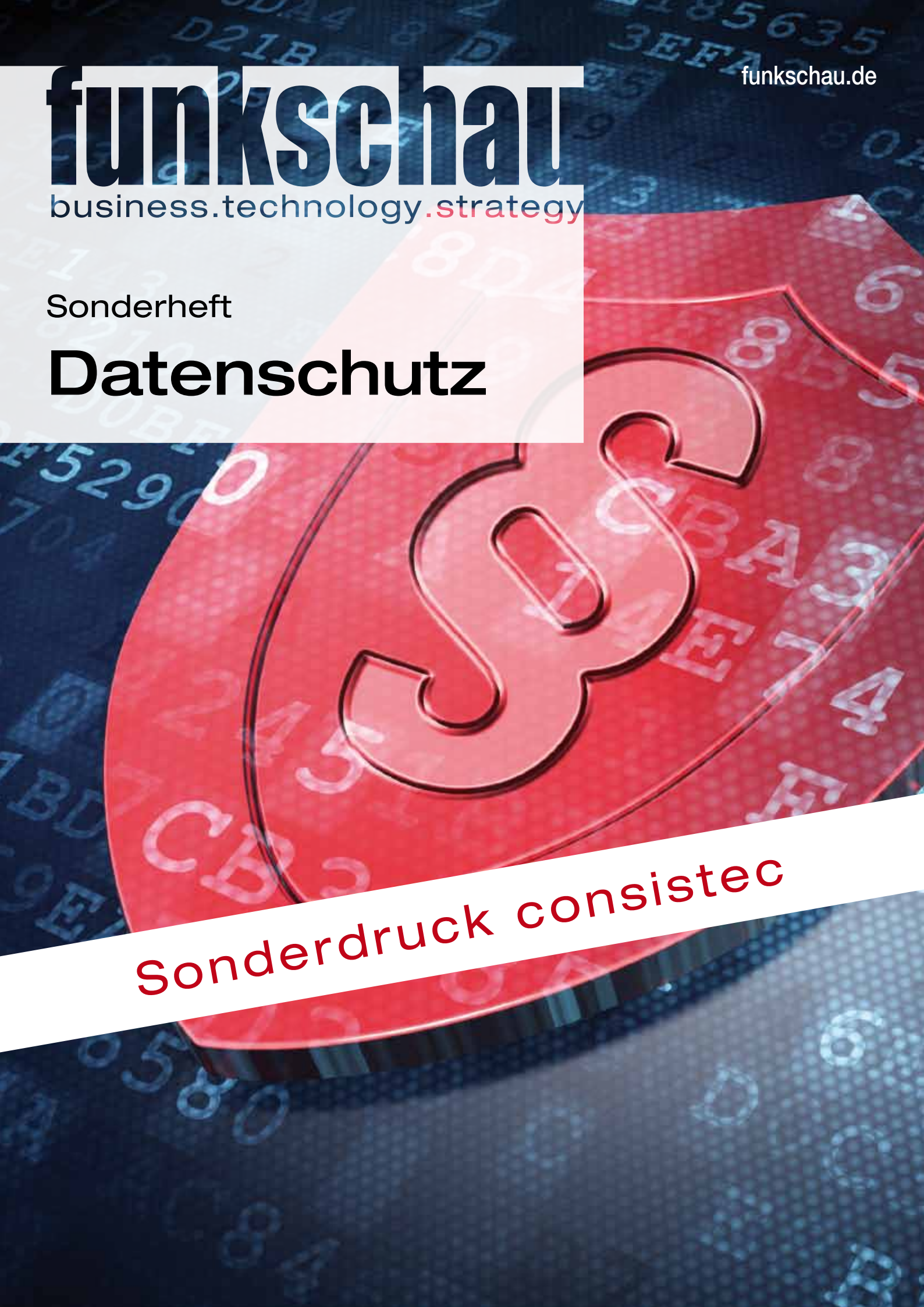
funkschau
business.technology.strategy

funkschau.de

Sonderheft

Datenschutz



Sonderdruck consistec

würdiger. Durch die zunehmende Vernetzung mit Lieferanten, Partnern und Kunden, die Automatisierung von Fertigungsprozessen sowie die intensive Auswertung von geschäftsrelevanten Daten kommt es zu einer immer stärkeren Kopplung des Unternehmenserfolgs an die Leistungsfähigkeit und Zuverlässigkeit der IT-Systeme sowie geschäftskritischen Anwendungen. Folglich haben viele Unternehmen ein vitales Interesse daran, ihre IT-Systeme zu überwachen: Monitoring und Optimierung von Performance und Qualität, Reduzierung von Ausfallzeiten, schnellere Fehlerlokalisierung und -behebung sowie Reduzierung von Risiken durch Schadsoftware, Spionage oder Sabotage.

Spannungsfelder

Im Kontext der EU-DSGVO oder dem BDSG entsteht so ein Spannungsfeld zwischen Datensicherheit, IT-Sicherheit, IT-Performance und Datensparsamkeit (§5, Abs. 1c der DSGVO). Dieses Dilemma ist in vielen Unternehmen nicht oder nur unzureichend gelöst.

Die unter den zuvor genannten Gesichtspunkten notwendigen Tracing- und Monitoring-Systeme stellen in Abhängigkeit der verwendeten Technologie ein unterschiedlich großes aber potenzielles Risiko zum Datenmissbrauch dar. Ein Beispiel, das in der Praxis sehr oft vorzufinden ist, ist die bei der Fehlersuche und -analyse weit verbreitete und hilfreiche Kombination aus IT-Mitarbeiter, Notebook, Wireshark (leistungsfähige Open Source Software zum Mitschneiden und Dekodieren von Netzwerkdaten) und konfigurierten Mirror-Ports an zentralen Switchen. Der anlassbezogene und zeitbegrenzte Einsatz legitimiert unter Datenschutzgesichtspunkten zwar die zuvor geschilderte Vorgehensweise, in der Praxis kann aber keine Einschränkung auf die zur Fehleranalyse tatsächlich relevanten Netzwerkdaten erfolgen. Insofern können IT-Mitarbeiter sehr schnell und sehr einfach mit vielen personenbezogenen Daten, die in den Netzwerkdaten enthalten sind, in Kontakt kommen.

Ebenfalls gängig und mit Risiken verbunden ist die Weitergabe von Trace-Files (Mitschnitt von Netzwerkdaten) an Lieferanten von IT-Systemen zur Fehleranalyse oder im Rahmen der Inbetriebnahme neuer Systeme und Anlagen. Auch hier kann es schnell dazu kommen, dass personenbezogene Daten oder kritische Infrastrukturinformationen unbeabsichtigt in die falschen Hände geraten.

Zwei Ansätze

Zum Schutz von personenbezogenen Daten oder anderen schützenswerten Informationen, wie beispielsweise Infrastrukturinformationen in Netzwerkdaten, gibt es neben dem Ansatz, den Zugriff erst gar nicht zu ermöglichen, zwei grundlegende Ansätze: Anonymisierung und Pseudonymisierung.

Bei der Anonysmisierung wird der Bezug der erhobenen Daten zu einer Person unwiederbringlich verworfen. Dies kann durch einfaches Entfernen von zuordenbaren Merkmalen erreicht werden oder aber durch Ersetzen mit einer vollkommen zufällig gewählten Alternative. Dies bietet die größtmögliche Sicherheit, hat aber auch teils gravierende Nachteile. Hauptproblem ist, dass Daten dadurch für ihren

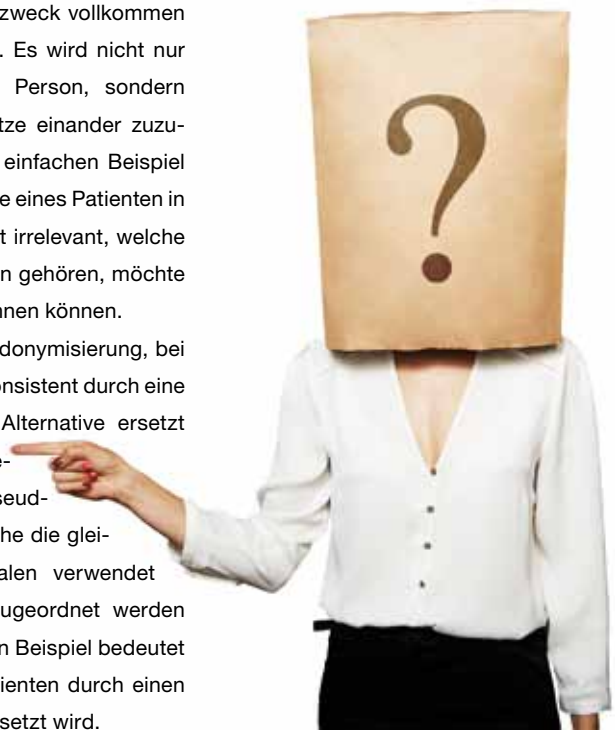
ursprünglichen Verwendungszweck vollkommen unbrauchbar werden können. Es wird nicht nur verhindert, die Daten einer Person, sondern auch, verschiedene Datensätze einander zuzuordnen. Dies kann an einem einfachen Beispiel verdeutlicht werden: Der Name eines Patienten in einer medizinischen Studie ist irrelevant, welche Befunde zum selben Patienten gehören, möchte man aber sinnigerweise erkennen können.

Die Alternative ist die Pseudonymisierung, bei der zuordenbare Merkmale konsistent durch eine scheinbar zufällig gewählte Alternative ersetzt werden. Vorteil dieses Vorgehens ist, dass unabhängig pseudonymisierte Datensätze, welche die gleiche Ersetzung von Merkmalen verwendet haben, weiterhin einander zugeordnet werden können. Im zuvor verwendeten Beispiel bedeutet das, dass der Name des Patienten durch einen eindeutigen, neuen Namen ersetzt wird.

Pseudonymisierung versucht die gegensätzlichen Anforderungen von Datenschutz und verbleibendem, für die Auswertung notwendigem Informationsgehalt für eine konkrete Anwendung zu optimieren. Wichtig ist es dabei zu verstehen, dass es absolut gesehen keine beste Pseudonymisierung gibt. Zur Erläuterung greifen wir wieder auf das Beispiel mit Patientendaten zurück: Das Geburtsdatum eines Patienten lässt Rückschlüsse auf die Person zu. Je nach Kontext kann diese Information verschieden geschützt werden. Spielt das Alter keine Rolle, lässt sich diese Information entfernen oder durch ein zufälliges Alter ersetzen. Ist die Information hingegen relevant, kann das Alter in Jahren oder auch nur in Altersklassen in den pseudonymisierten Daten verbleiben.

Ein brauchbares Pseudonymisierungssystem muss sich auf die konkreten Anforderungen der Anwendung anpassen lassen, sodass es den optimalen Mittelweg zwischen Datenschutz und Nützlichkeit der Daten finden kann. Insbesondere ist notwendig, dass es diese Entscheidung für jede in den Daten enthaltene Information erlaubt und nützliche Pseudonymisierungsprimitive für jede Art von Merkmal anbietet.

Eine weitere Herausforderung bei Pseudonymisierung ist es, die schützenswerten Merkmale aus einem Datensatz zu extrahieren und nach Modifikation wieder reibungslos zu integrieren. Es ist zum Beispiel nur schwer zu garantieren, dass alle schützenswerten Informationen in einem Prosatext erkannt werden. Genauso



PSEUDONYMISIERUNG KANN UNTER PRAKTISCHEN GESICHTSPUNKTEN NUR VERLÄSSLICH FUNKTIONIEREN, WENN DIE KODIERUNG VON INFORMATIONEN BEKANNT IST.

schwer ist es, nach Pseudonymisierung von einzelnen Bestandteilen des Textes wieder einen flüssigen Prosatext zu erzeugen. Bei einem Gedicht mit festem Reimschema kann es schwer bis fast unmöglich sein, auch pseudonymisiert noch das Reimschema zu erfüllen.

Pseudonymisierung kann unter praktischen Gesichtspunkten nur verlässlich funktionieren, wenn die Kodierung von Informationen bekannt ist und von den Datensätzen auch eingehalten wird. Dieses Problem würde bei der bereits zuvor angeführten Patientenstudie auftreten, wenn man bei einem Patienten das Geburtsdatum in ein falsches Feld, das nicht pseudonymisiert wird, einträgt oder das Geburtsjahr in römischen Ziffern schreibt. Ebenso problematisch wäre die Erwähnung des Geburtsjahres in Freitextform beim Befund.

Die Anwendung von Pseudonymisierung im eingangs aufgeführten Beispiel ‚Monitoring von IT-Systemen' für Bereiche, die gerne von Unternehmen bei der Risikoabschätzung im Kontext der EU-DSGVO übersehen werden, ist besonders anspruchsvoll. Die Pseudonymisierung von Netzwerkdaten steht vor all den zuvor im Einzelnen erläuterten Herausforderungen: Netzwerkdaten sind extrem heterogen, es gibt unzählige Protokolle, die dekodiert und schließlich wieder kodiert werden müssen. Viele Informationen tauchen an vielen verschiedenen Stellen auf, das System muss sie zuverlässig überall erkennen und konsistent verfremden.

Zudem ist es oft erforderlich, Netzwerkdaten erst aus vielen Netzwerkpaketen zusammenzusetzen, bevor sich der Inhalt voll verstehen lässt. Eine Netzwerkkorrespondenz in diesem Licht so zu verfremden, dass sie für den Beobachter wieder einen gültigen Austausch von Paketen darstellt, ist äußerst kompliziert. In Kombination mit dem Anspruch, das Maß an Verfremdung dem notwendigen Maß an Erhalt von Informationen anpassen zu können, ergibt sich ein extrem anspruchsvolles Themenfeld, das es bei erfolgreicher Umsetzung erlaubt, das Ausbalancieren zwischen Datensicherheit und -minimierung stärker Richtung IT-Sicherheit und praxistauglichen Prozessen bei der Inbetriebnahme, bei der Wartung oder der Fehleranalyse von IT-Systemen zu verschieben.

Thomas Sinnwell ist CEO R&D bei Consistec Engineering & Consulting