Stellungnahme des vhw sachsen zum Entwurf der SächsPersAnGDVO

Der Entwurf stellt recht detailliert Vorgehensweisen und Kostenabschätzungen dar. Der Kernpunkt aus der Sicht des Datenschutzes ist die Pseudonymisierung laut § 3, Abs. (3), da hier entsprechend § 2 personenbezogene Daten verarbeitet werden.

Mir erschließt sich nicht klar, warum hier Geburtsdatum und Personalnummer als Grundlage für die Erstellung des Datensatzkennzeichens verwendet werden. Es ist allgemein bekannt, dass gerade das Geburtsdatum und auch die vielen Mitarbeitern zugängliche Personalnummer nur eine schwache Pseudonymisierung ermöglichen. Klar, hier soll ein Hash-Verfahren genutzt werden, um diese beiden Eingaben zu kombinieren und damit eine Rückverfolgung deutlich zu erschweren.

Allerdings halte ich die Vorgabe „Das dafür zu verwendende Verfahren wird einheitlich von der Staatskanzlei vorgegeben." für bedenklich. Die Einheitlichkeit ist natürlich sinnvoll. Allerdings sollte hier ein Hash-Verfahren nach dem aktuellen Stand der Technik verwendet werden und die Vorgabe nicht durch die Staatskanzlei, sondern durch den Datenschutzbeauftragten des Freistaats Sachsen erfolgen. Sonst besteht nämlich klar ein potentieller Interessenskonflikt. Die Verfahren und auch die Hardware entwickeln sich nämlich ständig weiter. Was vor 10 Jahren noch als sicher galt, ist es heute häufig nicht mehr.

Um die Schwäche des Datenpaares Geburtsdatum und Personalnummer zu verdeutlichen, schätzen wir einmal die Größe des Datenraums sowie die sich daraus ergebende Zeit für eine vollständige Suche ab. Als Alter eines aktiv Beschäftigten kommen 18 bis 67 Jahre in Frage, das sind 50 mögliche Geburtsjahre. Ein Jahr hat nun 365 oder 366 Tage. In einer Abschätzung nach oben liefert das also 50 x 400 = 20.000 mögliche Geburtsdaten. Eine Personalnummer ist bei mir 7-stellig, das gibt 10.000.000 Möglichkeiten. Multipliziert mit der Zahl der in Frage kommenden Geburtsdaten liefert dies 200.000.000.000 = 2×10^{11} Möglichkeiten. Dies entspricht in etwa nur der Sicherheit eines 6-stelligen Passworts, wenn Groß- und Kleinbuchstaben sowie Dezimalziffern und noch zwei Sonderzeichen – insgesamt 64 Zeichen – verwendet werden, da der Logarithmus von 200 Milliarden zur Basis 64 in etwa 6,3 beträgt. Es ist mit einer *Brute-Force*-Attacke (vollständiges Durchprobieren) bereits auf aktueller Rechentechnik in weniger als einer Minute zu knacken.

Deshalb ist es sinnvoll, den Suchraum durch das Anfügen einer dritten Eingabe, bekannt unter dem Begriff *Salt*, signifikant zu vergrößern. Dieses *Salt* wird bei jeder Pseudonymisierung zufällig erzeugt und dient dann als dritte Eingabe für das Hash-Verfahren. Wählt man z. B. 6 zufällige Großbuchstaben als *Salt*, so wird der Suchraum 26^6 mal so groß und man erhält somit in etwa 300 Mio. als Multiplikator der Anzahl der Möglichkeiten. Damit steigt die Sicherheit gegen De-Pseudonymisierung auf in etwa eine Passwortlänge von 11, was als derzeit ausreichend angesehen wird. Das Knacken würde mit aktueller Rechentechnik mehrere Jahre dauern.

Zusammenfassend sind also meine Empfehlungen die folgenden:

1) Nicht die Staatskanzlei, sondern der Landesdatenschutzbeauftragte sollte das zu verwendende Hash-Verfahren festlegen.
2) Das Verfahren ist regelmäßig dem algorithmischen und technischen Fortschritt anzupassen.
3) Als dritte Eingabe für das Hash-Verfahren neben Geburtsdatum und Personalnummer ist ein mindestens 6-stelliges, zufällig generiertes *Salt* zu verwenden.

Nur so kann nach meiner Darlegung eine angemessene Pseudonymisierung sichergestellt werden.

….....Prof. Dirk Müller.....
Landesvorsitzender des vhw sachsen
Dresden, am 27.06.2019