

Big Data – Zeig mir deine Daten und ich sag dir, wer du bist

DaSou-Podcast-Folge mit Sharleen Betker

Karina Filusch: Hallo und herzlich willkommen beim DaSou Podcast. Ich bin Karina Filusch, Datenschutzanwältin und externe Datenschutzbeauftragte. In jeder Folge sprechen wir mit einer Expertin oder einem Experten über Datensouveränität, abgekürzt DaSou. Schön, dass du wieder dabei bist und reinhörst. Abonniere doch unseren Podcast, wenn er dir gefällt und hinterlass uns auch gerne eine Bewertung, darüber freuen wir uns sehr. Heute möchten wir über Big Data sprechen. Dafür ist Sharleen Betker bei mir. Sie ist ein echter Management-Nerd und hat in ihrem Wirtschaft- und Recht-Studium angefangen, sich für Datenschutz zu interessieren, weil das Thema für sie ganz klar ein Zukunftsthema ist. Aktuell berät sie Unternehmen und Behörden dabei, wie sie mit Managementsystem ihre Informationssicherheit und ihren Datenschutz steuern und verbessern können. Liebe Sharleen, eines deiner Herzentemen ist auch Big Data. Big Data hört man sehr oft und eigentlich weiß man gar nicht so richtig, was sich dahinter verbirgt. Was ist das denn?

Sharleen Betker: Big Data an sich ist erst mal nur die Bezeichnung für ein Phänomen bzw. einen Versuch, eine bestimmten Art von Informationen, die eine bestimmte Art von Merkmalen haben, miteinander zu kopieren. Das heißt, als Big Data werden solche Informationen bezeichnet, die eine große Menge an Daten sind, die auch unterschiedlich strukturiert sind, z.B. durch unterschiedliche Formate, und außerdem in Echtzeit ausgewertet werden. Es gibt auch Definitionen, die sagen, dass dazu auch gehört, dass es unterschiedliche Datenqualitäten innerhalb des Datenbestandes gibt. Für alle, die sich das besser merken wollen, gibt es die Bezeichnung „vier V“ bzw. „four V“. Das wären dann volume, variety, velocity und veracity.

Karina Filusch: Wenn diese vier V's erfüllt sind, dann hat man Big Data?

Sharleen Betker: Genau, also im Prinzip sagt man, wenn mindestens eine so große Menge an Daten hat, die dann meistens im Terrabyte-Bereich oder im Zettabyte-Bereich liegt, kann man von Big Data sprechen.





Karina Filusch: Warum ist Datenschutz und Big Data so schwer auf einen Nenner zu bringen?

Sharleen Betker: Grundsätzlich können Big Data zunächst unterschiedliche Daten sein. Das heißt, Big Data können auch Sachdaten sein, wie der Jurist sagen würde, also Daten ohne Personenbezug. Es können natürlich auch personenbezogene Daten sein und damit hat dann der Datenschutz seinen Auftritt. Der offensichtlichste Konflikt zwischen Big Data und Datenschutz ist der, dass Big Data, wie der Name schon sagt, natürlich eine große Datenmenge bezeichnet, während der Datenschutz natürlich versucht, Daten zu minimieren. Der Zweck von Big Data ist im Grunde genommen der Erkenntnisgewinn. Man möchte anhand der Analyse der Big Data Erkenntnisse gewinnen und die liegen meistens außerhalb von dem, was wir als Menschen verarbeiten können. Dadurch kann man auch Tatsachen offenlegen, denen wir uns gar nicht bewusst sind oder die eigentlich geheim bleiben sollten. Wenn personenbezogene Daten analysiert werden, kann es dazu führen, dass Verhaltensstrukturen oder Merkmale von uns aufgedeckt werden, obwohl wir das gar nicht wollten oder von denen wir gar nicht wissen, dass wir sie haben. Dadurch entsteht auch die Gefahr der Manipulation, sodass diese Daten, die uns nicht bewusst sind oder die geheim bleiben sollten, gegen uns verwendet werden können. Gegen diese Gefahr hat die DSGVO die Grundsätze der Zweckbindung und der Verhältnismäßigkeit etabliert. Aus Sicht der Unternehmen besteht der Reiz an Big Data gerade darin, dass man anhand der Rohdaten unterschiedliche Erkenntnisse gewinnen kann - je nachdem, unter welcher Fragestellung man die Daten auswertet. Das heißt, Innovation und Kundenwünsche könnten durch die Daten gefunden oder bestätigt werden. Ob das tatsächlich so ist, hängt allerdings von der Datenqualität und der Qualität der Auswertungsmethode ab. Hinzu kommt außerdem, dass sich der Aufwand für die Datenerhebung durch Mehrfachnutzung natürlich auch relativiert. Das heißt, je mehr Fragestellungen ich anhand der Daten habe, umso mehr lohnen sich die Kosten, die ich am Anfang hatte, um die Daten überhaupt zu sammeln. Das heißt, die Motivation der Unternehmen ist groß, die Rohdaten für unterschiedliche Fragestellungen zu nutzen, die ihnen zu dem Zeitpunkt der Datenerhebung vielleicht gar nicht bekannt sind. An diesem Punkt kommt die DSGVO ins Spiel. Demgegenüber steht der Grundsatz der Zweckbindung. Damit wir als Betroffene eine Entscheidung treffen können, ob und inwieweit wir der Verarbeitung zustimmen, muss vor der Verarbeitung klar sein, wofür die Daten ausgewertet werden. Sonst weiß ich gar nicht, wozu ich zustimme. Das heißt, ich als Unternehmen bin eigentlich dazu gezwungen, schon vorher möglichst eindeutig zu sagen, wofür ich die Daten verarbeite. Hinzu kommt, dass der Grundsatz der Verhältnismäßigkeit z.B. die Speicherbegrenzung mit einschließt. Diese schließt sich nahtlos an das Dilemma an.





Rechtsanwältin Karina Filusch, LL.M. | Friedrichstraße 95, 10117 Berlin | T: 030 219 11 555

Wer Daten sammelt, um Fragen zu einem unbestimmten Zeitpunkt in der Zukunft zu klären, der möchte sie natürlich möglichst lange speichern. Das ist der Grundkonflikt zwischen Big Data und Datenschutz.

Karina Filusch: Du hast gerade Unternehmen angesprochen, die das als Geschäftsmodell praktizieren. Wie genau funktioniert das?

Sharleen Betker: Wenn ich Geschäftsmodell sage, dann meine ich zunächst nur grundlegend das abstrakte Konzept, um Unternehmen zu beschreiben. Es gibt nicht so etwas wie das „Big-Data-Geschäftsmodell“, sodass man sagen kann „Okay, wenn du folgende Merkmale erfüllt hast, dann hast du ein Big-Data-Geschäftsmodell“. Das ist mehr als Herangehensweise oder Versuch zu betrachten, solche Unternehmen zu gruppieren und zu beschreiben. Wenn ich Big-Data-Geschäftsmodell sage, dann meine ich Unternehmen, die ihre Dienstleistungen oder ihr Produkt mithilfe von Big Data erbringen und bei denen die Beschaffenheit der Informationen untrennbar mit dem Geschäft verbunden ist. Ein Beispiel wäre Facebook als Plattform selbst. Denn bei Facebook ist es so, dass sie versuchen, Netzwerkeffekte zu nutzen. Sie haben unterschiedliche Arten von Daten. Es gibt Texte, Videos, Bilder und so weiter. Es gibt auch unterschiedliche Datenqualitäten. Beispielsweise kann man sich nie sicher sein, ob die Daten korrekt angegeben sind. Das wird nicht geprüft. Man kann von den Interessen von Personen, die sie bei Facebook angeben, darauf schließen, was sie interessiert. Das geht auch, wenn ich mir ein Profil angucke und schaue, was der Nutzer zuletzt gelikt hat. Das heißt, ich habe auch in unterschiedlichen Schichten Informationen über diese Person.

Karina Filusch: Jetzt hast du ein ganz konkretes Beispiel schon genannt: Facebook. Dort erleben wir Big Data. Wo erleben wir Big Data noch im Alltag?

Sharleen Betker: Es gibt unterschiedliche Möglichkeiten, mit Big Data Geschäftsmodelle zu errichten. Man könnte beispielsweise nach der Intensivität der Verarbeitung abgrenzen. Man könnte sagen, dass das Bereitstellen von Rohdaten ein Big Data Geschäftsmodell ist. Wenn man Daten nur analysiert und visualisiert, dann wäre das eine andere Stufe. Die höchste Stufe der Datenverarbeitung wäre diejenige, bei der man Entscheidungsfindungen trifft. Im Alltag ist es fast schon schwierig zu sagen, wo keine Big Data stattfindet. Das liegt daran, dass die großen Unternehmen wie Facebook und Amazon Algorithmen benutzen, um uns bestimmte Produkte anzubieten. Viele Teile der Werbung und des Marketings sind darauf übergegangen, Kundengruppen über





Rechtsanwältin Karina Filusch, LL.M. | Friedrichstraße 95, 10117 Berlin | T: 030 219 11 555

Algorithmen ausfindig zu machen. Dort muss man davon ausgehen, dass Big Data im Spiel ist.

Karina Filusch: Zum Beispiel bei einer Payback-Karte oder Kundenkarte, so würde ich mir das vielleicht vorstellen, oder?

Sharleen Betker: Genau. Je nachdem, wie viele Daten da verarbeitet werden, würde man dann von Big Data sprechen.

Karina Filusch: Big Data sammelt also ganz viele Daten. Was ist an Big Data nützlich oder scheint zumindest auf den ersten Blick nützlich zu sein?

Sharleen Betker: Mit der richtigen Datenqualität, Fragestellung und auch Auswertungsmethode können Informationen dazu führen, dass wir bessere Entscheidungen treffen und vor allem auch mit weniger Annahmen arbeiten müssen. Datenschutz und Big Data sind nicht prinzipiell unvereinbar. Mit dem richtigen Konzept können Daten auch beide Bedürfnisse erfüllen, nämlich das nach Antworten und das nach Datenschutz. Dafür muss ich mir zuvor die Frage stellen: Welche Informationen brauche ich eigentlich wofür? Und wie kann ich personenbezogene Daten bestmöglich reduzieren? Es muss meine Motivation sein, zu versuchen, personenbezogene Daten zu schützen. Kritisch wird es dort, wo personenbezogene Daten bewusst und ausschließlich zur Gewinnerzielung eingesetzt werden, weil ich sie dann natürlich auch nicht reduzieren kann. Beispielsweise habe ich im Marketing die Motivation so viel wie möglich über die Kunden und Kundinnen herauszufinden. Je individueller mein Marketing ist, umso kosteneffizienter ist es und vielleicht auch umso gewinnbringender, weil ich die richtige Zielgruppe anspreche. Hier liegt eines der Probleme im Zusammenhang mit Big Data: Man hat die Möglichkeit viele Informationen herauszufinden und einen möglichst hohen Nutzen damit zu erzielen, obwohl es eigentlich nicht erstrebenswert ist, Personen dermaßen zu erforschen, um das Wissen für geschäftliche Interessen auszunutzen.

Karina Filusch: Denkst du, dass es den meisten Nutzer*innen überhaupt bekannt ist, dass mit ihren Daten Geld verdient wird? Werden sie darüber transparent informiert?

Sharleen Betker: Ich finde die Frage richtig gut. Meines Erachtens fallen einem bei längerer Überlegung drei heiße Eisen auf. Das Erste ist die Außendarstellung des





Unternehmens. Ich glaube, wenn man Leute auf der Straße fragen würde: „Was ist der Geschäftszweck von Facebook?“, dann würden die meisten antworten „Facebook ist ein soziales Netzwerk“. Das ist aber tatsächlich gar nicht das, womit Facebook sein Geld verdient, und das wissen wir im Hinterkopf auch alle. Facebook verdient sein Geld eigentlich als sogenannte multi sided platform. Das heißt, dass Facebook Kundinnen und Kunden mit Werbeträgern zusammenbringt. Wenn man Facebook Ads oder Facebook Advertisement auf Google aufruft, dann merkt man auch, dass Facebook eine vollkommen andere Außenkommunikation in Richtung Unternehmen hat. Auf der anderen Seite gibt es auch die normale Login-Seite, mit der man sich einloggen würde. Ich glaube, dass vielen gar nicht bewusst ist, dass das der Hauptkern des Geschäftsmodells ist.

Karina Filusch: Was sind die anderen Seiten?

Sharleen Betker: Das Zweite wären für mich die Sachen, die die DSGVO auch etabliert hat, nämlich die Datenschutzerklärung und die betroffenen Rechte. Es ist interessant, wie die umgesetzt sind. Meiner Meinung nach handelt es sich an sich um ein großartiges Mittel zur Transparenz. Ich finde es auch gut, dass Unternehmen zu ihrem Glück gezwungen werden, indem sie eine gewisse Transparenz gegenüber ihren Kunden schaffen müssen. Allerdings kann man eine gute Datenschutzerklärung oder auch eine schlechte Datenschutzerklärung schreiben. Davon hängt die Transparenz auch ab und so genau kann das in der DSGVO gar nicht vorgegeben werden. Außerdem finde ich, dass Vergleichsmöglichkeiten Transparenz schaffen könnten. Für mich bedeutet das, dass ich, wenn mir nur eine Option zur Verfügung steht, davon ausgehe, dass diese Option den besten Standardwert hat. Ich finde, dass häufig nicht genug auf Sicherheitsmaßnahmen eingegangen wird. Ein aktuelles Beispiel, bei dem es sich zwar nicht um ein Geschäftsmodell handelt (beziehungsweise nur in einem Fall), sind die Corona-Warn-App und die Luca-App. Die eine App ist ein Geschäftsmodell. Bei der anderen App kann man sich darüber streiten. So kann man die Corona-Warn-App im Prinzip als Geschäftsmodell analysieren, aber grundsätzlich will die Corona-Warn-App kein Geld verdienen. Dementsprechend würde ich das als Unternehmen herausnehmen. Die Corona-Warn-App hat von Anfang an ein ganz klares Datenschutzdesign mit dem dezentralen Ansatz gehabt, was man auch jetzt im Nachhinein immer noch nachvollziehen kann. Für mich ist die Qualität der Sicherheit auch eine transparente Information, die ich zu einem Produkt oder zu einer Dienstleistung bekomme. Bislang haben wir noch keine Systeme, die mir zeigen können, wie die Datenqualität und auch der Umgang des Unternehmens mit meinen Daten sind. Ist er sicherer als andere Unternehmen? Gibt es Verbesserungen? Ich glaube, dass man das als Verbraucher und





Rechtsanwältin Karina Filusch, LL.M. | Friedrichstraße 95, 10117 Berlin | T: 030 219 11 555

Verbraucherin gar nicht so genau nachvollziehen kann, wenn man nicht in dem Unternehmen selbst tätig war und mal gesehen hat, wie andere Unternehmen das machen. Das wird nicht nach außen getragen und da fehlt glaube ich noch die Transparenz und sowas wie ein Gütesiegel.

Karina Filusch: Denkst du an sich, dass mit Big Data Grenzen überschritten werden?

Sharleen Betker: Kein Jura-Podcast ohne die übliche Jura Floskel: Es kommt darauf an. Es kommt darauf an, ob mit Big Data Sachdaten analysiert, beziehungsweise ausgewertet werden oder ob das personenbezogene Daten sind. Wenn es um personenbezogene Daten geht, würde ich sagen, dass insbesondere bei Big-Data-Modellen ein sehr hohes Risiko dafür besteht, dass Grenzen überschritten werden. Insbesondere Modelle, in denen die Motivation besteht, möglichst viel über eine Person zu wissen. Ich denke, dass man wenig falsch machen kann, wenn man sich an das hält, was die DSGVO von einem fordert und sich die Motivation zu eigen macht, den Datenschutz möglichst von Anfang an zu benennen. Wenn man die Grundsätze der DSGVO einhält, dann schützt man die Daten vor bestimmten Risiken. Sollte dann ein Schaden eintreten ist dieser meist gering. So verhält es sich insbesondere, wenn man keine sensiblen personenbezogenen Daten verarbeitet hat oder wenn man die sensiblen personenbezogenen Daten mit Sicherheitsmaßnahmen wie der Pseudonymisierung oder technischen Maßnahmen geschützt hat. Ein Restrisiko wird immer bestehen bleiben. Meiner Meinung nach liegt die Verantwortung, dieses Risiko zu reduzieren bei den Unternehmen. Das Unternehmen trägt auch die Verantwortung, den Schaden für die Betroffenen zu reduzieren.

Karina Filusch: Gibt es denn für dich gesetzliche Grauzonen, um mit Daten Geld zu verdienen?

Sharleen Betker: Ich würde sagen, dass die Grauzone dort anfängt, wo man etwas nicht direkt verbieten kann bzw. wo es einen Interpretationsspielraum für die Umsetzung gibt. Die DSGVO ist seit 2018 in Kraft und trotzdem hat es immer noch keiner geschafft, Facebook in die Schranken zu weisen. Dementsprechend gibt es v.a. dort eine Grauzone, wo es an Konsequenzen für Unternehmen fehlt, wenn diese sich nicht an den Datenschutz halten. Leider gibt es auch noch viele Grauzonen bei der Umsetzung der DSGVO. Die DSGVO legt die Grundsätze dessen fest, was für uns Datenschutz bedeutet. Es gibt auch das Standard-Datenschutz-Modell, mit dem man versucht, technische und organisatorische Maßnahmen für die Unternehmen darzustellen und ihnen





Rechtsanwältin Karina Filusch, LL.M. | Friedrichstraße 95, 10117 Berlin | T: 030 219 11 555

Hilfestellungen zu geben. Das reicht allerdings noch lange nicht aus, um in der Praxis dafür zu sorgen, dass die Unternehmen sorgsamer mit ihren Daten umgehen.

Karina Filusch: Mich würde noch interessieren: Weißt du zufällig, wie es in anderen Ländern aussieht? Jetzt haben wir über die DSGVO, also über Europa, gesprochen. Weißt du da zufällig Näheres?

Sharleen Betker: Facebook ist ein gutes Beispiel, weil das Unternehmen auch in den USA und weltweit operiert. Man kann bereits sehen: Die anderen Länder versuchen sich an der EU zu orientieren, weil wir sowas wie einen Standard geschaffen haben, was Datenschutz angeht und der Vorreiter auf diesem Gebiet sind.

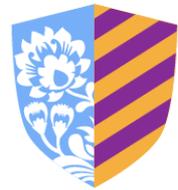
Karina Filusch: Man hört es immer mal wieder, dass die DSGVO in anderen Ländern kopiert wird, weil sie so gut ist. Wir beschwerten uns, dabei schützt die DSGVO doch eigentlich nur unsere Rechte und normiert etwas, was eigentlich schon vorher Gesetz oder Norm war. Was ist denn für dich persönlich DaSou?

Sharleen Betker: Für mich bedeutet Datensouveränität vor allem, dass ich volle Kontrolle über meine Daten habe und gleichzeitig, dass ich auch eine Auswahl habe. Insbesondere bei sozialen Medien möchte ich persönlich aktiv Kontrolle darüber haben, welche Teile von mir wie bewertet werden. Ich möchte falsche Daten korrigieren können und ich will nicht nur wissen, wie die Verarbeitung funktioniert, sondern auch die Entscheidungen, die darauf basieren, verstehen. Ich persönlich würde gerne in meine Darstellung aktiv eingreifen können. Auch in der nicht-digitalen Welt ist es so, dass das, was ich in der Öffentlichkeit (und dazu gehören für mich auch Unternehmen) preisgebe bzw. nicht preisgebe in meiner Verantwortung liegt. Gleiches gilt für die Art und Weise wie ich mich präsentiere. Ich finde, dass dies sich auch im Digitalen widerspiegeln muss. Zudem finde ich es wichtig, dass Unternehmen uns nicht dazu zwingen können, Dinge geheim zu halten oder Ansichten von uns zu verfälschen, weil das immer noch unsere Selbstentfaltung ist.

Karina Filusch: Ich finde besonders schön, wie du das dargestellt hast. Liebe Sharleen, vielen lieben Dank, dass du da warst und dass du dir so viel Zeit genommen hast, um uns zu erklären, was Big Data eigentlich bedeutet. Lieben Dank!

Sharleen Betker: Sehr gerne, es hat sehr viel Spaß gemacht!





Rechtsanwältin Karina Filusch, LL.M. | Friedrichstraße 95, 10117 Berlin | T: 030 219 11 555

Karina Filusch: DaSou ist eine Produktion der Kanzlei Filusch. Mehr Infos findet ihr auf unserer Website www.dasou.law. Der Jingle wurde komponiert von Mauli, die Idee zu DaSou hatte Axel Jürs, das Cover hat Hélène Baum erstellt, beraten wurden wir von Susan Stone. Editiert wurde der Podcast von Christoph Hinners.

