

# DS-GVO: Geänderte Vorgaben für die Wissenschaft

## Was sind die neuen Rahmenbedingungen und welche Fragen bleiben offen?

Open Access  
Big Data  
Genetische Daten  
Sensible Daten  
Biobanken

■ In diesem Artikel werden die aus der geänderten EU-Datenschutzgrundverordnung (DS-GVO) resultierenden Änderungen im Hinblick auf die Wissenschaft ausgewertet. Insbesondere in den Blick genommen werden dabei allgemeine Änderungen (Zweckbindungsregelung, Anonymisierungs- und Pseudonymisierungsvorgaben), Forschung mit sensiblen Daten, insbesondere im Kontext der individualisierten Medizin und der genetischen Forschung. Hier wird argumentiert, dass sich die Forschung in diesem Feld auch mit der neuen DS-GVO in einer Grauzone bewegt. Weiterhin wird auf das im wissenschaftlichen Kontext relevante Feld der „Open Data“, d.h. auf die wissenschaftliche Nachnutzung von in einem wissenschaftlichen Kontext erhobenen Daten in anderen Forschungsprojekten, und Big Data-Forschung, insbesondere auch im Hinblick auf Biobanken, eingegangen.

■ In this article the changes in regards to scholarship as a result of the changed EU-Data Protection Basic Regulation (DS-GVO) will be analyzed. In particular, a special focus will be on the general changes (rules regarding earmarking, anonymization and pseudonymization provisions), research with sensitive data, in particular in the context of individualized medicine and genetic research. It will be argued that research also is in a gray area with the new DS-GVO. Furthermore, the article will consider “open data”, which is a relevant field in the context of scholarship, i.e. the scholarly re-use of data collected in another research project, and Big Data Research, in particular in regards to bio-banks.

Lesedauer: 11 Minuten

## I. Einleitung

Für wissenschaftliche Forschung gilt, dass – wie auch der *Rat für Sozial- und Wirtschaftsdaten (RatSWD)* in einer Erklärung i.R.d. Konzeption der neuen DS-GVO forderte – die persönlichen Daten von Bürgern vor Missbrauch geschützt bleiben müssen. Dies sei für die empirische Forschung essenziell, denn Teilnehmer an wissenschaftlichen Untersuchungen müssten sicher sein, dass ihre Anonymität gewahrt wird. Ansonsten könne es zu einem Vertrauensverlust und damit auch zu einer geringeren Akzeptanz wissenschaftlicher Forschung kommen.<sup>1</sup>

Dies betrifft vor allem die Teilnahmebereitschaft. Denn wer möchte z.B. Informationen über seine politische Orientierung preisgeben oder an medizinischen Untersuchungen zu Forschungszwecken teilnehmen, wenn er nicht sicher sein kann, dass diese Informationen vertraulich bleiben und nur zu Forschungszwecken verwendet werden? Insofern sind die datenschutzrechtlichen Rahmenbedingungen für die Wissenschaft von großem Belang.

Zum einen, was jedwede herkömmliche qualitative und quantitative empirische Forschung mit Personendaten, seien sie sozialwissenschaftlicher, psychologischer oder medizinischer Art, betrifft. Zum anderen aber auch, was die großen neuen Felder „individualisierte“ Medizin, Open Data und Big Data-Ansätze in der Forschung betrifft.

Am 15.12.2015 hat sich der *Europäische Rat* auf die neue DS-GVO geeinigt.<sup>2</sup> Im Mai 2016 soll sie nach der Verabschiedung im *Europäischen Parlament* offiziell in Kraft treten und ab Anfang bis Mitte 2018 nach der Integration in die Gesetzgebungen der Mitgliedstaaten angewendet werden. Sie löst dann die bisher gültige Europäische Datenschutzrichtlinie (DS-RL) aus dem Jahr 1995<sup>3</sup> ab und versucht den aktuellen Herausforderungen, die aus Globalisierung, neuen Technologien wie Cloud Computing und Big Data resultieren, einen angemessenen datenschutzrechtlichen Rahmen zu geben.

Damit gelten dann für die gesamte Europäische Union einheitliche Regelungen. Dies war bislang nicht der Fall, denn die alte

Richtlinie wurde in den Mitgliedstaaten sehr unterschiedlich umgesetzt. Mit der neuen Verordnung, die seit 2012 vorbereitet und nun abgestimmt wurde, wird es erstmals einheitliche EU-weite Vorgaben geben, die auch für die Forschung gelten.

## II. Erweiterung der Zweckbindung

Eine wesentliche Änderung der neuen DS-GVO besteht darin, dass die Zweckbindungsregelung differenziert und erweitert wird. In der DS-RL hieß es noch, dass die Verarbeitung personenbezogener Daten dem „angestrebten Zweck zu entsprechen, dafür erheblich zu sein und nicht darüber hinauszugehen“ habe, wobei „die Zwecke ... eindeutig und rechtmäßig sein und bei der Datenerhebung festgelegt“<sup>4</sup> werden müssen.<sup>5</sup> Die neue Verordnung ermöglicht hingegen die Einwilligung in „Bereiche“ wissenschaftlicher Forschung: „Oftmals kann der Zweck der Datenverarbeitung für Zwecke der wissenschaftlichen Forschung zum Zeitpunkt der Datenerhebung nicht vollständig angegeben werden. Daher sollten betroffene Personen ihre Einwilligung für bestimmte Bereiche wissenschaftlicher Forschung geben können, wenn die anerkannten ethischen Standards der wissenschaftlichen Forschung eingehalten werden. Die betroffenen

<sup>1</sup> RatSWD, Stellungnahme des Rates für Sozial- und Wirtschaftsdaten zu ausgewählten Punkten des aktuellen Entwurfs der EU-Datenschutz-Grundverordnung (DSGVO) v. 24.2.2015, abrufbar unter: [http://www.ratswd.de/dl/RatSWD\\_Stellungnahme\\_EUDSGVO.pdf](http://www.ratswd.de/dl/RatSWD_Stellungnahme_EUDSGVO.pdf); alle in den Fußnoten angegebenen URLs wurden zuletzt abgerufen am 14.3.2016.

<sup>2</sup> Diesem Artikel ist die derzeitige offizielle deutsche Arbeitsfassung v. 28.1.2016 zu Grunde gelegt; vgl. *Rat der Europäischen Union*, Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten und zum freien Datenverkehr (Datenschutz-Grundverordnung – DS-GVO) [erste Lesung] – Politische Einigung. 5455/16. 28.01.2016., abrufbar unter: [http://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CONSIL:ST\\_5455\\_2016\\_INIT&from=EN](http://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CONSIL:ST_5455_2016_INIT&from=EN).

<sup>3</sup> RL 95/46/EG des Europäischen Parlaments und des Rates v. 24.10.1995 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten und zum freien Datenverkehr (DS-RL), abrufbar unter: <http://eur-lex.europa.eu/legal-content/DE/TXT/?uri=celex%3A31995L0046>.

<sup>4</sup> Erwägungsgrund 28 der DS-RL.

<sup>5</sup> Abschn. I, Art. 6 Abs. 1 DS-RL.

Personen sollten Gelegenheit erhalten, ihre Einwilligung nur für bestimmte Forschungsbereiche oder Teile von Forschungsprojekten in dem vom verfolgten Zweck zugelassenen Maße zu erteilen.“<sup>6</sup>

In einer ersten Stellungnahme wertete der *RatSWD* dies als eine für die Wissenschaft positive Veränderung.<sup>7</sup> In Einverständniserklärungen für Teilnehmer wissenschaftlicher Studien könnten damit bestimmte Bereiche festgelegt werden, die auch über eine spezifische Forschungsfrage hinausgehen. Werden darüber hinausgehende Nutzungen in Erwägung gezogen, müssten Teilnehmer jedoch weiterhin vorab informiert werden.<sup>8</sup>

### III. Anonymisierung und Pseudonymisierung

Auch die im Vorfeld länger diskutierte Frage, welche Anforderungen an die Anonymisierung wissenschaftlicher Daten zu stellen sind, ist nun geklärt. Entgegen dem ursprünglichen Entwurf wird eine vollkommene Anonymisierung nicht mehr ausdrücklich gefordert. Eine solche Vorgabe wäre etwa bei der Durchführung von Längsschnittstudien sehr problematisch gewesen, da neue Erhebungswellen nicht mehr mit bereits bestehenden Daten von Teilnehmern hätten in Verbindung gebracht werden können.<sup>9</sup>

Tatsächlich wird die Pseudonymisierung von Teilnehmerinformationen neben der Anonymisierung nach der neuen Regelung als ein mögliches alternatives Mittel zur völligen Anonymisierung angesehen, sofern gewährleistet ist, dass „die Bestimmung von betroffenen Personen nicht oder nicht mehr möglich ist.“<sup>10</sup> Allerdings muss dabei abgewogen werden, ob eine Person tatsächlich nicht mehr bestimmbar ist, und hierfür sollten auch „alle objektiven Faktoren, wie die Kosten der Identifizierung und der dafür erforderliche Zeitaufwand, herangezogen werden, wobei sowohl die zum Zeitpunkt der Verarbeitung verfügbare Technologie als auch die technologische Entwicklung“ berücksichtigt werden sollen.<sup>11</sup>

Dies entspricht im Wesentlichen der bereits im deutschen Datenschutzrecht verwendeten *Maxime*, die keine absolute, sondern bloß eine faktische Aufhebung des Personenbezugs fordert. § 3 Abs. 6 BDSG definiert als „Anonymisieren das Verändern personenbezogener Daten derart, dass die Einzelangaben über persönliche oder sachliche Verhältnisse nicht mehr oder

nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmbar natürlichen Person zugeordnet werden können.“<sup>12</sup> Bei der Festlegung der jeweiligen Anonymisierungsverfahren muss dementsprechend den konkreten Risiken der De-Anonymisierung Rechnung getragen werden, wobei auch wirtschaftliche und technologische Faktoren zu berücksichtigen sind. Dies betrifft auch die schnelle technologische Entwicklung, die dazu führen kann, dass ein Verfahren, das gegenwärtig eine ausreichende Anonymität der Daten garantiert, ggf. in wenigen Jahren diesen Schutz nicht mehr gewährleistet. Insofern müssen die Anonymisierungsverfahren eine Art „Schutzreserve“ beinhalten, die zukünftigen Risiken vorbeugt.

### IV. Forschung mit sensiblen Daten in der Grauzone

Die personalisierte Medizin ist eines der großen Forschungsgebiete, das ohne sehr sensible Daten nicht auskommt. Sie wird durch die Forschungsprogramme der EU und auch innerhalb Deutschlands mit Forschungsgeldern in großem Umfang unterstützt. Z.B. wird personalisierte Medizin und der „verstärkte Einsatz von Datenbanken und elektronischen Patientenakten als Datenquellen für Versuche und Wissenstransfer“ im europaweiten Forschungsprogramm Horizon 2020 gefördert<sup>13</sup> und in zahlreichen neuen Ausschreibungen adressiert.<sup>14</sup> Auch die *Bundesregierung* fördert in ihrem Gesundheitsrahmenprogramm die personalisierte, individualisierte Medizin.<sup>15</sup>

Dabei ist personalisierte Medizin vielfach auf genetische Analysen angewiesen, die es ermöglichen, abzuschätzen, welche Therapien bei Menschen auf Grund ihrer genetischen Disposition voraussichtlich anschlagen werden oder nicht. Die DS-GVO definiert in Art. 4 Abs. 10 genetische Daten wie folgt: „genetische Daten‘ [sind] personenbezogene Daten jedweder Art zu den ererbten oder erworbenen genetischen Merkmalen eines Menschen, die eindeutige Informationen über die Physiologie oder die Gesundheit dieses Menschen liefern und insbesondere aus der Analyse einer biologischen Probe des betreffenden Menschen gewonnen wurden.“<sup>16</sup>

Das bedeutet – jedenfalls wenn sie detailliert erfasst werden –, dass sie an sich nicht anonymisierbar sind, da die enthaltenen Informationen per se personenbezogen sind.<sup>17</sup> Genetische Daten sind einzigartig und selbst wenn alle anderen sie direkt identifizierenden Merkmale entfernt werden, sind sie im Prinzip wieder einer Person zuzuordnen.<sup>18</sup> Genomforschung und damit aber auch wiederum die personalisierte Medizin ist damit datenschutzrechtlich riskant. Nach Art. 9 DS-GVO dürfen besonders sensible Daten – soweit keine ausdrückliche Einwilligung des Betroffenen vorliegt – nur im öffentlichen Interesse oder im Interesse des Einzelnen erhoben und verarbeitet werden, z.B. im Kontext von wissenschaftlichen Forschungsprojekten und „Studien, die im öffentlichen Interesse im Bereich der öffentlichen Gesundheit durchgeführt werden“.

Wie sie gehalten und verwendet werden, muss damit strengen Regeln unterliegen. Die Mitgliedstaaten können aber auch „weitere Bedingungen – einschließlich Beschränkungen – in Bezug auf die Verarbeitung von genetischen Daten, biometrischen Daten oder Gesundheitsdaten“ beibehalten oder einführen.<sup>19</sup>

### V. Open Data und Big Data

Gleichzeitig sind die Aussichten darauf, durch die Zusammenführung großer Datenbestände aus unterschiedlichen Quellen neue Erkenntnisse zu gewinnen, groß und die Erwartungen hoch. „Wissenschaftler wollen“, so wird es in der digitalen Agenda der *Bundesregierung* beschrieben, „unkompliziert wissenschaftliche Informationen austauschen und über Länder-

<sup>6</sup> Erwägungsgrund 25aa der DS-GVO (o. Fußn. 2).

<sup>7</sup> *RatSWD*, PM: EU-Datenschutz: Drohende Gefährdung der empirischen Sozial- und Wirtschaftsforschung abgewendet, Berlin 23.12.2015, abrufbar unter: <http://www.ratswd.de/pressemitteilung/23122015>.

<sup>8</sup> Erwägungsgrund 49 der DS-GVO (o. Fußn. 2), sowie ebenda Art. 14, Abschn. 1b.

<sup>9</sup> *RatSWD* (o. Fußn. 1).

<sup>10</sup> Art. 83 Abs. 1 DS-GVO (o. Fußn. 2).

<sup>11</sup> Erwägungsgrund 23 der DS-GVO (o. Fußn. 2).

<sup>12</sup> *BMJV*, BDSG, 25.2.2015, S. 5.

<sup>13</sup> *EU-Kommission*, Vorschlag für Beschluss des Rates über das spezifische Programm zur Durchführung des Rahmenprogramms für Forschung und Innovation „Horizont 2020“ (2014-2020). Kom(2011)811 endg.; 2011/0402 (CNS) 30.11.2011, S. 60, abrufbar unter: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2011:0811:FIN:de:PDF>.

<sup>14</sup> *EU-Kommission*, Horizon 2020: Work Programme 2016-2017. European Commission Decision C (2015) 13.10.2015, S. 7 ff., abrufbar unter: [www.http://ec.europa.eu/research/participants/data/ref/h2020/wp/2016\\_2017/main/h2020-wp1617-health\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/wp/2016_2017/main/h2020-wp1617-health_en.pdf).

<sup>15</sup> *BMBF*, Rahmenprogramm Gesundheitsforschung der Bundesregierung, 2010, S. 5, abrufbar unter: [http://www.gesundheitsforschung-bmbf.de/\\_media/Gesundheitsforschungsprogramm.pdf](http://www.gesundheitsforschung-bmbf.de/_media/Gesundheitsforschungsprogramm.pdf).

<sup>16</sup> Art. 4 Abs. 10 DS-GVO (o. Fußn. 2).

<sup>17</sup> Vgl. auch von *Hardenberg*, ZD 2014, 117.

<sup>18</sup> Vgl. dazu z.B. auch *Arning u.a.*, Deutsches Ärzteblatt 108, 2011; *Gymrek u.a.*, „Identifying personal genomes by surname inference“, *Science*, 339, 2013, 321-324, doi: 10.1126/science.1229566.

<sup>19</sup> Erwägungsgrund 42a der DS-GVO (o. Fußn. 2).

grenzen hinweg zusammenarbeiten.“<sup>20</sup> Dem liegt die plausible Annahme zu Grunde, dass die Aussagekraft steigt, wenn Fragestellungen studienübergreifend verfolgt werden können. Und es ist im Sinne der Wissenschaft, wenn Forschung, die mit öffentlichen Mitteln gefördert wurde, auch anschließend für andere wissenschaftliche Fragestellungen zur Verfügung steht. Ein zumindest für andere Wissenschaftler offener Zugang (Open Access) auch zu Forschungsdaten (Open Data) wird deshalb von verschiedenen Institutionen gefördert und mittlerweile sogar gefordert.

Die *EU-Kommission* will den offenen Zugang zu wissenschaftlichen Daten durch eine Experimentalphase einleiten.<sup>21</sup> Auch die Vorgaben für deutsche Forschungsprogramme des *BMBF* und der *DFG* fordern zunehmend einen Open Access nicht nur für publizierte Ergebnisse, sondern auch für die erhobenen Daten.

In Richtung von Big Data gehen außerdem Ansätze, wissenschaftliche Daten mit Daten z.B. der Krankenkassen oder der Sozialversicherung zu verknüpfen. In der DS-GVO findet die Verknüpfung von Daten mit medizinischen und sozialwissenschaftlichen Registern Erwähnung. Hier wird in der Verordnung davon ausgegangen, dass es einen großen Mehrwert hinsichtlich der Erkenntnisse über Krankheiten oder soziale Umstände gibt und „daher personenbezogene Daten zu wissenschaftlichen Forschungszwecken verarbeitet werden [können], wobei sie angemessenen Bedingungen und Garantien unterliegen [müssen], die im Recht der Mitgliedstaaten oder im Unionsrecht festgelegt sind.“<sup>22</sup>

Zur Zusammenführung und dem i.S.d. Open Data Zugänglich-Machen von wissenschaftlichen Daten in Repositorien werden in der DS-GVO keine Aussagen gemacht. Sofern die Daten vollständig anonymisiert oder in einer Weise pseudonymisiert sind, die eine Bestimmung der Personen nicht möglich macht, wird es hier vermutlich keine Risiken für Teilnehmer an wissenschaftlichen Studien und auch keinen Dissens zu datenschutzrechtlichen Vorgaben geben. Technische Maßnahmen sollen „sicherstellen, dass personenbezogene Daten grundsätzlich nicht ohne Eingreifen einer natürlichen Person einer unbestimmten Zahl von natürlichen Personen zugänglich gemacht werden.“<sup>23</sup>

Ein vollständig automatisierter Zugriff auf Massendaten oder deren Auswertung mit einer „Joker-Anfrage“, bei der wie bei einer Rasterfahndung eine Vielzahl von Datensätzen selektiert werden können, die bestimmten Merkmalen entsprechen, ist ausgeschlossen. Ungeklärt ist allerdings, wie mit dem Open Data-Konzept für genetische Daten umgegangen werden kann. Schon jetzt werden, um Aussagen über genetische Dispositionen zu erhalten, zunehmend Daten aus verschiedenen Forschungsprojekten in internationalen Biodatenbanken<sup>24</sup> zusammengeführt,<sup>25</sup> denn erst dann lassen sich Muster und Zusammenhänge überhaupt erkennen. Dabei geht es nicht mehr nur um medizinische Fragestellungen, sondern auch um Analysen, die eine sozialwissenschaftliche Relevanz haben.<sup>26</sup> Doch das Problem besteht darin, dass genetische Daten eben auch nach der Erwägung des *Europäischen Rats* persönliche Daten sind,<sup>27</sup> prinzipiell reidentifizierbar sind und sich die Frage stellt, inwiefern sie überhaupt in Biobanken mit weitreichender Nutzung durch verschiedene Anwender überführt werden sollten. Zwar könnte mit entsprechend gestalteten Einverständniserklärungen

eine Überführung der Daten in große Biobanken durch die Teilnehmer erlaubt werden, doch ist unklar, wie freiheitliche Grundrechte von Teilnehmern dort tatsächlich geschützt werden können. Denn Biobanken selbst haben – wie etwa die UK-Biobank – teilweise keine Beschränkung für kommerzielle Nutzung. Und wie gut die Daten vor Reidentifizierung und Verknüpfung mit weiteren Informationen geschützt sind, ist offen.

## VI. Ausblick

Insgesamt räumt die DS-GVO mit Art. 83 den Mitgliedstaaten ein, dass sie „Garantien und Ausnahmen“ für die Wissenschaft gesetzlich festschreiben können. Diese betreffen etwa Einschränkungen im Auskunftsrecht (Art. 15), dem Recht auf Berichtigung (Art. 16), dem Recht auf Einschränkung der Verarbeitung (Art. 17a) und dem Widerspruchsrecht (Art. 19).

Die gemachten Vorgaben können also weiter und enger gefasst und präzisiert werden. Die Mitgliedstaaten können auch die genauen Bestimmungen zur Nutzung besonders sensibler Daten regeln. So heißt es im Erwägungsgrund 42a: „Den Mitgliedstaaten sollte gestattet werden, weitere Bedingungen einschließlich Beschränkungen in Bezug auf die Verarbeitung von genetischen Daten, biometrischen Daten oder Gesundheitsdaten beizubehalten oder einzuführen.“<sup>28</sup>

Angesichts der beschriebenen Risiken wäre es sinnvoll, wenn durch die Gesetzgebung der Mitgliedsländer hier präzise Regelungen getroffen werden. Wie die Garantien und Ausnahmen und die spezifischen Regelungen dann im Detail aussehen werden, ist derzeit allerdings noch offen.



Dr. phil. Katrin Schaar

ist wissenschaftliche Mitarbeiterin am Institut für Psychologie, Humboldt-Universität zu Berlin, Projekt BASE-II.

**20** Bundesregierung, „Digitale Agenda – Bildung, Forschung, Wissenschaft, Kultur und Medien“, abrufbar unter: <[https://www.digitale-agenda.de/Webs/DA/DE/Handlungsfelder/5\\_BildungForschung/bildung-forschung\\_node.html](https://www.digitale-agenda.de/Webs/DA/DE/Handlungsfelder/5_BildungForschung/bildung-forschung_node.html)>;

**21** Press Release: Scientific data: Open access to research results will boost Europe's innovation capacity, v. 17.7.2012, S. 1, abrufbar unter: [http://europa.eu/rapid/press-release\\_IP-12-790\\_en.htm](http://europa.eu/rapid/press-release_IP-12-790_en.htm); „A novelty in Horizon 2020 is the Pilot on Open Research Data which aims to improve and maximise access to and reuse of research data generated by projects. While certain Work Programme parts and calls have been explicitly identified to participate in the Pilot on Open Research Data, individual projects funded under the other Work Programme parts and calls can choose to participate in the Pilot on a voluntary basis. Participating projects will be required to develop a Data Management Plan (DMP), in which they will specify what data the project will generate, whether and how it will be exploited or made accessible for verification and reuse, and how it will be curated and preserved. Further guidance on the Pilot Open Research Data and Data Management is available on the Participant Portal“; *EU-Kommission* (o. FuBn. 14), S. 6.

**22** Erwägungsgrund 125aa der DS-GVO (o. FuBn. 2), S. 72.

**23** Art. 23 Abs. 2 DS-GVO (o. FuBn. 2).

**24** Biobank UK, vgl. <<http://www.ukbiobank.ac.uk/>>;

**25** Resource for Genetic Epidemiology Research on Adult Health and Aging (GERA), abrufbar unter: [http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs000674.v1.p1](http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000674.v1.p1).

**26** Vgl. z.B. C. A. Rietveld u.a., GWAS of 126,559 Individuals Identifies Genetic Variants Associated with Educational Attainment, *Science* 340, 2013, 1467-1471, doi: 10.1126/science.1235488.

**27** Vgl. o. FuBn. 2.

**28** Erwägungsgrund 42 der DS-GVO (o. FuBn. 2).