

Lisa Hönegger

Das Teilen und Archivieren von Daten in den Sozialwissenschaften

Handbuch Repositorienmanagement, Hg. v. Blumesberger et al., 2024, S. 549–572

<https://doi.org/10.25364/978390337423229>



Dieses Werk ist lizenziert unter einer Creative Commons Namensnennung 4.0 International Lizenz, ausgenommen von dieser Lizenz sind Abbildungen, Screenshots und Logos.

Lisa Hönegger, Universität Wien, Universitätsbibliothek, lisa.hoenegger@univie.ac.at |

ORCID iD: 0000-0001-6530-7343

Zusammenfassung

Die Ansprüche der Open-Science-Bewegung hin zu FAIRen Forschungsdaten stellen Forschende der Sozialwissenschaften und Archive vor eine große Herausforderung: der Umgang mit personenbezogenen Daten und die daraus entstehende Spannung zwischen Datenteilen und Datenschutz. Dieser Beitrag setzt sich mit dem Bedarf an Forschungsdatenmanagement (FDM)-Services im sozialwissenschaftlichen Bereich auseinander und thematisiert unterschiedliche Maßnahmen, die das Datenteilen unter Einhaltung rechtlicher und forschungsethischer Anforderungen möglich machen. Anhand von allgemeinen datenschutzrechtlichen Grundlagen, dem internationalen Vergleich von FDM-Praxen und den etablierten Prozessen bei AUSSDA – The Austrian Social Science Data Archive wird dargelegt, wie Datenteilen und Datenschutz vereinbar sind.

Schlagwörter: Sozialwissenschaft; Datenschutz; Open Science; Daten teilen; Datenarchiv; Forschungsdatenmanagement; personenbezogene Daten; sensible Daten; Österreich

Abstract

Sharing and Archiving Data in the Social Sciences

The demands of the open-science movement towards FAIR research data pose a major challenge to researchers in the social sciences and to archives: the handling of personal data and the resulting tension between data sharing and data protection. This contribution addresses the need for research data management (RDM) services in the social sciences and addresses different measures that make data sharing possible while complying with legal and research ethics requirements. Based on general data protection principles, an international comparison of RDM practices, and the established processes at AUSSDA – The Austrian Social Science Data Archive, this contribution shows how data sharing and data protection are compatible.

Keywords: Social science; data protection; open science; data sharing; data archive; research data management; personal data; sensitive data; Austria

1. Einleitung

Die Anforderung an Forschende, Forschungsdaten über den gesamten Forschungsprozess adäquat zu managen, steigt kontinuierlich an – Forschungsdatenmanagement (FDM) etabliert sich als Teil des wissenschaftlichen Prozesses und der guten wissenschaftlichen Praxis. Ein für viele Forschende dabei relativ neuer Aspekt ist neben der Publikation der Forschungsergebnisse auch die Veröffentlichung der zugrundeliegenden Forschungsdaten, welcher von immer mehr Fördergebern durch dezidierte Fördervorgaben zu Open bzw. FAIR¹ Data verpflichtend wird. Auch wissenschaftliche Fachzeitschriften können mittels Publikationsrichtlinien ein ähnliches Ziel verfolgen, ebenso wie die eigene Forschungsinstitution beispielsweise durch eine Policy zu Forschungsdatenmanagement. Daten sollen FAIR – also auffindbar, zugänglich, interoperabel und nachnutzbar – gemacht werden, um so für mehr Transparenz in der Wissenschaft zu sorgen und eine Nachnutzung der Daten durch andere Forschende zu ermöglichen. Der österreichische Fonds zur Förderung der wissenschaftlichen Forschung (FWF) inkludiert in seiner Open-Access Policy die Verpflichtung zu „Open Access to Research Data“ und die FAIR Prinzipien.² Auch die erst kürzlich verabschiedete Open-Science Policy Austria³ legt den Fokus darauf, dass Forschungsdaten im Einklang mit den FAIR-Prinzipien verfügbar gemacht werden sollen. Die Deutsche Forschungsgemeinschaft (DFG) hat die Forderung der Veröffentlichung bzw. Zugänglichmachung von Forschungsdaten bereits in den Kodex „Leitlinie zur Sicherung guter wissenschaftlicher Praxis“ aufgenommen und spezifiziert, dass „soweit dies möglich und zumutbar ist, die den Ergebnissen zugrunde liegenden Forschungsdaten, Materialien und Informationen [...] verfügbar zu machen“ sind.⁴ Auch auf europäischer Ebene werden offene Wissenschaft und offene bzw. FAIRe Daten gefordert, u. a. von der Europäischen Kommission in „The EU’s open science policy“.⁵

Um diesen Ansprüchen gerecht zu werden, sind Forschende auf Repositorien bzw. Archive angewiesen, die die nötigen Infrastrukturen liefern, um Daten nicht nur langfristig und sicher zu verwahren, sondern diese auch zu veröffentlichen. Neben der Unterstützung, um Daten entsprechend den FAIR-Prinzipien aufzubereiten,

1 FAIR Principles von GO FAIR: <https://www.go-fair.org/fair-principles/>

2 Siehe <https://www.fwf.ac.at/en/about-us/what-we-do/open-science/open-access-policy/open-access-policy-for-research-data>

3 BMBWF, BMDW, BMK: Open Science Policy Austria: <https://www.bmbwf.gv.at/Themen/HS-Uni/Hochschulgovernance/Leitthemen/Digitalisierung/Open-Science/Open-Science-Policy-Austria.html>

4 Deutsche Forschungsgemeinschaft (2019), S. 9.

5 https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/open-science_en

brauchen Forschende vor allem auch Informationen und Beratung zu den rechtlichen Aspekten der Datenveröffentlichung. Hier treten Fragen in den Vordergrund, ob, in welcher Form oder wie lange die Daten überhaupt veröffentlicht werden dürfen bzw. welchem Personenkreis unter welchen Bedingungen und mit welchen Nachnutzungsrechten die Daten zur Verfügung gestellt werden können. Genau dieser Schritt der Veröffentlichung weckt erwartungsgemäß viele Unsicherheiten bei Forschenden, die sich mit komplexen rechtlichen und auch ethischen Fragestellungen auseinandersetzen müssen. Besonders, wenn es sich bei den Forschungsdaten um personenbezogene Daten handelt, ist zusätzliche Sorgfalt geboten, damit der Datenschutz gewahrt wird.

Einerseits gibt es also verschiedenste, auch rechtlich relevante Anforderungen, die Forschende zum Teilen der Daten verpflichten. Andererseits ist es oft gar nicht so leicht festzustellen, wie oder ob Daten überhaupt geteilt werden dürfen. Die Gründe gegen das Teilen von Daten können vielfältig sein und je nach Datenart und -inhalt variieren, z. B. um Urheberrecht, Nutzungsrechte oder potenzielle Verwertungsinteressen nicht zu verletzen. Diese Aspekte gilt es zwar bei allen Veröffentlichungen zu beachten, in diesem Beitrag liegt der Fokus aber auf dem Umgang mit personenbezogenen Daten und den datenschutzrechtlichen Einwänden bei Veröffentlichungen in den Sozialwissenschaften. Sozialwissenschaftliche Daten sind häufig personenbezogene Daten – wie zum Beispiel Umfragedaten, Interviewdaten, Beobachtungsdaten, aber auch Register- und Verwaltungsdaten⁶ – und ihre Verarbeitung ist dann auch datenschutzrechtlichen Gesetzen unterworfen, allen voran der Datenschutzgrundverordnung (DSGVO)⁷. Diese Daten stehen in diesem Beitrag im Fokus und dienen als Beispiel, auch wenn für personenbezogene Daten anderer wissenschaftlicher Disziplinen die gleichen datenschutzrechtlichen Regeln gelten und die hier behandelten Aspekte ebenfalls Gültigkeit haben können.

Die Ambivalenz von Open Data und Datenschutz entsteht nicht nur durch diverse rechtliche oder vertragliche Anforderungen, sondern bei der Veröffentlichung von Forschungsdaten stehen auch forschungsethische Interessen im Spannungsverhältnis miteinander. Forschende sind moralisch verpflichtet, ihre Forschungssubjekte (und damit auch die zugehörigen Daten) zu schützen:

6 Für einen Überblick zur Bandbreite von sozialwissenschaftlichen Daten siehe u. a. Perry, A.; Reker, J. (2018).

7 Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung) unter <https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:32016R0679&from=DE>. Siehe auch Leitfaden der österreichischen Datenschutzbehörde zur Verordnung unter <https://www.dsb.gv.at/recht-entscheidungen/gesetze-in-oesterreich.html>

Verstöße gegen den Datenschutz widersprechen der Handlungslogik und den Interessen wissenschaftlicher Forschung, da sie unethisch wären, zu einem Vertrauensverlust in der Bevölkerung führen und damit die Forschung von ihrer empirischen Basis abschneiden könnten.⁸

Andererseits ist es ein Anliegen, Zugriff zu vollständigen und umfassenden Forschungsdaten zu erhalten und Daten in demselben Zustand der Wissenschaft zur Verfügung zu stellen. Einige Maßnahmen, die grundsätzlich das Datenteilen und deren Nachnutzung ermöglichen, wie z. B. die Anonymisierung oder Pseudonymisierung⁹, lassen sich aber nur schwer mit dem vollständigen Erhalt der vorhandenen Informationen vereinbaren.¹⁰ Die Veröffentlichung von Daten in den Sozialwissenschaften erfordert daher in der Praxis oft auch eine Abwägung der Interessen der Beteiligten und jenen von Open Science bzw. des Erhalts des wissenschaftlichen Wertes von Forschungsdaten.

Mit der Diskussion von relevanten rechtlichen Grundlagen zum Datenschutz, international praktizierten Standards und den etablierten Prozessen bei AUSSDA – The Austrian Social Science Data Archive sollen folgende Fragen behandelt werden: Welche Aspekte gilt es grundsätzlich bei der Verarbeitung von personenbezogenen Daten zu beachten? Wie können Forschende bei der datenschutzkonformen Veröffentlichung von Forschungsdaten unterstützt werden? Wie können Repositorien selbst mit personenbezogenen Daten umgehen? In Kapitel 2 werden die theoretischen, rechtlichen Grundlagen zu personenbezogenen Daten erläutert, die den nachfolgend beschriebenen Prozessen zu Grunde liegen. Kapitel 3 beschäftigt sich mit einer Einordnung sozialwissenschaftlicher Daten in ihrer Sensibilität. Kapitel 4 stellt dar, wie international mit dem Begriff Anonymität umgegangen wird und wie sich diese Auslegungen auf die wissenschaftliche Praxis und auf das Forschungsdatenmanagement auswirken. Kapitel 5 beschreibt auf Basis all dieser Überlegungen die Prozesse und den Umgang mit personenbezogenen Forschungsdaten beispielhaft bei AUSSDA und zeigt, dass sozialwissenschaftliche Daten unter sorgfältiger Einhaltung datenschutzrechtlicher Vorgaben, ethischer Überlegungen und wissenschaftlicher Standards geteilt werden können.

8 Rat für Sozial- und Wirtschaftsdaten (RatSWD) (2015), S. 2.

9 Siehe Abschnitt 5.2 für eine Begriffsbestimmung.

10 Lauber-Rönsberg, A. (2021), S. 100.

2. Personenbezogene Forschungsdaten

Die DSGVO schützt personenbezogene Daten, d. h. Informationen, die sich auf eine identifizierbare Person beziehen.¹¹ Da sozialwissenschaftliche Forschung oftmals auf der Verarbeitung von personenbezogenen Daten basiert, ist der Umgang mit personenbezogenen Daten eine große Herausforderung im Forschungsdatenmanagement in dieser Disziplin. Sofern Daten anonym sind, d.h. dass die Re-Identifizierung von Personen nicht mehr möglich ist oder überhaupt nie möglich war, findet die DSGVO keine Anwendung und die Daten können (aus datenschutzrechtlicher Sicht) ohne Einschränkungen verarbeitet und auch in Repositorien veröffentlicht werden. Pseudonyme Daten hingegen gelten nach wie vor als personenbezogene Daten und unterliegen weiterhin den Anforderungen der DSGVO. Pseudonym sind Daten, die erst nach Hinzuziehen von zusätzlichen Informationen eine Re-Identifizierung zulassen, sofern diese zusätzlichen Informationen separat gespeichert und nicht frei zugänglich sind.¹² Bei sozialwissenschaftlichen Daten ist es oftmals gar nicht so einfach festzustellen, ob Daten anonym oder personenbezogen (pseudonymisiert oder nicht-pseudonymisiert) sind. Dies spielt aber eine große Rolle für die Verarbeitung solcher Daten und ganz besonders für deren Veröffentlichung, da es aus datenschutzrechtlicher Sicht keine Einwände gegen eine Veröffentlichung ohne Zugangsbeschränkungen oder anderen Sicherheitsvorkehrungen von anonymen Daten gibt, bei lediglich pseudonymen oder nicht-pseudonymisierten personenbezogenen Daten hingegen schon.

Personenbezogene Daten ergeben sich entweder aus direkten oder indirekten Identifikatoren. Direkte Identifikatoren, wie z. B. der Name, die genaue Adresse oder die vollständige IP-Adresse sind relativ gut erkennbar und klar als personenbezogene Informationen klassifizierbar. Indirekte Identifikatoren jedoch – und da ergibt sich die Schwierigkeit für sozialwissenschaftliche Daten – ergeben sich aus der Fülle der verfügbaren (spezifischen) Information zu einem Individuum und ermöglichen dadurch eine Re-Identifizierung. Als Beispiel können hier Umfragedaten dienen, die u. a. Informationen zu Alter, Geschlecht, Wohnort und Beruf enthalten. Für sich alleine stehend wären diese Informationen keine personenbezogenen Daten, gemeinsam können sie es aber sein, sofern eine Rückführung auf einzelne Personen damit möglich wird. Eine Re-Identifizierung könnte potenziell möglich sein, wenn z. B. die Informationen „Beruf“ und „Arbeitsort“ verfügbar sind und so die Rektorin einer örtlichen Grundschule identifiziert wird oder wenn die Variablen „Wohnort“ und „Herkunft“ verknüpft werden und damit eine Person mit

11 Feiler, L.; Forgó, N. (2017), S. 2f.

12 Art. 4 (5) DSGVO

senegalesischer Abstammung in einer Kleinstadt erkannt wird.¹³ Ob es sich also um personenbezogene Daten handelt, kann oftmals nur in der Betrachtung der Gesamtheit der Daten festgestellt werden.

Um einen richtigen Umgang mit diesen Daten zu gewährleisten, ist die Unterscheidung in personenbezogene und nicht-personenbezogene Daten also nötig. Im Zweifelsfall sollte davon ausgegangen werden, dass es sich um personenbezogene Daten handelt, die geschützt werden müssen und nur unter bestimmten Bedingungen verfügbar gemacht werden dürfen.¹⁴ Die DSGVO spezifiziert weitere Kriterien, die für diese Entscheidung herangezogen werden können – so

sollten alle Mittel berücksichtigt werden, die [...] wahrscheinlich genutzt werden, um die natürliche Person direkt oder indirekt zu identifizieren. [...] [Dafür] sollten alle objektiven Faktoren, wie die Kosten der Identifizierung und der dafür erforderliche Zeitaufwand, herangezogen werden, wobei die zum Zeitpunkt der Verarbeitung verfügbare Technologie und technologische Entwicklungen zu berücksichtigen sind.¹⁵

Das ist im sozialwissenschaftlichen Bereich insofern relevant, als dass eine Re-Identifizierung durch indirekte Identifikatoren häufig nicht gänzlich ausgeschlossen werden kann, ein Teilen dieser Daten aber trotzdem möglich ist, nämlich mittels geeigneter Sicherheitsmaßnahmen unter Berücksichtigung des Risikos, des Aufwandes für eine Re-Identifizierung und des möglichen Schadens von Personen (siehe dafür Kapitel 4 dieses Beitrages).

3. Sensible Forschungsdaten in den Sozialwissenschaften

Die DSGVO reguliert und schützt grundsätzlich alle personenbezogenen Daten, liefert aber gleichzeitig auch einige Ausnahmen und Möglichkeiten zur Datenverarbeitung (insbesondere der Archivierung und Verfügbarmachung) von personenbezogenen Daten für wissenschaftliche Zwecke, sofern geeignete Garantien für die Sicherheit der Daten bestehen.¹⁶ Diese Garantien können mit passenden technischen und organisatorischen Maßnahmen erfüllt werden, die sich am jeweiligen Risiko der Datenverarbeitung für betroffene Personen orientieren. Dabei soll die Eintrittswahrscheinlichkeit, wie auch die Größe des Risikos oder des denkbaren Schadens beachtet werden.¹⁷

13 Beispiele siehe Meyermann, A.; Porzelt, M. (2014), S. 6.

14 Lauber-Rönsberg, A. (2021), Anm. 10, S. 100.

15 Erwägungsgrund 26 DSGVO

16 Art. 89 DSGVO

17 Art. 32 DSGVO und Erwägungsgründe 74-78 DSGVO

Für die Arbeit von Repositorien und die Veröffentlichung oder Zurverfügungstellung von personenbezogenen Forschungsdaten ist das von großer Bedeutung, da mit diesen Ausnahmen und Möglichkeiten einerseits eine rechtliche Grundlage für die Verarbeitung dieser Daten geschaffen wurde, gleichzeitig aber i. S. der DSGVO Maßnahmen gefordert werden, die Daten von Individuen zu schützen. Für eine Analyse der Risiken und der Festlegung geeigneter, darauf aufbauender Schutzmaßnahmen ist darauf zu achten, dass u. a. eine „unbefugte Offenlegung von oder unbefugter Zugang zu personenbezogenen Daten“ keinen „physischen, materiellen oder immateriellen Schaden“ hervorrufen.¹⁸ Forschende und Repositorien müssen also Maßnahmen setzen, die sicherstellen, dass Daten gemäß dem Risiko einer Verletzung des Datenschutzes und dem potenziellen Schaden nach einer Verletzung desselben verarbeitet werden – besonders, wenn der Verarbeitungsschritt die Veröffentlichung dieser Daten ist. Gewisse festgelegte personenbezogene Daten sind darüber hinaus per Gesetz unter dem Begriff „*besondere Kategorien personenbezogener Daten*“ besonders schützenswert, weil sie als „besonders sensibel“ eingestuft werden und „erhebliche Risiken für die Grundrechte und Grundfreiheiten“ von Personen darstellen können.¹⁹ Das sind Daten zur politischen Meinung, ethnischen Herkunft, religiösen Überzeugung, Gewerkschaftszugehörigkeit, sowie genetische und biometrische Daten wie auch Gesundheitsdaten und Daten zum Sexualleben.²⁰

Sozialwissenschaftliche Forschungsdaten müssen nicht per se personenbezogen sein und damit der DSGVO unterliegen, sind es aber häufig und können zum Teil auch besonders sensible personenbezogene Daten sein. Bei der Aufbereitung und Veröffentlichung dieser Daten ist es daher wichtig, diverse Risiken und potenzielle Schäden zu beachten und die Veröffentlichung, inkl. Schutzmaßnahmen, an diesen Aspekten festzumachen. Dabei können keine pauschalen Einteilungen anhand von wissenschaftlichen Disziplinen oder Methoden vorgenommen werden – eine Einzelbetrachtung und -entscheidung ist immer notwendig.

Das Risiko einer Verletzung des Datenschutzes bei personenbezogenen Forschungsdaten ist vor allem bei großen Datenmengen oder hohem Informationsgehalt besonders hoch. Das kann bei sozialwissenschaftlichen Umfragen durchaus gegeben sein, dem kann aber angemessen mit Aufbereitungs- und Zugangsmaßnahmen von Repositorien begegnet werden. Im Gegensatz dazu stellt z. B. die Verarbeitung von Registerdaten eine größere Herausforderung dar, weshalb dafür wieder strengere Maßnahmen nötig sind, denen z. B. mit der Schaffung eines Austrian-

18 Erwägungsgrund 83 DSGVO

19 Erwägungsgrund 51 DSGVO

20 Art. 9 DSGVO

Micro-Data Centers bei der Statistik Austria nachgekommen werden soll.²¹ Registerdaten sind Daten in Verzeichnissen oder Datenbanken mit Informationen über die Bevölkerung, die gesetzlich vorgesehen sind.²² Dabei steht das Eindämmen des unbefugten Zugriffs durch nicht berechnigte Personengruppen bzw. die Verhinderung der Nutzung für nicht berechnigte Zwecke im Vordergrund.²³

Ein weiteres Risiko stellen Daten dar, die sich nicht oder nicht gut anonymisieren/pseudonymisieren lassen, wie z. B. genetische oder biometrische Daten, für die die DSGVO auch weitere Einschränkungen oder Bedingungen durch nationale Gesetzgebung zulässt.²⁴ Das Bundesministerium für Arbeit, Soziales, Gesundheit und Konsumentenschutz (BMASGK) hielt in einer Stellungnahme zum Schutz sensibler Daten fest, dass genetische Daten eine eindeutige Zuordnung zu Individuen ermöglichen und mit dem Vorliegen von Referenzdaten damit eine Re-Identifizierung ermöglicht wird, weshalb eine Anonymisierung solcher Daten praktisch nicht erfolgen kann.²⁵ Für die Sozialwissenschaften sind diese Kategorien von Daten eher seltener Bestand, wobei auch „Gesichtsbilder“, also Fotografien von Personen, als biometrische Daten bezeichnet werden.²⁶ Diese werden aber nur dann als besonders sensible Daten eingestuft, wenn ihre Verarbeitung sich spezieller Technik bedient und der Identifizierung von Personen dienen soll, was wiederum einen eher seltenen Vorgang für den Bereich der sozialwissenschaftlichen Forschung darstellt. Allerdings gibt es auch im sozialwissenschaftlichen Bereich Forschungsdaten, wo eine Anonymisierung/Pseudonymisierung nicht gut möglich ist bzw. als nicht sinnvoll erachtet wird. Ein Beispiel wären Daten aus dem Bereich der Biographieforschung, da hier durch die Datendichte zu einzelnen Personen eine Anonymisierung nicht ohne eine Verfälschung oder einen erheblichen Verlust an Information erreicht werden kann.²⁷

Die Größe des möglichen Schadens für Studienteilnehmer:innen bei Eintritt einer Datenschutzverletzung hat eine Auswirkung darauf, welche Maßnahmen als geeignet im Umgang mit personenbezogenen Daten gelten können. Solche Schäden können beispielsweise eine Diskriminierung, ein finanzieller Verlust, die Rufschädi-

21 Rat FTE (2021)

22 <https://www.bmbwf.gv.at/Themen/Forschung/Forschung-in-%C3%96sterreich/Strategische-Ausrichtung-und-beratende-Gremien/Leitthemen/Registerforschung.html>

23 Für einen Überblick zur Relevanz von Registerdaten für die sozialwissenschaftliche Forschung siehe Oberhofer, H. et al. (2019), S. 494-504.

24 Art. 9 (4) DSGVO

25 BMASGK (2019), S. 8.

26 Art. 4 (14) DSGVO

27 Siouti, I. (2018), S. 6f.

gung oder andere Aspekte sein, die zu einem wirtschaftlichen oder sozialen Nachteil für die betroffenen Personen führen.²⁸ Für den Umgang mit diesen Daten ist es dabei durchaus relevant, ob Betroffene „nur“ mit einem finanziellen Nachteil konfrontiert wären, mit einer Diskriminierung (z. B. am Arbeitsplatz auf Grund von unbefugtem Zugriff auf Gesundheitsdaten) oder der Bedrohung bzw. Verfolgung auf Grund der unbefugten Offenlegung der politischen Meinung (z. B. in Regionen mit eingeschränkter Presse- oder Meinungsfreiheit).²⁹

Unabhängig vom Dateninhalt gelten Kinder gemäß DSGVO grundsätzlich als besonders schutzwürdig, weshalb dieser Umstand bei einer Veröffentlichung von Forschungsdaten, die personenbezogene Informationen über Kinder enthalten, in der Risiko- und Schadensabwägung berücksichtigt werden muss.³⁰ Sollten also z. B. sozialwissenschaftliche Umfragen unter Schüler:innen oder internationale Leistungsvergleichsstudien von Schüler:innen Forschungsgegenstand sein, ist bei einer Datenveröffentlichung besondere Sorgfalt angebracht.

Es kann also festgestellt werden, dass Forschungsdaten im Bereich der Sozialwissenschaften in vielen Fällen personenbezogene Daten enthalten, mitunter auch besonders sensible, die einen noch höheren Schutz benötigen. Wie hoch das Risiko einer Datenschutzverletzung und die möglichen Folgeschäden für die Betroffenen sind, muss vor einer Veröffentlichung immer abgewogen und mit geeigneten Maßnahmen begleitet werden. Welche Daten liegen vor, personenbezogene und besonders sensible? Wie wurden diese Daten erhoben – wurden sie vertraulich im Rahmen einer Umfrage generiert, in der Anonymität zugesichert wurde? Wie sollen diese veröffentlicht werden, pseudonymisiert/anonymisiert oder mit direktem Personenbezug? Wie hoch ist das Risiko, dass sich betroffene Personen re-identifizieren lassen oder dass Unbefugte Zugriff erhalten? Wie hoch wäre der damit einhergehende Schaden? Diese und ähnliche Fragen können dazu dienen, die Sensibilität und damit auch das verbundene Risiko einer Veröffentlichung in Repositorien zu eruieren.

28 Erwägungsgründe 75 und 85 DSGVO

29 Für weitere Beispiele zur Einordnung siehe auch Harvard Information Security (2020).

30 Erwägungsgrund 75 DSGVO

4. Der Umgang mit „Anonymität“ im europäischen Vergleich – „faktische Anonymität“ und „statistical disclosure control“

Die rechtliche Grundlage für den Umgang mit personenbezogenen Daten sind in Deutschland, genau wie in Österreich, die DSGVO und darüber hinaus weitere nationale Gesetze, die sich auf die Öffnungsklauseln der DSGVO stützen. Eine solche Öffnungsklausel gibt es speziell für die Wissenschaft durch Artikel 89 der DSGVO. Über die DSGVO hinaus regelt Deutschland den Datenschutz durch das Bundesdatenschutzgesetz oder die Landesdatenschutzgesetze.³¹ Für die wissenschaftliche Verwendung von personenbezogenen Forschungsdaten wurde auch mit dem deutschen Bundesstatistikgesetz von 1987 eine wichtige Grundlage geschaffen, die wesentliche Vorteile für die Wissenschaft und die wissenschaftliche Nutzung und Weitergabe von Daten hatte – die Aufnahme des Begriffes der „faktischen Anonymität“, wie von der European Science Foundation definiert. Die faktische Anonymität (oder auch relative Anonymität) steht in dieser Diskussion im Gegensatz zur „absoluten Anonymität“, die nur dann besteht, wenn jegliche theoretische Möglichkeit der Re-Identifizierung ausgeschlossen ist. Schon damals wurde festgehalten, dass es keine hundertprozentige Sicherheit gibt, dass ein Datensatz nicht und niemals de-anonymisiert werden kann, selbst wenn noch so viele Maßnahmen zur Vorbeugung getroffen werden. Das Konzept der faktischen Anonymität basiert auf dieser Annahme und spezifiziert, dass Daten auch noch (faktisch) anonym sind, wenn eine Re-Identifizierung nur durch einen unverhältnismäßig großen Aufwand erreicht werden kann.³² Auch im Bundesdatenschutzgesetz-alt (welches bis zur Einführung der DSGVO Gültigkeit hatte) wurde am Prinzip der faktischen Anonymität festgehalten, in § 3 Abs. 6.³³

Mit Einführung der DSGVO wurde aber wieder am Konzept der „faktischen Anonymität“ gerüttelt, das für die datenbasierte Forschung und auch für eine Veröffentlichung von Daten im Sinne von Open oder FAIR Data äußerst relevant war. Die DSGVO enthält eine Definition, die an das Verständnis der absoluten Anonymität angelehnt wird, wenn sie festhält, dass Daten nur dann anonym sind, wenn sie gar nicht oder nicht mehr auf Personen zurückgeführt werden können.³⁴ Diese Ansicht beruht rein auf Auslegungen, da die DSGVO selbst keine Unterscheidung zwischen absoluter und faktischer Anonymität trifft. Die DSGVO spezifiziert aber auch, dass

31 Lauber-Rönsberg, A. (2021), Anm. 10, S. 98f.

32 Wirth, H. (1992), S. 7f.

33 Watteler, O.; Ebel, T. (2019), S. 65.

34 Erwägungsgrund 26 DSGVO

Faktoren, wie der für eine Re-Identifizierung nötige Zeitaufwand und die möglichen Mittel, eine Rolle spielen – was wiederum an die Definition der faktischen Anonymität angelehnt ist.³⁵ Dass diese Unterscheidung für die wissenschaftliche Praxis vor allem im Bereich der Sozialwissenschaften wichtig ist, geht auch aus der Stellungnahme des Rates für Sozial- und Wirtschaftsdaten (RatSWD) zur DSGVO aus 2015 hervor. Darin wird die Definition der Anonymität dahingehend kritisiert, dass diese nicht mehr der faktischen entspräche, wie es für die Praxis notwendig wäre:

Es ist nicht ausgeschlossen, dass sich in der Rechtspraxis das Konzept der absoluten Anonymisierung durchsetzt. [...] Absolut anonymisierte Daten sind für wissenschaftliche Analysen kaum geeignet, da ihr Informationsgehalt für fast alle Forschungsfragen nicht mehr ausreichend ist [...]. Aus diesem Grunde war die sukzessive Ablösung des Konzepts der absoluten Anonymität durch das Konzept der relativen Anonymität im BStatG, BDSG und Sozialgesetzbuch ein ganz erheblicher Fortschritt für die empirische Forschung in Deutschland, ohne dass hierdurch die datenschutzwürdigen Belange der Bürgerinnen und Bürger beeinträchtigt worden wären.³⁶

Dass das Konzept der faktischen (relativen) Anonymität trotz DSGVO noch nicht aus der Praxis verschwunden ist, sieht man auch anhand des Umgangs der deutschen Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder. Bei faktisch anonymen Daten soll die Möglichkeit der Identifikation von Individuen nahezu ausgeschlossen werden, dabei aber der Informationsgehalt in den Daten erhalten bleiben. Es wird auch festgehalten, dass für eine faktische Anonymität nicht nur der Datensatz an sich betrachtet werden muss, sondern auch weitere Maßnahmen, wie die Zugangsbedingungen zu den Daten, eine Rolle spielen.³⁷

Die Diskussion im deutschen Forschungsdatenmanagement (FDM) im sozialwissenschaftlichen Bereich dient in einem ersten Schritt der Sichtbarmachung einer grundlegenden Herausforderung im Verständnis von Anonymität bei Forschungsdaten, wonach es eine absolute Anonymität praktisch nicht gibt. In der FDM-Praxis geht es also folglich nicht darum, eine absolute Anonymität zu erreichen, sondern viel mehr um die Risikominimierung bei der Veröffentlichung von faktisch anonymisierten Forschungsdaten.

Der britische Datenservice UKDS bedient sich zu diesem Zweck dem Konzept „Statistical Disclosure Control (SDC)“. „Disclosure“ erfolgt, sobald veröffentlichte Daten genutzt werden, um unbekannte (sensible) personenbezogene Informationen über

35 Watteler, O.; Ebel, T. (2019), Anm. 33, S. 65f.

36 RatSWD (2015), Anm. 8, S. 4.

37 <https://www.forschungsdatenzentrum.de/de/anonymitaet>

Individuen herauszufinden und diese gegebenenfalls auch zu veröffentlichen. Das Ziel von SDC ist es, dieses Risiko so weit wie möglich und nötig zu minimieren, während gleichzeitig so viele Informationen wie möglich erhalten bleiben. „Although SDC aims to prevent the re-identification of a data subject, this is a ‘risk minimisation’ strategy rather than a ‘risk elimination’ one. Few would wholeheartedly claim to have removed disclosure risk entirely”.³⁸ Abgesehen davon, dass nicht gewährleistet werden kann, dass Datensätze zu einem gewissen Zeitpunkt und in sich geschlossen absolut anonym sind, inkludiert SDC auch Aspekte, die gar nicht in der Kontrolle der Datenbearbeiter:innen liegen und trotzdem eine Auswirkung auf das Re-Identifizierungsrisiko haben können. So spezifiziert auch die DSGVO, dass zukünftige technologische Entwicklungen bei der Anonymisierung bedacht werden sollen,³⁹ während gleichzeitig immer mehr persönliche Daten von Individuen offen zugänglich werden, die das Risiko wiederum erhöhen können:

Trying to eliminate the risk of statistical disclosure entirely would ignore changes in how we think about our data and our appetite for disclosing confidential information about ourselves; and we may ultimately fail in SDC as a result. Instead, risk minimisation takes account of outside conditions that we cannot control.⁴⁰

Einen eindeutigen und unumstrittenen Weg im datenschutzkonformen Umgang mit personenbezogenen Forschungsdaten gibt es grundsätzlich nicht, auch nicht durch die Schaffung eines einheitlichen europäischen Rahmens mit der DSGVO. Es scheint aber klar, dass eine absolute Anonymität von sozialwissenschaftlichen Daten nie zu hundert Prozent sichergestellt werden kann,⁴¹ weshalb für den praktischen Umgang mit diesen Daten auch andere Faktoren beachtet werden müssen, darunter vor allem das schon erwähnte potenzielle Risiko einer Re-Identifizierung inklusive des damit einhergehenden möglichen Schadens. Um das Risiko zu minimieren, müssen angemessene Schutzmaßnahmen eingerichtet werden. Im nachfolgenden Kapitel 5 soll aufgezeigt werden, wie AUSSDA als sozialwissenschaftliches Repositorium mit (sensiblen) personenbezogenen Daten umgeht, welche Maßnahmen zum Schutz von Individuen bestehen und empfohlen werden, um das Re-Identifizierungsrisiko zu minimieren, während der wissenschaftliche Wert bestmöglich erhalten bleibt.

38 Griffiths, E. et al. (2019), S. 20.

39 Erwägungsgrund 26 DSGVO

40 Griffiths, E. et al. (2019), Anm. 31, S. 20.

41 Siehe auch Stam, A.; Kleiner, B. (2020), S. 4f.

5. Der Umgang mit personenbezogenen Daten beim Austrian Social Science Data Archive (AUSSDA)

Bisher wurden die rechtlichen Grundlagen erörtert, auf Basis derer die Veröffentlichung von personenbezogenen Daten, speziell im Bereich der Sozialwissenschaften, grundsätzlich stattfinden kann, und wie personenbezogene Daten definiert, erkannt und in ihrer Schutzwürdigkeit eingeordnet werden können. Die etablierten Prozesse der Datenveröffentlichung beim österreichischen sozialwissenschaftlichen Archiv AUSSDA basieren auf diesen Grundlagen und Auslegungen – die Umlegung dieser Überlegungen in die Veröffentlichung von Datensätzen mit hoher Heterogenität und Komplexität ist allerdings stets Gegenstand von Diskussion, Anpassung und Weiterentwicklung.

5.1. Die rechtliche Grundlage – Einwilligung oder Ausnahme für wissenschaftliche Zwecke

Laut DSGVO braucht es für jede Verarbeitung von personenbezogenen Daten eine entsprechende Rechtsgrundlage. Gemäß Art. 6 DSGVO kann eine dieser Grundlagen die Einwilligung der betroffenen Person in die Verarbeitung für einen oder mehrere bestimmte Zwecke sein. Bei sensiblen Daten oder auch „besonderen Kategorien personenbezogener Daten“ besteht grundsätzlich ein Verarbeitungsverbot, außer eine der genannten Ausnahmen gemäß Artikel 9 DSGVO trifft zu. Eine dieser Ausnahmen ist ebenfalls die (ausdrückliche) Einwilligung. Die Einwilligung ist eine der wichtigsten Grundlagen für eine rechtmäßige Datenverarbeitung in den Sozialwissenschaften für Forschende. Ob bei einer Online-Umfrage oder bei einem Interview, die Teilnehmer:innen werden gebeten, davor in die Erhebung und Verarbeitung ihrer Daten einzuwilligen. Natürlich gibt es auch einige Szenarien, in denen eine Einwilligung nicht eingeholt werden kann, z. B., wenn öffentlich zugängliche Daten genutzt werden, wie bei Forschung im Bereich der sozialen Medien. Für eine rechtmäßige Verarbeitung solcher Daten muss eine andere Rechtsgrundlage gemäß Art. 6 DSGVO gegeben sein.

Wie genau eine Einwilligungserklärung aussehen muss, hängt wiederum vom Forschungsvorhaben ab und lässt sich nicht pauschal beantworten. Dafür gibt es bereits einige Vorlagen von verschiedensten wissenschaftlichen Einrichtungen zur freien Verfügung – die aber wiederum sehr genau an die Gegebenheiten für jede Situation und jedes Projekt angepasst werden müssen. Damit eine Einwilligung grundsätzlich gültig ist, muss sie gemäß DSGVO folgende Aspekte beinhalten: Sie muss freiwillig, mit einer aktiven Handlung, für einen konkreten Zweck, informiert und unmissverständlich gegeben werden und leicht widerrufbar sein. Das kann

mündlich, schriftlich oder auch elektronisch erfolgen.⁴² Es empfiehlt sich aus Gründen der Nachvollziehbar- und Beweisbarkeit, die Einwilligung zu dokumentieren.⁴³ Zumindest ist aber auch (nicht unbedeutend für Online-Umfragen) geregelt, dass eine gültige Einwilligung auch mittels Anklicken eines Kästchens erfolgen kann.⁴⁴

Forschende, die planen, personenbezogene Daten für ihr Projekt zu verarbeiten, und diese Informationen nicht aus öffentlichen Quellen beziehen, sondern direkt von ihren Forschungssubjekten, benötigen im Regelfall deren Einwilligung. Zusätzlich ist das auch ethisch der beste Weg, da damit gleichermaßen eine umfassende und verständliche Aufklärung zur geplanten Datenverarbeitung stattfindet und den Teilnehmer:innen wichtige Informationen über ihre Rechte gegeben werden. Die Einwilligungserklärung soll so formuliert sein, dass die Teilnehmer:innen wissen, welche ihrer Daten gesammelt werden und was damit weiter passiert, bis zu einer allfälligen Löschung oder eben Archivierung und Publikation ihrer Daten.⁴⁵ Dieser letzte Aspekt ist oftmals noch nicht Bestandteil solcher Einwilligungserklärungen und auch vieler Vorlagen, die frei zugänglich sind, da die Veröffentlichung von Daten selbst ein Prozess ist, der sich erst als Teil der wissenschaftlichen Praxis etabliert. Hier empfiehlt es sich, den Blick auf Formulierungen zu lenken, die einer Archivierung und Publikation klar entgegenstehen, und diese gegebenenfalls zu vermeiden, wie z. B., dass Daten nicht außerhalb des Projektteams weitergegeben werden; dass Daten nur in aggregierter Form veröffentlicht werden; oder, dass Daten nach Projektende gelöscht werden. Das alles ist im Falle einer Archivierung gemäß den FAIR-Prinzipien nicht zutreffend, sofern es nicht einen guten (datenschutzrechtlichen) Grund gibt, dass die Daten nicht mit Personen außerhalb des Projektteams geteilt, dass die Daten nur aggregiert veröffentlicht oder gelöscht werden. Selbst wenn Archive auf Basis anderer Grundlagen das Recht haben, diese Daten zu archivieren, wäre das forschungsethisch bedenklich, da den Teilnehmer:innen etwas Anderes zugesagt wurde.

Für ein Archiv ist das Stützen der Verarbeitungsprozesse auf die Einwilligung als rechtliche Grundlage gewissermaßen schwieriger, da die Daten dort nicht erhoben werden und kein Kontakt zu den Teilnehmer:innen besteht. Es kann auch nicht überprüft werden, ob eine rechtsgültige Einwilligung an die Forschenden bzw. Datenerheber:innen abgegeben wurde. AUSSDA stützt sich für die Archivierung und

42 Erwägungsgrund 32 DSGVO

43 Erwägungsgrund 42 DSGVO

44 Erwägungsgrund 32 DSGVO

45 Siehe auch Verbund Forschungsdaten Bildung (n. d.)

Zurverfügungstellung daher nicht auf die Einwilligung, sondern auf die Ausnahmen, die u. a. für personenbezogene Daten zu im öffentlichen Interesse liegenden Archivzwecken, zu wissenschaftlichen oder historischen Forschungszwecken oder zu statistischen Zwecken mit Art. 89 DSGVO geschaffen wurden.⁴⁶ Für Österreich ist das weiter in § 7 des Datenschutzgesetzes geregelt, wonach für die schon genannten wissenschaftlichen Zwecke personenbezogene Daten verarbeitet werden dürfen, die entweder öffentlich zugänglich sind oder pseudonymisierte Daten sind.⁴⁷ Das Forschungsorganisationsgesetz ermöglicht die Arbeit von Repositorien darüber hinaus mit der Berechtigung zum Sammeln und Archivieren von Forschungsdaten, um einen optimalen Zugang zu schaffen.⁴⁸

AUSSDA und andere Archive sind also grundsätzlich dazu berechtigt, personenbezogene Forschungsdaten zu verarbeiten und diese unter Einhaltung bestimmter Maßnahmen zur Einhaltung der Rechte und Freiheiten von betroffenen Personen auch zu archivieren und zu veröffentlichen. Eine dieser Maßnahmen, sowohl in der DSGVO als auch im österreichischen Datenschutzgesetz festgehalten, ist die Pseudonymisierung der Daten.⁴⁹

5.2. Anonymisierung/Pseudonymisierung

Anonyme oder anonymisierte Daten sind von der DSGVO ausgenommen und können ohne weitere Auflagen aus datenschutzrechtlicher Sicht verarbeitet werden.⁵⁰ Dass sozialwissenschaftliche Daten in den seltensten Fällen anonym sind oder (absolut) anonymisiert werden können (siehe Kapitel 4), stellt das sozialwissenschaftliche Forschungsdatenmanagement vor Herausforderungen. Die Pseudonymisierung hingegen ist ein Konzept, das in der DSGVO spezifiziert ist und als Maßnahme angewandt werden kann, personenbezogene Daten geschützt zu verarbeiten. Daten sind pseudonym, wenn es ohne Hinzuziehen zusätzlicher Informationen nicht mehr möglich ist, Individuen zu identifizieren, solange diese zusätzlichen Informationen gesondert aufbewahrt werden und Unbefugte keinen Zugriff darauf haben.⁵¹ Der Vorgang der Pseudonymisierung entspricht dabei auch dem Grundsatz der Datenminimierung der DSGVO, nachdem Daten dem Zweck angemessen und nur auf

46 Hönegger, L. et al. (2020), S. 3.

47 § 7 des Bundesgesetzes zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten (Datenschutzgesetz – DSG) StF: BGBl. I Nr. 165/1999.

48 § 2f des Bundesgesetzes über allgemeine Angelegenheiten gemäß Art. 89 DSGVO und die Forschungsorganisation (Forschungsorganisationsgesetz – FOG) StF: BGBl. Nr. 341/1981 idF BGBl. Nr. 448/1981. <https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=10009514>

49 Hönegger, L. et al. (2020), Anm. 46, S. 4.

50 Erwägungsgrund 26 DSGVO

51 Art. 4 (5) DSGVO

das für diesen Zweck notwendige Maß beschränkt verarbeitet werden dürfen.⁵² Daten sollen also nicht grundlos gesammelt und gespeichert werden, sondern nur, wenn es einen legitimen Grund dafür gibt.

AUSSDA geht bei an das Archiv übermittelten Daten davon aus, dass es sich in den meisten Fällen um pseudonymisierte Daten handelt, da nicht ausgeschlossen werden kann, dass zusätzliche, getrennt aufbewahrte Informationen vorhanden sind, die eine Re-Identifizierung ermöglichen können (z. B. bei den Datenerheber:innen). Solange der Personenbezug theoretisch von allen Personen, die zusätzliche Informationen haben (z. B. die Datenerheber:innen) wiederhergestellt werden kann, sind die Daten nicht anonym, sondern pseudonym.⁵³ Zur Aufbewahrung von Zusatzinformationen (z. B. Identifizierungsschlüssel) gibt es unterschiedliche Praxisen und Aufbewahrungsfristen. AUSSDA veröffentlicht derzeit Daten, die zumindest pseudonymisiert sind, d. h. die Zusatzinformationen (die für eine Re-Identifizierung nötig werden) stehen potenziellen Nachnutzer:innen in weiterer Folge nicht zur Verfügung. Damit Daten aber erfolgreich pseudonymisiert werden, dürfen sie ohne solche Zusatzinformationen eine Re-Identifizierung nicht ermöglichen. Dabei folgen wir wieder, wie bisher dargelegt, dem Konzept der „faktischen Anonymisierung“, d. h. der Datensatz, der bei AUSSDA veröffentlicht wird, wird auf ein Re-Identifizierungsrisiko geprüft und den Datengeber:innen Empfehlungen hin zur (faktischen) Anonymisierung gegeben. Diese Empfehlungen sind z.B. im quantitativen Bereich das Löschen von einzelnen Variablen (z. B. Name, Geburtsdatum oder E-Mail-Adresse); das Zusammenfassen in Gruppen (z. B. Angabe von Altersgruppen statt Alter); oder im qualitativen Bereich das Ersetzen von Information durch Pseudonyme (u. a. Name, Herkunft, Wohnort).

Datenprüfungen werden bei Dateneingängen von AUSSDA durchgeführt. Diese Prüfungen werden nach gleichen Prinzipien, allerdings abhängig von Datentyp und möglichem Dateninhalt (inklusive Risikoabschätzung), durchgeführt. Die Voraussetzung für eine Datenübernahme und Veröffentlichung durch AUSSDA ist aber immer, dass Datengeber:innen den Datengewinnungsprozess offenlegen, damit das Risiko bewertet werden kann. Die Datentypen, die derzeit bei AUSSDA veröffentlicht werden, gliedern sich in numerische Daten und Textdaten. Für eine praktische Anleitung zu Anonymisierungsschritten bei AUSSDA für numerische Daten (vor allem Umfragedaten) siehe die AUSSDA-Data-Deposit Guideline⁵⁴. Diese kann jedoch, wie der Name bereits verrät, immer nur als Richtlinie gesehen werden – die kon-

52 Art. 5 (1) lit. c DSGVO

53 Watteler, O.; Kinder-Kurlanda, K. E. (2015), S. 517.

54 Butzlaff, I. (2020), S. 10f.

kreten, für eine Anonymisierung notwendigen Schritte hängen immer von weiteren Faktoren ab (wie z. B. Datenmenge und Informationsgehalt) und müssen kontextbezogen betrachtet, angepasst und nach Ermessen umgesetzt werden. Eine vollständige Liste oder ein vollständig beschriebener Vorgang der Anonymisierung kann nicht pauschal für Forschungsdaten festgelegt werden. Eine besondere Herausforderung stellen dabei die stetig neu entstehenden Datenarten und -typen dar, die immer wieder neue Anonymisierungsschritte erfordern.

Die kritische Frage für Forschende, aber auch für Archive ist, ab welchem Zustand Daten als anonymisiert/pseudonymisiert betrachtet werden können. Da die verbreitete Auffassung im Umgang mit sozialwissenschaftlichen Daten ist, dass eine Re-Identifikation niemals hundertprozentig ausgeschlossen werden kann (siehe Kapitel 4), gibt es dafür keine klare Trennlinie. Vielmehr geht es um Risikominimierung und darum, die Balance zwischen Datenschutz und Wissenserhalt für die Forschung zu finden. AUSSDA agiert bei der Datenkontrolle nach denselben Prinzipien und betrachtet Daten dann als anonymisiert/pseudonymisiert, sobald nur mehr mit unverhältnismäßigem Aufwand eine Re-Identifikation durch Dritte stattfinden könnte. Dabei orientiert sich das Archiv an den oben dargelegten, international vergleichbaren Standards im sozialwissenschaftlichen Forschungsdatenmanagement, bewertet Datensätze in ihrem Kontext und setzt technische und organisatorische Maßnahmen in Abwägung von Datenschutz und Risiko.⁵⁵ Sollte der denkbare Schaden nach einer (nur mit unverhältnismäßigem Aufwand möglichen) Re-Identifikation dennoch groß sein, nutzt das Archiv weitere Maßnahmen zum Schutz der Daten – die Zugangskontrolle und die Einschränkung der Nachnutzung.

5.3. Zugangskontrollen und Lizenzen

Nach einer Prüfung der Daten und Betrachtung des Hintergrunds der Datenerhebung wird von AUSSDA (oftmals gemeinsam mit Datengeber:innen) eine Einschätzung des Risikos vorgenommen. Da wir davon ausgehen, dass eine Re-Identifizierung niemals komplett ausgeschlossen werden kann, müssen das Risiko im unwahrscheinlichen, aber möglichen Fall einer Datenschutzverletzung beachtet und Maßnahmen zum weiteren Schutz gesetzt werden. Diese Maßnahmen sind bei AUSSDA die Zugangskontrolle und die Einschränkung der Nachnutzung durch die Vergabe von Lizenzen. Damit folgt AUSSDA international etablierten Empfehlun-

55 Watteler, O.; Ebel, T. (2019), Anm. 33, S. 67.

gen im Umgang mit personenbezogenen, sensiblen Daten. “In case you are archiving sensitive data you should always go for a restricted license or closed access”.⁵⁶

AUSSDA hat drei verschiedene Zugangsbedingungen etabliert, die die unterschiedlichen Risiken adressieren gemäß der AUSSDA-Access Policy.⁵⁷ Darüber hinaus gibt es grundsätzlich zwei verschiedene Arten von Lizenzen, offene Lizenzen (Open Access – OA) und Lizenzen für die wissenschaftliche Nutzung (Scientific Use File – SUF).

Wenn Daten absolut anonym sind bzw. kein Risiko aufweisen, können diese ganz offen zur Verfügung gestellt werden. Diese Daten können im Online-Katalog⁵⁸ eingesehen und direkt heruntergeladen werden. Die Daten sind mit einer Open-Access-Lizenz versehen und können daher auch ohne Einschränkungen, aber mit der Verpflichtung zur Quellenangabe (in der Bibliografie) für alle Zwecke und von allen Personen nachgenutzt werden.

Sind die Daten nicht absolut anonym, sondern so bearbeitet, dass sie als „faktisch anonym“ zu verstehen wären, werden Daten eingeschränkter zur Verfügung gestellt, sowohl hinsichtlich der Zugangsbedingungen, als auch der Nachnutzungslicenz. AUSSDA vergibt in diesen Fällen eine Scientific-Use-Lizenz, die regelt, dass Daten nur für wissenschaftliche Zwecke verwendet werden dürfen. Diese Einschränkung ist möglich, weil die DSGVO und auch andere relevante Materiegesetze (z. B. das Datenschutzgesetz oder das Forschungsorganisationsgesetz) Ausnahmen für die wissenschaftliche Nutzung von personenbezogenen Daten erlauben. Gleichzeitig wird diese Einschränkung als nötig erachtet, um einer missbräuchlichen Verwendung solcher Daten vorzubeugen. Wie schon zu Beginn erwähnt, ist der datenschutzkonforme Umgang und die Sicherung von Daten nicht nur aus rechtlicher Sicht nötig, sondern wird auch aus forschungsethischer Sicht gefordert. Forschende sind darauf angewiesen, dass Teilnehmer:innen auf den korrekten Umgang mit ihren Daten vertrauen und die weitergegebenen Daten ausreichend geschützt werden. Forschende sind dazu auch in Richtlinien zur guten wissenschaftlichen Praxis verpflichtet.⁵⁹

Sobald Daten nur mehr für wissenschaftliche Zwecke verwendet werden dürfen und dementsprechend mit einer SUF Lizenz lizenziert werden, wird wiederum je

56 OpenAIRE Sensitive data guide. Storing sensitive data: <https://www.openaire.eu/sensitive-data-guide>

57 Kaczmirek, L.; Hönegger, L. (2019)

58 AUSSDA Dataverse: <https://data.aussda.at/>

59 Siehe z. B. Rat für Sozial- und Wirtschaftsdaten (RatSWD) (2017), S. 18.

nach Risiko differenziert, wie diese Daten zugänglich gemacht werden. Daten, die ein geringes Risiko nach einer potenziellen Re-Identifikation haben, können nach einem (institutionellen) Login direkt eingesehen und heruntergeladen werden. Die Nachnutzungslizenz (die die Nutzer:innen der Daten rechtlich bindet) spezifiziert dabei, dass Daten nur für wissenschaftliche Zwecke genutzt werden dürfen.⁶⁰ Wird das Risiko als größer eingeschätzt, wird eine zusätzliche Zugangshürde eingebaut. Forschende können solche Daten zunächst nicht einsehen und auch nicht herunterladen, sondern müssen mittels eines Formulars den genauen Zweck darlegen, weshalb sie Zugriff auf die Daten benötigen. Dabei wird von AUSSDA geprüft, ob ein Zugriff auf diese Daten für das dargestellte Forschungsvorhaben nötig ist und das Ansuchen legitim ist. Dieser bei AUSSDA restriktivste Zugang schränkt das Missbrauchspotenzial deutlich ein und wird daher als strengste Maßnahme für besonders schutzwürdige Daten gesetzt.

Als sozialwissenschaftliches Repositorium setzen wir uns aktiv für Open Science und FAIRe Daten ein, weshalb wir bei allen technischen und organisatorischen Maßnahmen nach dem Prinzip „so offen wie möglich, so geschlossen wie nötig“⁶¹ agieren. Wenn aus datenschutzrechtlicher Perspektive keine Einschränkungen nötig sind, werden diese nicht umgesetzt. Gibt es aber begründete Bedenken, werden die Daten mit Zugangsbeschränkungen geschützt. Diese Maßnahmen werden stets unter Interessensabwägung aller Beteiligten getroffen und zielen darauf ab, datenbasierte Forschung unter Einhaltung der nötigen Sicherheitsstandards zu ermöglichen. Damit orientiert sich AUSSDA auch an europäischen Forschungsvorgaben zum Datenmanagement.⁶²

6. Fazit

Damit Forschende ihre Daten entweder ganz öffentlich oder zumindest für die wissenschaftliche Nutzung zur Verfügung stellen können, braucht es neben den Infrastrukturen wie Archiven bzw. Repositorien vor allem auch eine Beratung und Unterstützung zur Klärung der rechtlichen Fragen und ethischen Bedenken. Der Schutz von personenbezogenen Daten ist unbedingt notwendig, gleichzeitig muss aber auch der wissenschaftliche Wert der Daten erhalten bleiben und zugänglich sein. Die rechtlichen Grundlagen, die das Datenteilen ermöglichen, sind äußerst komplex und ihre Umsetzung in der Praxis oftmals nicht eindeutig geregelt, was Forschende vor enorme Herausforderungen stellt. Die Praxis des Datenteilens, wie

60 Für den vollen Lizenztext siehe <https://aussda.at/vertraege-und-lizenzen-bei-aussda/wissenschaftlicher-zweck/>

61 European Commission (2021), S. 34.

62 Ebd., Anm. 61, S. 64.

von der Open-Science-Bewegung gefordert⁶³, ist mit vielen Hürden konfrontiert, wie z. B. mangelnde Ressourcen für die Aufbereitung der Daten oder mangelnde Anreize für das Datenteilen. Ein für Wissenschaftler:innen aber ebenfalls relevanter Faktor, der das Datenteilen verhindert, sind bestehende rechtliche Einschränkungen und Unsicherheiten im Bereich Datenschutzverletzungen.⁶⁴ Um die Praxis des Datenteilens bestmöglich zu unterstützen, brauchen Forschende Ansprechpersonen, um solche rechtlichen Unsicherheiten aus dem Weg zu räumen und datenschutzkonform ihre Daten zur Verfügung zu stellen:

Angesichts der Komplexität der rechtlichen Rahmenbedingungen sind Wissenschaftlerinnen und Wissenschaftler auf eine unterstützende Infrastruktur der Forschungseinrichtung angewiesen, um ein sachgerechtes und rechtskonformes FDM betreiben zu können.⁶⁵

AUSSDA ist als Repositorium, das sozialwissenschaftliche Daten veröffentlicht, bemüht, rechtliche Fragen zur Archivierung und Zurverfügungstellung von personenbezogenen Daten zu klären – durch umfangreiche Recherchen, Rücksprachen mit Datenschutzbeauftragten und Gutachten von unabhängigen Rechtsberater:innen – damit Forschende, die ihre Daten zur Verfügung stellen, um die Risiken und Möglichkeiten Bescheid wissen. In der Praxis geht es vorwiegend darum, diese Risiken, aber gleichzeitig auch die Möglichkeiten zu erkennen, einzuschätzen und danach Maßnahmen zu setzen, die ein Datenteilen mit dem nötigen Schutz ermöglichen. Dabei sucht AUSSDA gemeinsam mit Forschenden nach Lösungen, wie, aber nicht ob Daten geteilt werden können. Im Vordergrund steht, ein Bewusstsein für Datenschutz zu schaffen, das eine offene Wissenschaft nicht einschränkt, sondern nachhaltig ermöglicht.

Bibliografie

Bauer, Bruno; Ferus, Andreas; Gorraiz, Juan; Gründhammer, Veronika; Gumpenberger, Christian; Maly, Nikolaus; Mühlegger, Johannes Michael; Preza, José Luis; Sánchez Solís, Barbara; Schmidt, Nora; Steineder, Christian (2015): Forschende und ihre Daten. Ergebnisse einer österreichweiten Befragung – Report. Version 1.2., S. 51-53.
<https://doi.org/10.5281/zenodo.32043>

BMASGK (2019): Schutz sensibler Daten. Position der Gesundheitssektionen VIII und IX des BMASGK. <https://www.sozialministerium.at/dam/jcr:99d679b8-3676-455e-abfc->

63 <https://www.bmbwf.gv.at/Themen/HS-Uni/Hochschulgovernance/Leitthemen/Digitalisierung/Open-Science.html>

64 Bauer, B. et al. (2015), S. 51-53.

65 Lauber-Rönsberg, A. (2021), Anm. 7, S. 111.

- a6ec27c4d64d/Schutz_sensibler_Daten_-_Position_der_Gesundheitssektionen_VIII_und_IX_.pdf (abgerufen am 19.04.2023)
- Butzlaff, Iris (2020): Data Deposit Guideline (Public version) v2.0. Information for Data Depositors. Wien: AUSSDA – The Austrian Social Science Data Archive. https://aussda.at/fileadmin/user_upload/p_aussda/Documents/Data-Deposit-Guideline_SUF_v2_0.pdf
- Deutsche Forschungsgemeinschaft (2019): Guidelines for Safeguarding Good Research Practice. Code of Conduct. <https://doi.org/10.5281/zenodo.3923602>
- European Commission (2021): Horizon Europe (HORIZON) Programme Guide. Version 1.3, November. https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/programme-guide_horizon_en.pdf (abgerufen am 18.04.2023)
- Feiler, Lukas; Forgó, Nikolaus (2017): EU-DSGVO. EU-Datenschutz-Grundverordnung. Kurzkomentar. Wien: Verlag Österreich.
- Griffiths, Emily; Greci, Carlotta; Kotrotsios, Yannis; Parker, Simon; Scott, James; Welpton, Richard; Wolters, Arne; Woods, Christine (2019): Handbook on Statistical Disclosure Control for Outputs. https://ukdataservice.ac.uk/app/uploads/thf_data-report_aw_web.pdf (abgerufen am 18.04.2023)
- Harvard Information Security (2020): Data Security Levels. Research Data Examples Quick Reference Guide. <https://security.harvard.edu/files/it-security/files/rdslexamples.pdf> (abgerufen am 18.04.2023)
- Hönegger, Lisa; Butzlaff, Iris; Kaczmirek, Lars (2020): Bereitstellung personenbezogener Daten durch AUSSDA. Gesetzliche Grundlage. Wien: AUSSDA – The Austrian Social Science Data Archive. https://aussda.at/fileadmin/user_upload/p_aussda/Documents/Bereitstellung_personenbezogener_Daten.pdf (abgerufen am 18.04.2023)
- Kaczmirek, Lars; Hönegger, Lisa (2019): Access Policy. Open Access and Restricted Access Procedures in the AUSSDA Repository. Vienna: AUSSDA – The Austrian Social Science Data Archive. https://aussda.at/fileadmin/user_upload/p_aussda/Documents/AUSSDA_Access_Policy_v2_0.pdf
- Lauber-Rönsberg, Anne (2021): Rechtliche Aspekte des Forschungsdatenmanagements. In: Putnigs, Markus; Neuroth, Heike; Neumann, Janna (Hg.): Praxishandbuch Forschungsdatenmanagement. Berlin: Walter de Gruyter, S. 89-114. <https://doi.org/10.1515/9783110657807-005>
- Meyermann, Alexia; Porzelt, Maike (2014): Hinweise zur Anonymisierung von qualitativen Daten. In: forschungsdaten bildung informiert 1.
- Oberhofer, Harald; Schwarz, Gerhard; Strassnig, Michael (2019): Registerforschung. Verwaltungs- und Statistikdaten für die Wissenschaft. In: Mitteilungen der VÖB 72 (2): Open Science, S. 494–504. <https://doi.org/10.31263/voebm.v72i2.3154>
- Perry, Anja; Recker, Jonas (2018): Sozialwissenschaftliche Forschungsdaten langfristig sichern und zugänglich machen. Herausforderungen und Lösungsansätze. Was sind sozialwissenschaftliche Daten. In: Das offene Bibliotheksjournal 5 (2). <https://www.o-bib.de/bib/article/view/2018H2S106-122/6392>

- Rat FTE (2021): Registerdatenforschung und Implementierung eines Austrian Micro-Data Centers. Fact Sheet. <https://repository.fteval.at/id/eprint/667>
- Rat für Sozial- und Wirtschaftsdaten (RatSWD) (2017): Forschungsethische Grundsätze und Prüfverfahren in den Sozial- und Wirtschaftswissenschaften. https://www.konsortswd.de/wp-content/uploads/RatSWD_Output9_Forschungsethik.pdf (abgerufen am 18.04.2023)
- Rat für Sozial- und Wirtschaftsdaten (RatSWD) (2015): Stellungnahme des Rates für Sozial- und Wirtschaftsdaten zu ausgewählten Punkten des aktuellen Entwurfs der EU-Datenschutz-Grundverordnung (DSGVO). 24.02. https://www.konsortswd.de/wp-content/uploads/RatSWD_Stellungnahme_EUDSGVO.pdf (abgerufen am 18.04.2023)
- Siouti, Irini (2018): Forschungsethik in der Biografieforschung. Herausforderungen im Forschungsfeld der politischen Partizipation. In: Forum: Qualitative Sozialforschung 19 (3), 28.
- Stam, Alexandra; Kleiner, Brian (2020): Data Anonymization. Legal, Ethical, and Strategic Considerations. In: FORS Guide Nr. 11, Version 1.0. Lausanne: Swiss Centre of Expertise in the Social Sciences FORS. <https://doi.org/10.24449/FG-2020-00011>
- Verbund Forschungsdaten Bildung (n. d.): FAQ „Muss ich eine Einwilligung zur späteren Archivierung und Nachnutzung der Daten einholen?“ <https://www.forschungsdaten-bildung.de/faq> (abgerufen am 18.04.2023)
- Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung). <https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:32016R0679&from=DE> (abgerufen am 18.04.2023)
- Watteler, Oliver; Kinder-Kurlanda, Katharina E. (2015): Anonymisierung und sicherer Umgang mit Forschungsdaten in der empirischen Sozialforschung. In: DuD – Datenschutz und Datensicherheit 8.
- Watteler, Oliver; Ebel, Thomas (2019): Datenschutz im Forschungsdatenmanagement. In: Jensen, Uwe; Netscher, Sebastian; Weller, Katrin (Hg.): Forschungsdatenmanagement sozialwissenschaftlicher Umfragedaten. Grundlagen und praktische Lösungen für den Umgang mit quantitativen Forschungsdaten. Opladen, Berlin, Toronto: Verlag Barbara Budrich. <https://shop.budrich.de/wp-content/uploads/2019/01/4.-Datenschutz-im-Forschungsdatenmanagement-Oliver-Watteler-und-Thomas-Ebel.pdf> (abgerufen am 18.04.2023)
- Wirth, Heike (1992): Die faktische Anonymität von Mikrodaten. Ergebnisse und Konsequenzen eines Forschungsprojektes. In: ZUMA Nachrichten 16 (30). <https://nbn-resolving.org/urn:nbn:de:0168-ssaoar-209679>

Lisa Hönegger ist stellvertretende Leitung von AUSSDA, dem österreichischen sozialwissenschaftlichen Datenarchiv, mit Standorten an den Universitäten Wien, Graz, Linz und Innsbruck. Sie arbeitet seit 2018 an der Universitätsbibliothek Wien

und beschäftigt sich hauptsächlich mit Forschungsdatenmanagement und Open Science im österreichischen und europäischen Kontext.