www.jusletter-it.eu

Peter Parycek / Johann Höchtl / Bettina Rinnerbauer

Zur Datenschutzrechtskonformität von Big Data Analysen der Verwaltung

Klassifikation von Verwaltungsdaten als Big Data, datenschutzrechtliche Herausforderungen der Analyse und ein möglicher Lösungsansatz

Government data is a valuable resource for analytic systems. The advent of big data enables new methodologies, processes and techniques to further improve predictive results towards evidence-based policy making for the benefit of the society at large. The legal framework on the one hand is an enabler for data analytics and on the other hand restricts potential use-cases. This article presents the principles of Big Data, discusses administrative registers in respect to Big Data, gives a succinct overview on the personal data protection act in Austria, analyses the methodology of the register-based census as a blueprint for Big Data analytics and concludes with a novel approach towards Big Data analytics using encrypted data on a peer-to-peer network in line with the current personal data protection act.

Category: Scientific Articles
Region: Austria
Field of law: Data Protection; Data Security; Big Data, Open Data Open Government; E-Government; IT-Governance

Citation: Peter Parycek / Johann Höchtl / Bettina Rinnerbauer, Zur Datenschutzrechtskonformität von Big Data Analysen der Verwaltung, in: Jusletter IT 24 September 2015

ISSN 1664-848X, http://jusletter-it.weblaw.ch, Weblaw AG, info@weblaw.ch, T +41 31 380 57 77 EDITIONS WEBLAW

Inhaltsübersicht

1. Einleitung[1]

[Rz 1] Im Zuge des Big Data Trends hat sich die Diskussion in Literatur und Judikatur in den letzten 3 Jahren verstärkt.[2] Der vorliegende Beitrag setzt den Schwerpunkt auf die rechtliche Analyse der Auswertung von Daten, die im Rahmen der Tätigkeit der Verwaltung auf Grund der Gesetze anfallen und stellt die Frage, inwieweit Datenanalysen und insbesondere Big Data Analysen zulässig sein könnten.

[Rz 2] Der Ausschöpfung des Potentials von Datenauswertungen in der Verwaltung steht das Grundrecht des Datenschutzes gegenüber, das jedermann den Anspruch auf Geheimhaltung der ihn betreffenden personenbezogenen Daten gewährt, soweit ein schutzwürdiges Interesse daran besteht.[3] Die rechtliche Qualifikation als personenbezogenes Datum ist daher im gegebenen Zusammenhang von besonderer Relevanz. Da für den Erkenntnisgewinn der Verwaltung als geeignete Datenquellen insbesondere elektronische Register in Betracht kommen, welche unter anderem personenbezogene

1 Zu deskriptiver, präskriptiver und prediktiver Analyse vgl. Jeffrey D. Camm / James J. Cochran / Michael J. Fry / Jeffrey W. Ohlmann / David R. Anderson / Dennis J. Sweeney / Thomas A. Williams, Essentials of Business Analytics, South-Western College Publishing, 2014, 5 ff.

2 Beispielhaft genannt seien Schwerpunkt-Ausgabe «Big Data», in: Jusletter IT 21. Mai 2015; Rolf H. Weber, Big Data: Sprengkörper des Datenschutzrechts?, in: Jusletter IT 11. Dezember 2013; Lukas Feiler / Siegfried Fina, Datenschutzrechtliche Schranken für Big Data, MR 2013, 303; Viktoria Haidinger, Relevante Themenbereiche in der Judikatur zu Big Data, Dako 2015/40 (71).

3 § 1 Bundesgesetz über den Schutz personenbezogener Daten (Datenschutzgesetz 2000 – DSG 2000) BGBl I 165/1999 i.d.F. BGBl I 83/2013 (im Folgenden: DSG).

Daten enthalten, ist es nötig, im Vorfeld eines derartigen Vorhabens für die Einhaltung des österreichischen Datenschutzgesetzes (DSG) Sorge zu tragen. In diesem Beitrag werden Voraussetzungen erläutert, die das DSG an den Umgang mit personenbezogenen Daten stellt.

[Rz 3] Der Beitrag kontextualisiert in Abschnitt 2 das Themenfeld Big Data und Analysen aus datenschutzrechtlicher Perspektive, diskutiert Verwaltungsregister als Quellen für Big Data und erläutert Big Data Konzepte, die im Verlauf der weiteren Untersuchung essentiell sind. Abschnitt 3 dient der Motivation indem mögliche Anwendungsfelder von Big Data in der öffentlichen Verwaltung ausgehend von bestehenden, traditionellen Verfahren vorerst im Blickwinkel von Big Data betrachtet werden und darauf basierend neue Methoden und Verfahren dargestellt werden, ohne Herausforderungen und Hürden zu verschweigen. Die rechtliche Auseinandersetzung in Abschnitt 4 nähert sich den kontroversiell diskutierten Begriffen personenbezogene vs. nicht personenbezogene Daten, erläutert Anonymisierung und Pseudnoymisierung und nennt rechtlich zulässige Anwendungsbereiche umfangreicher, auch mitunter personenbezogener Auswertungen. Abschnitt 5 setzt sich detailliert mit der Methode der Registerzählung als eine rechtlich zulässige Methode statistischer Auswertung auseinander, um in Abschnitt 6 ein neuartiges Konzept der Datenanalyse zu präsentieren, welches geeignet erscheint, eine neue Sichtweise auf das Themenfeld Big Data und Datenschutz zu werfen und die rechtliche Diskussion durch eine neue Betrachtungsweise zu bereichern. Mit Abschnitt 7 schließlich endet die kritische Auseinandersetzung des aktuellen Themas und gibt Ausblick auf weitere erforderliche Betrachtungen und mögliche Handlungsalternativen.

2. Begriffsdefinitionen und Big Data Grundlagen

[Rz 4] Um sich dem komplexen, neuartigen und sowohl aus (insbesondere datenschutz-)rechtlicher als auch aus technischer Sicht intensiv diskutierten Themengebiet von Big Data und dessen möglichen Anwendungsbereichen in der öffentlichen Verwaltung zu nähern, ist eine thematische Abgrenzung, Definition relevanter Begriffe und Darlegung essentieller Big Data Grundlagen erforderlich, denen sich dieser Abschnitt widmet.

2.1. Begriffsdefinitionen: Register, Verwaltungsdaten

[Rz 5] Das Bundesstatistikgesetz definiert «Öffentliche Register» als Register, die auf Grund bundesgesetzlicher Bestimmungen der öffentlichen Einsicht unterliegen.[4] Die Bedeutung des Begriffes «**Register**» wurde – soweit ersichtlich – durch den Gesetzgeber nicht festgelegt.

[Rz 6] Im Alltag wird unter «Register» ein Verzeichnis verstanden, dessen Inhalt oft alphabetisch gereiht ist und etwa in einer Bibliothek Verwendung finden kann. Ein Register des öffentlichen Bereichs hat keine eindeutige Struktur, sondern kann vielmehr verschiedene Formen annehmen. Merkmale, die *öffentliche elektronische Register* kennzeichnen, sind insbesondere die Datensammlung in strukturierter Art. bezogen auf eine bestimmte Materie, kombiniert mit Attributen zur eindeutigen Identifizierung und eine gesetzlich festgelegte Aktualisierung.[5]

---

[4] § 3 Z 18 Bundesgesetz über die Bundesstatistik (Bundesstatistikgesetz 2000), BGBl I 163/1999 i.d.F. BGBl I 40/2014.

[5] Arthur Winter, Zentrale Registerlösungen im föderalen Bundesstaat des 21. Jahrhunderts, in: Erich Schweighofer / Franz Kummer (Hrsg.), Europäische Projektkultur als Beitrag zur Rationalisierung des

[Rz 7] **Verwaltungsdaten** sind gemäß Bundesstatistikgesetz Daten, die bei Stellen in Wahrnehmung von bundes- oder landesgesetzlich übertragenen Aufgaben oder in Vollziehung unmittelbar anwendbarer gemeinschaftsrechtlicher Vorschriften angefallen sind.[6]

[Rz 8] Davon ausgehend, können Register unter Verwaltungsdaten subsumiert werden.

2.2. Öffentliche Register als Quelle für Big Data Auswertungen

[Rz 9] Es gibt keine allgemein gültige Definition für **Big Data**. Dennoch herrscht weitgehend Einigkeit darüber, dass sich Big Data durch Größe (Volume), Geschwindigkeit (Velocity) und Verschiedenartigkeit bzw. Unstrukturiertheit (Variety) auszeichnet[7]. In der Folge wurden mit Variability, Veracity, Visualisation und Value weitere «V's»[8] dieser Klassifikation hinzugefügt. Eine Meta-Definition beschreibt Big Data als «*data whose size forces us to look beyond the tried-and-true methods that are prevalent at that time*»[9]

[Rz 10] Attribute wie «Größe» und «Geschwindigkeit» sind nicht exakt definiert und unterliegen laufenden Veränderungen, wodurch die Überprüfung hinsichtlich Register zu keinem eindeutigen Ergebnis führt: Wie groß «Big Data» nun ist, d.h. ob es sich hier z.B. um Gigabytes, Terabytes oder Petabytes handelt, steht nicht fest[10]. Auch die Geschwindigkeit von Netzwerken, denen im Zusammenhang mit der Verarbeitung von großen Datenmengen eine wichtige Rolle zukommt, verdoppelt sich alle 6 Monate[11], sodass dieses Merkmal zur Klassifikation von Registern als Big Data keine allgemeingültige Aussagekraft besitzt.

[Rz 11] Einfacher ist die Feststellung, dass öffentliche elektronische Register tendenziell strukturierte Daten enthalten[12], während Big Data durch Unstrukturiertheit charakterisiert ist. Untersu-

---

Rechts. Tagungsband des 14. Internationalen Rechtsinformatik Symposions IRIS 2011 bzw. Jusletter IT 24. Februar 2011 (Rz 1–2).

6 § 3 Z 17 Bundesstatistikgesetz.

7 Douglas Laney, 3D Data Management: Controlling Data Volume, Velocity, and Variety, 949. Stamford, Connecticut: META Group, 2001 (http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf [alle Internetquellen zuletzt besucht am 17. September 2015]). Vgl. auch Lukas Feiler / Siegfried Fina, Datenschutzrechtliche Schranken für Big Data, MR 2013, 303 (303); Andrew McAfee / Erik Brynjolfsson, Big Data: The Management Revolution, Harvard Business Review, October 2012, 61 (62–63); Astrid Epiney, Big Data und Datenschutzrecht: Gibt es einen gesetzgeberischen Handlungsbedarf?, in: Jusletter IT 21. Mai 2015 (Rz 6).

8 Eileen Mc Nulty, Understanding Big Data: The Seven V's, 2014 (http://dataconomy.com/seven-vs-big-data/).

9 Adam Jacobs, The Pathologies of Big Data, Communications of the ACM Vol 52 Nr. 8 (August 2009), 44 (https://queue.acm.org/detail.cfm?id=1563874). Als «*a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data by enabling high velocity capture, discovery, and/or analysis*» wird Big Data von John Gantz / David Reinsel, The Digital Universe in 2020: Big Data, Bigger Digital Shadows and Biggest Growth in the Far East, IDC IVIEW, December 2012, 9 (http://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf) definiert.

10 Vgl. zu dieser Entwicklung etwa Martin Hilbert / Priscila López, The World's Technological Capacity to Store, Communicate, and Compute Information, Science 332, 60 (2011) (http://www.uvm.edu/{~}pdodds/files/papers/others/2011/hilbert2011a.pdf); Adam Jacobs, The Pathologies of Big Data, Communications of the ACM Vol 52 Nr. 8 (August 2009), 44 (https://queue.acm.org/detail.cfm?id=1563874).

11 George Gilder, Metcalfe's law and legacy, Forbes ASAP 13 (1993).

12 Arthur Winter, Zentrale Registerlösungen im föderalen Bundesstaat des 21. Jahrhunderts, in: Erich Schweighofer / Franz Kummer (Hrsg.), Europäische Projektkultur als Beitrag zur Rationalisierung des Rechts. Tagungsband des 14. Internationalen Rechtsinformatik Symposions IRIS 2011 bzw. Jusletter IT 24. Februar 2011 (Rz 1–2).

chungen sagen weiters voraus, dass der Großteil neu generierter Daten unstrukturiert sein wird[13], automatisierte Sensorerhebungen tragen zu dieser Vielfalt bei. In Zukunft könnte eine ansteigende Menge an unstrukturierten Verwaltungsdaten die Feststellung der Strukturiert- oder Unstrukturiertheit von Registern erschweren.

[Rz 12] Eine abschließende Aussage, ob Register als Big Data einzuordnen sind kann nicht getroffen werden. Werden Aspekte wie der Verwendungszweck[14], die Möglichkeit der Zusammenschau mehrere Verwaltungsregister, zeitserienorientierte Auswertungen und die Verwendung verwaltungsexterner Daten mit in die Betrachtung miteinbezogen, ist eine Würdigung von Registern im Kontext von Big Data jedenfalls indiziert.

2.3. Big Data Methoden

[Rz 13] Die im Umfeld von Big Data Analysen diskutierten Methoden, Verfahren und technologischen Rahmenbedingungen unterscheiden sich nur unwesentlich von jenen des Decision Support oder der Business Intelligence und umfassen unter anderem folgende wichtige Komponenten:

[Rz 14] **Data Mining** ist das intelligente, häufig automatisierte Aufspüren und die Extraktion von interessanten und relevanten Mustern und Zusammenhängen in großen Datenbeständen. Die große Herausforderung in diesem Zusammenhang ist die Verarbeitung unstrukturierter, sprachlicher Texte wie sie in E-Mails oder Dokumenten aus Textverarbeitungssystemen enthalten sind, aber auch intonierte Sprache und Videos. Die zusätzliche Komplexität des Erkennens von Gefühls- oder Stimmungslagen wird als **Sentiment Mining** bezeichnet.

[Rz 15] **Maschinelles Lernen** bedeutet das automatisierte Erkennen von Zusammenhängen in Datenbeständen. Es ist ein mächtiges und wichtiges Werkzeug das den Wissensarbeiter sowohl bei der Aufbereitung von Daten, als auch deren Auswertung unterstützt. Während Data Mining noch ein großes anwendungsspezifisches Wissen des Wissensarbeiters voraussetzt, soll der Analyst beim maschinellen Lernen mit Hilfe von Algorithmen bei der Identifikation neuer Erkenntnisse vollständig unterstützt werden.

[Rz 16] **Genetische Algorithmen** sind eine Reaktion auf die wachsende Komplexität, die an Auswertungen gerichtet werden. Die wachsende Zahl an Eingabeparametern zur Steigerung der Entscheidungsqualität führt zunehmend dazu, dass Algorithmen mit den traditionellen Mitteln der Informatik nicht mehr formuliert werden können. Ein Wissensarbeiter assistiert und trainiert den genetischen Algorithmus, indem er Ergebnisse des Algorithmus nach deren Ergebnisqualität reiht wodurch der Algorithmus zwischen guten und schlechten Ergebnissen unterscheiden kann und in der weiteren Ausführung selbsttätig die Berechnungsvorschrift ändert. Genetische Algorithmen bilden eine Teilmenge der Algorithmen des maschinellen Lernens.

[Rz 17] **Künstliche Intelligenz** ist der Versuch menschenähnliches Verhalten, vor allem im kognitiven Bereich, durch Computersysteme zu emulieren. Im Bereich der Vorhersage menschlichen Verhaltens kommt ihnen eine besondere Bedeutung zu, da menschliches Verhalten nicht immer rational und begründbar und dadurch schwierig in Algorithmen zu formulieren ist. Diese Unschärfe

---

13 John Gantz / David Reinsel, The Digital Universe in 2020: Big Data, Bigger Digital Shadows and Biggest Growth in the Far East, IDC IVIEW, December 2012, 9 (http://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf).

14 Adam Jacobs, The Pathologies of Big Data, Communications of the ACM Vol 52 Nr. 8 (August 2009), 36 (https://queue.acm.org/detail.cfm?id=1563874).

und Irrationalität wird durch Methoden der künstlichen Intelligenz berücksichtigt.

[Rz 18] **Interaktive Visualisierungen** erfüllen zwei wesentliche Funktionen: Einerseits unterstützen sie den Wissensarbeiter beim Erkennen von Datenmustern, wobei die Interaktivität das Verändern einzelner oder mehrerer Auswertungsparameter wie den Beobachtungszeitraum oder die Granularität der Datenauflösung bedeuten. In dieser Funktion ergänzen oder komplementieren sie Algorithmen des maschinellen Lernens, da die Intuition und Erfahrung des Wissensarbeiters noch nicht ausreichend in Algorithmen abgebildet werden kann. Andererseits sind interaktive Visualisierungen ein wichtiges Instrument der Ergebnisvermittlung an Entscheider.

[Rz 19] **Streaming Verfahren** sind eine technologische Weiterentwicklung des Datenverarbeitungsprozesses, der lange Zeit vom statischen Denken (Input, Verarbeitung, Output) geprägt war. Bei der stapelorientierten Verarbeitung wird aus der Datenbasis ein Abzug erstellt und gelangt anschließend zur Auswertung, in Zwischenschritten besteht die Möglichkeit, Konsistenz-, Fehler-, und Vollständigkeitsprüfungen durchzuführen. Im Streaming Verfahren werden Daten nicht zuerst gesammelt um sie zu verarbeiten, sondern in oder nahe Echtzeit in den Modellen, in welchen sie verwendet werden, aktualisiert womit Ergebnisse sehr zeitnahe oder in Echtzeit zur Verfügung stehen.

[Rz 20] **Verteilte Datenhaltung und Datenverarbeitung** wird vor allem durch eine **Cloud-Infrastruktur** ermöglicht. Daten werden aus Performance- und Datensicherheitsgründen verteilt und redundant gehalten und örtlich möglichst dort ausgewertet, wo sie auch gespeichert sind.

[Rz 21] Die genannten Komponenten können in unterschiedlicher Intensität von der öffentlichen Verwaltung in Big Data Anwendungsfeldern genutzt werden. Mögliche Anwendungsfelder und daraus erwachsende Herausforderungen und Hürden widmet sich der folgende Abschnitt.

3. Big Data Anwendungsfelder in der Verwaltung

[Rz 22] Dieser Abschnitt behandelt bestehende Konzepte der Business Intelligence und des Decision Support und deren Reinterpretation im Licht von Big Data, präsentiert gänzlich neue Möglichkeiten, die durch den technologischen Fortschritt ermöglicht wurden, und diskutiert Herausforderungen und Hürden des Einsatzes der neuen Perspektiven.

3.1. Traditionelle Analyseverfahren im Licht von Big Data

[Rz 23] Bekannte Analyseverfahren erhalten im Kontext der durch Big Data eröffneten Möglichkeiten erweiterte Perspektiven. Dieser Abschnitt behandelt traditionelle Analyseverfahren, die im Zuge von Business Intelligence und des Decision Supports bereits Gegenstand des wissenschaftlichen Diskurses waren[15] und durch Big Data Methoden neue Aktualität, erweiterten Funktionsumfang und breitere Anwendungsbereiche erfahren.

[Rz 24] **Simulationen** modellieren vereinfachte, komplexe Zusammenhänge der Realität. Unter dem Aspekt von Big Data wird es möglich, die darunterliegenden Modelle besser an reale Gegebenheiten anzupassen und unter Ausnutzung gesteigerter Rechnerkapazitäten, wie sie von verteil-

15 Thomas H. Davenport, Big data at work: dispelling the myths, uncovering the opportunities. Boston Massachusetts: Harvard Business Press, 2014.

ten Cloud-Infrastrukturen bereitgestellt werden, zu wesentlich schnelleren Ergebnissen bis hin zu Echtzeitsimulationen zu gelangen. Einsatzbeispiele wären optimierte Verkehrsleitsysteme, die neben bestehenden Sensordaten wie der Verkehrsdichte, aktuelle Wetterdaten und Bewegungsdaten in Mobilfunkzellen angemeldeter Verkehrsteilnehmer miteinbezieht.

[Rz 25] **Prognosen** liefern Aussagen über Eintrittswahrscheinlichkeiten der Zukunft. Ein Prognosemodell wird durch die Anzahl der Parameter bestimmt und die Prognosegüte steigt mit der Anzahl der berücksichtigten Variablen. Mit zunehmender Multidimensionalität stoßen traditionelle Prognoseverfahren an ihre Grenzen und werden durch innovative Big Data Methoden, wie genetische Algorithmen, ergänzt. Weiters kann die berücksichtigte Datenmenge als Grundlage einer Prognose durch verteilte Datenhaltung und intelligente Datenzugriffsverfahren, wie sie von Big Data Infrastruktur unterstützt wird, gesteigert werden. Anwendungsgebiete in der Verwaltung wären beispielsweise in der Analyse historischer Veränderungen in Flächenwidmungen, die etwa Grünland in urbanes Gebiet verwandelten und die damit einhergehende Gefahr von Überschwemmungen, Hangrutschungen und Vermurungen veränderten, welche Einfluss auf zukünftige Widmungsverfahren haben könnten.

[Rz 26] **Frühwarnsysteme** helfen Entscheidern, rasch und frühzeitig auf drohende Gefahren zu reagieren. Sie vereinen häufig Simulationen und Prognosen und können Gefahrenpotentiale durch Visualisierungen veranschaulichen. Frühwarnsysteme können in vielfältiger Weise in der öffentlichen Verwaltung zum Einsatz kommen und umfassen, je nach ihrer Zielrichtung, wenige bis viele Komponenten des Big Data Methodensatzes: Vermeintlich isolierte Einzelereignisse, wie der Wegzug einer einzelnen oder weniger Personen aus einer in seiner Entwicklung ansonsten stabilen Gemeinde kann der Indikator einer tiefgreifenden Veränderung sein, auf welche frühzeitig reagiert werden sollte.

[Rz 27] Zugeschnittene, **personalisierte Services und Dienstleistungen** steigern die Zufriedenheit von Bürgerinnen und Bürgern mit Verwaltungseinheiten, verminderter Suchaufwand fördert die nationale Wohlfahrt, indem Serviceempfänger rascher jene Dienstleistungen konsumieren, für die sie berechtigt sind und senken zudem die Bearbeitungskosten innerhalb der Verwaltung. Die Personalisierung von Services ist ein informationstechnisch aufwändiger Prozess, der es erfordert, die individuellen Präferenzen einer Person aus bestehenden Angaben wie seiner Lebenslage, Geschlecht, Alter, Wohnort oder bereits konsumierten Dienstleistungen zu extrahieren und mit den Eigenschaften anderer Servicekonsumenten abzugleichen. Eine wesentliche Verbesserung der anzunehmenden konsumierten Services kann des Weiteren erfolgen, indem öffentlich bzw. allgemein verfügbare Informationen über eine bestimmte Person, beispielsweise aus sozialen Netzwerken, in die Erstellung der personalisierten Sicht auf das Serviceangebot mit einbezogen werden. Aus dem Big Data Methodensatz kommen dazu Algorithmen des maschinellen Lernens zum Einsatz, die eine Clusterbildung der möglichen Servicekonsumenten vornehmen und aufgrund der informationsintensiven Datenauswertung eine verteilte Datenverarbeitung erfordern. Ein mögliches Anwendungsszenario wäre ein personalisiertes Serviceangebot der Verwaltung, das aufgrund der in der Vergangenheit in Anspruch genommenen Services mögliche nächste zu konsumierende Services vorhersieht. Der Abschluss des Kaufes eines (entsprechend gewidmeten) Grundstücks könnte auf eine Bauabsicht hindeuten und Muster für Bauanträge und notwendige Schritte für die Inanspruchnahme von Fördermöglichkeiten könnten im Portfolio des möglichen Serviceangebotes höher priorisiert werden. Die Lebenssituation von Bürgern kann Einfluss auf die Gestaltung von Formularen nehmen und somit zu einem personalisierten Serviceerlebnis beitragen. Die Lebenslage Student könnte auf typische Ausgaben hindeuten, die wiederum in Steuererklärungsformularen («die fünf häufigsten

Ausgabegrößen») berücksichtigt werden könnten.

[Rz 28] Big Data Analysen ermöglichen automatisierte oder teilautomatisierte, **einzelfallbezogene Entscheidungen**[16]. Hier werden aus der gesamten Datenmenge einer bestimmten Person Muster erstellt und mit Plausibilitätsprüfungen verschnitten. Unter dem Gesichtspunkt der Kriminalitäts- und Betrugsbekämpfung sind einzelfallbezogene Entscheidungen in den Blickwinkel der Betrachtung gerückt. Aus den in Abschnitt 2.3 genannten Methoden könnten durch virtuelle Zusammenlegung der Datenbasis, Auswertung von Einkünften und daraus erwachsenden steuerrechtlichen Verpflichtungen, Doppelförderungen reduziert und der Hinterziehung von Steuern und Abgaben entgegengewirkt werden. Während einzelfallbezogenen Entscheidungen auf Grund des aktuellen Austeritätsprinzips vor allem auf Einnahmenmaximierung abzielt, ist der mögliche Anwendungsbereich nicht darauf beschränkt und könnte gleichermaßen auf die optimale Verteilung von Fördermitteln angewandt werden.

3.2. Big Data erweiterte Anwendungsgebiete

[Rz 29] Einen theoretischen Rahmen zur Kontextualisierung der Aktivitäten der öffentlichen Hand spannt der Policy-Cycle auf. Er ist ein generisches Modell, das die Lebenszyklen öffentlichen Handelns, deren Implikationen und Umsetzungen zu illustrieren versucht und Phasen oder Abschnitte im Umsetzungsprozess identifiziert, die einer eingehenderen Betrachtung zugeführt werden können. Abbildung 1 stellt den von Nachmias / Felbinger (1982) vorgeschlagen ursprünglichen Policy-Cycle[17] wie das von Höchtl / Parycek / Schöllhammer (2015) im Lichte der Möglichkeiten von Big Data Analytics überarbeitete Modell[18] dar.

[Rz 30] Im evidenzbasierten Entscheidungsprozess treffen Entscheidungsträger ihre Beschlüsse auf Grundlage von Prognosen, die nicht nur auf empirischen Daten, sondern auf einer Zusammenschau mehrerer geeigneter Datensätze basieren. Die Analyse miteinander verknüpfter Daten kann zu neuen Erkenntnissen führen. Unter diesem Aspekt ist die Möglichkeit, durch Streaming-Verfahren zu schnelleren Prognose und Simulationsergebnissen zu gelangen, von besonderem Interesse. Durch laufende Analyse und Auswertung gewonnene Informationen könnten in vielen Bereichen öffentlichen Handelns zur Optimierung interner Prozesse, zur Steigerung von Effektivität und Effizienz beitragen. Reaktionen auf zukünftig möglicherweise auftretende Ereignisse können früher bis sofort erfolgen.

[Rz 31] Externe und interne Daten können zur Evaluierung ex post genauso wie zur Analyse der Anforderungen an ein bestimmtes Projekt a priori oder zur Erhebung der Meinungen in Bezug auf konkrete Fragestellungen genutzt werden. So kann die Entscheidung, welche Phasen der Implementierung einer Maßnahme im Policy-Cycle vorrangig bearbeitet werden soll, aus der Auswertung von Foren und Blog Posts mit Hilfe von Data Mining ermittelt werden.

[Rz 32] Eine online-Partizipationsplattform wurde in Österreich zuletzt etwa von der Stadt Wien

16 Bei der Anstellung dieser Überlegungen sind rechtliche Rahmenbedingungen wie insbesondere die des § 49 DSG zu berücksichtigen, siehe Kapitel 7.

17 David Nachmias / Claire Felbinger, Utilization in the Policy Cycle: Directions for Research, Review of Policy Research, Volume 2, Issue 2, November 1982, 300–308.

18 Eingereicht zur Publikation im Special Issue «Open Data Value and Theory» des Journals of Organizational Computing and Electronic Commerce (JOCEC).

im Zuge der Erarbeitung der Digitalen Agenda Wien genutzt[19] und wird aktuell vom Bundesrat zur kollektiven Erarbeitung eines Grünbuches herangezogen.[20] Auf diese Weise wird die aktive Kommunikation zwischen den Adressaten der Services und der Verwaltung intensiviert und können Services anhand individueller, konkreter Vorschläge überarbeitet oder gänzlich neu gestaltet werden. Die derart gesammelten Beiträge wurden in einer aufwändigen offline-Phase durch ein Experten- und Bürgergremium bearbeitet, gegliedert und für die weitere Abstimmung vorbereitet. In solchen Situationen können Methoden des Text Minings sinnverwandte Beiträge identifizieren und durch Algorithmen des maschinellen Lernens gliedern. Die zeitnahe Verfügbarkeit von Ergebnissen evidenzbasierter Indikatoren des Policy-Cycles ermöglicht eine andauernde Evaluierung der geplanten Umsetzung ohne erst konkret erwartete Ergebnisse abwarten zu müssen und somit wertvolle Zeit in der ex-post Adaption zu verlieren. Auf diese Weise kann die Zufriedenheit mit angebotenen Serviceleistungen gesteigert werden.

Abbildung 1: Darstellung des Policy Cycle. Links in Anlehnung an Nachmias / Felbinger (1982, 305). Rechts: Der Policy Cylce bei permanenter Evaluierung durch Big Data Analytics.

[Rz 33] Die genannten Potentiale erklären das Interesse an der Auswertung vorhandener Daten, denen Herausforderungen im Bereich des dazu notwendigen Wissens, der Arbeitsorganisation, der kulturellen Bereitschaft zur Zusammenarbeit und der Abkehr von Silodenken gegenüberstehen.

3.3. Herausforderungen und Hürden

[Rz 34] Die wachsende Unstrukturiertheit der Daten, die erforderliche Interdisziplinarität bei Datenauswertungen, ein notwendiges Maß an Risikobereitschaft zu schnellen guten aber nicht notwendigerweise perfekten Entscheidungen, das Miteinbeziehen von Daten aus anderen Abteilungen sowie öffentlich verfügbare, qualitativ aber ungesicherte Daten, öffnet für die Verwaltung neue Herausforderungen und Hürden.

[Rz 35] Bei der Verwendung von Konzepten, die durch neue Technologien entstehen, finden sich etablierte Organisations- und Gesellschaftsstrukturen oft vor Barrieren gestellt wieder. Die technologische Entwicklung verläuft oft schneller als sich Organisationen verändern können und technologische Entwicklungen werden zur Risikominimierung bewusst abgewartet, womit die Gefahr, technologisch überholt zu werden, in Kauf genommen wird.

[Rz 36] Hürden, die überwunden werden müssen, um mit innovativen Ansätzen zu arbeiten, wurden bereits mehrfach untersucht. In Bezug auf Open Data, definierbar z.B. als alle gespeicherten Daten, die ohne jegliche Einschränkung in Bezug auf Nutzung und Weiterverbreitung im öffentlichen Interesse zur Verfügung gestellt werden können,[21] wurden rechtliche von politischen, gesellschaftlichen,

19 www.digitaleagenda.wien.

20 www.besserentscheiden.at.

21 Christian P. Geiger / Jörn von Lucke, Open Government Data, in: Peter Parycek / Manuel J. Kripp / Noella Edelmann (Hrsg.), CeDEM11 Proceedings of the International Conference for E-Democracy and Open Government, Edition Donau-Universität Krems, Krems., 2011, 183–194 (184), mit Bezug auf Chris-

wirtschaftlichen, institutionellen, operativen und technischen Barrieren unterschieden.

[Rz 37] Eine Big Data Analysen inhärente Eigenschaft ist hohe Geschwindigkeit. Staatliche Hierarchie und Entscheidungsstrukturen, die teilweise Informationsflüsse der papierbasierten Verwaltung widerspiegeln, sind mit den vergleichsweise rasant möglichen Ergebnissen von Big Data Analysen und damit ermöglichten zeitnahen Entscheidungen nur bedingt vereinbar. Daher kann neben der Erforderlichkeit der Untersuchung der rechtlichen und technischen Voraussetzungen von Big Data Analysen die Etablierung veränderter organisatorischer Rahmenbedingungen erforderlich sein, um das Potential der Auswertung großer Datenmengen ausschöpfen zu können.

[Rz 38] Kulturell ist ein gewisses Maß an Fehlerakzeptanz erforderlich. Ein Wesensmerkmal von Big Data ist die Erkenntnis, dass ein Mehr an Daten, selbst wenn diese von ungesicherten Quellen stammen und potentiell Fehler enthalten, zu qualitativ besseren Ergebnissen führt, als wenige gesicherte, hochqualitative Daten. Dieser Argumentationslinie liegt die Vermutung zu Grunde, dass mögliche Verluste an Exaktheit auf der Mikro-Ebene durch Erkenntnisgewinn auf der Makro-Ebene ausgeglichen werden.[22] Blindes Vertrauen in Big Data Auswertungen ist dabei nicht angebracht. «Google-Flu-Trends» wurde als Beispiel zur Untermauerung der Big Data Fähigkeiten ins Rennen geführt und sollte die Ausbreitung von Grippewellen durch Analyse von Suchbegriffen betreffend Symptome und Medikamente in die Google Suchmaschine vorhersagen. Analysen belegten allerdings, dass die Voraussagen von Grippewellen nur sehr unzutreffend waren und faktisch durch das Zahlenmaterial nicht belegt werden können.[23] Qualitativ hochwertige Daten sind höher zu priorisieren als mehr, aber qualitativ fragwürde Daten.

[Rz 39] Technische Hürden können im Bereich von Datenheterogenität, der zu verarbeitenden Datenmenge und Anforderungen zur zeitnahen Herbeiführung von Ergebnissen identifiziert werden,[24]die erweiterte Infrastrukturen mit neuen Verfahren erfordern. Verteilte Datenhaltung und die Fähigkeit von IT-Systemen, in Echtzeit anfallenden Rohdaten anstatt aufbereitete Daten in einem Data Ware House zu verarbeiten, sind Beispiele neuer Erfordernisse für deren Konfiguration und Betrieb zusätzliches Wissen notwendig ist. Erstanschaffungskosten sowie Aufwendungen im Bereich von Schulungen und Weiterbildung sind ökonomische Herausforderungen, die zudem in Zeiten wirtschaftlicher Ressourcenknappheit zu argumentieren sind.

[Rz 40] Operativ kann fehlendes Wissen angeführt werden. Um Big Data und damit verbundene Verfahren bestmöglich nutzen zu können, müssen eine Vielzahl an Wissensbereich abgedeckt werden die sich ausgehend von mathematisch/statistischen Fertigkeiten über technische Kenntnisse bis hin zu organisatorischen Fähigkeiten erstrecken und vor allem auch «business acumen» umfassen. Der Big Data Wissensarbeiter besitzt somit als deutliche Abgrenzung zum Analysten, ein Gespür für wirtschaftliche Erfordernisse und seine Aktivitäten sind an der verfolgten Gesamtstrategie ausgerichtet. Die genannten Wissensbereiche können auf Grund ihrer Breite nur unzureichend von

---

tian P. Geiger / Jörn von Lucke, Open Government Data – Frei verfügbare Daten des öffentlichen Sektors, Gutachten für die Deutsche Telekom AG zur T- City Friedrichshafen, Version vom 3. Dezember 2010, Deutsche Telekom Institute for Connected Cities, Zeppelin University GmbH, Friedrichshafen 2010. http://www.zeppelin-university.de/deutsch/lehrstuehle/ticc/TICC-101203-OpenGovernmentData-V1.pdf.

22 Kenneth Cukier / Viktor Mayer-Schönberger, Big Data: A Revolution That Will Transform How We Live, Work, and Think, 1st edition. Boston: Eamon Dolan/Houghton Mifflin Harcourt, 2013.

23 Kaiser Fung, Google Flu Trends' Failure Shows Good Data > Big Data, Harvard Business Review (https://hbr.org/2014/03/google-flu-trends-failure-shows-good-data-big-data).

24 Nrusimham Ammu / Mohd Irfanuddin, Big Data Challenges, International Journal of Advanced Trends in Computer Science and Engineering, Vol. 2, No.1, 2013, 613–615 (614).

einer Person abgedeckt werden, was abteilungsübergreifende Zusammenarbeit sowie Koordination durch das neue Rollenbild des Chief Data Officers sinnvoll erscheinen lässt[25].

[Rz 41] Menschliche Hürden können im Bereich Widerstand vor Veränderung und befürchteter Machtverlust identifiziert werden. Evidenz- und wirkungsbasierte Verwaltungsführung hat als ein Ziel Verwaltungshandeln auf Basis von Fakten zu steuern, das den gestalterischen Freiraum von Entscheidungen einschränkt, der als Machtverlust empfunden werden könnte.

[Rz 42] Der wirksame Einsatz von Big Data Analysen erfordert somit ein ganzheitliches Rahmenwerk, das rechtliche, institutionelle bzw. organisatorische, wirtschaftliche und technische Faktoren berücksichtigt. Die als Einleitung zu diesem Abschnitt genannten Anwendungsfelder wie Simulation, Prognose oder Frühwarnsysteme haben unterschiedliche Relevanz in Bezug auf den Anwendungsbereich des DSG. Diesen speziellen datenschutzrechtlichen Anforderungen, welchen eine Analyse und Auswertung von Verwaltungsdaten zu entsprechen hat, widmet sich der folgende Abschnitt.

4. Rechtmäßige Verwendung von Daten

[Rz 43] Je nach Kategorie der Daten differieren die datenschutzrechtlichen Anforderungen, die an die Verwendung der Daten gestellt werden. Gemäß § 1 Abs. 1 DSG hat jedermann einen Anspruch auf Geheimhaltung der ihn betreffenden personenbezogenen Daten, soweit ein schutzwürdiges Interesse daran besteht. Daraus folgt einerseits, dass sich der Anwendungsbereich des DSG nur auf jene Daten erstreckt, die einen Personenbezug aufweisen und andererseits, dass geprüft werden muss, inwieweit ein schutzwürdiges Interesse an der Geheimhaltung besteht.

[Rz 44] Im Folgenden werden zunächst nicht personenbezogene Daten von personenbezogenen Daten abgegrenzt. Auf diese Unterscheidung aufbauend werden sodann die Anforderungen beleuchtet, die das DSG an die Verwendung der jeweiligen Art der Daten[26] stellt.

4.1. Nicht personenbezogene Daten

[Rz 45] Innerhalb der Menge der nicht personenbezogenen Daten können Daten, die zu keinem Zeitpunkt einen Personenbezug aufwiesen (z.B. Wetterdaten) von solchen Daten unterschieden

25 Yang Lee / Stuart Madnick / Richard Wang / Forea Wang / Hongyun Zhang, A cubic framework for the chief data officer : succeeding in a world of big data, 2014, MIS Quarterly Executive, 13(1), 1–13.

26 In der Literatur ist auch die Differenzierung der Daten einerseits nach den enthaltenen Informationen (sensible oder nicht sensible Daten) und andererseits nach der Art der Verknüpfung mit Identifizierungsmerkmalen (direkt personenbezogene oder indirekt personenbezogene oder anonymisierte Daten) vertreten, siehe Alexander Hönel / Nicolas Raschauer / Wolfgang Wessely, Datenschutzrechtliche Fragestellungen im Zusammenhang mit klinischen Prüfungen, RdM 2006/76.

werden, die von niemandem[27] auf einen Betroffenen rückführbar sind[28] (=**anonymisierte**[29] **Daten**). Diesen wurde der Personenbezug durch einen Bearbeitungsschritt genommen.

[Rz 46] Liegen anonymisierte Daten vor, so ist der Betroffene nicht mehr identifizierbar.[30]

[Rz 47] Mangelnde Rückführbarkeit ist gegeben, sofern der Personenbezug nicht hergestellt werden kann und somit weder personenbezogene, noch indirekt personenbezogene Daten vorliegen.[31]

[Rz 48] Wenn Daten im Sinn der obigen Ausführungen nicht auf den Betroffenen rückführbar sind, ist das Bestehen eines schutzwürdigen Interesses an der Geheimhaltung ausgeschlossen. Betroffener ist derjenige, dessen Merkmale diese Daten enthalten bzw. gemäß § 3 Z 3 DSG jede vom Auftraggeber[32] verschiedene natürliche oder juristische Person oder Personengemeinschaft, deren Daten verwendet werden.

[Rz 49] Die Bedeutung von «nicht personenbezogene Daten» wird durch das DSG nicht festgelegt.

[Rz 50] § 3 Z 15 Bundesstatistikgesetz definiert nicht personenbezogene Daten dann als vorliegend, wenn die Identität der Betroffenen mit Mitteln, die vernünftigerweise angewendet werden können, nicht mehr bestimmt werden kann.

[Rz 51] Das DSG enthält zwar keine Definition nicht personenbezogener Daten, jedoch eine Legaldefinition personenbezogener Daten, die den Anwendungsbereich des DSG vorgibt. Diese Definition wird im Folgenden zur Bestimmung des Begriffes «nicht personenbezogene Daten» herangezogen.

[Rz 52] Gemäß § 4 Z 1 DSG sind Daten dann personenbezogen, wenn es sich um Angaben über Betroffene handelt, deren Identität bestimmt oder bestimmbar ist. Im Umkehrschluss sind daher nicht personenbezogene Daten, als Gegenteil von personenbezogenen Daten, jene Daten über Betroffene, deren Identität nicht bestimmbar ist.

[Rz 53] Bestimmbar ist die Identität Betroffener, wenn sie mit Mitteln bestimmt werden kann, «*die vernünftigerweise entweder vom Verantwortlichen für die Verarbeitung oder von einem Dritten zur Identifizierung angewandt werden können, die also weder ihrer Art. noch ihrem Aufwand nach vollkommen ungewöhnlich sind.*»[33] Bilddaten sind beispielsweise bestimmbare personenbezogene

---

27 Anmerkung: gemäß Richtlinie 95/46/EG, Erwägungsgrund 26 mit Mitteln, die vernünftigerweise angewendet werden können; nach Walter Dohr / Hans J. Pollirer / Ernst M. Weiss / Rainer Knyrim, Kommentar zum Datenschutzrecht, Manz 2014, Anm. 2 zu §4 legt Erwägungsgrund 26 der RL «strengere Maßstäbe» an. Für die Beurteilung der Wahrscheinlichkeit der Identifizierung ist auch das Wissen des Empfängers der Daten maßgeblich, siehe Viktoria Haidinger, Der Weg von personenbezogenen zu anonymen Daten, Dako 2015/34, 56.

28 Alexander Hönel / Nicolas Raschauer / Wolfgang Wessely, Datenschutzrechtliche Fragestellungen im Zusammenhang mit klinischen Prüfungen, RdM 2006/76.

29 Daten, die keinen Personenbezug mehr aufweisen, werden auch anonymisierte Daten genannt, siehe etwa Rainer Knyrim, Big Data: datenschutzrechtliche Lösungsansätze, Dako 2015/35, 60.

30 Walter Dohr / Hans J. Pollirer / Ernst M. Weiss / Rainer Knyrim, Kommentar zum Datenschutzrecht, Manz 2014, Anm. 2 zu § 4.

31 Walter Dohr / Hans J. Pollirer / Ernst M. Weiss / Rainer Knyrim, Kommentar zum Datenschutzrecht, Manz 2014, Anm. 9 zu § 1.

32 Unter Auftraggeber ist im Wesentlichen derjenige zu verstehen, der die Entscheidung getroffen hat, die Daten zu verwenden (das ist gemäß § 4 Z 8 DSG jede Art. der Handhabung von Daten), unabhängig davon, ob derjenige die Daten selbst verwendet, oder damit einen Dienstleister beauftragt (§ 4 Z 4 DSG). Dienstleister ist wer die Daten nur zur Herstellung eines ihm aufgetragenen Werkes verwendet (§ 4 Z 5 DSG).

33 Alexander Hönel / Nicolas Raschauer / Wolfgang Wessely, Datenschutzrechtliche Fragestellungen im Zusammenhang mit klinischen PrüfungenRdM 2006/76, FN 26–29. Siehe auch Klaus M. Simonic / Günther Gell, Datenschutz-Policy, Version 1.1 vom 1. September 2001, Institut für Medizinische Informatik, Statistik und Dokumentation LKH/Universitätsklinikum Graz, 7 (https://www.medunigraz.at/imi/de/projects/DS-Policy-FL-V1-1.pdf).

Daten i.S.d. § 4 Z 1 DSG.[34]

[Rz 54] Die Komplexität der Frage, welche Mittel «vernünftig» und welche «vollkommen ungewöhnlich» sind, wurde wiederholt thematisiert. Etwa wird im Hinblick auf den Aufwand der Bestimmung der Identität von Betroffenen ausgeführt, dass dieser unter Berücksichtigung der Mittel der IT bei Big Data in der Regel verhältnismäßig sei.[35]

[Rz 55] Welche Folgen hätte die Identifizierung einer Person unter Heranziehung eines Mittels, das vernünftigerweise nicht angewendet werden kann und sowohl seiner Art als auch seinem Aufwand nach völlig ungewöhnlich ist? Wenn die Rückführung auf die Identität des Betroffenen nur mit derartigen Mitteln vorgenommen werden kann, kann dann der Ausschluss dieser Daten aus dem Anwendungsbereich des DSG angenommen werden, weil die Identität der Betroffenen dann nicht mehr «bestimmbar» i.S.d. § 4 Z 1 DSG ist?

[Rz 56] An die Auswertung nicht personenbezogener Daten knüpft das DSG keine Anforderungen.

4.2. Personenbezogene Daten

[Rz 57] Angaben über Betroffene, deren Identität bestimmt oder bestimmbar ist, sind personenbezogene Daten (§ 4 Z 1 DSG). Eine Teilmenge[36] personenbezogener Daten sind indirekt personenbezogene Daten. Für einen Auftraggeber, Dienstleister oder Empfänger der Übermittlung sind nur indirekt personenbezogene Daten solche, auf Grund derer der jeweilige Auftraggeber, Dienstleister oder Empfänger einer Übermittlung mit rechtmäßigen Mitteln die Identität der/des Betroffenen nicht bestimmen kann (§ 4 Z 1 DSG).

[Rz 58] Personenbezogene Daten sind Angaben über einen Betroffenen, woraus dessen Identität entweder «direkt ersichtlich (‹bestimmt›)»[37] oder «ohne besonderen zusätzlichen Aufwand ‹bestimmbar›» ist.[38]

[Rz 59] Die Datenschutzkommission[39] hielt 2013 fest, dass die Qualifikation von Daten als indirekt personenbezogen es erfordere, dass der Verwender der Daten die Identität der Betroffenen nicht bestimmen könne, wenn er nicht rechtlich verpönte Mittel wie Einbruch, Zwang oder Bestechung einsetzt, um jenes Instrument zu erlangen, womit die Re-Identifizierung möglich würde. Die Vor-

---

34 Siehe z.B. jeweils im Zusammenhang mit Videoüberwachung Bescheid der Datenschutzkommission vom 21. Januar 2009, GZ K121.425/0003-DSK/2009; Bescheid der Datenschutzkommission vom 12. Mai 2010, K202.094/0004-DSK/2010; Bescheid der Datenschutzkommission vom 13. Dezember 2013, K202.128/0004-DSK/2013.

35 Rolf H. Weber / Dominic Oertly, Aushöhlung des Datenschutzes durch De-Anonymisierung bei Big Data Analytics?, in: Jusletter IT 21. Mai 2015, 3 unter Hinweis auf Bruno Baeriswyl, Big Data zwischen Anonymisierung und Re-Individualisierung, in: Rolf H. Weber / Florent Thouvenin (Hrsg.), Big Data und Datenschutz – Gegenseitige Herausforderungen, ZIK Band 59, Zürich 2014, 46; siehe auch Klaus M. Simonic / Günther Gell, Datenschutz-Policy, Version 1.1 vom 1. September 2001, Institut für Medizinische Informatik, Statistik und Dokumentation LKH/Universitätsklinikum Graz, 8 (https://www.medunigraz.at/imi/de/projects/DS-Policy-FL-V1-1.pdf).

36 Indirekt personenbezogene Daten fallen ebenfalls unter «personenbezogene Daten», wofür der eindeutige Wortlaut (indirekt «personenbezogen») sowie aus gesetzessystematischer Sicht die Anführung der indirekt personenbezogenen Daten in § 4 Z 1 DSG spricht.

37 Walter Dohr / Hans J. Pollirer / Ernst M. Weiss / Rainer Knyrim, Kommentar zum Datenschutzrecht, Manz 2014, Anm. 2 zu § 4.

38 Walter Dohr / Hans J. Pollirer / Ernst M. Weiss / Rainer Knyrim, Kommentar zum Datenschutzrecht, Manz 2014, Anm. 2 zu § 4.

39 § 61 Abs. 9 DSG normiert, dass die Datenschutzbehörde mit Ablauf des 31. Dezember 2013 an die Stelle der früheren Datenschutzkommission tritt.

aussetzung einer «ausreichenden faktischen (technisch-organisatorischen) Absicherung der Daten gegen die Möglichkeit missbräuchlicher Re-Identifikation» werde von § 4 Z 1 DSG vorgegeben. Diese missbräuchliche Re-Identifikation solle für den Verwender «praktisch nicht möglich» sein.[40] Im Umkehrschluss bedeutet dies, dass es sich nicht um indirekt personenbezogene Daten handelt, wenn die Re-Identifikation zwar nur mit rechtswidrigen Mitteln, aber praktisch mit überschaubarem Aufwand durchgeführt werden könnte.

[Rz 60] Indirekt personenbezogen sind verschlüsselte Daten, die der Verwender der Daten nur durch Nutzung ihm rechtlich nicht zustehender Mittel z.B. dadurch, dass er sich illegal den Entschlüsselungscode verschafft, entschlüsseln kann. Pseudonymisierung (d.h. statt des Identifikationsmerkmals – wie z.B. des Namens – wird ein anderes Identifikationsmerkmal wie eine mehrstellige Aneinanderreihung von Buchstaben und Zahlen angeführt) führt zur Erzeugung von indirekt personenbezogenen und nicht von anonymisierten Daten. Unter Verwendung des Schlüssels kann der jeweilige Betroffene identifiziert werden.[41]

[Rz 61] Indirekt personenbezogenene Daten können daher auch als pseudonymisierte[42] Daten kategorisiert werden.

[Rz 62] Als pseudonymisiert werden personenbezogene Datensätze etwa bezeichnet, nachdem sie um die primären Identifikationsmerkmale, mittels derer Betroffene direkt identifiziert werden können (z.B. Name, Sozialversicherungsnummer) und die sekundären Identifikationsmerkmale, die in Zusammenschau zur Identifikation Betroffener führen können (etwa Adresse, Alter, Geschlecht), reduziert wurden.[43]

[Rz 63] Andererseits hat auch die Teilmenge der **sensiblen Daten**, die nur natürliche Personen betreffen können und die auf Grund ihres Inhalts besonders schutzwürdig sind, spezielle Eigenschaften. Sensible Daten sind Daten natürlicher Personen über deren rassische und ethnische Herkunft, politische Meinung, Gewerkschaftszugehörigkeit, religiöse oder philosophische Überzeugung, Gesundheit oder ihr Sexualleben (§ 4 Z 2 DSG).

[Rz 64] Um von personenbezogenen Daten durch Analyse Gebrauch machen zu können, ist die Verwendung der Daten nötig. Die grundlegenden Bedingungen, unter welchen personenbezogene Daten verwendet werden können, werden von § 1 Abs. 1 und 2 DSG geregelt. Deren Verwendung ist in folgenden Fällen zulässig:

- allgemeine Verfügbarkeit (siehe Kapitel 4.2.1),
- mangelnde Rückführbarkeit der Daten auf den Betroffenen (siehe Kapitel 4.1),
- lebenswichtiges Interesse des Betroffenen (siehe Kapitel 4.2.2),
- Zustimmung des Betroffenen (siehe Kapitel 4.2.3),
- Wahrung überwiegend berechtigter Interessen eines anderen
- Liegt der letztgenannte Fall in der Form eines Eingriffs einer staatlichen Behörde (in das Grundrecht) vor, so darf dies nur auf Grund von Gesetzen, an die besondere, genau umschriebene Anforderungen gestellt werden, geschehen (siehe Kapitel 4.2.4).
- Genehmigung der Datenschutzbehörde (siehe Kapitel 4.2.5).

---

40 Datenschutzkommission, Datenschutzbericht 2009, 34 (https://www.dsb.gv.at/DocView.axd?CobId=40344).

41 Walter Dohr / Hans J. Pollirer / Ernst M. Weiss / Rainer Knyrim, Kommentar zum Datenschutzrecht, Manz 2014, Anm. 2 zu § 4.

42 Rainer Knyrim, Big Data: datenschutzrechtliche Lösungsansätze, Dako 2015/35, 59.

43 Viktoria Haidinger, Der Weg von personenbezogenen zu anonymen Daten, Dako 2015/34, 56 und 59.

4.2.1. Allgemeine Verfügbarkeit

[Rz 65] Die Verfassungsbestimmung § 1 DSG gewährleistet jedermann einen Anspruch auf Geheimhaltung der ihn betreffenden personenbezogenen Daten, soweit ein schutzwürdiges Interesse vorliegt. Das Bestehen eines solchen ist für jene Fälle ausgeschlossen, in denen Daten bereits allgemein verfügbar sind (§ 1Abs. 1 DSG).

[Rz 66] Allgemeine Verfügbarkeit setzt die Zugänglichkeit der Daten für jedermann[44] im Zeitpunkt der Verwendung der Daten voraus. Beispiele hierfür sind Daten öffentlicher Register und sonstige öffentlich abrufbare Informationsquellen wie etwa solche aus dem Internet. In einer öffentlichen Verhandlung vorgekommene Daten werden ohne qualifizierte massenmediale Berichterstattung oder Berichterstattung im Internet nicht als allgemein verfügbar gewertet.[45]

[Rz 67] Allgemein verfügbar sind demnach insbesondere frei zugängliche Daten des Social Media Bereichs.

[Rz 68] Im Zuge einer wissenschaftlichen Untersuchung wurden in diesem Zusammenhang bereits Rückschlüsse aus dem Verhalten von Einwohnern im Social Media Netzwerk Twitter auf die regionale Arbeitslosigkeit in Spanien getroffen.[46]

[Rz 69] Der Nutzung des Potentials der Auswertung allgemein verfügbarer Daten wie etwa von öffentlich einsehbaren Social Media-Beiträgen durch die Verwaltung steht daher aus datenschutzrechtlicher Sicht jedenfalls dann nichts entgegen, wenn der Nutzer bei Eingabe seiner Daten vorhersehen konnte, dass er damit einer Veröffentlichung zustimmt. Nach dem Wortlaut des § 1 DSG führt auch eine allgemeine Verfügbarkeit der Daten gegen den Willen des Nutzers zum Verlust des Geheimhaltungsanspruchs. In der Literatur wird vertreten, dass nur die rechtlich zulässige Veröffentlichung einen Verlust des Geheimhaltungsanspruchs zur Folge haben kann.[47] Die Forcierung solcher Methoden kann insbesondere zusätzliche Informationsquellen für Vorhersagen von Entwicklungen[48] und zusätzliche Indikatoren für deren Erklärung schaffen.

[Rz 70] Es ist festzuhalten, dass sowohl bei allgemeiner Verfügbarkeit der Daten, als auch bei mangelnder Rückführbarkeit auf einen Betroffenen daher kein Geheimhaltungsanspruch besteht.

Abbildung 2: Einteilung der Daten in personenbezogene, indirekt personenbezogene, sensible und nicht personenbezogene Daten[49]

---

44 Und nicht bloß Zugang für einen begrenzten Personenkreis besteht, siehe OGH vom 28. November 2013, 6Ob165/13b m.w.N. Zum Nichtvorliegen allgemeiner Verfügbarkeit bei Offenbarung der Daten an einen begrenzten Personenkreis siehe auch OGH vom 3. September 2002, 11Os109/01.

45 OGH vom 24. November 2014, 17Os40/14g (17Os41/14d).

46 Alejandro Llorente / Manuel Garcia-Herranz / Manuel Cebrian / Esteban Moro, Social Media Fingerprints of Unemployment. PLoS ONE 10(5): e0128692. doi:10.1371/journal.pone.0128692.

47 Walter Dohr / Hans J. Pollirer / Ernst M. Weiss / Rainer Knyrim, Kommentar zum Datenschutzrecht, Manz 2014, Anm. 8 zu § 1.

48 Zur Prognosefähigkeit siehe auch Alejandro Llorente / Manuel Garcia-Herranz / Manuel Cebrian / Esteban Moro, Social Media Fingerprints of Unemployment. PLoS ONE 10(5): e0128692. doi:10.1371/journal.pone.0128692

49 Unterscheidung personenbezogene – indirekt personenbezogene Daten in Anlehnung an Alexander Hönel / Nicolas Raschauer / Wolfgang Wessely, Datenschutzrechtliche Fragestellungen im Zusammenhang mit klinischen PrüfungenRdM 2006/76, FN 26–29. Siehe auch Klaus M. Simonic / Günther Gell, Datenschutz-Policy, Version 1.1 vom 1. September 2001, Institut für Medizinische Informatik, Statistik und Dokumentation LKH/Universitätsklinikum Graz, 7 (https://www.medunigraz.at/imi/de/projects/DS-Policy-

[Rz 71] Abbildung 2 veranschaulicht die Abgrenzung der legaldefinierten Begriffe personenbezogener Daten (§ 4 Z 1 DSG), nur indirekt personenbezogener Daten (§ 4 Z 1 DSG) und sensibler Daten (§ 4 Z 2 DSG) von nicht personenbezogenen Daten.

4.2.2. Lebenswichtiges Interesse

[Rz 72] Lebenswichtiges Interesse eines Betroffenen liegt vor, wenn durch die Nichtverwendung der Daten das körperliche Überleben gefährdet wäre.[50]

[Rz 73] Im Kontext der optimierten Entscheidungsfindung wird die Verwendung von personenbezogenen Daten zur Analyse und Auswertung im lebenswichtigen Interesse des Betroffenen nicht den Regelfall darstellen. Es sind jedoch auch Anwendungsfälle dafür denkbar. Es wäre beispielsweise zu prüfen, ob nach dem Auftreten bzw. Bekanntwerden einer ansteckenden Krankheit aus einem betroffenen Land eingereiste Personen sowie deren Angehörige innerhalb der Inkubationszeit kontaktiert werden könnten. Aus getätigten Aussagen in öffentlichen Netzwerken sowie den in Apotheken bezogenen, frei erhältlichen Medikamenten zur Behandlung von Symptomen könnten etwa Rückschlüsse auf den Ursprung der Krankheit gezogen werden und geeignete Vorsichtsmaßnahmen in der Umgebung dieser Personen entwickelt werden. Lebenswichtiges Interesse besteht hier sowohl für die von einer Krankheit betroffenen Person selbst als auch für deren Umgebung.

[Rz 74] Insbesondere für das Auffinden und die Identifizierung von Abgängigen im bereits eingetretenen Katastrophenfall, normiert § 48a DSG spezielle Regelungen. Die potentiell mögliche Früherkennung bevorstehender Katastrophen durch Datenanalyse wird im Rahmen des DSG nicht geregelt. Vorstellbar wäre etwa die Schaffung eines mit anonymisierten Daten laufenden Frühwarnsystems, wobei technisch sichergestellt ist, dass mit Eintritt des (Katastrophen-)Falles die Re-Identifizierung im lebenswichtigen Interesse des Betroffenen möglich ist. Diese Ansätze sind weiterer detaillierter rechtlicher und technischer Prüfung zu unterziehen.

4.2.3. Zustimmung

[Rz 75] Gerechtfertigt ist die Verwendung personenbezogener Daten weiters mit der Zustimmung des Betroffenen. Unter Zustimmung versteht § 4 Z 14 DSG die gültige, insbesondere ohne Zwang abgegebene Willenserklärung des Betroffenen, dass er in Kenntnis der Sachlage für den konkreten Fall in die Verwendung seiner Daten einwilligt (§ 4 Z 14 DSG). Dies kommt grundsätzlich in Betracht.

[Rz 76] Weitere Voraussetzung bei der Verwendung von Daten ist die Einhaltung der in § 6 Abs. 1 DSG festgelegten Grundsätze.

[Rz 77] Der im vorliegenden Kontext der Big Data Analyse besonders relevante Grundsatz ist jener, dass Daten gemäß § 6 Abs. 1 Z 2 DSG nur für **festgelegte, eindeutige und rechtmäßige Zwecke** ermittelt und – außer für wissenschaftliche oder statistische Zwecke nach §§ 46, 47 DSG – nicht in einer mit diesen Zwecken unvereinbaren Weise weiterverwendet werden dürfen. Bereits

FL-V1-1.pdf) und Richtlinie 95/46/EG, Erwägungsgrund 26 sowie § 4 Abs. 1 DSG.

50 Andreas Lehner / Friedrich Lachmayer, Datenschutz im Verfassungsrecht, in: Lukas Bauer / Sebastian Reimer (Hrsg.), Handbuch Datenschutzrecht (2009) Facultas, 101 im Kontext gerechtfertigter gesetzlicher Regelungen zur Verwendung von Daten.

die Ermittlung von (personenbezogenen[51]) Daten ist daher an im Vorhinein festgelegte, eindeutige und rechtmäßige Zwecke gebunden.

[Rz 78] Insbesondere der Big Data Analyse ist der Gedanke immanent, dass nicht nur Antworten auf konkrete Fragen erzielt werden, sondern dass Daten (meistens) für einen bestimmten Zweck anfallen und erst später analysiert und ausgewertet werden sollen. Aber auch die Verwaltung hat bei in Auftrag gegebener Analyse von Daten das Interesse zu Ergebnissen zu gelangen, die zur Effizienzsteigerung bei der Erfüllung ihrer gesetzlichen Pflichten beitragen sowie im Rahmen der gestalterischen Freiheit zu Schaffung innovativer und bürgerfreundlicher Services führen. In beiden Fällen können die genauen Ziele – und damit der Zweck –unbekannt sein und sich erst nach der Auswertung zeigen.

[Rz 79] Knyrim führt aus, der **Grundsatz der Zweckbindung** stehe der Idee von Big Data diametral entgegen und hält zutreffend fest, dass eine Zustimmung Betroffener nur für jene Situationen geeignet ist, «*wenn zu Beginn der ursprünglichen Datenverarbeitung bereits klar ist, in welcher Form genau die Daten durch wen für welche Big Data Anwendungen weiterverwendet werden sollen.*» Es ist eine Zustimmung zur exakten und abschließenden Anführung sowohl des ursprünglichen Zwecks als auch der Weiterverwendung erforderlich.[52]

[Rz 80] Auch Feiler/Fina weisen auf die schwierige Vereinbarkeit vieler Big Data Anwendungen mit dem Grundsatz der Zweckfestlegung hin.[53]

[Rz 81] Weiters wird argumentiert, dass im Bereich der Hoheitsverwaltung durch die bestehende Über- und Unterordnung «*im Zweifel nicht von der Freiwilligkeit (und damit Wirksamkeit) einer Zustimmung ausgegangen werden*» kann.[54]

[Rz 82] Die erforderliche Einholung einer Zustimmung a priori im Sinn einer Einwilligung des Betroffenen in Kenntnis der Sachlage für den konkreten Fall, wird in der Regel eine große Hürde darstellen. Der Zweck der Verarbeitung müsste im Zeitpunkt der Ermittlung der Daten bekannt sein und hinreichend genau konkretisiert werden können. Eine Zustimmung zur Verwendung personenbezogener Daten für Big Data Analysen wird daher nach dem aktuellen Stand der Technik in den meisten Fällen impraktikabel oder schier unmöglich sein.[55]

4.2.4. Gesetzliche Regelung

[Rz 83] Eine weitere Option, wie im Einklang mit dem DSG personenbezogene Daten durch Verwaltungsbehörden verwendet werden können, ist die Schaffung einer gesetzlichen Grundlage, die an die Erfüllung der im Folgenden beschriebenen Voraussetzungen gebunden ist.

[Rz 84] Derartige Bestimmungen haben die Anforderungen des § 1 Abs. 2 DSG zu erfüllen. Dies bedeutet, soweit die Verwendung der Daten nicht im lebenswichtigen Interesse (siehe 4.2.2) oder

---

51 Gemäß § 4 Z 1 werden Daten und personenbezogene Daten als Angaben über Betroffene, deren Identität bestimmt oder bestimmbar ist, definiert. Diese Zumessung der gleichen Bedeutung ergibt sich daraus, dass der Anwendungsbereich des DSG nur personenbezogene Daten umfasst.

52 Rainer Knyrim, Big Data: datenschutzrechtliche Lösungsansätze, Dako 2015/35, 60.

53 Lukas Feiler / Siegfried Fina, Datenschutzrechtliche Schranken für Big Data, MR 2013, 303 (303)

54 Rundschreiben des BKA-Verfassungsdienst zur legistischen Gestaltung von Eingriffen in das Grundrecht auf Datenschutz, GZ BKA-810.016/0001-V/3/2007.

55 In diesem Sinn stellen auch Rolf H. Weber / Dominic Oertly, Aushöhlung des Datenschutzes durch De-Anonymisierung bei Big Data Analytics?, in: Jusletter IT 21. Mai 2015, die Ungeeignetheit der Einwilligung für Big Data Anwendungen in der Form der aktuellen Regelung fest.

mit Zustimmung (siehe 4.2.3) des Betroffenen erfolgt, sind Beschränkungen des Anspruchs auf Geheimhaltung nur zur Wahrung überwiegender berechtigter Interessen eines anderen zulässig, und zwar bei Eingriffen einer staatlichen Behörde nur auf Grund von **Gesetzen**, an die folgende kumulative Anforderungen gestellt werden:

- Notwendigkeit der Gesetze aus in Art. 8 Abs. 2 der Europäischen Konvention zum Schutze der Menschenrechte und Grundfreiheiten (EMRK)[56] genannten Gründen
- Vorsehung der Verwendung von Daten, die ihrer Art nach besonders schutzwürdig sind nur zur Wahrung wichtiger öffentlicher Interessen
- Gleichzeitige Festlegung angemessener Garantien für den Schutz der Geheimhaltungsinteressen der Betroffenen festlegen
- Beachtung des Verhältnismäßigkeitsgrundsatzes (Eingriff in das Grundrecht nur in der gelindesten, zum Ziel führenden Art)

[Rz 85] Die oben angesprochenen, **in Art. 8 Abs. 2 EMRK genannten Gründe** sind Gründe der nationalen Sicherheit, der öffentlichen Ruhe und Ordnung, der Verteidigung der Ordnung und Verhinderung von strafbaren Handlungen, des Schutzes der Gesundheit und Moral anderer und des Schutzes der Rechte und Freiheiten anderer. Bezogen auf mit Big Data Analysen verfolgte Zielsetzungen kommen in erster Linie etwa das wirtschaftliche Wohl des Landes und der Schutz der Gesundheit anderer in Betracht.

[Rz 86] Notwendig im Sinn des Verhältnismäßigkeitsgrundsatzes ist nach der Judikatur des VfGH beispielsweise die Erhebung von Wirtschaftsdaten für Zwecke der Wirtschaftsforschung und der Wirtschaftspolitik für das wirtschaftliche Wohl eines Landes, wie der BKA-VD unter Bezugnahme auf VfSlg. 12.228/1989 ausführt.[57]

[Rz 87] Für den Fall des Vorliegens einer gesetzlichen Ermächtigung oder Verpflichtung zur Verwendung von Daten normiert § 8 Abs. 1 Z 1 DSG, dass schutzwürdige Geheimhaltungsinteressen bei der Verwendung von nicht-sensiblen Daten nicht verletzt sind.

[Rz 88] Insbesondere sind schutzwürdige Geheimhaltungsinteressen dann nicht verletzt, wenn die Verwendung der Daten gemäß § 8 Abs. 3 Z 1 DSG für einen Auftraggeber des öffentlichen Bereichs eine wesentliche Voraussetzung für die Wahrnehmung einer ihm gesetzlich übertragenen Aufgabe ist.

[Rz 89] Diese Regelung erfordert eine genaue Umschreibung der Aufgaben des Auftraggebers des öffentlichen Bereichs durch das Gesetz und dass «*klare Rückschlüsse auf damit verbundene Datenverwendungen möglich sind (…) Wichtig ist, dass die Zusammenschau der in den Materiengesetzen enthaltenen Regelungen mit den allgemeinen Grundsätzen über die Verwendung von Daten gemäß Art. 2 (= §§ 4 ff) DSG 2000 eine im Auslegungsweg ermittelbare, hinreichend präzise Regelung darstellt. Dies freilich nur unter der Voraussetzung, dass die sich daraus ergebenden Grenzen der Datenerhebung und -verwendung § 1 Abs. 2 letzter Satz zufolge nach Maßgabe des Verhältnismäßigkeitsgrundsatzes bestimmt werden, sodass «der Eingriff in das Grundrecht nur in der gelindesten, zum Ziel führenden Art vorgenommen» wird. (vgl. VfGH 15. Juni 2007, G 147/06 ua zur «Section Control»). Das Absehen von einer «ausdrücklichen gesetzlichen Ermächtigung» wird insbesondere dann zulässig sein, wenn eine genaue Determinierung der zu verwendenden*

56 Europäische Menschenrechtskonvention, BGBl 210/1958 i.d.F. BGBl III 30/1998.

57 Rundschreiben des BKA-Verfassungsdienst zur legistischen Gestaltung von Eingriffen in das Grundrecht auf Datenschutz, GZ BKA-810.016/0001-V/3/2007.

Datenarten gar nicht möglich ist.»[58]

[Rz 90] Ihrer Art nach besonders schutzwürdig sind sensible Daten (z.B. politische Meinungen). Solche dürfen nur zur Wahrung wichtiger öffentlicher Interessen verwendet werden. Als wichtiges öffentliches Interesse gewertet wurde beispielsweise bei der Verwendung von Gesundheitsdaten, dass sich Wissenschaftler mit den im Zentrum der betreffenden Forschung stehenden Fragen noch beinahe gar nicht auseinandergesetzt hatten und die Ergebnisse potentiell einerseits für die Wissenschaft wertvoll sein könnten und andererseits neues, potentiell lebensrettendes Wissen für die Notfallmedizin und das Rettungswesen daraus abgeleitet werden könnte.[59]

[Rz 91] Ein Gesetz, das den Anforderungen des § 1 Abs. 2 DSG i.V.m. Art. 8 Abs. 2 EMRK entspricht, hat für jedermann vorhersehbar festzulegen, unter welchen Voraussetzungen Auskünfte über geschützte Daten für die Wahrnehmung konkreter Verwaltungsaufgaben erforderlich sind.[60] «*Der Gesetzgeber muss somit nach den Vorgaben des §1 Abs2 DSG2000 eine materienspezifische Regelung in dem Sinn vorsehen, dass die Fälle zulässiger Eingriffe in das Grundrecht auf Datenschutz konkretisiert und begrenzt werden (VfSlg 18.643/2008).*»[61]

[Rz 92] Anhand des Beispiels einer des Gesundheitstelematikgesetzes (GTelG)[62] soll demonstriert werden, wie der Gesetzgeber diese Voraussetzungen im Einzelnen erfüllen kann.

[Rz 93] Als legitimer Zweck gemäß Art. 8 Abs. 2 EMRK wird der Schutz der Gesundheit angeführt. In den Erläuterungen wird auf Art. 8 Abs. 4 Richtlinie 95/46/EG referenziert, wonach es Mitgliedstaaten offen steht, bei Vorliegen eines wichtigen öffentlichen Interesses Ausnahmen des Grundsatzes, dass die Verarbeitung sensibler personenbezogener Daten untersagt sein soll, zu schaffen. Als wichtiges öffentliches Interesse werden Verbesserungen im öffentlichen Gesundheitsbereich angeführt, die durch die Elektronische Gesundheitsakte (ELGA) verwirklicht werden können, wie z.B. höhere Qualität durch schnelleren Zugang zu medizinischen Informationen. Als angemessene Garantien werden die gesetzlich festgelegten Teilnehmerrechte (§ 16 GTelG) sowie eine beratende Ombudsstelle angeführt. Zusätzlich ist der Zugriff auf die Gesundheitsdaten in der Hinsicht eingeschränkt, als nur für Zwecke der Gesundheitsversorgung während eines Behandlungs- oder Betreuungsverhältnisses auf Daten zugegriffen werden darf. Dabei wird auf den Verhältnismäßigkeitsgrundsatz sowie auf § 9 Z 12 DSG verwiesen, wonach schutzwürdige Interessen bei der Verwendung sensibler Daten u.a. dann als nicht verletzt gelten, wenn die Verwendung der Daten zum Zweck der Gesundheitsversorgung erforderlich ist und etwa durch ärztliches Personal vorgenommen wird. [63]

4.2.5. Genehmigung der Datenschutzbehörde

[Rz 94] Für bestimmte Verwendungszwecke wie Zwecke der Wissenschaftlichen Forschung und Statistik, auf die im Folgenden eingegangen wird, sind Sonderbestimmungen vorgesehen.

---

58 Rundschreiben des BKA-Verfassungsdienst zur legistischen Gestaltung von Eingriffen in das Grundrecht auf Datenschutz, GZ BKA-810.016/0001-V/3/2007.

59 Bescheid der Datenschutzkommission vom 14. April 2010, K202.086/0007-DSK/2010.

60 VfGH vom 28. November 2001, B2271/00.

61 VfGH vom 12. März 2013, G76/12.

62 Bundesgesetz betreffend Datensicherheitsmaßnahmen bei der Verwendung elektronischer Gesundheitsdaten (Gesundheitstelematikgesetz 2012 – GTelG 2012), BGBl I 111/2012 i.d.F. BGBl I 83/2013.

63 ErlRV 1936 BlgNR XXIV. GP, 4–5 (http://www.parlament.gv.at/PAKT/VHG/XXIV/I/I_01936/fname_271569.pdf).

[Rz 95] Gemäß § 46 Abs. 1 DSG dürfen Daten für Zwecke wissenschaftlicher oder statistischer Untersuchungen vom Auftraggeber verwendet werden, wenn die **Ergebnisse nicht personenbezogene Daten** sein sollen. Diesfalls dürfen Daten verwendet werden, die öffentlich zugänglich sind, Daten, die für den Auftraggeber nur indirekt personenbezogen sind oder etwa solche Daten, die für andere Zwecke zulässigerweise vom Auftraggeber der Untersuchung ermittelt wurden.

[Rz 96] Nach Feiler/Fina ist die in § 6 Abs. 1 Z 2 HS 2 i.V.m. § 46 DSG geregelte Ausnahme vom Zweckbindungsgrundsatz bei privaten Auftraggebern von Big Data Anwendungen nur dann anwendbar, wenn Daten für punktuelle wissenschaftliche oder statistische Erhebungen verwendet werden.[64]

[Rz 97] Dies bedeutet für die Verwaltung, dass etwa statistische Untersuchungen mit öffentlich zugänglichen Daten, mit Daten, die vom konkreten Auftraggeber zulässigerweise ermittelt wurden und mit Daten, die für den Auftraggeber nur indirekt personenbezogen sind, durchgeführt werden dürfen.

[Rz 98] **Personenbezogene Ergebnisse** dürfen im Einklang mit den Voraussetzungen des § 46 Abs. 2 DSG erzielt werden.

[Rz 99] § 46 Abs. 2 DSG regelt die Bedingungen der Genehmigung durch die Datenschutzbehörde betreffend Daten, die nicht unter § 46 Abs. 1 DSG fallen, etwa weil die Ergebnisse personenbezogen sein sollen oder weil die Verwendung der Daten nicht öffentlich zugängliche Daten, die nicht vom Auftraggeber ermittelt wurden, zum Gegenstand hat.

[Rz 100] Gemäß § 46 Abs. 1 und Abs. 2 DSG dürfen Daten gemäß besonderen gesetzlichen Vorschriften oder mit Zustimmung des Betroffenen oder mit Genehmigung der Datenschutzbehörde verwendet werden. Diese Genehmigung ist gemäß § 46 Abs. 3 DSG zu erteilen, sofern

1. die Einholung von Zustimmungen der Betroffenen unmöglich ist oder einen unverhältnismäßigen Aufwand verursacht und
2. ein öffentliches Interesse an der Verwendung besteht und
3. die fachliche Eignung des Antragstellers glaubhaft gemacht wird.

[Rz 101] Im Fall der Heranziehung sensibler Daten werden strengere Anforderungen gestellt. Das Vorliegen eines **wichtigen** öffentlichen Interesses und dass diejenigen Personen, die die Daten verwenden entweder hinsichtlich des Gegenstandes der Untersuchung einer gesetzlichen Verschwiegenheitspflicht unterliegen oder dass deren Verlässlichkeit glaubhaft ist, wird vorausgesetzt (§ 46 Abs. 3 DSG). Ebenso wird eine Erklärung des über die Datenbestände Verfügungsbefugten über die Zurverfügungstellung dieser gefordert (§ 46 Abs. 3a DSG).

[Rz 102] Der direkte Personenbezug ist unverzüglich zu verschlüsseln, wenn in einzelnen Phasen die Arbeit mit nur indirekt personenbezogenen Daten erfolgen kann. Sobald der Personenbezug für die Zwecke wissenschaftlicher oder statistischer Arbeit nicht mehr notwendig ist, ist er gänzlich zu beseitigen (§ 46 Abs. 5 DSG). Dies entspricht auch dem Grundsatz des § 6 Abs. 1 Z 5 DSG, wonach Daten nur solange in personenbezogener Form aufbewahrt werden dürfen, als dies für die Erreichung der Zwecke, für die sie ermittelt wurden erforderlich ist. Längeres Aufbewahren ist nur auf Grund besonderer gesetzlicher Regelungen zulässig.

[Rz 103] Der Antrag auf Genehmigung des Filmens (Speicherung und Auswertung von Bilddaten) von Personen an Haltestellen sowie in bestimmten öffentlichen Verkehrsmitteln zum Zweck der

64 Lukas Feiler / Siegfried Fina, Datenschutzrechtliche Schranken für Big Data, MR 2013, 303 (304).

wissenschaftlichen Untersuchung von Fahrgastwechselzeiten und der darauf folgenden Analyse der inneren Ausstattung der Verkehrsmittel wurde insbesondere aus folgenden Gründen gemäß § 46 Abs. 2 Z 3 i.V.m. Abs. 3 DSG mit Auflagen genehmigt:[65]

[Rz 104] Die Voraussetzungen des § 46 Abs. 1 (insbesondere Z 3, da direkt personenbezogene Daten übermittelt werden sollen) und Abs. 2 Z 1 und Z 2 DSG lägen nicht vor.

[Rz 105] Die Zustimmung der Betroffenen könne mangels deren Bestimmtheit nicht im Vorfeld eingeholt werden.

[Rz 106] Es seien – insofern als etwa die Verwendung von Krücken aufgezeichnet werden soll, woraus Rückschlüsse auf Verletzungen oder Krankheiten gezogen werden können – auch sensible Daten von der Ermittlung betroffen. Als wichtiges öffentliches Interesse werde die Zufriedenheit der Fahrgäste und die Zuverlässigkeit des Fahrbetriebes von öffentlichen Verkehrsmitteln gewertet. Schließlich sei die fachliche Eignung der antragstellenden Universität glaubhaft gemacht worden.

[Rz 107] Das Vorliegen eines wichtigen öffentlichen Interesses kann sich etwa auch daraus ergeben, «*dass die Rechtsordnung dem Thema «Integration» in mehrfacher Hinsicht einen hohen Stellenwert einräumt (…), die von der Studie angestrebten statistischen Aussagen derzeit nicht vorhanden sind und konkrete Finanzierungs- bzw. Unterstützungszusagen von zumindest drei mit dem Thema befassten staatlichen bzw. staatlich finanzierten Einrichtungen (Magistraten bzw. Ämter der Landesregierungen) vorliegen.*»[66]

[Rz 108] Diesen Ausführungen folgend, kann das Vorliegen eines (wichtigen) öffentlichen Interesses bei wissenschaftlichen oder statistischen Untersuchungen der staatlichen Verwaltung in vielen Fällen zum Tragen kommen. Zu Recht führt Knyrim aus, dass öffentliches Interesse bei rein privatwirtschaftlichen Big Data Anwendungen tendenziell nicht dargelegt werden kann und in weiterer Folge die Genehmigung i.S.d. § 46 Abs. 2 DSG eher für Big Data Anwendungen für die öffentliche Verwaltung, für öffentlich geförderte Forschungsprojekte oder für Forschung im öffentlichen Interesse in Betracht kommen wird.[67]

4.3. Zusammenfassender Überblick

[Rz 109] Im Folgenden gelangt zur Darstellung, welche Daten beispielsweise für welches Anwendungsgebiet von Big Data Analysen grundsätzlich geeignet wären. Unter der Voraussetzung der Erfüllung der in Kapitel 4.2.4 beschriebenen Erfordernisse wäre für jeden Anwendungsbereich die Erlassung einer gesetzlichen Regelung ein geeignetes Mittel, um Daten DSG-gemäß zu analysieren und auszuwerten.

Tabelle 1: Big Data Anwendungsbereiche und in Betracht kommende, nach DSG geeignete Daten

Anwendungsbereich	Erläuterung	in Betracht kommende Daten

65 Bescheid der Datenschutzkommission vom 12. Mai 2010, K202.094/0004-DSK/2010.

66 Bescheid der Datenschutzkommission vom 7. September 2006, K202.047/0009-DSK/2006.

67 Rainer Knyrim, Big Data: datenschutzrechtliche Lösungsansätze, Dako 2015/35, 60–61.

Simulation	Simulationen werden auf aggregierte Größen angewandt, der Personenbezug ist dabei häufig nicht relevant	z.B. nicht personenbezogene Daten, allgemein verfügbare personenbezogene Daten
Prognosen	Prognosen geben Aufschluss über Eintrittswahrscheinlichkeiten zukünftiger Ereignisse und haben häufig einen aggregierten Bezug, der die Identifikation einzelner Personen zur Erzielung relevanter Ergebnisse wenig sinnvoll erscheinen lässt.	z.B. nicht personenbezogene Daten, allgemein verfügbare personenbezogene Daten
Frühwarnsysteme	Frühwarnsysteme vereinen häufig Simulationen und Prognosen und können darüber hinausgehend gesetzte Aktivitäten einzelner Personen in die Auswertung miteinbeziehen. Die Handlung einer Einzelperson kann Trends erkennen lassen, auf welche Entscheider global reagieren müssen.	z.B. allgemein verfügbare personenbezogene Daten,in Einzelfällen auch personenbezogene Daten im lebenswichtigen Interesse Betroffener
personalisierte Services und Dienstleistungen	Aktionen eines Kollektiv werden analysiert und dienen als Vorlage um Services für eine Einzelperson zu optimieren. Die der Analyse zu Grunde liegenden Daten benötigen keinen Personenbezug, Ziel der Analyse ist aber eine einzelne Person oder eine Personengruppe.	z.B. allgemein verfügbare personenbezogene Daten, personenbezogene Daten mit Zustimmung der Betroffenen
einzelfallbezogene Entscheidungen	Einzelfallbezogene Entscheidungen haben besitzen die höchste Intensitätsstufe der Auswertung personenbezogener Daten. Als Datenmaterial der Auswertung werden personenbezogene Daten herangezogen und für eine bestimmte Person möglicherweise Aktionen setzen zu können.	insbesondere gesetzliche Regelung (§ 49 DSG)

[Rz 110] Gemäß den erörterten datenschutzrechtlichen Vorgaben bestimmt sich die bei der Analyse von Verwaltungsdaten angezeigte Vorgehensweise nach der Art und dem Ursprung der zu verwendenden Daten sowie nach dem Wesen der zu erzielenden Ergebnisse. Zusammenfassend wird festgehalten, dass die Nutzung allgemein verfügbarer Daten oder solcher Daten ohne Personenbezug

keinen Beschränkungen des DSG unterliegt (siehe Kapitel 4.2.1). Die weitere Be- bzw. Verarbeitung ursprünglich allgemein verfügbarer Daten kann jedoch weitere Rechtsfragen aufwerfen. Hinzuweisen ist in diesem Zusammenhang exemplarisch auf die durch den EuGH vorgenommene Klassifizierung der Tätigkeit einer Suchmaschine als Verarbeiten von personenbezogenen Daten im Sinn von Art. 2 lit b Richtlinie 95/46/EG.[68]

[Rz 111] Allerdings ist festzustellen, dass ähnlich wie die Definition von Big Data, die im Bereich der Datengröße sowie der Fähigkeit Datenmengen rasch zu bearbeiten durch den technologischen Fortschritt laufend angepasst werden muss (vgl. 2.2, Öffentliche Register als Quelle für Big Data Auswertungen), die Möglichkeit der Re-Identifizierung von Personen aus anonymisierten Daten durch höhere Rechenleistung und neuartige Analysemethoden gegeben sein kann. Die Art und Weise der Rückführung sowie der damit verbundene Aufwand sind somit Parameter, die im Zeitverlauf neu bewertet und durch Adaption von Anonymisierungsverfahren berücksichtigt werden müssen. Das Problem der Re-Identifikation anonymisierter Daten wurde durch die Analyse von Kreditkartentransaktionen beschrieben. Durch die Miteinbeziehung von öffentlich verfügbaren Daten sowie legal beschaffbarer weiterer Metadaten abseits des zur Analyse bereitgestellten Auswertungsmaterials, konnten anonymisierte Kreditkartentransaktionen, die für einen Zeitraum von drei Monaten Forschern bereitgestellt wurden, 90% der beteiligten Personen re-identifiziert und somit eindeutig zugeordnet werden.69

[Rz 112] Die von Verwaltungseinheiten in Auftrag gegebene Analyse von Verwaltungsdaten wird in der Regel nicht im lebenswichtigen Interesse der Betroffenen gelegen sein, in Einzelfällen ist dies jedoch denkbar (siehe Kapitel 4.2.2). Die Zustimmung aller Betroffenen zur Verwendung der Daten wird eher selten zum Tragen kommen können (siehe Kapitel 4.2.3).

[Rz 113] Schließlich kommt eine gesetzliche Regelung, die Verwendung von Daten i.S.d. § 46 Abs. 1 DSG in der dort beschriebenen Weise (siehe Kapitel 4.2.4) oder eine Genehmigung der Datenschutzbehörde nach § 46 Abs. 2 Z 3 i.V.m. Abs. 3 DSG in Betracht (4.2.5).

[Rz 114] Nachfolgend wird zunächst auf die Methode der Registerzählung eingegangen. Es wird ausgeführt, wie die Bundesanstalt Statistik Österreich durch die Verwendung des Konzepts des bereichsspezifischen Personenkennzeichens Amtliche Statistik (bPK-AS) anonymisierte Daten zur statistischen Auswertung nutzt. Daran anschließend wird mögliches unausgeschöpftes Potential im Hinblick auf die Analyse der Registerdaten aufgezeigt. Darauf folgt die Vorstellung einer zweiten Methode, die für die Analyse von Verwaltungsdaten im Einklang mit dem geltenden Datenschutzrecht geeignet sein könnte.

5. Methodische Umsetzung durch Registerzählung

[Rz 115] Volkszählungen sind u.a. wichtig für politische und wirtschaftliche Entscheidungen, haben wahlrechtliche Implikationen sowie Auswirkungen auf die Aufteilung gewisser finanzieller Ressourcen unter den Gebietskörperschaften und dienen der für die volkswirtschaftliche Gesamtrechnung erforderlichen Bestandsaufnahme der bestehenden Daten. Große Teile der Daten, die für die Registerzählung gesammelt werden sollten, waren vor der Normierung der Registerzählung bereits vorhanden. Hiermit ist vor allem das vom BMI als Betreiber und Dienstleister für die Meldebehörden

68 Urteil des EuGH C-131/12 vom 13. Mai 2014, Google Spain SL, Google Inc. / Agencia Española de Protección de Datos, Mario Costeja González.

zu führende Zentrale Melderegister gemeint. Weiters sind sowohl die von der Bundesanstalt Statistik Österreich zu führenden Register (Gebäude- und Wohnungsregister, Bildungsstandregister, Unternehmensregister) als auch andere, in Vollziehung gesetzlicher Vorschriften gesammelte Daten angesprochen, etwa solche betreffend das Ausmaß der Beschäftigung, wie sie beim Arbeitsmarktservice Österreich oder bei den Sozialversicherungsträgern vorliegen.Die für die Verknüpfung dieser Daten notwendige gesetzliche Grundlage der registerbasierten Volks-, Arbeitsstätten-, Gebäude- und Wohnungszählung ist das Bundesgesetz über die Durchführung von Volks-, Arbeitsstätten-, Gebäude- und Wohnungszählungen (Registerzählungsgesetz).[69]

[Rz 116] Dieses wurde auf Grund von Verordnung (EG) Nr. 763/2008 über Volks- und Wohnungszählungen, ABl. Nr. L 218 vom 13. August 2008, geändert.[70]

[Rz 117] Für jedes Jahrzehnt ist eine Volks-, Arbeitsstätten-, Gebäude- und Wohnungszählung angeordnet. Zwischenzählungen dürfen von der Bundesregierung 5 Jahre nach der letzten Zählung angeordnet werden (§ 1 Abs. 1 und 2 Registerzählungsgesetz).

[Rz 118] Im Sinn der verfassungsrechtlichen Grundsätze der Sparsamkeit, Wirtschaftlichkeit und Zweckmäßigkeit, ist eine Zwischenzählung nur gerechtfertigt, «*wenn mit einer wesentlichen Änderung des letzten Zählergebnisses zu rechnen ist.*»[71]

5.1. Erhebung anonymisierter Daten

[Rz 119] Die Daten werden gemäß § 4 Registerzählungsgesetz nicht mit Namen der Betroffenen, sondern durch Heranziehung bereichsspezifischer Personenkennzeichen (bPK) gemäß § 9 Bundesgesetz über Regelungen zur Erleichterung des elektronischen Verkehrs mit öffentlichen Stellen (E-GovG)[72] erhoben. Zur Berechnung des bPK ist es zunächst erforderlich, dass aus der ZMR-Zahl, die natürlichen Personen im Rahmen des Zentralen Melderegisters[73] zugewiesen wird, die Stammzahl in einer durch eine starke Verschlüsselung gesicherten Weise abgeleitet wird (§ 6 Abs. 2 E-GovG). BPKs werden grundsätzlich durch nicht umkehrbare Ableitungen aus der Stammzahl gebildet (§ 13 Abs. 1 E-GovG). Für jeden Bereich staatlicher Tätigkeit, dem eine Datenanwendung zugehörig ist, werden eigene bPK verwendet (§ 9 Abs. 2 E-GovG). Dadurch soll sichergestellt werden, dass eine Verknüpfung aller Datensätze über einen Betroffenen verhindert wird. Dies ist ein Beispiel von «privacy by design» in der öffentlichen Verwaltung.

[Rz 120] Die Durchführung der Erhebung ist in § 6 Registerzählungsgesetz im Wesentlichen wie folgt geregelt: Die Inhaber von gemäß Registerzählungsgesetz miteinzubeziehenden Verwaltungsdaten haben auf Verlangen der Bundesanstalt Statistik Österreich bei der Stammzahlenregisterbehörde (Datenschutzbehörde) die Erzeugung der bPK für den betreffenden staatlichen Tätigkeitsbereich

69 ErlRV 1193 BlgNR XXII. GP 3–5 (http://www.parlament.gv.at/PAKT/VHG/XXII/I/I_01193/fname_051935.pdf); Bundesgesetz über die Durchführung von Volks-, Arbeitsstätten-, Gebäude- und Wohnungszählungen (Registerzählungsgesetz), BGBl I 33/2006 i.d.F. BGBl I 125/2009.

70 ErlRV 320 BlgNR XXIV. GP 1, http://www.parlament.gv.at/PAKT/VHG/XXIV/I/I_00320/fname_165758.pdf

71 ErlRV 1193 BlgNR XXII. GP 6, http://www.parlament.gv.at/PAKT/VHG/XXII/I/I_01193/fname_051935.pdf

72 Bundesgesetz über Regelungen zur Erleichterung des elektronischen Verkehrs mit öffentlichen Stellen (E-Government-Gesetz – E-GovG), BGBl I 10/2004 i.d.F. BGBl I 83/2013.

73 Natürlichen Personen, die nicht im Zentralen Melderegister eingetragen, steht eine Eintragung im Ergänzungsregister offen.

sowie die Erzeugung des bPK-AS (bereichsspezifisches Personenkennzeichen Amtliche Statistik, siehe § 15 Bundesstatistikgesetz) zu beantragen. Verschlüsselte bPK-AS sind von Inhabern der Verwaltungsdaten aufzubewahren. Die Bundesanstalt Statistik Österreich ist zur Aufbewahrung der bPK-AS und der verschlüsselten bPK des Tätigkeitsbereichs verpflichtet.

[Rz 121] Dieser vollständig anonymisierte Datenbestand wird mit genauso anonymisierten Datenbeständen von Inhabern anderer Verwaltungsdaten verknüpft, womit Datenschutz besser sichergestellt werden kann im Vergleich zur papierbasierten Volkszählung, wobei es durchaus zu Situationen kommen kann, in denen die auswertenden Personen der gleichen Gemeinde angehören wie die teilnehmenden Personen und diesen nahe stehen.[74]

[Rz 122] Kurz zusammengefasst generiert die Bundesanstalt Statistik Österreich einen asymmetrischen Schlüssel, dessen öffentlichen Teil sie an die Stammzahlenregisterbehörde übersendet. Der Inhaber der Verwaltungsdaten beantragt auf Verlangen der Bundesanstalt Statistik Österreich sowohl das bPK für seinen staatlichen Tätigkeitsbereich als auch das verschlüsselte bPK-AS. Die Beantragung des bPK-AS kann durch die Bundesanstalt Statistik Österreich deshalb nicht erfolgen, weil zu dessen Berechnung das Geburtsdatum und der Name der/des Betroffenen nötig ist und die Bundesanstalt Statistik Österreich die verwendeten Daten ohne Namen der Betroffenen speichert. Diese übermittelt der Inhaber der Verwaltungsdaten in der Folge gemeinsam mit den Daten der/des Betroffenen ohne Namen an die Bundesanstalt Statistik Österreich. Sie entschlüsselt schließlich das bPK-AS und erhält somit den eindeutigen Ordnungsbegriff der Statistik.[75]

Abbildung 3: Durch bPK und bPK-AS anonymisierte Datenbestände der Registerzählung[76]

[Rz 123] Zur Qualitätssicherung werden Basisdaten (§ 4 Registerzählungsgesetz) mit sogenannten Vergleichsdaten auf Vollständigkeit und Übereinstimmung verglichen. Etwa bei Widersprüchlichkeit der Daten kann zusätzlich zur Abklärung der Bundesanstalt Statistik Österreich mit dem Inhaber der Verwaltungsdaten unter Anführung des jeweils verschlüsselten bPK und bPK-AS auch eine Befragung der Betroffenen erforderlich sein. Diesfalls hat der Inhaber der Verwaltungsdaten auf Verlangen der Bundesanstalt Statistik Österreich den Namen und die Adresse der Betroffenen bekannt zu geben (§ 5 Abs. 1, 2 und Abs. 5 i.V.m. § 6 Abs. 2 Bundesstatistikgesetz).

[Rz 124] Die Anlage zum Registerzählungsgesetz bestimmt die Erhebungsmerkmale. Beispielsweise werden gemäß 1.1, 1.4 bis 1.7 der genannten Anlage die Wohnadresse des Hauptwohnsitzes (Definition: § 1 Abs. 7 Meldegesetz [MeldeG]), die Adresse der Kontaktstelle der Obdachlosen, das Geburtsdatum, das Geschlecht und die Staatsangehörigkeit erhoben. Diese Merkmale werden gemäß § 4 Abs. 1 Z 1 Registerzählungsgesetz durch Beschaffung von Verwaltungsdaten von den Meldebehörden erhoben. Das Ergebnis wird anschließend mit den Daten der Sozialversicherungsträger, der Krankenfürsorgeanstalten der Länder und Gemeinden, der Kammern für freie Berufe,

74 http://statistik.gv.at/web_de/frageboegen/registerzaehlung/weitere_informationen/faq/055947.html.

75 ErlRV 1193 BlgNR XXII. GP 13–14 (http://www.parlament.gv.at/PAKT/VHG/XXII/I/I_01193/fname_051935.pdf); siehe auch § 13 Bundesstatistikgesetz.

76 In Anlehnung an ErlRV 1193 BlgNR XXII. GP 14 (http://www.parlament.gv.at/PAKT/VHG/XXII/I/I_01193/fname_051935.pdf).

mit den einschlägigen Daten der Bundesanstalt Statistik Österreich betreffend natürliche Personen (Schul- und Hochschulstatistik, Bildungsstandregister, Gebäude- und Wohnungsregister), mit den Daten der Abgabenbehörden des Bundes (Steuerregister) und den Daten des Arbeitsmarktservice Österreich sowie mit Daten der zentralen Zulassungsevidenz, des Familienbeihilfenregisters, des Zentralen Fremdenregisters, des Betreuungsinformationssystems, des Asylwerberinformationssystems, der Sozialhilfeträger der Länder, der Dienstbehörden und der die Dienstgeberfunktion wahrnehmenden Verwaltungsstellen des Bundes und der Länder verglichen (§ 5 Abs. 1 Registerzählungsgesetz).

[Rz 125] Es können 15 maßgebliche Registerbereiche, unterteilt in die unten erstgenannten Basisregister und die darauffolgenden Vergleichsregister, unterschieden werden:[77]

1. Zentrales Melderegister
2. Daten des Hauptverbandes der Sozialversicherungsträger
3. Steuerdaten
4. Daten des Arbeitsmarktservices
5. Bildungsstandregister
6. Schul- und Hochschulstatistik
7. Gebäude- und Wohnungsregister
8. Unternehmens- und Land- und forstwirtschaftliches Register
9. Fremdenregister
10. Dienstgeberdaten des Bundes und der Länder
11. Sozialhilfedaten der Länder
12. Familienbeihilferegister
13. Zivildienerdatei
14. Präsenzdienerdatei
15. Zentrale Zulassungsevidenz

[Rz 126] Auf diese Weise werden zu erhebende Merkmale durch Miteinbeziehung von 15 Datenquellen in einem auf Konsistenz geprüft, während andererseits die Ergebnisse in tieferer Ebene ausgewertet werden können. Dadurch können sie für beliebige Gebiete, auch regional, unabhängig von Verwaltungsgrenzen, gut genutzt werden.[78]

5.2. Laufende Auswertung nach dem Verfahren der Registerzählung

[Rz 127] Aus rechtlicher Sicht ist festzuhalten, dass gemäß § 4 Bundesstatistikgesetz die Organe der Bundesstatistik jene Statistiken zu erstellen haben, die durch internationalen, unmittelbar wirksamen Rechtsakt, durch Bundesgesetz oder – unter den in § 4 Abs. 3 Bundesstatistikgesetz näher konkretisierten Voraussetzungen – durch Verordnung angeordnet sind. § 1 Registerzählungsgesetz regelt den Grundsatz der Durchführung der Zählung an der Wende eines jeden Jahrzehnts und ermächtigt die Bundesregierung zur Anordnung im 5-Jahres-Rhythmus, falls die Auswirkungen voraussichtlich eine Änderung der Entsendung von Mitgliedern in den Bundesrat haben. Die

77 http://statistik.gv.at/web_de/frageboegen/registerzaehlung/weitere_informationen/faq/055947.html#index15.

78 http://statistik.gv.at/web_de/frageboegen/registerzaehlung/weitere_informationen/faq/055947.html#index15.

Periodizität der statistischen Erhebung (Zeitabstände der Datenerhebung[79]) hat die verfassungsrechtlichen Grundsätze der Sparsamkeit, Wirtschaftlichkeit und Zweckmäßigkeit zu berücksichtigen (siehe Kapitel 5.1). Die nach geltender Rechtslage geregelte Periodizität, der durch die Bundesanstalt Statistik Österreich durchzuführenden Statistiken variiert unter anderem von «laufend» über monatlich, vierteljährlich, jährlich bis hin zu einem angeordneten 10-Jahres-Zyklus. Die nächste Volks-, Arbeitsstätten-, Gebäude- und Wohnungszählung 2021 wird aktuell vorbereitet.[80]

[Rz 128] Unter der Annahme, dass die durch laufende Auswertung gewonnene Information die dabei anlaufenden Kosten bei weitem überwiegen würde und unter der Voraussetzung der Normierung laufender Auswertungen aus den Gründen, die zur Schaffung des Registerzählungsgesetzes führten (siehe Kapitel 5.1), könnte Datenanalyse in Echtzeit rechtlich ermöglicht werden. Gemäß Richtlinie 95/46/EG[81], 2. Erwägungsgrund, stehen «*(…) Datenverarbeitungssysteme (…) im Dienste des Menschen; sie haben, (…) zum wirtschaftlichen und sozialen Fortschritt, zur Entwicklung des Handels sowie zum Wohlergehen der Menschen beizutragen.*» In diesem Sinn könnten laufende Zusammenführungen unterschiedlicher Daten im Sinn einer beständigen Evaluierung gesetzter Maßnahmen für eine breitere Entscheidungsgrundlage und ein gesteigertes Wissen über die Bedürfnisse der Bevölkerung, insbesondere in Bezug auf Dienstleistungen der Verwaltung, in Betracht gezogen werden. Festzuhalten ist, dass im Hinblick auf die erwähnten Fragen Forschungsbedarf besteht.

[Rz 129] Um zu einer laufenden Evaluierung, jedenfalls aber zu einer höheren Frequenz als es die Registerzählung vorsieht, zu gelangen, müssten die technischen Maßnahmen geschaffen werden, um von einem stapelorientierten Verfahren zu einem Streaming-Verfahren zu gelangen. Ob und in wie weit das Streaming-Verfahren mit den rechtlichen Rahmenbedingungen der Registerzählung bzw. der Bereichsabgrenzung, die Grundlage des bPK-Konzepts ist, in Einklang gebracht werden kann, ist allerdings nicht Gegenstand der weiteren Betrachtungen.

[Rz 130] Neben den technischen Maßnahmen zur Steigerung der Analysefrequenz wird es des Weiteren erforderlich sein, auch den Inhalt der Analysen zu hinterfragen. Für die Registerzählung wird ein zeitlicher Rahmen von 10 Jahren vorgegeben, während Zwischenzählungen frühestens nach 5 Jahren möglich sind (vgl. Abschnitt 5). Der Erhebungsgegenstand der Registerzählung lässt vermuten, dass kurzfristigere Erhebungsintervalle nicht durch die zu erwartenden Veränderungen in den Ergebnissen zu rechtfertigen sind. Kürzere bis laufende Erhebungsintervalle müssen somit durch die zu erwartenden Ergebnisse begründbar sein, um dem Grundsatz der Sparsamkeit und Verhältnismäßigkeit Genüge zu tun. Andererseits steht dieser Grundsatz im Widerspruch zu dem Big Data Prinzip, wonach Auswertungen auch ohne Vorliegen eines konkreten Zwecks zu positiven Wirkungseffekten führen können. In Abschnitt 4.2.3 Zustimmung widmeten wir uns diesem Spannungsfeld, in welchem der Grundsatz der Ermittlung von Daten für einen festgelegten, eindeutigen und rechtmäßigen Zweck und der Grundgedanke von Big Data zueinander stehen.

[Rz 131] Konkret könnten etwa Infrastrukturmaßnahmen – sei es betreffend den öffentlichen Verkehr oder die Nahversorgung – mit Hilfe der Ergebnisse aus der Analyse von Daten über Pendlerströme zielgerichteter getroffen werden. Pendler oder Schichtarbeiter benötigen flexible Öffnungs-

79 § 3 Z 12 Bundesstatistikgesetz.

80 Vgl. Anlage II zu Bundesgesetz über die Bundesstatistik (Bundesstatistikgesetz 2000) BGBl I 163/1999 i.d.F. BGBl I 40/2014.

81 Richtlinie 95/46/EG des Europäischen Parlaments und des Rates vom 24. Oktober 1995 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten und zum freien Datenverkehr, ABl L 281 vom 23. November 1995, 31.

zeiten von Ämtern und Geschäften. Junge Familien werden in der Zukunft möglicherweise einen Kindergartenplatz, ältere Menschen einen Seniorenplatz benötigen sowie öffentliche Infrastrukturmaßnahmen, die ihnen das Altwerden in der gewohnten Umgebung ermöglichen etc. Für Eltern Jugendlicher könnte etwa der Anteil der unter 15-Jährigen in den Gemeinden interessant sein, um an die Politik jugendförderliche Maßnahmen heranzutragen.[82] Insbesondere für alleinstehende ältere Menschen könnte der Anteil der Senioren von Interesse sein. Zur Erkennung von Trends und Korrelationen und als aktueller Service für Bürgerinnen und Bürger könnten etwa nachstehende Daten laufend erhoben und beinahe in Echtzeit analysiert werden:

- die Anzahl und das Angebot für Personen, die sich in Karenz befinden;
- die Anzahl der Arbeitsstätten mit qualifizierten Stellen einer bestimmten Fachrichtung;
- der Standort von Schulen mit bestimmtem pädagogischem Angebot und Daten zum Lehrkörper sowie über Schüler;
- Daten zu Studierenden und Lehrenden an Universitäten mit speziellem Lehrangebot oder einschlägiger Forschung;
- die Anzahl gewisser Gebäudearten;
- das Verhältnis des Eigentums zur Miete der Gebäude und Wohnungen in einer bestimmten Region;
- verfügbare Infrastruktur in einer Region;
- das Verhältnis der Meldungen eines Hauptwohnsitzes zur Meldung eines Nebenwohnsitzes in einer Region.

[Rz 132] Würden Daten wie diese z.B. wöchentlich aufbereitet und leicht zugänglich bzw. barrierefrei vorliegen, könnten insbesondere auch interessierte Einwohner einer Region die Entwicklung ihres Gebiets – beinahe in Echtzeit – mitverfolgen. Ein kürzeres Erhebungsintervall wäre des Weiteren im Sinn der wirkungsorientierten Verwaltungsführung der rationalen Planungsqualität von Entscheidungsträgern förderlich.

[Rz 133] Angesichts einer Begrenzung auf 15 – strukturierte, siehe Kapitel 5.1 – Datenquellen, ist anzunehmen, dass es sich bei einer Verschneidung von diesen ausgewählten Registern (derzeit noch) nicht um Big Data Analyse handelt. Die Analyse von Daten auf wöchentlicher Basis wäre bei Adaptierung des Registerzählungsgesetzes möglich, kostengünstiger wäre bei weiterer Granularisierung des Erhebungsintervalls der Umstieg auf ein Streaming-Auswertungsverfahren. Die daraus entstehenden Ergebnisse könnten der Bevölkerung fast in Echtzeit zur Verfügung gestellt werden und potentiell Innovationen in Gesellschaft, Wirtschaft und Verwaltung fördern.

6. Big Data Analysen verschlüsselter Daten

[Rz 134] Die im vorhergehenden Abschnitt genannte Methode erfordert das wirksame Zusammenspiel mehrerer Institutionen. Dieser Abschnitt erläutert einen weiteren Ansatz zur Lösung des Problems der Auswertung insbesondere solcher großen Datenbestände, die einen Personenbezug enthalten. Wäre es möglich, Daten auch dann sinnvoll auszuwerten, wenn diese verschlüsselt und in Teile geteilt sind, wobei bewusst Redundanzen herbeigeführt werden? Wäre es möglich, dass dadurch Daten entstehen, welche aus dem Anwendungsbereich des DSG ausgenommen sind? Das

[82] Gemeindetabelle Österreich (http://www.statistik.at/web_de/statistiken/menschen_und_gesellschaft/bevoelkerung/volkszaehlungen_registerzaehlungen_abgestimmte_erwerbsstatistik/index.html).

nachfolgend beschriebene Konzept könnte dazu führen, dass die wissenschaftliche Auseinandersetzung mit den Möglichkeiten von Big Data Auswertungen ursprünglich personenbezogen gewesener Daten unter einem gänzlich neuen, bedeutsamen Blickwinkel und in Einhaltung des DSG erfolgen könnte. Dieses charmante Gedankenexperiment wurde theoretisch gelöst: Die sinnbringende Analyse verschlüsselter Daten, wobei die Analyse selbst lediglich auf Teilen dieser Daten ausgeführt wird. Ausführungen betreffend der technische Funktionsweise und eine erste datenschutzrechtliche Würdigung sind Inhalt dieses Abschnitts.

6.1. Enigma als möglicher Lösungsansatz für Big Data Analytics in der Verwaltung

[Rz 135] Im 2015 veröffentlichten Beitrag *Enigma: Decentralized Computation Platform with Guaranteed Privacy* wird eine praktikable Implementierung eines theoretisch gelösten Problems vorgestellt: die Durchführung von Auswertungen auf verschlüsselte Daten. Die Besonderheit an der von den Entwicklern benannten Enigma-Methode ist, dass zur Wahrung des Geheimnisses die Analysedaten verschlüsselt werden. Bevor wir auf die Analyseprozesse von Enigma näher eingehen, widmen wir uns zur Steigerung des Verständnisses kurz den wesentlichen Architekturelementen.

[Rz 136] **Homomorphe Verschlüsselung** bedeutet, dass der Träger eines Geheimnisses (welcher in der Terminologie des DSG in der Regel der Auftraggeber sein wird) dieses in verschlüsselter Form mit anderen teilen kann, diese darauf mathematische Operationen ausführen können, ohne selbst den Inhalt der Nachricht verstehen zu können und ein Ergebnis in einer Art.und Weise erzielen, dass für den Träger des Geheimnisses nach Entschlüsselung einen Sinn ergibt[83].

[Rz 137] Eine **Blockchain** ist eine auf mehrere Speicherplätze replizierte – z.B. von einem Netzwerk mehrerer Computer gebildete – Datenbank. Diese enthält anwachsende Datenbestände, die gegen Fälschung und Änderungen durch die Netzwerkteilnehmer, auch Nodes genannt, gesichert sind. Im Kontext von Enigma kann als Node ein einzelner, sich an einer Analyse Beteiligender (oder präziser: sein IT-System, das mit dem Netzwerk verbunden ist) bezeichnet werden.

[Rz 138] Neben der Blockchain[84] verwendet das Enigma-Netzwerk noch eine **verteilte, hashbasierte Datenbank**, auf der verschlüsselte als auch unverschlüsselte Daten gespeichert werden können. Während die Blockchain auf jedem der teilnehmenden Nodes repliziert ist, hat jeder Node-Teilnehmer nur einen Teil der verteilten Datenbank.

[Rz 139] **Smart Contracts** sind Ausführungsvorschriften, die von einem Computersystem nach vorheriger Vereinbarung durchgeführt werden und deren Existenz, Vorbedingungen oder auch Ergebnisse von den Netzwerkteilnehmern bestätigt werden[85].

[Rz 140] **Bitcoin** ist eine Form virtueller Währung bei der die Geldeinheiten, Bitcoins, dezentral von den Netzwerkteilnehmern erzeugt und verwaltet werden. Dieses Netzwerk wird aus Teilnehmern gebildet, die einen Bitcoin-Client ausführen und über das Internet miteinander verbunden sind[86].

---

83 Craig Gentry, Fully homomorphic encryption using ideal lattices, STOC. Vol. 9. 2009.

84 Christian Decker / Roger Wattenhofer, Information propagation in the Bitcoin network, in Peer-to-Peer Computing (P2P), 2013 IEEE Thirteenth International Conference on, 2013, S. 3.

85 Vitalik Buterin, A next-generation smart contract and decentralized application platform, White Paper, 2014. S. 1.

86 Satoshi Nakamoto, Bitcoin: A peer-to-peer electronic cash system, (2008): 28. http://www.cryptovest.co.uk/resources/Bitcoin%20paper%20Original.pdf.

6.2. Berechnungsabläufe im Enigma-Netzwerk

[Rz 141] Prinzipien im Enigma Netzwerk sind, dass keiner, außer der Träger eines Geheimnisses, jemals Zugang zu den vollständigen Daten hat: Jeder Teilnehmer am Enigma-Netzwerk verfügt nur über einen scheinbar willkürlichen und bedeutungslosen Teil der Daten, ähnlich einer Menge an Puzzlesteinen, die keinen Rückschluss auf das gesamte Bild erlauben. Das weitere wichtige Funktionsprinzip von Enigma ist die Möglichkeit, verteilt Berechnungen (Analysen) auf verschlüsselte Daten durchführen zu können, ohne Zugang zu den Rohdaten zu haben.

[Rz 142] Während die Lösung des Problems der Analyse verschlüsselter Daten bereits theoretisch erfolgte,[87] konnten Implementierungen keine praktische Verwendung finden, da die erforderliche Berechnungszeit bis zu einer Million höher war als im Fall unverschlüsselter Daten. Durch eine Reihe von Optimierungen gegenüber der strikten homomorphen Verschlüsselung konnte der Berechnungsaufwand im Enigma Netzwerk wesentlich verkürzt werden und ist rund 100-fach höher gegenüber dem unverschlüsselten Fall, was deren Anwendung als Grundlage für Big Data Auswertungen in der öffentlichen Verwaltung sinnvoll erscheinen lässt.

[Rz 143] Der Auftraggeber einer Analyse im Enigma-Netzwerk veröffentlicht im Fall der Personenbezogenheit von Daten diese unter Nutzung eines Algorithmus, der an das Prinzip der homomorphen Verschlüsselung angelehnt ist. Der Enigma-Interpreter nutzt die Blockchain zur Koordination der verteilten Analyse. Zweck der Blockchain ist, dass die Tatsache der Übernahme einer Berechnungsoperation sowie die Bekanntgabe eines (verschlüsselten) Endergebnisses über die Blockchain erfolgt und von Netzwerkteilnehmern durch ein kryptographisches Verfahren (Signatur) bestätigt wird. Außerdem werden in der Blockchain die Verweise auf die eigentlichen, verschlüsselten Berechnungsdaten abgelegt, die aus Sicherheitsgründen und auf Grund ihrer Größe unter Verwendung einer verteilten Hash-Datenbank auf den Nodes des Enigma Netzwerks verteilt liegen.

[Rz 144] Der Datenanalyst erhält einen Teil der Berechnungsvorschrift der Gesamtauswertung, die auch Verweise auf die dazu erforderlichen verschlüsselten Daten in der verteilten hashbasierten Datenbank enthält. Auf Basis dieser Informationen führt der von ihm betriebene Teil des Enigma Netzwerks jene Rechenoperationen aus, die in seinem «Private Contract» verlangt wurden, den er durch Teilnahme am Netzwerk eingegangen ist. «Private Contract» wird insofern als eine Weiterentwicklung des «Smart Contracts» verstanden, als ein Private Contract auch geheimzuhaltende Informationen enthalten kann. Zur Erhöhung der Datensicherheit werden die Rechenoperationen bewusst teilweise redundant ausgeführt. Das Ergebnis der Berechnungen macht für den einzelnen Teilnehmer keinen Sinn. Er gibt das Ergebnis in der Folge an den Auftraggeber zurück, der sie zusammensetzen und als einziger Akteur sinnvoll verwenden kann.

[Rz 145] Um die Sicherheit im Enigma-Netzwerk zu erhöhen, wurde eine Reihe von organisatorischen Maßnahmen ergriffen. An der Teilnahme am Enigma-Netzwerk berechtigt sind Nodes, nachdem sie eine Menge an Bitcoins treuhänderisch an einen Berechnungskoordinator überwiesen haben.[88] Diese Summe kann vom Teilnehmer einer Analyse im Enigma-Netzwerk zurückgefordert werden. Würde ein Analyseteilnehmer gegen seinen Private Contract verstoßen, wie beispielsweise eine Rechenoperation verfrüht abbrechen oder falsche Ergebnisse retournieren, werden ihm Bitcoins entzogen und diese unter den vertragstreuen Teilnehmern aufgeteilt. Dies ist möglich, da manipulierte Aktionen vom System erkannt werden. Sinkt das «Bitcoinkonto» eines Teilnehmers

87 Craig Gentry, Fully homomorphic encryption using ideal lattices, STOC. Vol. 9. 2009, 169–178.

88 https://bitcoin.org/en/.

unter einen gewissen Wert, so bekommt er keine Rechenoperationen mehr zugeteilt. Auch der Ausschluss eines Teilnehmers aus dem Netzwerk ist möglich. In Fällen vertragsgemäßer Rechenleistung durch Teilnehmer, werden diese nach näher konkretisierten Maßstäben in Bitcoins bezahlt.

[Rz 146] In diesem Zusammenhang ist es wichtig festzuhalten, dass das Funktionieren des Engima-Netzwerks als Grundlage von Analyseoperationen unabhängig von der digitalen Währung Bitcoin ist: Jedes andere Treuhandservice, ob elektronisch oder nicht, könnte organisatorisch zur Teilnahme am Netzwerk sowie zur Wahrung der vertraglichen Treue bei der Übernahme von Analyseaufgaben beitragen.

6.3. Rechtliche Schlussfolgerungen

[Rz 147] Indem in Enigma jeder einzelne Node nur einen bedeutungslosen, scheinbar willkürlichen Teil der Daten hat, wäre die Identität der einzelnen Betroffenen für die Nodes weder «direkt ersichtlich (‹bestimmt›)»[89], noch «ohne besonderen zusätzlichen Aufwand ‹bestimmbar›».[90]

[Rz 148] Die richtige Lieferung von Ergebnissen kann überprüft werden, während die betroffenen Daten geheim bleiben.[91] Es liegt nahe, dass es sich hier um die Nutzung nicht personenbezogener Daten handelt.

[Rz 149] Vorausgesetzt wird, dass ein direkter Personenbezug nicht vorliegen wird.

[Rz 150] Zu prüfen ist daher, ob nicht doch indirekter Personenbezug festgestellt werden kann oder ob hier wirklich von der Verwendung anonymisierter Daten, d.h. nicht personenbezogener Daten, ausgegangen werden kann.

[Rz 151] Dazu wird unter Verweis auf 4.2.1 festgehalten, dass Daten von Erwägungsgrund 26 der Richtlinie 95/46/EG festgehalten wird, bei der Beurteilung, ob eine Person bestimmbar ist, «sollten alle Mittel berücksichtigt werden, die vernünftigerweise entweder von dem Verantwortlichen für die Verarbeitung oder von einem Dritten eingesetzt werden könnten, um die betreffende Person zu bestimmen.» Hingegen verlangt § 4 Z 1 DSG für Klassifizierung als nur indirekt personenbezogene Daten, dass der Verwender der Daten die Identifizierung eines Betroffenen mit rechtlich zulässigen Mitteln nicht durchführen kann.

[Rz 152] Mit rechtlich zulässigen Mitteln kann der einzelne teilnehmende Datenanalyst bzw. der Node die Identifizierung Betroffener nicht durchführen. Dazu müsste er sich rein theoretisch auf illegalem Weg zunächst den Entschlüsselungscode beschaffen. Dies könnte dazu führen, dass er in Bezug auf seinen Teil der Daten eine Identifizierung oder Re-Identifizierung erreichen kann.

[Rz 153] Die Rückführung auf die Identität des Betroffenen könnte diesfalls nur mit Mitteln vorgenommen werden, die ihrem Aufwand nach völlig ungewöhnlich und absolut unverhältnismäßig sind, d.h. vernünftigerweise nicht ergriffen werden können. Indirekt personenbezogene Daten liegen für den in Prüfung stehenden Auftraggeber oder Dienstleister oder Empfänger der Übermittlung vor, wenn dieser mit rechtmäßigen Mitteln die Identität der/des Betroffenen nicht bestimmen kann (§ 4

89 Walter Dohr / Hans J. Pollirer / Ernst M. Weiss / Rainer Knyrim, Kommentar zum Datenschutzrecht, Manz 2014, Anm. 2 zu § 4.

90 Walter Dohr / Hans J. Pollirer / Ernst M. Weiss / Rainer Knyrim, Kommentar zum Datenschutzrecht, Manz 2014, Anm. 2 zu § 4.

91 Guy Zyskind / Oz Nathan / Alex «Sandy» Pentland, Enigma: Decentralized Computation Platform withGuaranteed Privacy, (2015), 2.

Z 1 DSG). Die Datenschutzkommission hat dies insofern präzisiert, als rechtlich verpönte Mittel wie Einbruch, Zwang oder Bestechung herangezogen werden müssten, um den Schlüssel zu erlangen, womit die Re-Identifizierung möglich würde. Die Voraussetzung einer «ausreichenden faktischen (technisch-organisatorischen) Absicherung der Daten gegen die Möglichkeit missbräuchlicher Re-Identifikation» werde von § 4 Z 1 DSG vorgegeben. Diese missbräuchliche Re-Identifikation solle für den Verwender «praktisch nicht möglich» sein.[92]

[Rz 154] Dem Konzept des Private Contract folgend, der die Aufgabe der Nodes vorschreibt, ist die Beurteilung, dass einzelne Nodes eine Re-Identifizierung und somit Rückführung der Daten auf die Identität von Betroffenen mit rechtmäßigen Mitteln nicht ausführen können, naheliegend. Der Aufwand, der betrieben werden müsste, um die Identität einzelner Betroffener zu bestimmen, kann unter Nutzung von Enigma jedoch nur mit Mitteln vorgenommen werden, die ihrer Art und ihrem Aufwand nach völlig ungewöhnlich und absolut unverhältnismäßig sind, d.h. vernünftigerweise nicht ergriffen werden können, weil

- die Analysedaten in einer Form verschlüsselt sind, dass ein derart extremes Ausmaß an Rechenleistung zusammengeschaltet werden müsste, das aktuell in absehbarer Zukunft weltweit nicht aufgebracht werden könnte;
- kein einzelner Node über den Gesamtdatensatz verfügt;
- Redundanzen bewusst eingebaut werden;
- nur die Beschreibungen der Daten bekannt sind;
- nur der Auftraggeber der Analyse einen Gesamtüberblick über den Gegenstand der Analyse hat;

[Rz 155] Daher könnten die Daten ab dem Zeitpunkt, zu dem sie noch außerhalb des Netzwerks dem Prinzip der homomorphischen Verschlüsselung folgend verschlüsselt wurden, im Sinn der Richtlinie 95/46/EG für die Nodes als nicht personenbezogene Daten angesehen werden. Der Mangel an Personenbezug der Daten hätte zur Folge, dass das DSG nicht zur Anwendung käme und alle möglichen Analysen und Auswertungen durchgeführt werden könnten.

[Rz 156] Selbst im Fall einer technischen und rechtlichen Widerlegung dieser Annahme, hätte Enigma entscheidende Vorteile. Unter der Annahme des Bestehens eines indirekten Personenbezuges, wären folgende rechtliche Überlegungen anzustellen.

[Rz 157] Bei der Überlassung von Daten an einen Dienstleister, ist sicherzustellen, dass dieser ausreichende Gewähr für eine rechtmäßige und sichere Datenverwendung bietet und ist mit diesem auch eine dahingehende Vereinbarung zu treffen, von deren Einhaltung sich Auftraggeber zu überzeugen haben (§ 10 DSG).

[Rz 158] Das Treffen einer derartigen Vereinbarung im Enigma-Netzwerk wirkt auf den ersten Blick schwierig, da die Identität der tatsächlich teilnehmenden Dienstleister von Vornherein nicht bestimmt ist. Deren Identität ist jedoch über deren Anmeldung und die Weiterleitung einer Summe an «Bitcoins» an die Blockchain bestimmbar.

[Rz 159] Den Teilnehmern wird der Zugang zu verschlüsselten Datenteilen nur nach Akzeptieren einer Vereinbarung mit dem Auftraggeber ermöglicht (im Rahmen von Enigma «Private Contract» genannt), weshalb es naheliegt, jeden Teilnehmer selbst als Dienstleister im Sinn des DSG zu sehen.

[Rz 160] Dass eine den Datenschutz sichernde Vereinbarung (siehe oben zu § 10 DSG) jedoch auf

92 Datenschutzkommission, Datenschutzbericht 2009, 34 (https://www.dsb.gv.at/DocView.axd?CobId=40344).

Grund der technischen Gestaltung anders abgebildet werden könnte (bis hin zur Gewährleistung der Einhaltung der gesetzlichen Pflichten nach Art eines «Self Executing Contracts») wird im Folgenden erörtert.

[Rz 161] Bei der Überlassung von Daten an Dienstleister ist insbesondere die Gewährung einer rechtmäßigen, sicheren Datenverwendung zu beachten und sind entsprechende Vereinbarungen zu treffen sowie deren Einhaltung zu prüfen. Will ein Auftraggeber des öffentlichen Bereichs von dieser Möglichkeit Gebrauch machen, ist zu prüfen, ob eine Mitteilung an die Datenschutzbehörde zu erfolgen hat (§ 10 DSG).

[Rz 162] Den Dienstleister treffen zusätzlich die in der linken Spalte in Tabelle 2 wiedergegebenen Pflichten des § 11 Abs. 1 Z 1–6 DSG. Die rechte Spalte der Tabelle enthält die damit im Zusammenhang stehenden Auswirkungen einer Anwendung des Konzepts Enigma.

Tabelle 2: Pflichten des DSG und Sicherstellung von deren Einhaltung durch Enigma

Pflichten des DSG	Sicherstellung durch Enigma
Verwendung der Daten ausschließlich im Rahmen der Aufträge des Auftraggebers; insbesondere die Übermittlung der verwendeten Daten ohne Auftrag des Auftraggebers ist nicht zulässig	Sämtliche im Enigma-Netzwerk agierende Beteiligte verfügen nur über einen Teil der verschlüsselten Daten. Manche Teile werden auch mehrmals ausgewertet (Redundanz). Zugang zu den ursprünglichen Daten besteht auf Grund des unverhältnismäßigen Aufwands, der für die Entschlüsselung notwendig wäre, daher praktisch für keinen Akteur des Enigma-Netzwerks. Deshalb ist eine Übermittlung der ursprünglichen Daten aus technischen Gründen beinahe ausgeschlossen. Die Auswertung der verschlüsselten Datenteile ergibt alleine für den Auftraggeber Sinn. Eine andere Verwendung der verschlüsselten Teile ergäbe keinen Sinn. Die Daten können nicht zu anderen Nodes übertragen werden.
Einhaltung der Datensicherheitsmaßnahmen des § 14 DSG; insbesondere Heranziehung nur solcher Mitarbeiter für die Dienstleistung, die sich dem Dienstleister gegenüber zur Einhaltung des Datengeheimnisses verpflichtet haben oder einer gesetzlichen Verschwiegenheitspflicht unterliegen	Die Einhaltung einer Verschwiegenheitspflicht ist durch Enigma obsolet. Kein Akteur des Enigma-Netzwerks verfügt über die ursprünglichen Daten, weil diese einerseits durch eine hoch komplexe Verschlüsselung gesichert sind und andererseits in Teile geteilt und redundant sind.

Hinzuziehung weiterer Dienstleister nur mit Billigung des Auftraggebers; Benachrichtigung des Auftraggebers von der beabsichtigten Hinzuziehung so rechtzeitig, dass er dies untersagen kann;	Das Enigma-Netzwerk funktioniert auch unter der Annahme, dass die Teilnehmer a priori unbekannt sind. Das Interesse, das der Auftraggeber regelmäßig an dieser Verständigung haben wird, wird sein Vertrauen in die Fähigkeiten und die Einhaltung der Datensicherheitsmaßnahmen durch den zusätzlichen Dienstleister sein. Auf Grund der technischen Gestaltung von Enigma, wird bereits technisch sichergestellt, dass gewisse Datensicherheitsmaßnahmen eingehalten werden. Es wäre theoretisch möglich, den Auftraggeber vor jeder neuen Teilnahme zu benachrichtigen. Da der Teilnehmer (nach dem derzeitigen technischen Konzept) nicht eindeutig identifiziert ist, wird dies dem Auftraggeber jedoch nicht die angestrebte Möglichkeit der Prüfung der Fähigkeiten bzw. der Einschätzung seines Vertrauens in die Einhaltung der Datensicherheitsmaßnahmen durch den Dienstleister bieten. Der abzuschließende Private Contract bietet zumindest eine vertragliche Absicherung der Fähigkeiten des Nodes. Auf Nichteinhaltung des Vertrages kann mit Entzug von Bitcoins oder Ausschluss aus dem Netzwerk reagiert werden. Daher wird der Node in der Regel in eigenem Interesse keine Durchführung von Rechenoperationen zusagen, die er nicht ausführen kann.

Sofern dies nach der Art. der Dienstleistung in Frage kommt: Schaffung der notwendigen technischen und organisatorischen Voraussetzungen für die Erfüllung der Auskunfts-, Richtigstellungs- und Löschungspflicht des Auftraggebers im Einvernehmen mit dem Auftraggeber;	Das Recht auf Auskunft, Richtigstellung und Löschung besteht gemäß § 1 Abs. 3 DSG soweit personenbezogene Daten zur automationsunterstützten (oder manuellen) Verarbeitung bestimmt sind. Im Enigma-Netzwerk sind nur verschlüsselte, redundante Datenteile einzelnen Nodes bekannt, weshalb eine Auskunftspflicht oder Richtigstellungspflicht nach der Art. der Dienstleistung eher nicht in Betracht kommen dürfte. Die Erfüllung der Löschungspflicht wäre im konkreten Fall eher nicht relevant, weil die ausgewerteten Datenteile für den einzelnen Node keinen Sinn ergeben, sondern lediglich für den Auftraggeber.
Übergabe aller Verarbeitungsergebnisse und Unterlagen, die Daten enthalten, nach Beendigung der Dienstleistung an den Auftraggeber oder in dessen Auftrag Aufbewahrung oder Vernichtung der genannten Ergebnisse und Unterlagen;	Da die Ergebnisse und die hoch komplex verschlüsselten Datenteile für das Enigma-Netzwerk nicht sinnvoll sind, ist die Gefahr der missbräuchlichen Verwendung (der die Vernichtung der Daten vorbeugen soll) als sehr, sehr gering einzustufen. Um die Leistung jedes Akteurs des Netzwerks sichtbar zu machen, ist es nötig, dass dieser sie an den Auftraggeber übermittelt. Dies ist auch erforderlich, um die entsprechende Menge an Bitcoins für die ausgeführte Leistung zu erhalten. Daher wird eine Übermittlung aller Ergebnisse an den Auftraggeber im eigenen Interesse des Nodes sein.
Zurverfügungstellung jener Informationen, die zur Kontrolle der Einhaltung der unter § 11 Z 1 bis 5 DSG genannten Verpflichtungen (für den Auftraggeber) notwendig sind	Die Einhaltung von Datensicherheitsmaßnahmen kann im Rahmen eines Protokolls dokumentiert werden.

[Rz 163] Die Verarbeitung von Daten (das Überlassen der Daten zwischen Dienstleister und Auftraggeber und das Vergleichen, Verändern, Verknüpfen etc. durch den Dienstleister) ist nach § 7 Abs. 1 DSG an die Voraussetzungen gebunden, dass Zweck und Inhalt der Datenanwendung von den gesetzlichen Zuständigkeiten oder rechtlichen Befugnissen des Auftraggebers gedeckt sind und die schutzwürdigen Geheimhaltungsinteressen der Betroffenen nicht verletzen. Der Inhalt der Datenanwendung, die die Analyse und Auswertung von Verwaltungsdaten zum Ziel hat, ist – unter der Annahme, dass es sich um in Vollziehung der Gesetze angefallene Daten handelt – wohl von den rechtlichen Befugnissen des Auftraggebers gedeckt. Ob der durch den Auftraggeber verfolgte Zweck von den rechtlichen Befugnissen umfasst ist, wird im Einzelfall zu prüfen sein. Bei der Ver-

wendung von indirekt personenbezogenen Daten gelten schutzwürdige Geheimhaltungsinteressen gemäß § 8 Abs. 2 DSG als nicht verletzt. Zulässig kann eine Datenanwendung nur sein, wenn Eingriffe in das Grundrecht verhältnismäßig sind und im Einklang mit § 6 DSG vorgegangen wird (§ 7 Abs. 3 DSG). Dies bedeutet vor allem, dass der Grundsatz der Ermittlung von Daten für eindeutig festgelegte Zwecke und das Verbot der Weiterverwendung dieser Daten in einer mit diesem Zweck unvereinbaren Weise zum Tragen kommt.

[Rz 164] Bei der Ermittlung der Daten zum Zweck des Handelns in Vollziehung der Gesetze wird der Zweck der weiteren Verwendung in der Form der Analyse und Auswertung in der Regel noch nicht bekannt sein. Die heranzuziehenden Daten werden in den meisten Fällen schon ermittelt worden sein. Es ist durchaus denkbar, dass eine Datenanalyse und Auswertung, die unter Verwendung von indirekt personenbezogenen Daten Zwecken des Erkenntnisgewinns zur Förderung einer bürgerzentrierten Verwaltung dienen soll, regelmäßig mit dem ursprünglich verfolgten Zweck nicht unvereinbar sein wird.

[Rz 165] Grundsätzlich sind Datenanwendungen von Auftraggebern vor Aufnahme der Anwendung der Datenschutzbehörde zwecks Registrierung im Datenverarbeitungsregister zu melden und dürfen nach der Meldung unmittelbar betrieben werden (§§ 17 Abs. 1, 18 Abs. 1 DSG). Sind insbesondere indirekt personenbezogene Daten in einer Datenanwendung enthalten, so entfällt die Meldepflicht des Auftraggebers gemäß § 17 Abs. 2 Z 3 DSG.

[Rz 166] Die Vorabkontrolle, d.h. Prüfung der Datenschutzbehörde nach § 20 DSG ist beispielsweise bei Datenanwendungen, die in einem Informationsverbundsystems durchgeführt werden sollen (§ 18 Abs. 2 Z 4 DSG) verpflichtend. Kennzeichnend für dieses ist nach der in § 4 Z 13 DSG enthaltenen Legaldefinition die gemeinsame Verarbeitung von Daten in einer Datenanwendung durch mehrere Auftraggeber und die gemeinsame Benützung der Daten in der Art, dass jeder Auftraggeber auch auf jene Daten im System Zugriff hat, die von den anderen Auftraggebern dem System zur Verfügung gestellt wurden. Dies kann auf die Auswertung unter Zuhilfenahme von Enigma nicht zutreffen, zumal die Daten nicht in unverschlüsselter Weise in das Enigma-Netzwerk gelangen, eine gemeinsame Nutzung von Daten nicht Ziel der Analyse ist und Auswertungen einzig für den jeweiligen Auftraggeber sinnvoll sind. Daher liegt hier kein Informationsverbundsystem im Sinn des § 4 Z 13 DSG vor.

7. Diskussion

[Rz 167] Dieser Beitrag beschäftigt sich nicht umfassend mit allen Hürden, auf die die Verwaltung bei der Umsetzung eines Projekts zur Analyse von Daten großen Umfangs stoßen kann. Vielmehr wird der Schwerpunkt auf relevante datenschutzrechtliche Rahmenbedingungen und einen neuen technischen Lösungsansatz gelegt. In diesem Rahmen nicht behandelt wurden beispielsweise die von Kaisler et al. aufgeworfenen Herausforderungen bezüglich Quantität und Qualität, die die Nutzung von Big Data mit sich bringt, etwa Entscheidungen betreffend die Auswahl der Daten, die Verlässlichkeit und Richtigkeit der Daten oder die Bestimmung der richtigen Menge der Daten, um eine Einschätzung oder Vorhersage einer spezifischen Wahrscheinlichkeit eines Ereignisses zu erzielen.[93] Hinzuweisen ist weiters etwa auf das Verbot automatisierter Einzelentscheidungen im

93 Stephen Kaisler / Frank Armour / J. Alberto Espinosa / William Money, Big Data: Issues and Challenges Moving Forward in System Sciences (HICSS), 2013 46th Hawaii International Conference, 995

Sinn von § 49 Abs. 1 DSG[94] und die in § 49 Abs. 2 und 3 DSG konkretisierten Bedingungen, die im Fall der Vorsehung ausschließlich automationsunterstützt erzeugter Entscheidungen einzuhalten sind.

[Rz 168] Ohne dass auf das Ergebnis der Entscheidung («Willensäußerungen, die Lebenschancen des Betroffenen schmälern») abgestellt würde, also auch wenn es für den Betroffenen vorteilhaft wäre, ist ausschließlich computerunterstützte Subsumption verboten.[95]

[Rz 169] Es ist damit zu rechnen, dass sich in diesem Zusammenhang zukünftiger Regelungsbedarf ergeben wird und Diskussionen zu führen sein werden.

[Rz 170] Die angestellten Überlegungen lassen erkennen, dass einzelne Register tendenziell nicht unter den Begriff Big Data subsumiert werden sollten. Die Zusammenschau von mehreren Verwaltungsregistern, angereichert um Sensordaten und verwaltungsexterne Daten, deren Verwendung letztlich ein bedeutendes Wesensmerkmal von Big Data Analysen darstellt, rechtfertigt jedoch die Würdigung von Verwaltungsregistern unter dem Blickwinkel von Big Data. Verwaltungsinterne Daten und öffentliche Daten stehen der Verwaltung potentiell zur Analyse zur Verfügung und könnten im Einklang mit datenschutzrechtlichen Anforderungen als Ausgangsbasis für Big Data Analysen weiter genutzt werden.

[Rz 171] Die vorgestellten Lösungsansätze bergen das Potential in sich, mit vorhandenen Ressourcen entlang des Policy Cycle durch Prognosen bessere Entscheidungen zu treffen, die zur laufenden Evaluierung herangezogen werden können (vgl. Abschnitt 3, Big Data Anwendungsfelder in der Verwaltung, laufende Evaluierung als Betriebsmodell des Big Data unterstützen Policy Cycles), besser auf zukünftig eintretende Ereignisse reagieren zu können und Bürgerservices sowie verwaltungsinterne Prozesse und Strategien zu optimieren.

[Rz 172] Enigma ist eine erste Implementierung eines Verfahrens, das Auswertungen auf Basis verschlüsselter Daten ermöglicht. Das technische Konzept Enigma bietet eine sehr hohe Sicherheit in Bezug auf die Geheimhaltung der Ursprungsdaten. Die Teilnehmer haben zu keinem Zeitpunkt Zugang zu den der Analyse zu Grunde liegenden Daten, die Analyse wird auf mehrere Netzwerkteilnehmer (Nodes) – teils redundant – aufgeteilt und die Analyseergebnisse sind nur für den Auftraggeber sinnvoll.

[Rz 173] Dieses Konzept ist kognitiv herausfordernd und in einer Weise an Servicenutzer zu kommunizieren, die das Vertrauen in den Umgang der Verwaltung mit ihren Daten stärkt. Der Befürchtung, es könnten sich für Betroffene negative Konsequenzen aus Auswertungen ergeben, ist entgegenzutreten. Dennoch muss festgehalten werden, dass das Netzwerk als solches in der Konzeptionsphase ist, der Berechnungsaufwand für Analysen aktuell 100-fach höher ist als im Fall der Auswertung nicht verschlüsselter Daten ist und Erfahrungswerte gänzlich fehlen. Enigma und dessen mögliche großflächiger Einsatz wirft unzweifelhaft eine Reihe von Fragen auf, denen sich die Verwaltung aktuell, wenn auch in einem anderen Kontext, stellt. Mit Veröffentlichung ausführlicher und exakter Beschreibungen der in der Verwaltung vorhandenen Daten, nicht aber der Daten selbst, so sie personenbezogen sind, könnten Wettbewerbe veranstaltet werden, deren Ziel die Er-

(998) (http://www.computer.org/csdl/proceedings/hicss/2013/4892/00/4892a995.pdf).

94 Unter § 49 Abs. 1 DSG fällt auch der Fall in welchem automationsunterstützt verarbeitete Daten am Ende durch die Mithilfe eines Menschen, des Systemprüfers, ausgewertet wird, siehe UFS Wien vom 9. November 2005, RV/2350-W/02.

95 Walter Dohr / Hans J. Pollirer / Ernst M. Weiss / Rainer Knyrim, Kommentar zum Datenschutzrecht, Manz 2014, Anm. 3–4 zu § 49.

stellung von Auswertungen ist. Die neuartigste und für die Verwaltung sinnhafteste Auswertung könnte dabei prämiert werden. Derart erstellte Algorithmen könnten auf kollaborativen Plattformen ähnlich zu jenen, wie sie bei der gemeinsamen Erstellung von Quellcode einem Open Source Ansatz folgend Verwendung finden, veröffentlicht werden. Dritten stehen somit jene Algorithmen zur Verfügung, die in der Verwaltung selbst als Grundalge für statistische Auswertungen, Prognosen und Analysen herangezogen werden, können von diesen ausgewertet, explorativ adaptiert und verbessert werden. Die Offenlegung von Algorithmen, stellt einen weiteren Evolutionsschritt von Offenheit und Transparenz in der Verwaltung dar. Zudem hat die Offenlegung eine Rechtsschutzfunktion, wie aus § 49 Abs 3 DSG hervorgeht: Bei automatisierten Einzelentscheidungen ist dem Betroffenen nach der genannten Bestimmung auf Antrag der logische Ablauf der automatisierten Entscheidungsfindung in allgemein verständlicher Form darzulegen. Gemeint ist hier die Programmlogik[96], sohin der Algorithmus. Das Entwicklerteam von Enigma plant den Client frei zur Verfügung zu stellen, womit jeder mit Internetzugang Teilnehmer am Enigma Netzwerk werden kann. Aufwändige Berechnungen der Verwaltung können so sicher und ohne Bindung an einen Cloudanbieter weltweit verteilt werden.

[Rz 174] Neue Technologien beherbergen das Potential für grundlegende Veränderungen. Diese Errungenschaften sollen aber einer kritischen Betrachtung unterzogen werden, ohne sich der Möglichkeit zu berauben, diese in den Dienst der Gesellschaft zu stellen, um – wie im 2. Erwägungsgrund der Richtlinie 95/46/EG festgehalten – den wirtschaftlichen und sozialen Fortschritt zu fördern. «*Ziel von Ansätzen muss es sein, das Recht der informationellen Selbstbestimmung der Betroffenen weit möglichst zu schützen und sich gleichzeitig nicht systematisch gegen neue Systeme zu stellen und damit etwa technologisch ins Hintertreffen zu geraten.*»[97] Es ist deshalb unumgänglich, sich im Bewusstsein der Herausforderungen und möglicher Gefahren (wie z.B. mit Fragen der Re-Identifikation vor dem Hintergrund der technischen Entwicklung) intensiv mit der Möglichkeit der Auswertung heterogener Daten zu befassen. Die Untersuchung der Anforderungen, die an unterschiedliche Nutzungsmethoden zu stellen sind, ist essentiell, um die Voraussetzungen einer Heranziehung neuer Technologien zunächst für die Optimierung bestehender und zukünftig vielleicht für mögliche neue Aufgabengebiete der Verwaltung zu schaffen. Gute Praxis, die im Rahmen von Forschungsprojekten entsteht, wäre dazu geeignet, dieses notwendige Vertrauen zum selbstverständlichen Technologieeinsatz aufzubauen. Dieser Beitrag soll als Zusammenfassung relevanter Überlegungen dienen und als Konzept möglicher komplementärer Lösungswege verstanden werden.

Univ-Prof. Mag. Dr. Peter Parycek, MAS ist Leiter des Department für E-Governance in Wirtschaft und Verwaltung an der Donau-Universität Krems.
Dr. Johann Höchtl ist Wissenschaftlicher Mitarbeiter im Zentrum für E-Governance der Donau Universität Krems.
Mag. Bettina Rinnerbauer ist Wissenschaftliche Projektmitarbeiterin im Zentrum für E-Governance, Donau-Universität Krems.
Die Autoren danken Christof Tschohl für die fachliche Diskussion.

96 Walter Dohr / Hans J. Pollirer / Ernst M. Weiss / Rainer Knyrim, Kommentar zum Datenschutzrecht, Manz 2014, Anm. 12 zu § 49.

97 Max von Schönfeld, Daten – das neue Öl?!, in: Jusletter IT 21. Mai 2015, 12.