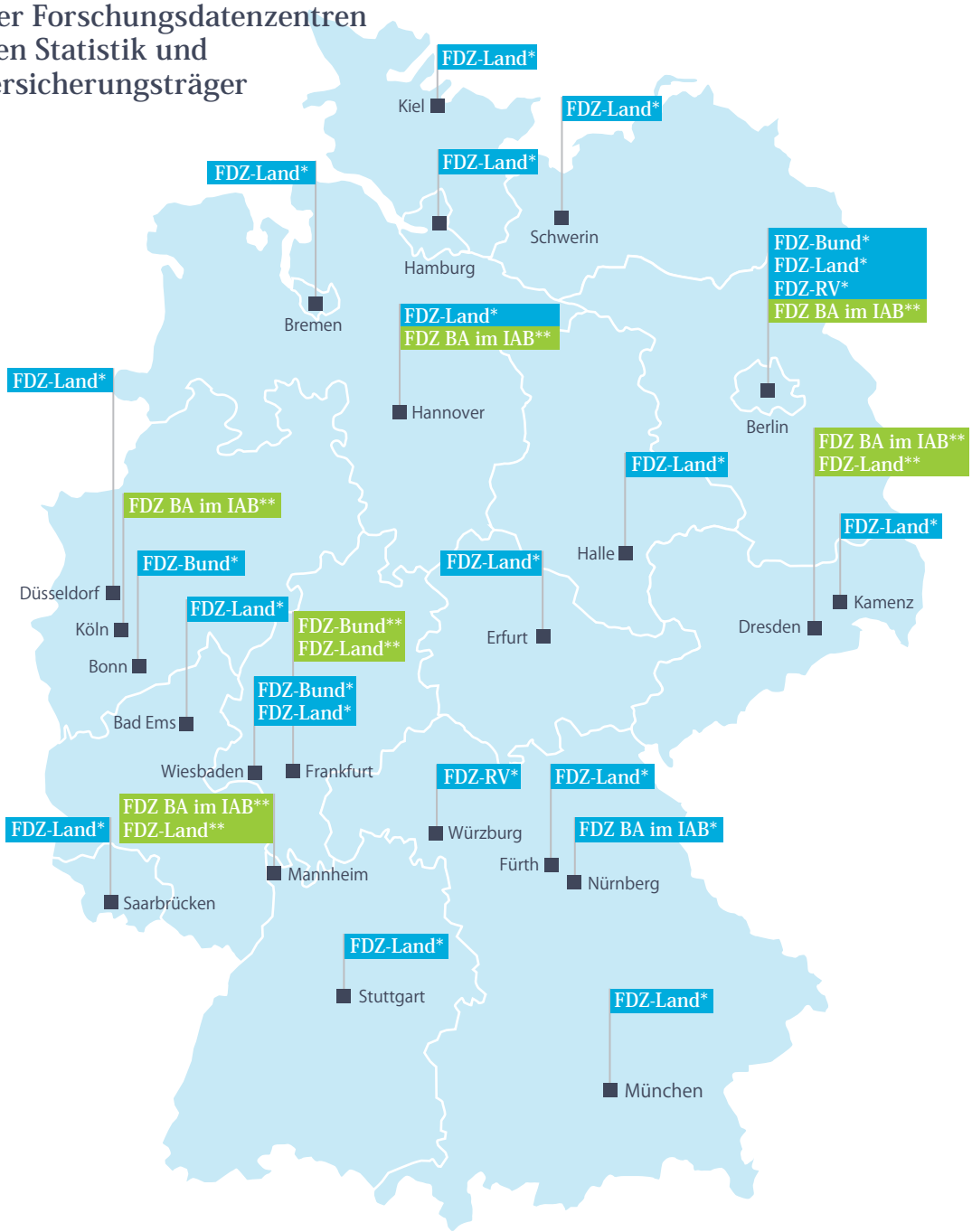
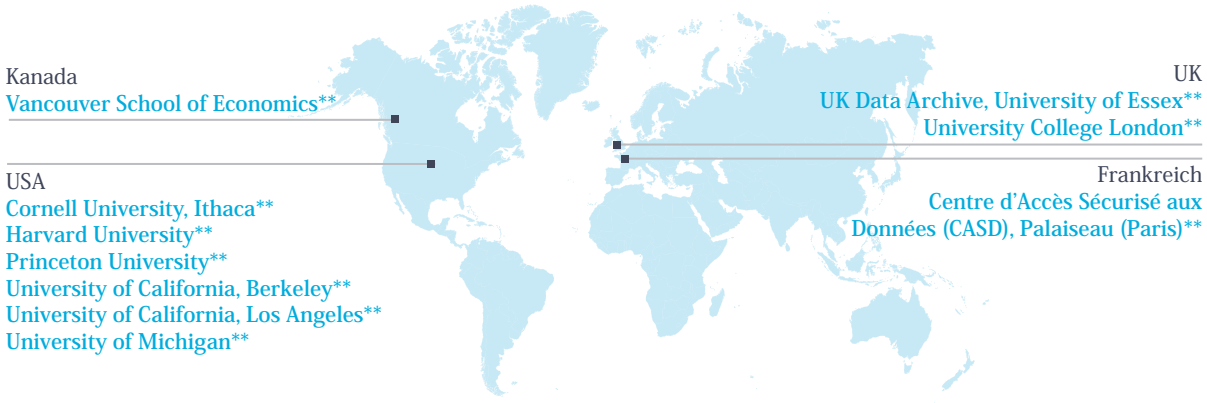


Abb. 4b:

Standorte der Forschungsdatenzentren der amtlichen Statistik und der Sozialversicherungsträger



Internationale Standorte (FDZ BA im IAB)



Anhang 3: Glossar / Definitionen

Campus File: Für die universitäre Lehre bestimmte und im Vergleich zu Scientific Use Files noch stärker anonymisierte Forschungsdatensätze.

Remote Access Client: Eine Software, die auf dem für die Nutzer zugänglichen Computer (Zugangspunkt) läuft und eine Verbindung zum Server herstellt.

Gastwissenschaftsarbeitsplatz (GWAP): Speziell gesicherte Arbeitsplätze in den FDZ, an denen Forschende auf sensiblere Daten zugreifen können. Typische Merkmale dieser Gastarbeitsplätze sind zum Beispiel, dass sie keinen unkontrollierten Netzzugang besitzen und ein lokales Abspeichern der Daten nicht möglich ist.

Mikrodaten: Daten aus statistischen Erhebungen, die sich direkt einzelne Erhebungsobjekte beziehen bzw. beziehbar sind, z. B. auf individuelle Personen oder einzelne Unternehmen. Diese Daten können besonders schützenswert sein, so dass entweder der Zugang für Forschende stärker kontrolliert werden muss oder vor der Zugänglichmachung geeignete (und rechtlich gebotene) Anonymisierungsverfahren anzuwenden sind.

Public Use File (PUF): So stark anonymisierte Forschungsdatensätze, dass keine Nutzungseinschränkungen bestehen und eine Weitergabe der Daten auch außerhalb der wissenschaftlichen Forschung möglich ist.

Sozialdaten: Gemäß SGB X §67 Absatz 2 Satz 1 sind Sozialdaten „personenbezogene Daten (Artikel 4 Nummer 1 der Verordnung (EU) 2016/679), die von einer in §35 des Ersten Buches genannten Stelle im Hinblick auf ihre Aufgaben nach diesem Gesetzbuch verarbeitet werden."

Remote Access: Oberbegriff für den Zugriff von einem lokalen Computer (Client) auf einen entfernten Computer (Server). Die Kommunikation zwischen Client und Software kann über eine verschlüsselte Verbindung erfolgen. Hier verstanden als der Zugriff aus der Ferne auf einen Server der Analysesoftware und Forschungsdaten bereitstellt.

Remote Desktop: Mit Hilfe eines Remote Access' wird die Benutzeroberfläche des Servers auf den Bildschirm des lokalen Clients übertragen. Hier würde dies bedeuten, dass Forschende auf dem Server die Analysesoftware starten und mit den Forschungsdaten arbeiten als würden sie lokal vorliegen (z. B. browsen). Über eine geeignete Konfiguration kann verhindert werden, dass die Daten auf dem lokalen Client gespeichert werden können. Das lokale Zugangsgerät dient lediglich dazu, mit dem Datenserver zu kommunizieren.

Remote Execution: Datenzugangsweg, bei dem die Forschenden die Daten und die Zwischenergebnisse nicht unmittelbar einsehen können, sondern ein Analyseskript schreiben, das über eine Schnittstelle an das FDZ übermittelt wird. Diese Skripte schreiben die Forschenden gewissermaßen blind an ihren eigenen Arbeitsplätzen (mit Hilfe eines eventuell verfügbaren Strukturdatensatzes, können die Syntax aber nicht die Inhalte geprüft werden). Sie übermitteln die Analyseskripte an das FDZ, das diese auf die Originaldaten anwendet. Abhängig von der Datensensibilität und/oder Rechtsgrundlage werden die Ergebnisdateien im FDZ geprüft (Outputkontrolle), bevor sie an die Forschenden übermittelt werden.

Scientific Use File (SUF): Forschungsdatensätze die faktisch anonymisiert sind, aber dennoch ein hohes Analysepotenzial bieten.

Mitwirkende bei der Erstellung

Mitglieder der AG:

Dr. Heike Wirth *(Vorsitz der AG)*
GESIS – Leibniz-Institut für Sozialwissenschaften, RatSWD

Prof. Dr. Ulrike Rockmann *(Ko-Vorsitz der AG)*
Senatsverwaltung für Inneres und Sport des Landes Berlin

Dana Müller
Institut für Arbeitsmarkt- und Berufsforschung

Dr. Jan Goebel
SOEP am Deutschen Institut für Wirtschaftsforschung Berlin

Tatjana Mika
Deutsche Rentenversicherung Bund

Beratend einbezogen:

Hans-Josef Fischer
Landesbetrieb Information und Technik Nordrhein-Westfalen (IT.NRW), RatSWD

Heike Habla
Statistisches Bundesamt, RatSWD

Bertram Raum
Experte für Datenschutzrecht

RatSWD Geschäftsstelle:

Marie Bormann

Dr. Mathias Bug

Dr. Tim Deeken

Impressum

Herausgeber:
Rat für Sozial- und Wirtschaftsdaten (RatSWD)
Rungestr. 9
10179 Berlin
office@ratswd.de
https://www.ratswd.de

Redaktion:
Marie Bormann, Dr. Mathias Bug

Gestaltung/Satz:
Claudia Kreutz

Icons:
made by Freepik from https://www.flaticon.com

Berlin, November 2019

RatSWD Output:
Die RatSWD Output Series dokumentiert die Arbeit des RatSWD in seiner 6. Berufungsperiode (2017–2020). In ihr werden seine Stellungnahmen und Empfehlungen veröffentlicht und auf diesem Weg einer breiten Leserschaft zugänglich gemacht.

Das diesem Bericht zugrunde liegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung (BMBF) unter dem Förderkennzeichen 01UW1802 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt, sofern nicht anders ausgewiesen, beim RatSWD.

doi: 10.17620/02671.42

Zitationsvorschlag:

RatSWD [Rat für Sozial- und Wirtschaftsdaten] (2019): Remote Access zu Daten der amtlichen Statistik und der Sozialversicherungsträger. RatSWD Output 5 (6). Berlin, Rat für Sozial- und Wirtschaftsdaten (RatSWD). https://doi.org/10.17620/02671.42.

Der Rat für Sozial- und Wirtschaftsdaten (RatSWD) berät seit 2004 die Bundesregierung und die Regierungen der Länder in Fragen der Forschungsdateninfrastruktur für die empirischen Sozial-, Verhaltens- und Wirtschaftswissenschaften. Im RatSWD arbeiten acht durch Wahl legitimierte Vertreterinnen und Vertreter der sozial-, verhaltens- und wirtschaftswissenschaftlichen Fachdisziplinen mit acht Vertreterinnen und Vertretern der wichtigsten Datenproduzenten zusammen.

Er versteht sich als institutionalisiertes Forum des Dialoges zwischen Wissenschaft und Datenproduzenten und erarbeitet Empfehlungen und Stellungnahmen. Der RatSWD engagiert sich für eine Infrastruktur, die der Wissenschaft einen breiten, flexiblen und sicheren Datenzugang ermöglicht. Solche Daten werden von staatlichen, wissenschaftsgetragenen und privatwirtschaftlichen Akteuren bereitgestellt. Der RatSWD hat 34 Forschungsdatenzentren akkreditiert, deren Kooperationen er fördert.

www.ratswd.de