JÖRN ERBGUTH

Datenschutzkonforme Verwendung von Hashwerten auf Blockchains

Wann sind kryptografische Hashwerte von personenbezogenen Daten selbst wieder personenbezogene Daten?

Privacy Enhancing Technology

Kryptografische Hashfunktionen gehören zur Privacy Enhancing Technology. Die mit ihnen errechneten Hashwerte sind eine Art digitaler Fingerabdruck für beliebige Daten. Werden Hashwerte auf einer unveränderlichen Blockchain abgelegt, lassen sich außerhalb der Blockchain abgelegte Objekte dauerhaft und verfälschungssicher validieren. Die technischen Eigenschaften von Hashfunktionen sorgen zudem dafür, dass die Informationsverarbeitung nur in eine Richtung – vom Objekt zum Hashwert, aber nicht umgekehrt – möglich ist. Richtig eingesetzt, können sie dazu beitragen, die Privatsphäre sicher zu schützen. Zugleich sind Anwender mit der meist recht pauschalen Aussage konfrontiert, dass der Hashwert eines personenbezogenen Datums wiederum ein personenbezogenes Datum sei. Dadurch entsteht ein hohes Maß an Rechtsunsicherheit. Der Aufsatz erörtert, ob Hashwerte personenbezogene Daten i.S.d. DS-GVO darstellen und differenziert dabei nach unterschiedlichen Verwendungsarten.

Lesedauer: 32 Minuten

I. Einleitung

Blockchain ist eine Technologie, die durch Offenheit, Transparenz und Unveränderlichkeit besticht. Jedoch stehen diesen Vorteilen auf den ersten Blick einige zentrale Prinzipien des Datenschutzes entgegen: Datenminimierung, Zugriffskontrolle und das Recht der betroffenen Person darauf, dass Daten, die nicht mehr benötigt werden, zu löschen sind. Dennoch kann die Blockchain-Technologie dazu beitragen, einen in Teilen sogar besseren Schutz der Privatsphäre zu gewährleisten als konventionelle Systeme.

Der Grund hierfür liegt darin, dass Privacy Enhancing Technology auf Blockchains dazu beiträgt, den Missbrauch von personenbezogenen Informationen technisch massiv zu erschweren oder effektiv auszuschließen. In Kombination mit verteilten Systemarchitekturen kann Blockchain-Technologie eine technische Zweckbindung von Daten ermöglichen. Diese verhindert Datenmissbrauch auch durch Stellen, denen in konventionellen Systemen nur wenige Hürden entgegengesetzt werden: Verantwortliche, Administratoren oder staatliche Dienste, aber auch Hacker, die einmal die Firewall überwunden haben, haben häufig freien Zugriff auf große Datenmengen. Doch lassen sich unveränderbare Blockchains tatsächlich mit Privacy Enhancing Technology DS-GVO-konform nutzen? Eine zentrale Frage ist dabei, wann Hashwerte[1] personenbezogene Daten darstellen und daher der DS-GVO unterliegen. Diese Frage hat die juristische Literatur bislang noch nicht intensiv analysiert. Der Beitrag fokussiert sich daher insbesondere auf die technischen Grundlagen und ihr Zusammenspiel mit der datenschutzrechtlichen Dogmatik.

II. Datenschutz durch Technikgestaltung

Als Ausdruck der Leitidee „Privacy by Design" verpflichtet Art. 25 Abs. 1 DS-GVO zu geeigneten technischen und organisatorischen Maßnahmen. Erwägungsgrund 78 verweist dazu auf die Grundsätze des Datenschutzes durch Technik. Ein technisches Mittel, um diesen Pflichten nachzukommen, ist die Verwendung von Privacy Enhancing Technology.[2] Damit kann der Datenschutz bereits ins technische Design des Systems eingebaut werden.[3] In der Folge fallen personenbezogene Daten teilweise erst gar nicht an oder lässt sich ihre Nutzung auf bestimmte Zwecke technisch beschränken. Auch Innentätern, wie etwa Administratoren, ist es dann unmöglich oder zumindest deutlich erschwert, bestimmte personenbezogene Daten zweckwidrig zu verarbeiten. Verletzt eine datenverarbeitende Stelle die Grundsätze des Art. 25 Abs. 1 DS-GVO, etwa indem sie Passwörter im Klartext speichert, können die Aufsichtsbehörden dies ahnden.[4] Die wohl prominentesten Vertreter der Privacy Enhancing Technology sind Verschlüsselung[5], Zero Knowledge Proofs (ZKPs)[6] und kryptografische Hashfunktionen. Der Beitrag beschränkt sich im Folgenden auf kryptografische Hashfunktionen, da diese im Kontext von Blockchains die größte Relevanz haben.

III. Kryptografische Hashfunktionen

Kaum ein Bereich des Einsatzes von Kryptografie kommt ohne kryptografische Hashfunktionen[7] aus. Sie sind nicht nur ein wesentlicher Bestandteil von Blockchains[8]. Auch die qualifizierte

1 Hashwerte werden im deutschen auch als Streuwerte bezeichnet. Diese Bezeichnung ist aber im Kontext der Blockchain wenig gebräuchlich, weshalb im vorliegenden Aufsatz durchgehend die ans Englische angelehnte Bezeichnung verwendet wird.

2 *Hansen* in: Simitis/Hornung/Spiecker gen. Döhmann, Datenschutzrecht, 1. Aufl. 2019, DS-GVO Art. 25 Rdnr. 16.

3 *Baumgartner/Gausling*, ZD 2017, 308, 309.

4 *LfDI Baden-Württemberg*, PM v. 22.11.2018, abrufbar unter: https://www.baden-wuerttemberg.datenschutz.de/lfdi-baden-wuerttemberg-verhaengt-sein-erstes-bussgeld-in-deutschland-nach-der-ds-gvo/.

5 Wenn Daten verschlüsselt werden, wird aus ihnen ein Ciphertext generiert, aus dem nur mit Hilfe des richtigen Schlüssels die ursprünglichen Daten wiederhergestellt werden können.

6 Bei Zero Knowledge Proofs (ZKPs) handelt es sich um ein technisches Verfahren, um eine bestimmte Tatsache zu beweisen, ohne dabei mehr als das zu Beweisende offen legen zu müssen. So verwendet die Kryptowährung Zcash ZKPs, um die Gültigkeit einzelner Transaktionen zu beweisen, ohne dass daraus der Inhalt oder die Parteien einer Transaktion ersichtlich sind. Anders, als etwa bei Bitcoin, können Außenstehende dann nicht transparent nachvollziehen, unter welchem Pseudonym eine Transaktion in welcher Höhe vorgenommen wurde, sondern sehen nur, dass eine gültige Transaktion stattgefunden hat. Ein anderer Anwendungsfall ist der Nachweis der Volljährigkeit, ohne dafür das konkrete Alter bzw. das Geburtsdatum offenlegen zu müssen. Zur technischen Funktionsweise s. *Delfs/Knebl*, Introduction to Cryptography, 3. Aufl. 2015, S. 121 f. Zur Verwendung s. *van Rijmenam*, How Zero Knowledge Proof Will Enable Trustless Transactions and Increase our Privacy, LinkedIn, abrufbar unter: www.linkedin.com/pulse/how-trustless-society-improve-our-privacy-mark-van-rijmenam/.

7 Im Folgenden wird nur noch von Hashfunktionen und Hashwerten gesprochen – gemeint sind jedoch stets kryptografische Hashfunktionen und kryptografische Hashwerte. In der Lit. findet sich auch die Bezeichnung „kryptologisch" statt „kryptografisch". Für eine Beschreibung von kryptografischen Hashfunktionen s. etwa *Delfs/Knebl* (o. Fußn. 6), S. 30 ff.

8 Für die Verkettung von Transaktionen und für die Verkettung von Blöcken verwenden Blockchains Hashfunktionen s. *Satoshi Nakamoto* (Pseudonym), Bitcoin: A Peer-to-Peer Electronic Cash System v. 1.11.2008, https://bitcoin.org/bitcoin.pdf.

elektronische Signatur, der nach Art. 25 Abs. 2 eIDAS-Verordnung die Rechtswirkung einer handschriftlichen Unterschrift zukommt, verwendet Hashfunktionen[9].

1. Grundprinzip

Hashfunktionen sind mathematische Funktionen, die sich nur in eine Richtung berechnen lassen, in die andere Richtung aber äußerst schwer oder gar nicht berechenbar sind. Ein Beispiel für solche Einwegfunktionen ist die Primfaktorzerlegung natürlicher Zahlen: Es ist sehr leicht, große Primzahlen (z.B. 99859 und 88007) miteinander zu multiplizieren. Es ist aber sehr schwer, aus dem Ergebnis (hier 8788291013) wieder die ursprünglichen Primzahlen zu errechnen.[10]

Im Gegensatz zur Verschlüsselung eines Datums ist das Ziel einer Hashfunktion nicht, das ursprüngliche digitale Objekt später rekonstruieren zu können, sondern es nur zu identifizieren. Deshalb reduzieren Hashfunktion die Datenmenge massiv. Das geschieht dadurch, dass Hashfunktionen recht kurze Zeichenfolgen – sog. Hashwerte – für beliebig große digitale Objekte errechnen. Typischerweise sind Hashwerte 256 Bit lang, was in der üblichen Base64-Schreibweise einer Zeichenfolge von 43 alphanumerischen Zeichen entspricht.

Im Gegensatz zu einfachen Hashfunktionen weisen „kryptografische Hashfunktionen" die Eigenschaft auf, dass sie „kollisionsresistent" sind. Das bedeutet, dass sich praktisch keine zwei Objekte finden lassen, die den gleichen Hashwert ergeben. Schon minimal unterschiedliche Objekte, die sich etwa nur durch ein Leerzeichen oder ein Pixel unterscheiden, haben komplett verschiedene Hashwerte. Hashwerte sind dadurch letztlich wie Fingerabdrücke für digitale Objekte: Mit einem Hashwert kann ein digitales Objekt identifizieret werden. Auf Grund der Eigenschaft als Einwegfunktion sagt aber umgekehrt der Hashwert selbst nichts über das digitale Objekt aus. Hashwerte werden daher gerne auch als Prüfsumme für digitale Objekte verwendet, da es praktisch ausgeschlossen ist, ein anderes digitales Objekt zu finden oder herzustellen, für welches die Hashfunktion den gleichen Hashwert ergibt.

2. Technische Sicherheitsrisiken

Damit Hashfunktionen tatsächlich kollisionsresistent und nicht umkehrbar sind, müssen aus technischer Sicht einige Risiken beachtet und durch geeignete Gegenmaßnahmen abgewendet werden:

a) Verfügbare Computerleistung

Dass die verfügbare Computerleistung stetig ansteigt, stellt ein potenzielles Risiko dar. Nach dem Mooreschen Gesetz[11] verdoppelt sich die Rechnerkapazität gängiger Computer etwa alle zwei Jahre. Damit sind in den nächsten 40 Jahren Leistungszuwächse um den Faktor eine Million zu erwarten. Dazu kommen spezielle Entwicklungen wie z.B. Bitcoin-Miner. Bei ihnen handelt es sich um Computer, die auf Hashing spezialisiert sind und dabei deutlich schneller als gewöhnliche Rechner sind. Je schneller Hashfunktionen ausführbar sind, desto höher ist die Gefahr, dass es möglich ist, das Ursprungobjekt zu „erraten". Sollen etwa 8-stellige Passwörter mit der gängigen Hashfunktion SHA-256 geschützt werden, so lässt sich ein solches Passwort mit einem etwa € 3.000,– teuren Bitcoin-Miner in weniger als vier Stunden durch Ausprobieren aus dem Hashwert ermitteln.[12] Ist das Ursprungsobjekt jedoch ausreichend variabel, d.h. mit genügend Entropie ausgestattet, lässt sich das Ausprobieren wirksam verhindern. So steigt die zum Erraten nötige Zeit bereits bei einem 12-stelligen Passwort auf etwa 100.000 Jahre.

b) Entropie

Während gescannte Daten eine hohe Entropie aufweisen, ist sie bei kurzen Textdaten teilweise sehr gering. Eine geringe Entropie bedeutet eine geringe Variabilität. Dadurch dauert ein Erraten durch Ausprobieren nicht lange. Um die ggf. unzureichende Entropie der zu hashenden Daten aufzustocken, können die Daten vor dem Hashen um einen zufälligen weiteren Wert ergänzt werden. Solche Hashes firmieren unter dem Fachbegriff „gesalzene Hashes". Falls der zusätzliche Wert geheim gehalten wird, dient er als Passwort. Man spricht dann auch von „gepfefferten Hashes".[13] Ohne dieses Passwort ist der Hashwert unbrauchbar und kann nicht mehr mit dem gehashten Objekt in Verbindung gebracht werden. Entropie in Form von Salz oder Pfeffer hat jedoch den Nachteil, dass diese zusätzliche Entropie vom eigentlichen Wert getrennt werden kann. D.h., dass die Kenntnis vom Salz- oder Pfefferwert das Erraten der eigentlichen Information ermöglichen kann. Dies ist z.B. bei gescannten Daten nicht der Fall, da diese von sich aus eine hohe Entropie beinhalten.

c) Hashfunktion wird unsicher

Eine kryptografische Hashfunktion kann sich als unsicher herausstellen, wenn alternative Berechnungsweisen entdeckt werden. Bei den Hashfunktionen MD-5 und SHA-1 ist dies bereits der Fall. Zwar lassen sich aus den Hashwerten (wegen der massiven Datenreduktion) auch bei einer unsicheren Hashfunktion keine umfangreichen Ursprungsobjekte rekonstruieren. Bei unsicheren Hashfunktionen ist es jedoch möglich, weitere Objekte zu finden, die den gleichen Hashwert ergeben. Dadurch leidet die Beweisfunktion z.B. von auf diesen Hashwerten basierenden elektronischen Signaturen.

d) Quantencomputer

Ein weiteres technisches Risiko für kollisionsfreie und unumkehrbare Hashfunktionen sind Quantencomputer. Sie rechnen zwar nicht unbedingt schneller als herkömmliche Rechner, aber anders; dadurch können sie bestimmte Aufgaben schnell berechnen, die bislang nur durch langwieriges Ausprobieren lösbar waren. Aktuell existieren – zumindest offiziell – noch keine Quantenrechner ausreichender Größe, um bestehende kryptografische Verfahren zu gefährden. Die Schätzungen darüber, wann diese verfügbar sein werden, schwanken jedoch stark.[14] Allerdings sind Quantenrechner bereits so gut erforscht, dass es möglich ist abzuschätzen, welche kryptografischen Verfahren durch sie unsicher werden und welche nicht.[15] So viel steht derzeit wohl fest: Die geläufigsten Hashfunktionen SHA-2 und SHA-3 sind erst bei sehr großen Quantencomputern gefährdet. Doch selbst dann leidet wiederum nur die Beweisfunktion. Denn auch mit Quantencomputern wird es nicht möglich sein, aus einem Hashwert von 256 Bit große Datenobjekte, die ausreichende Entropie haben, zu berechnen.

---

9 BeckOK VwVfG/*Rost*, 43. Ed., Stand: 1.4.2019, VwZG § 5 Rdnr. 75.

10 Die tatsächlich verwendeten Primzahlen sind deutlich größer.

11 Es ist kein Naturgesetz, sondern eine Beobachtung, die schon seit über 50 Jahren ziemlich genau zutrifft; vgl. *Wikipedia*, Mooresches Gesetz, abrufbar unter: https://de.wikipedia.org/wiki/Mooresches_Gesetz.

12 Es gibt etwa $7{,}2*10^{16}$ verschiedene 8-stellige Passwörter. Ein aktueller Bitcoin-Miner, etwa der Antminer S17 Pro-53TH/s, kann $5{,}3*10^{12}$ Kombinationen pro Sekunde durchprobieren und braucht daher weniger als 4 Stunden, um alle 8-stelligen Passwörter durchzuprobieren.

13 *Spacey*, Cryptography: Salt vs Pepper, Simplicable v. 26.11.2016, abrufbar unter: https://simplicable.com/new/salt-vs-pepper.

14 *IBM*, Quantencomputer: Der Beginn der kommerziellen Quanten-Ära, abrufbar unter: https://www.ibm.com/de-de/blogs/think/2018/02/23/quantencomputer/; TR Online, Code-knackende Quantencomputer kommen näher, heise online v. 7.6.2019, https://www.heise.de/tr/artikel/Code-knackende-Quantencomputer-kommen-naeher-4441160.html.

15 *Chen et al.*, Report on Post-Quantum Cryptography, NISTIR 8105, S. 2 https://nvlpubs.nist.gov/nistpubs/ir/2016/nist.ir.8105.pdf.

IV. Personenbezogene Daten

Da mit Hashfunktionen ein praktisch eindeutiger digitaler Fingerabdruck eines digitalen Objekts angefertigt werden kann, aus dem aber kein direkter Rückschluss auf das Ursprungsdatum möglich ist, stellt sich im datenschutzrechtlichen Kontext eine wichtige Frage: Ist der Hashwert eines personenbezogenen Datums selbst wieder als personenbezogenes Datum anzusehen? Die Frage ist schon deshalb von zentraler Bedeutung, da die DS-GVO nur bei personenbezogenen Daten Anwendung findet (Art. 2 Abs. 1 DS-GVO). Der Begriff „personenbezogene Daten" ist in Art. 4 Nr. 1 DS-GVO legaldefiniert: Er umfasst Informationen, die sich auf eine identifizierte oder identifizierbare Person beziehen.

Daher wird geprüft, ob sich ein Hashwert auf eine identifizierbare Person bezieht und ob aus dem Hashwert eine Information über eine betroffene Person ermittelbar ist:

1. Personenbezug

a) Relativer und absoluter Personenbezug

Bei der Identifizierbarkeit stellt sich die Frage, auf wen dabei abzustellen ist: Ist nur auf den Verantwortlichen abzustellen oder reicht eine Identifizierbarkeit durch beliebige Personen? Art. 4 Nr. 1 DS-GVO schränkt den Personenkreis nicht ein, auf den abzustellen ist. Auch Erwägungsgrund 26 spricht hier von Mitteln, die „der Verantwortliche" oder eine „andere Person" nutzen könnte. Darüber hinaus muss jedoch die handelnde Person die verfügbaren Mittel auch zur Identifizierung nutzen können. Dazu muss sie gemäß einer modifizierten relativen Theorie die eigentlichen Daten und die Identifizierungsmittel zusammenführen können.[16] Wenn also die Personen, die Zugriff auf die Daten haben können, auch indirekt keinen Zugriff auf die Identifizierungsmittel erhalten können, so liegt keine Personenbeziehbarkeit vor. Bei der Ablage von Hashwerten auf öffentlichen Blockchains sind die Hashwerte für jeden zugreifbar. Daher liegt eine Personenbeziehbarkeit bereits vor, wenn es Personen gibt, die Zugriff auf Identifizierungsmittel erhalten könnten.

b) Aufwand der Personalisierung

Um beurteilen zu können, ob sich ein Datum auf eine identifizierbare Person bezieht, sind laut Erwägungsgrund 26 alle Mittel zu berücksichtigen, die der Verantwortliche oder eine andere Person nach allgemeinem Ermessen wahrscheinlich nutzen. Der sehr abstrakt gehaltene Wortlaut gibt vor, dass der Aufwand der Identifizierung eine Rolle spielt, gibt aber keine klare Antwort auf die Frage, wie aufwändig eine Identifizierungsmöglichkeit konkret sein muss, damit sie nicht mehr unter diese Formulierung fällt. Aus dem risikobasierten Ansatz des europäischen Datenschutzrechts folgern *Specht/Müller-Riemenscheider,* dass eine rein theoretische Identifizierungsmöglichkeit jedenfalls nicht zu regulatorischen Konsequenzen führen dürfe.[17] So sieht der *EuGH* in der Rs. Breyer dann keine personenbezogenen Daten, wenn die Identifizierung nur mit „unverhältnismäßigem Aufwand" möglich ist.[18]

c) Zukünftige Entwicklungen

Erwägungsgrund 26 gibt vor, dass im Rahmen des Aufwands der Identifizierbarkeit nicht nur die Technologie, die zum Zeitpunkt der Verarbeitung verfügbar ist, sondern auch zukünftige Entwicklungen zu berücksichtigen sind. Dies wird als Pflicht interpretiert, die technologische Entwicklung zum Zeitpunkt der Verarbeitung zu berücksichtigen.[19] Insbesondere dürfe eine Identifizierung auch im Laufe einer langen Speicherdauer nicht zu erwarten sein.[20] Anstieg der Rechenleistung sowie die Verfügbarkeit von Quantencomputer sind deshalb proaktiv zu berücksichtigen.[21] Zusätzlich besteht die (ggf. schwer zu erfüllende) Pflicht, ein System bei unerwarteten Entwicklungen anzupassen. Sind die Informationen weitergegeben oder veröffentlicht worden, könnte man zudem eine Informationspflicht analog zu Art. 17 Abs. 2 DS-GVO erwägen.

Bei Hashwerten ist eine Offenlegung der Ursprungsdaten (jedenfalls soweit sie ausreichend umfangreich und variabel sind) bereits auf Grund der Reduktion der Datenmenge praktisch ausgeschlossen.[22] Im Vergleich zu unsicher gewordenen Verschlüsselungsverfahren, ist das direkte Risiko der Offenlegung personenbezogener Daten bei unsicher gewordenen Hashfunktionen vernachlässigbar.

2. Information

Für die Frage, ob personenbezogene Daten vorliegen, ist es nicht nur erforderlich, dass ein Personenbezug hergestellt werden kann, sondern auch dass es sich um eine „Information" über eine Person handelt. Die Schwelle für das Vorliegen einer Information ist niedrig. Im Volkszählungsurteil urteilte das *BVerfG*, es gäbe im Zeitalter der automatischen Datenverarbeitung kein belangloses Datum mehr.[23] Das war zukunftsweisend bis visionär, da die Möglichkeiten des Deep Learning und der Predicitve Analysis dem *BVerfG* wohl noch nicht bekannt waren.[24] Als Information gilt auch das, was nicht explizit in den Daten steht, sondern sich irgendwie, z.B. aus einem Kontext oder mittels statistischer Verfahren, ableiten lässt. Das ist selbst dann der Fall, wenn die Ableitung mit einer Unsicherheit verbunden bleibt.[25]

Dennoch gibt es Situationen, in denen es zweifelhaft ist, ob tatsächlich eine Information über eine betroffene Person vorliegt. Listet ein Buch etwa alle möglichen Geburtsdaten aller lebenden Menschen oder alle möglichen IP-Adressen (V4) auf, so lässt sich zwar ein Personenbezug zu den Nutzern dieser IP-Adressen oder den Menschen mit diesen Geburtsdaten herstellen. Allerdings ist dort weder direkt noch indirekt eine Information über eine spezifische Person abgelegt. Wenn man die Information (z.B. Geburtsdatum) bereits benötigt, um in einem Buch mit allen Geburtsdaten genau diese wiederzufinden und dort keine weitere Information zur über dieses Geburtsdatum identifizierten Person findet, so erhält man lediglich ein Echo der bereits vorhandenen Information. Damit liegt hier trotz Abdruck eines eine Person identifizierenden Datums (Geburtsdatum) keine Information über diese betroffene Person vor. I.E. führt daher ein Personenbezug nicht automatisch zur Klassifizierung als personenbezogenes Datum.

---

16 *Klar/Kühling* in: Kühling/Buchner, 2. Aufl. 2018, DS-GVO Art. 4 Nr. 1 Rdnr. 26 f.; *Simitis/Hornung/Spiecker gen. Döhmann* (o. Fußn. 2), Art. 4 Nr. 1 Rdnr. 60 ff.

17 *Specht/Müller-Riemenscheider*, ZD 2014, 71, 73 f.; zum selben Ergebnis kommt später der *EuGH* MMR 2016, 842 m. Anm. *Moos/Rothkegel*, Rdnr. 46; so auch *Moritz/Karg*, in Simitis/Hornung/Spiecker gen. Döhmann (o. Fußn. 2), Art. 4 Nr. 1 Rdnr. 6; BeckOK DatenschutzR/*Schild*, 28. Ed., Stand: 1.2.2019, Art. 4 Rdnr. 18; a.A. *Voitel*, DuD 2017, 686.

18 *EuGH* MMR 2016, 842 m. Anm. *Moos/Rothkegel*, Rdnr. 46; so auch *Österreichische Datenschutzbehörde*, E. v. 5.12.2018, DSB-D123.270/009-DSB/2018; die Datenschutzbehörde setzt zudem anonymisierte Daten einer Löschung gleich.

19 *Krügel*, ZD 2017, 455, 456.

20 *Art. 29-Datenschutzgruppe*, Stellungnahme 4/2007 zum Begriff „personenbezogene Daten", 20.6.2007, 01248/07/DE, WP136, S. 18.

21 Vgl. dazu die Ausführungen unter III.2.

22 Vgl. dazu die Ausführungen unter III.2.

23 *BVerfG* NJW 1984, 419, 422.

24 Diese Verfahren erlauben aus eigentlich belanglosen Daten hochkritische Schlüsse zu ziehen. *Schulte*, Predicitve Analytics: Moderne Schauermärchen v. 26.2.2019, abrufbar unter: https://www.informatik-aktuell.de/betrieb/kuenstliche-intelligenz/predictive-analytics-potentiale-und-grenzen.html.

25 *Klar/Kühling* (o. Fußn. 16), Rdnr. 10.

1. Name	2. Adresse	3. Geburtsdatum	4. Zeitraum des Bezugs bestimmter Sozialleistungen	5. Body-Mass-Index	6. Referenznr. der Untersuchungskohorte
Max Müller	Berlin	05.11.63	< 2 Jahre	15	QA5FRD4
Eva Schön	München	01.07.82	> 5 Jahre	14	2B48HFG
Ulrich Klein	Hamburg	13.12.77	< 2 Jahre	16	RC3URPQ
Herbert Schmidt	Köln	15.05.85	> 5 Jahre	18	SD289K9
Ulrike Groß	Stuttgart	12.12.70	< 2 Jahre	20	%E1FL7Q

Abb. 1: Beispiel der Art. 29-Datenschutzgruppe für die Beurteilung von Hashfunktionen. Der Hashwert wird aus den ersten drei Feldern gebildet, die anschließend gelöscht werden.

3. Art. 29-Datenschutzgruppe: WP 216

Die *Art. 29-Datenschutzgruppe* hat 2014 zu Anonymisierungstechniken Stellung bezogen.[26] Dabei hat sie ihren Blick auch auf Hashfunktionen gerichtet – allerdings lediglich als Mittel der „Pseudonymisierung". D.h. es wurden nur bestimmte Teile eines Datensatzes (z.B. Name, Adresse, Geburtsdatum) durch einen Hashwert (hier als Referenznummer bezeichnet) ersetzt (Abb. 1). Die übrigen Bestandteile der Datensätze wurden unverändert beibehalten. Dadurch ergeben sich drei Probleme:

a) Herausgreifen

Auch wenn sich aus den pseudonymisierten Daten selbst keine Rückschlüsse auf die Identität der Personen ergeben, kann man mit Namen, Adresse und Geburtsdatum den Hashwert (Referenznummer) berechnen und damit den in der Tabelle abgelegten Daten zuordnen. Der Hashwert sperrt hier die Identifikation nur in die eine Richtung – die Identifikation der Person ausgehend vom Datensatz. In die andere Richtung – dem Auffinden eines Datensatzes zu einer bestimmten Person – ist er durchlässig.

b) Verknüpfbarkeit

Wird zweimal die gleiche Information gehasht, so ergibt eine Hashfunktion immer den identischen Hashwert. Dabei handelt es sich um eine zentrale Eigenschaft von Hashfunktionen. Kommt ein solcher Hashwert dann an zwei Stellen zum Einsatz, ist es möglich, Verknüpfungen zwischen diesen beiden Einträgen herzustellen. *Facebook* nutzte diese Eigenschaft etwa, um E-Mail-Adressen für gezieltes Marketing (Facebook Custom Audience) abzugleichen.[27] Wenn Hashwerte so eingesetzt werden, dass dadurch ein Abgleich von Personen über deren gehashte E-Mail-Adressen erfolgt, so werden natürlich personenbezogene Daten ausgetauscht.[28]

c) Inferenz

Wenn zusammen mit einem Hashwert weitere Daten abgelegt werden, kann sich der Personenbezug auch aus letzteren ergeben. Im Fall des von der *Art. 29-Datenschutzgruppe* betrachteten Pseudonymisierungsansatzes werden bei jedem Eintrag einige Daten durch Hashwerte ersetzt und die übrigen Daten unverändert beibehalten. Die unverändert beibehaltenen Daten können es jedoch ermöglichen, durch einen Abgleich mit externen Daten, den Personenbezug herzustellen. Im Web-of-Trust-Fall wurden etwa pseudonymisierte Browserverläufe weitergegeben.[29] Bereits über die spezifische Kombination der aufgerufenen Web-Adressen mit auf Twitter geposteten Links lassen sich Personen eindeutig identifizieren. Vergleichbar pseudonymisierte Daten bleiben selbst dann personenbezogen, wenn man als Referenznummer keinen berechneten Hashwert, sondern einen zufälligen Wert wählt. Das Risiko der Identifizierung von Betroffenen durch Inferenz beruht daher in den meisten Fällen auf dem gewählten Pseudonymisierungsansatz und ist unabhängig von der Verwendung von Hashfunktionen.

4. Reaktion der Literatur

Soweit sich die wissenschaftliche Literatur mit Hashfunktionen beschäftigt, stützt sie sich lediglich auf das WP 216 der *Art. 29-Datenschutzgruppe*, ohne näher zu diskutieren, welche technischen Fallkonstellationen dort überhaupt in Frage standen.[30]

5. Blockchain Observatory Report on Blockchain and GDPR und neuere Literatur

Im Juni 2018 hat das *Blockchain Observatory* der *EU-Kommission* einen Report zum Thema Blockchain und DS-GVO veröffentlicht.[31] Dieser beschäftigt sich auch mit dem Thema Hashwerte. Neben technischen Risiken, die bei unsachgemäßer Verwendung auftreten können, unterscheidet das Papier auch nach Anwendungsfällen. Dienen die Hashwerte etwa dazu, komplexe Datensätze zu beglaubigen, so werden die Risiken als gering eingestuft. Neuere Literatur hat dies aufgegriffen und schlägt nun eine Einzelfallbetrachtung vor.[32]

6. Stellungnahme der CNIL

Die französische Aufsichtsbehörde (*CNIL*) hat als erste europäische Aufsichtsbehörde umfangreich zur Blockchain-Technolo-

26 *Art. 29-Datenschutzgruppe*, Stellungnahme 5/2014 zu Anonymisierungstechniken v. 10.4.2014, 0829/14/DE, WP216.

27 Bei Facebook Custom Audience wurden auf Seiten des Werbetreibenden und auf Seiten von Facebook die E-Mail-Adressen der Betroffenen gehashed. Über identische Hashwerte konnte identifiziert werden, wenn eine Person in der Kundendatei des Werbetreibenden auch Nutzer bei Facebook ist. Die Werbung wurde dann gezielt nur für diesen Nutzerkreis bei Facebook geschaltet. Damit wurde ein direkter Austausch der E-Mail-Adressen vermieden, aber trotzdem die personenbezogene Information übermittelt, welche Facebook-Nutzer auch in der Kundendatei der werbetreibenden Firma sind. Darüber hinaus sind E-Mail-Adressen zu kurz bzw. haben zu wenig Entropie, sodass aus Hashwerten von E-Mail-Adressen, die E-Mail-Adressen auch durch Ausprobieren rekonstruierbar sind; vgl. dazu auch *Bayerisches Landesamt für Datenschutzaufsicht*, PM v. 4.10.2017, abrufbar unter: https://www.lda.bayern.de/media/pm2017_07.pdf.

28 *BayVGH* ZD 2019, 43.

29 Ein Browser-Plugin zeichnete den Browserverlauf heimlich auf. Die aufgezeichneten Daten wurden ohne direkte Identifikation der Betroffenen weitergegeben und an Interessenten verkauft. *Andreas Dewes* und *Svea Eckert* kauften diese Daten verdeckt auf und demonstrierten auf dem 33C3-Kongress, wie etwa über Twitter öffentlich geteilte Web-Adressen ausreichen, um viele der Betroffenen zu identifizieren; s. Profilingvortrag zu #nacktimNetz beim 33C3, Netzpolitik.org, abrufbar unter: https://netzpolitik.org/2017/profilingvortrag-zu-nacktimnetz-beim-33c3/.

30 So etwa *Finck*, EDPL 2018, 17, 22 f.; *Moser*, The Application & Impact of the European General Data Protection Regulation on Blockchains, R3 Reports v. 15.2.2017, abrufbar unter: https://www.r3.com/wp-content/uploads/2018/04/GDPR_Blockchains_R3.pdf; *Ibanñez/O'Hara/Simperl*, On Blockchains and the General Data Protection Regulation, https://eprints.soton.ac.uk/422879/1/BLockchains_GDPR_4.pdf; *Jaccard/Tharin*, GDPR & Blockchain: the Swiss take, Jusletter IT v. 4.12.2018.

31 *EU-Blockchain Observatory*: Blockchain and the GDPR v. 16.10.2018, abrufbar unter: https://www.eublockchainforum.eu/sites/default/files/reports/20181016_report_gdpr.pdf.

32 *Finck*, Blockchain and the General Data Protection Regulation, Panel for the Future of Science and Technology (STOA), European Parliament, PE 634.445, abrufbar unter: https://www.europarl.europa.eu/stoa/en/document/EPRS_STU(2019)634445; *Möri*, Blockchain und Datenschutz, Jusletter IT, abrufbar unter: http://jusletter-it.weblaw.ch/issues/2019/23-Mai-2019/blockchain-und-daten_cbe632b9ad.html.

gie Stellung bezogen. Sie hat eine Rangfolge der Verfahren aufgestellt:[33]

- 1. Fall: Ein Commitment-Verfahren. Hierunter fallen auch Zero Knowledge Proofs
- 2. Fall: Ein Hashwert, der mit einem geheimen Passwort abgesichert ist (vgl. III.2.b) – gepfefferte Hashes
- 3. Fall: Verschlüsselung
- 4. Fall: Ein Hashwert ohne Passwort

Im Fall 2 und 3 werden Daten über ein Passwort zusätzlich abgesichert. Allerdings erscheint der Mehrwert des Passwortschutzes von Daten auf Blockchains fraglich. Wenn Passwort-geschützte Daten auf einer Blockchain liegen, sind die Passwörter nicht mehr modifizierbar. Gängige Sicherheitsstandards sehen dagegen vor, dass Passwörter änderbar sein müssen.[34] Zudem wird bei mit Passwort geschützten Hashwerten (gepfefferte Hashwerte) das Passwort stets für die Validierung benötigt. Daher muss es den gleichen Personen bekannt gegeben werden, wie das zu validierende digitale Objekt. Dann macht es bzgl. der Sicherheit nur noch einen geringen Unterschied, ob ein ausreichend vielfältiges Objekt gleichzeitig als eine Art Passwort wirkt oder ob ein zusätzliches Passwort gewählt wurde. Ein zusätzliches Passwort macht das Verfahren daher in vielen Fällen kaum sicherer, sondern nur komplizierter. Das beeinträchtigt die Nutzerfreundlichkeit.

Verschlüsselte personenbezogene Daten (Fall 3) bergen im Gegensatz zu Hashwerten die Gefahr, dass ein kompromittiertes Passwort dazu führt, dass die Ursprungsdaten für Dritte verfügbar werden. Die Prioritätenreihenfolge der *CNIL* ist daher nicht ganz nachvollziehbar. Verschlüsselte personenbezogene Daten sollten vielmehr auf Blockchains generell vermieden werden.

Die *französische Aufsichtsbehörde* geht lediglich beim Commitment-Verfahren (Fall 1) und beim gepfefferten Hashing (Fall 2) nach dem Löschen der externen Daten davon aus, dass die auf der Blockchain verbleibenden Daten vollständig anonymisiert sein können.[35] Dabei lässt die *CNIL* offen, ob sie damit die personenbezogenen Daten als gelöscht ansieht, wie es die *österreichische Aufsichtsbehörde*[36] in einem anderen Fall macht und der Wortlaut des Erwägungsgrunds 26 sowie die Ausführungen des *EuGH* in der Rs. Breyer[37] vermuten lassen würden.

Gemäß ihrer nicht ganz nachvollziehbaren Einordnung, sieht die *CNIL* bei einfachen Hashwerten lediglich eine Risikoverringerung und fordert eine Datenschutz-Folgenabschätzung, um sicher zu gehen, dass das Risiko hinnehmbar ist.[38]

7. Ergebnis

Ob Hashwerte von personenbezogenen Daten wiederum als personenbezogene Daten anzusehen sind, hängt davon ab, ob im konkreten Anwendungsfall aus dem Hashwert personenbezogene Informationen abgeleitet werden können. Zumindest dort, wo dies praktisch unmöglich ist, ist ein Hashwert nicht als personenbezogenes Datum anzusehen. Folgt man dem *EuGH*, liegt zudem auch dann kein personenbezogenes Datum vor, wenn die Identifikation nur mit einem unverhältnismäßigen hohen Aufwand möglich ist.

V. Differenzierung nach Anwendungsfällen

Um konkret bestimmen zu können, ob es sich bei Hashwerten um personenbezogene Daten handelt, geht kein Weg daran vorbei, verschiedene Anwendungsfälle zu unterscheiden. Den Ausgangsfall bildet ein Datenobjekt, welches ein personenbezogenes Datum darstellt und dessen Hashwert der Verantwortliche auf einer Blockchain abgelegt hat. Häufig dient dies zum Beweis, dass das gehashte Objekt nicht nachträglich manipuliert wurde:

1. Isolierter Hashwert, dessen Ursprungsobjekt sicher gelöscht wurde

Wenn der Nachweis via Blockchain nicht mehr benötigt wird und das gehashte Objekt selbst sicher gelöscht wurde, ist es nicht mehr möglich, das ursprüngliche Objekt mit Hilfe des Hashwerts zu rekonstruieren. Hat ein fachgerechtes Hashing stattgefunden, ist es auch ausgeschlossen, das dahinterstehende digitale Objekt zu erraten. I.E. besteht dann kein Risiko des Herausgreifens, der Verknüpfbarkeit oder der Inferenz. Die auf der Blockchain verbleibenden Hashwerte sind damit nicht mehr personenbezogen.

2. Isolierter Hashwert, dessen Ursprungsobjekt existiert

Ist das gehashte Objekt nicht gelöscht, sondern weiterhin verfügbar, kann aus ihm auch der Hashwert berechnet werden. Wer Zugriff auf das Objekt hat, kann dann auch den Hashwert auf der Blockchain lokalisieren. Damit lässt sich nachweisen, dass das Objekt seit dem Schreiben des entsprechenden Blocks nicht manipuliert wurde. Ein typisches Beispiel ist etwa die Validierung von Urkunden, wie etwa Universitätsdiplomen, über eine Blockchain (vgl. Abb. 2).[39]

Der Verantwortliche – also z.B. die Universität – berechnet für das Ursprungsobjekt – z.B. einem Universitätsdiplom – den Hashwert und schreibt ihn über einen ihm zuordnenbaren Blockchain-Account auf eine öffentliche Blockchain. Dadurch ist es möglich, den Hashwert mit der Universität und auch mit einem ungefähren Datum zu verbinden. Diese Angaben finden sich jedoch auch schon bereits in der Diplomurkunde.

Da der Hashwert nicht mehrfach verwendet wird, ist über den Hashwert eine „Verknüpfbarkeit" verschiedener Einträge auf der Blockchain nicht möglich. In der Regel ist auch eine „Inferenz" ausgeschlossen, d.h. aus den implizit aus dem Blockchain-Kontext vorhandenen Daten „Universität" und „ungefähres Datum" kann kein Personenbezug hergestellt werden. Allerdings könnte ein Herausgreifen möglich sein: Das Ursprungsobjekt – also im Beispiel das Diplom – dient als eine Art Schlüssel für den Eintrag auf der Blockchain. Mit ihm kann der Eintrag auf der Blockchain lokalisiert werden. Damit ist ein Herausgreifen möglich. Da über das Diplom der Eintrag auch mit der Person des Diplomanden verbunden ist, besteht ein Personenbezug des Hashwerts zum Diplomanden.

Fraglich ist jedoch, ob mit dem Hashwert auf der Blockchain auch eine Information über die betroffene Person verbunden ist. Ein Personenbezug alleine – ohne Information – führt noch nicht zum Vorliegen eines personenbezogenen Datums.[39] Der Eintrag auf der Blockchain teilt nichts mit, was nicht sowieso schon im Diplom steht. Gleichzeitig ist das Diplom mit dieser Information erforderlich, um den Eintrag zu identifizieren. Der Eintrag ver-

33 *CNIL*, La Blockchain, Premiers éléments d'analyse de la CNIL, 09/2018, S. 8, abrufbar unter: https://www.cnil.fr/sites/default/files/atoms/files/la_blockchain.pdf; englisch: https://www.cnil.fr/sites/default/files/atoms/files/blockchain.pdf.
34 S. etwa BSI-Grundschutz, M 2.11 Regelung des Passwortgebrauchs, abrufbar unter: https://www.bsi.bund.de/DE/Themen/ITGrundschutz/ITGrundschutzKatalog e/Inhalt/_content/m/m02/m02011.html.
35 *CNIL* (o. Fußn. 33), S. 9 f.
36 *Datenschutzbehörde* (o. Fußn. 18).
37 *EuGH* MMR 2016, 842, Rdnr. 46.
38 *CNIL* (o. Fußn. 33), S. 8.
39 Der Autor hat ein Pilotprojekt für Universitätsdiplome der Universität Genf implementiert, das hier als Beispiel dienen soll; vgl. dazu *Erbguth/Benkacem/Gessler/Burgi*, Certification of University Diplomas, 2019, abrufbar unter: https://erbguth.ch/slides/DiplomaPaper.pdf.
39 S. unter IV.2.

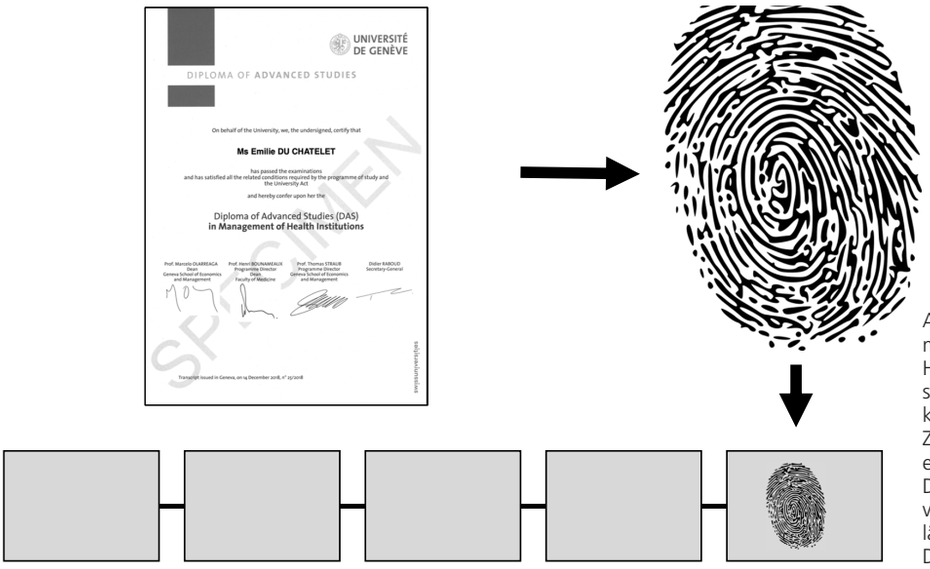

Abb. 2: Einsatz eines Hashwerts, um Diplome zu validieren. Aus dem Diplom wird ein Hashwert berechnet. Bei ihm handelt es sich um eine Prüfsumme, die sich bei der kleinsten Manipulation verändern würde. Zur Absicherung wird diese Prüfsumme auf einer Blockchain unverändert gespeichert. Damit lässt sich das Diplom unabhängig von der Universität validieren. Umgekehrt lässt sich aus dem Hashwert jedoch kein Diplom rekonstruieren.

mittelt daher keine Information über die betroffene Person, die nicht schon zur Herstellung des Personenbezugs zwingend erforderlich wäre. Man muss über den Hashwert dem System nachweisen, dass man das Diplom bereits hat und kann dann implizit auf Informationen schließen, die sowieso im Diplom stehen. Ohne im Besitz des Diploms zu sein, kann man den Hashwert nicht berechnen und damit den Eintrag auf der Blockchain auch nicht identifizieren. Wie im Buch mit allen möglichen Geburtsdaten[40] findet man nur die Information, die man zum Auffinden der Information bereits benötigte.

Der Eintrag auf der Blockchain gibt jedoch implizit noch eine weitere Information preis: Er bestätigt die Echtheit des Diploms. Bei einem unechten oder manipulierten Diplom würde sich der aus dem Diplom berechnete Hashwert nicht auf der Blockchain finden lassen. Ist deshalb die Validierungsmöglichkeit echter Diplome eine personenbezogene Information? Dagegen spricht, dass hier gleich einem Papier- und Tintengutachten nur die Prüfung stattfindet, ob das Dokument echt und unverfälscht ist. Daher ist auch hier der Hashwert nicht als personenbezogenes Datum zu betrachten.

3. Hashwert mit weiteren Daten auf einer Blockchain

In klassischen Pseudonymisierungsszenarien wird nur ein Teil der Daten gehasht und der Rest unverändert belassen, damit er unabhängig von der Person ausgewertet werden kann. Dies ist der Anwendungsfall, den die *Art. 29-Datenschutzgruppe* im WP 216[40] untersucht hatte.

Der Hashwert fungiert hier als eine Art Schlüssel bzw. ID, womit sich weitere Informationen, die sich auf der Blockchain befinden, den gehashten Daten zuordnen lassen (Herausgreifen). Finden bestimmte Hashwerte doppelt Verwendung, so liegt zusätzlich eine „Verkettungsmöglichkeit" vor. Eine Gefahr könnte sich auch durch die „Inferenz" der Daten, die neben den Hashwerten abgelegt sind, mit externen Daten ergeben.

Sind die verbleibenden Daten aber so ausgewählt, dass weder Verkettung noch Inferenz möglich sind, gibt es einen Weg, auch das Herausgreifen zu neutralisieren. „Neutralisieren" deshalb, da das Herausgreifen möglich bleibt, aber keine Information mehr vermittelt. Zur Illustration wird hier auf dem Beispiel der *Art. 29-Datenschutzgruppe* aufgesetzt (vgl. Abb. 1).

Im Ausgangsfall wurden aus Namen, Adresse und Geburtsdatum der Hashwert (von der WP 29 „Referenzwert" genannt) berechnet. Über diesen Hashwert konnte dann in der pseudonymisierten Tabelle (Spalten 4-6) der Zeitraum des Bezugs von Sozialleistungen und der Body-Mass-Index nachgeschlagen werden. Um dies zu verhindern, fließen in Abb. 3 nicht mehr nur Name, Adresse und Geburtsdatum in die Berechnung des Hashwerts ein. Vielmehr wird der Hashwert aus der kompletten Zeile sowie einem zusätzlichen zufälligen Wert (Salz[42]) berechnet. Als Hashwert kommt zudem ein sicheres Verfahren, wie etwa SHA-3 mit längeren Hashwerten, zum Einsatz. Wie im Ursprungsbeispiel (Abb. 1) werden anschließend die ersten drei Spalten ausgeblendet. Nun ist ein Herausgreifen des Eintrags nur möglich, wenn eine Person bereits über die komplette Information eines Datensatzes verfügt. Wegen der Verwendung des zufälligen „Salzes" lassen sich die Angaben auch nicht erraten.

Durch die „Neutralisierung" verhält es sich hier nun genauso wie im zweiten Fall der über eine Blockchain validierbaren Diplome: Ein Herausgreifen ist möglich und es kann damit ein Personenbezug hergestellt werden. Aber alle Informationen, die man durch das Herausgreifen erhält, sind bereits erforderlich, um überhaupt den Hashwert berechnen zu können. Wenn jedoch das Salz sowie einige Datenfelder bekannt sind, können durch Ausprobieren fehlende Datenfelder erraten werden. Um zu ver-

40 S. unter IV.2.
41 *Art. 29-Datenschutzgruppe* (o. Fußn. 26).
42 S. unter III.2.b).

1. Name	2. Adresse	3. Geburtsdatum	4. Zeitraum des Bezugs bestimmter Sozialleistungen	5. Body-Mass-Index	6. Salz	7. Hashwert
Max Müller	Berlin	05.11.63	< 2 Jahre	15	23DF6CB7F8023ADC	b601fdab42f72d7e2c9600787b0ce958a6093d1569655004838099173 4e28b01
Eva Schön	München	01.07.82	> 5 Jahre	14	78585ACBF234DA01	b9fb85d573c41ebf5d3eba262857711e35bcc364569cf6d1f33d024fae40a4e8
Ulrich Klein	Hamburg	13.12.77	< 2 Jahre	16	56FDEA786BC5EE9A	a335e0b763550a9a4e50ca1d8c4b34a962f132e32317c1478e4e7d9ca7e01f6c
Herbert Schmidt	Köln	15.05.85	> 5 Jahre	18	AB5552CCA3209BFA	c31cb0bd973f2a68c10fb1ff2c830048f0e6fadfe8b31bce83148d2247ba0621
Ulrike Groß	Stuttgart	12.12.70	< 2 Jahre	20	145AA0233BCB6D91	bc470b22fd8d63d343ea4deb88261889c6867c76cd2e887352e08615c6bc2710

Abb. 3: Hashwerte werden nun aus der gesamten Zeile berechnet. Die ersten drei Spalten werden sowohl danach als auch vorher gelöscht. Durch die Bildung des Hashwerts aus der gesamten Zeile, ist für eine Identifizierung bereits die komplette Information erforderlich, sodass eine Identifizierung keine weitere Information ergibt.

hindern, dass das Salz isoliert als eine Art Zugriffsschlüssel weitergegeben werden kann, empfiehlt es sich, den Hash z.B. über einen Scan mit hoher Entropie satt über die Kombination aus wenig komplexen Zahlen- oder Buchstabenwerten und Salz berechnen zu lassen. So lassen sich nach der Neutralisierung keine personenbezogenen Informationen mehr aus der jetzt anonymisierten Tabelle ableiten.

VI. Schlussbetrachtung

Je nach ihrer Verwendung sind Hashwerte personenbezogener Daten entweder selbst als personenbezogene Daten oder aber als anonyme Daten einzustufen. Das ist keine „Grauzone“[43]. Vielmehr sind Hashwerte ein technisches Hilfsmittel, welches in seinem Verwendungskontext betrachtet werden muss. Bei richtiger Verwendung, wie etwa der Validierung umfangreicher externer Daten, können Hashwerte von personenbezogenen Daten auf Blockchains DS-GVO-konform abgelegt werden.

Dieses Ergebnis ergibt sich unabhängig davon, ob man den Aufwand zur Herstellung des Personenbezugs berücksichtigt, oder – wie die französische Aufsichtsbehörde *CNIL* – auch dann noch einen Personenbezug annimmt, wenn die Personenbeziehbarkeit außergewöhnlich aufwendig, aber nicht völlig unmöglich erscheint. Die *CNIL* empfiehlt dabei – nicht ganz nachvollziehbar – den Einsatz gepfefferter Hashwerte, d.h. Hashwerte, die mit einem Passwort geschützt sind.

Bei später entdeckten Fehlern in den Algorithmen der Privacy Enhancing Technology bleiben die Verantwortlichen aber stets in der Verantwortung – sie müssen künftige Entwicklungen stets mitberücksichtigen. Damit führt das Schreiben von Daten auf unveränderliche Blockchains zu einem Haftungsrisiko. Gerade die massive Datenreduktion durch kryptografische Hashwerte bietet jedoch – richtig verwendet – eine Gewähr dafür, dass selbst dann keine personenbezogenen Informationen aus den Einträgen der Blockchain ableitbar sind, wenn eine Hashfunktion als solche nicht mehr sicher ist.

Jörn Erbguth
ist Diplom-Informatiker und Diplom-Jurist, berät zu Blockchain und Datenschutz in Genf. Er promoviert aktuell zum Thema Blockchain Governance und lehrt an verschiedenen Schweizer Universitäten.

Bedanken möchte ich mich für die fruchtvollen Diskussionen in der Gruppe DIN SPEC 4997 „Privacy by Blockchain Design“ und insbesondere bei Michael Kolain für sein umfangreiches und wertvolles Feedback.

43 *EU-Blockchain Observatory* (o. Fußn. 31), S. 21.