Begrüßung und Vorstellung

Vortragender:

Benjamin Raschke, Informatiker, ULD

Zuständigkeit: Analyse und Bewertung von Konzepten und Techniken zur Umsetzung datenschutzrechtlicher Anforderungen in den Bereichen „Data protection by design" und Data protection by default"

Kontakt:

ULD35@datenschutzzentrum.de
+49 431 988-1395

Übersicht des Vortrags

Wir beginnen mit einem kurzen Überblick über die rechtlichen Grundlagen und was wir unter „Data Protection by Design" und „Data Protection by Default" verstehen.

Wir diskutieren Maßnahmen und Werkzeuge – insbesondere für die Software-Entwicklung, an die sich der „by Default"-Ansatz richtet.

Im Anschluss präsentiere ich Überlegungen und Beispiele für den „by Default"-Ansatz und wir diskutieren verschiedene Szenarien.

Im Vortrag haben wir genug Zeit für Diskussionen und gerne auch Erfahrungsberichte aus dem Publikum.

Begriffsbestimmung

Der Begriff „Privacy by design" wurde von Ann Cavoukian geprägt. In den 90ern prägte sie das Konzept, für das sie sieben Grundprinzipien aufstellte (darunter „Privacy by Default"), und auf ihre Arbeit wird heute noch häufig Bezug genommen:

https://www.ipc.on.ca/wp-content/uploads/Resources/7foundationalprinciples.pdf

Die unterschiedlichen Denkweisen des Schutzes der Privatsphäre (privacy) bzw. des Datenschutzes (data protection) führen in den meisten Fällen zu den gleichen Ergebnissen. Die erste Denkweise eines „Privatsphäre-Schutzes" ist eher im US-amerikanischen Raum vertreten, während in Europa eher der „Datenschutz" vertreten wird.

Im Folgenden nutzen wir hauptsächlich den Begriff „Data Protection".

Die rechtliche Grundlage
in der Datenschutz-Grundverordnung (DS-GVO)

EU-Datenschutzgrundverordnung (deutsch)
EU-General Data Protection Regulation (englisch)

- Wichtig: Die Deutsche Übersetzung ist mangelhaft – bspw. beim Titel des Artikels 25
    - deutsch:
      Artikel 25: Datenschutz durch Technikgestaltung und durch datenschutzfreundliche Voreinstellungen
    - englisch:
      Article 25: Data protection by design and by default

- Absatz 1 betrifft die Anforderung „Data Protection by Design" und der Satz kann in drei Sinn-Abschnitte unterteilt werden
    - (rot) Zu berücksichtigende Rahmenbedingungen für die Anforderung, d.h. Abschätzung von Kosten, Aufwand und Risiko
    - (blau) Die Anforderung muss in der Planung und in der Durchführung berücksichtigt werden
    - (grün) Es müssen organisatorische und technische Maßnahmen ergriffen werden.

Die rechtliche Grundlage
in der Datenschutz-Grundverordnung (DS-GVO)

- Wichtig: Auch hier gibt es einen Übersetzungsfehler – das einschränkende Wort „grundsätzlich" gibt es nur in der deutschen Übersetzung

- In Absatz 2 wird die Anforderung „Data Protection by Default" beschrieben, ebenfalls unterschieden in drei Sinnabschnitte
  - (schwarz) Durch die Voreinstellungen muss ein möglichst hohes Datenschutz-Niveau gewährleistet werden
  - (dunkelrot) Dies gilt für alle Aspekte der Datenverarbeitung.
  - (grün) Es müssen technische und organisatorische Maßnahmen ergriffen werden.

Die rechtliche Grundlage
in der Datenschutz-Grundverordnung (DS-GVO)

In Zertifizierungsverfahren im Sinne der DS-GVO kann die Erfüllung der Anforderungen „Data Protection by Design“ und „Data Protection by Defaul“ nachgewiesen werden.

Konkret heißt das aber auch, dass der Nachweis dieser Anforderungen nicht ohne weiteres durch die Erfüllung anderer Normen (z.B. ISO, DIN) oder eigenständiger Siegel ohne DS-GVO-Bezug geführt werden kann.

Die Grundsätze des Datenschutzes

Die Datenschutzgrundsätze der DS-GVO findet man in Artikel 5. Sie sind in der Grafik rot geschrieben. Hinzu kommt noch das „Recht auf Löschung" (auch: Recht auf Vergessenwerden) in Artikel 17.

In Fachkreisen wurden in den letzten Jahren die Datenschutz-Gewährleistungsziele entwickelt, die zusammen mit dem Prinzip Datenminimierung ein theoretisches Fundament bilden.

Die Datenschutzgrundsätze der DS-GVO und die Gewährleistungsziele + Datenminimierung lassen sich gut ineinander überführen.

Grundsatzfrage von „Data Protection by Design“:
Wer soll es machen?

Artikel 25, Absatz 1

„Unter Berücksichtigung […] trifft der Verantwortliche […] geeignete technische und organisatorische Maßnahmen […].“

(siehe Folie 4)

Die DS-GVO richtet sich in diesem Punkt klar an die Verantwortlichen, nicht an die Herstellerinnen und Hersteller.

Die Herstellerinnen und Hersteller von Software (die eigentlich „Data Protection by Default” umsetzen müssten) können also nur auf dem Umweg adressiert werden.

Wie sollen „Data Protection by Design" und „Data Protection by Default" denn umgesetzt werden?

Im Erwägungsgrund 78 der DS-GVO für den Artikel 25 werden einige Hinweise zur Umsetzung gegeben:

- Wann?
  (hellblau) Die Anforderungen „Data Protection by Design" und „Data Protection by Default" müssen immer mit bedacht werden: Bei der Auswahl von Software, der Entwicklung, der Gestaltung und auch bei der Nutzung. Explizit wird erwähnt, dass man diese Anforderungen schon in öffentlichen Ausschreibungen berücksichtigen kann.
- Was?
  (dunkelrot) Sofern personenbezogene Daten verarbeitet werden – das heißt in allen Anwendungen, Diensten und Produkten.
- Wer?
  (dunkelblau) Neben den Verantwortlichen werden auch die Herstellenden erwähnt, die besonderen Einfluss auf die Entwicklung und Gestaltung haben (grün).

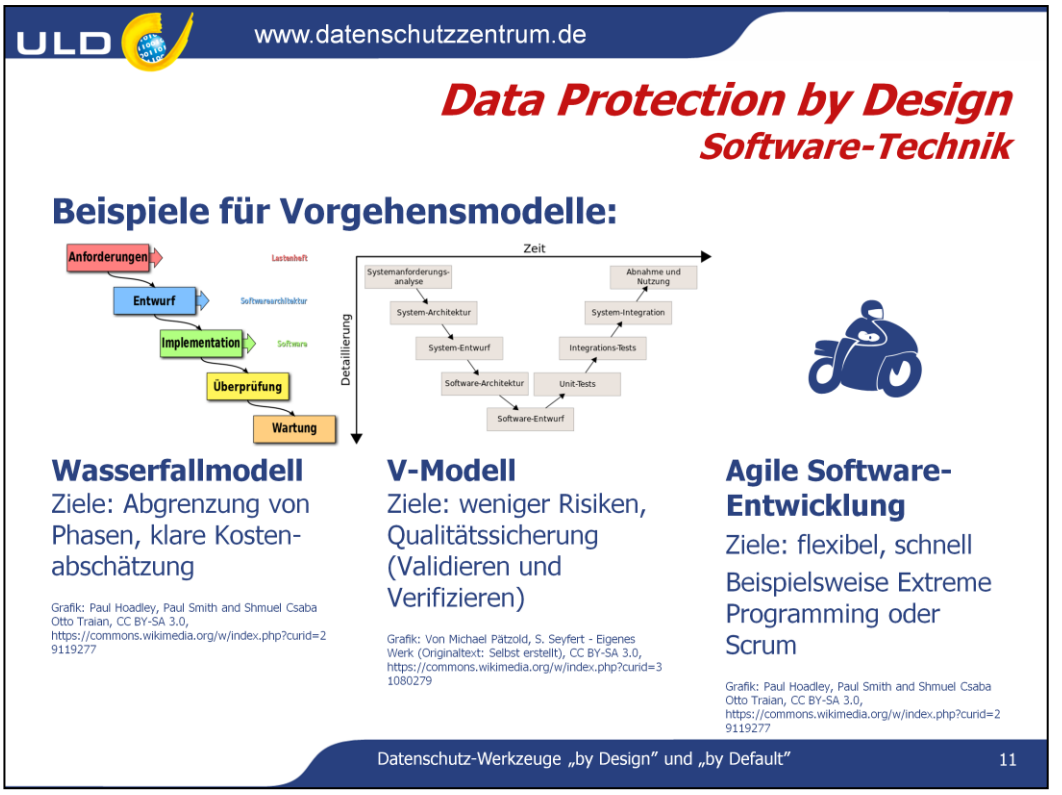
Eine rechtliche Verpflichtung, bei öffentlichen Ausschreibungen den Anspruch „Data Protection by Design and by Default" vorzuschreiben gibt es bisher leider noch nicht.

Software-Technik / Software engineering

In der zweiten Hälfte des 20. Jahrhunderts haben einzelne Software-Projekte eine Größe und Komplexität erreicht, dass sie nicht mehr ohne Hilfsmittel geplant und gesteuert werden konnten.

Deshalb hat sich das (Forschungs-)Gebiet der Software-Technik entwickelt, auf dem Methoden und Techniken entwickelt werden, um Software planbar und sicher zu entwickeln und komplexen Anforderungen an Software gerecht zu werden.

Da die Anforderung „Data Protection by Design" eine solch komplexe Anforderung ist, soll in Kürze in die Software-Technik eingeführt werden.



Vorgehensmodelle

Vorgehensmodelle dienen dazu, die Planung und Entwicklung von Software zu strukturieren und dadurch zu verbessern.

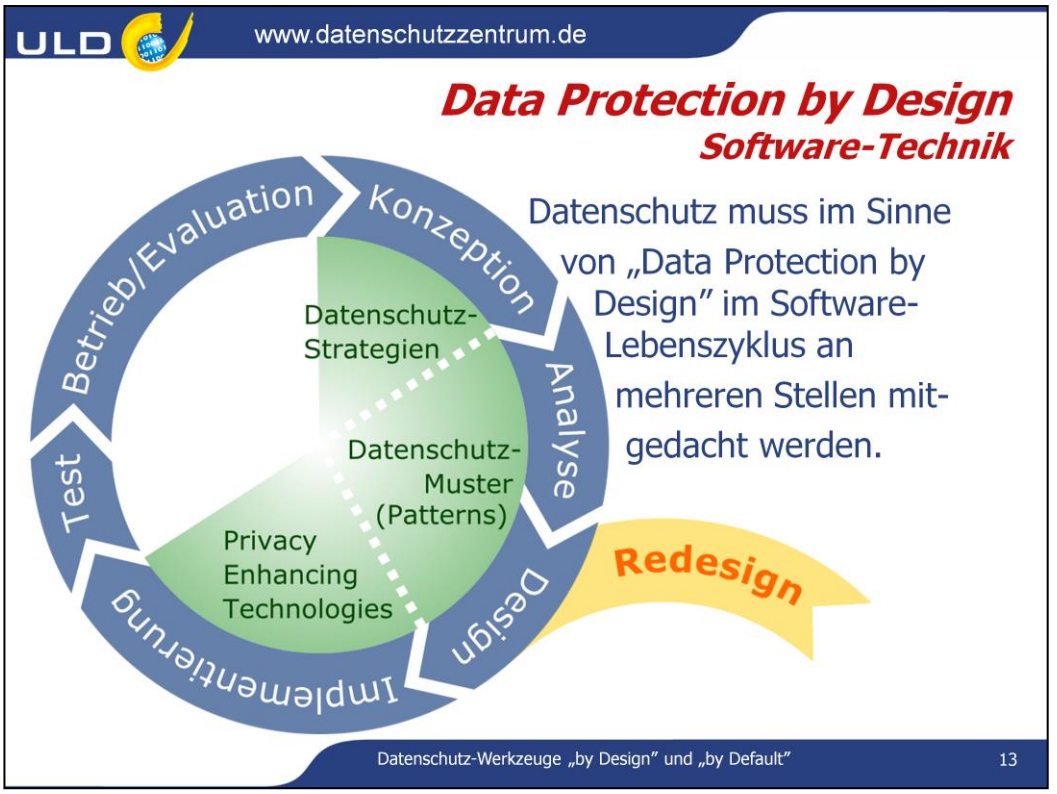
Neben den beispielhaft aufgeführten Vorgehensmodellen gibt es etliche mehr, wie beispielsweise Prototyping, Spiralmodell, iterative oder inkrementelle Vorgehensmodelle. Teilweise können diese auch miteinander kombiniert werden.

Das Wasserfallmodell ist ein relativ frühes Vorgehensmodell und dient im Wesentlichen der Abgrenzung von einzelnen Phasen des Entwicklungsprozesses.

Das V-Modell wurde entwickelt, um noch besser sicherzustellen, dass neben der Entwicklung auch die Qualitätssicherung einbezogen wird.

Zur Gruppe der Agilen Software-Entwicklung gehören verschiedene Vorgehensweisen wie z.B. Extreme Programming oder die Scrum-Methode. Der Agilen Software-Entwicklung ist gemein, dass sie prinzipiell auf Flexibilität und keine konkreten Vorfestlegungen setzen. Bei komplexen Anforderungen wie „Data Protection by Design" muss dabei sichergestellt werden, dass sie den ganzen Prozess über nicht aus dem Blick verloren werden.

Software-Technik - Diskussion



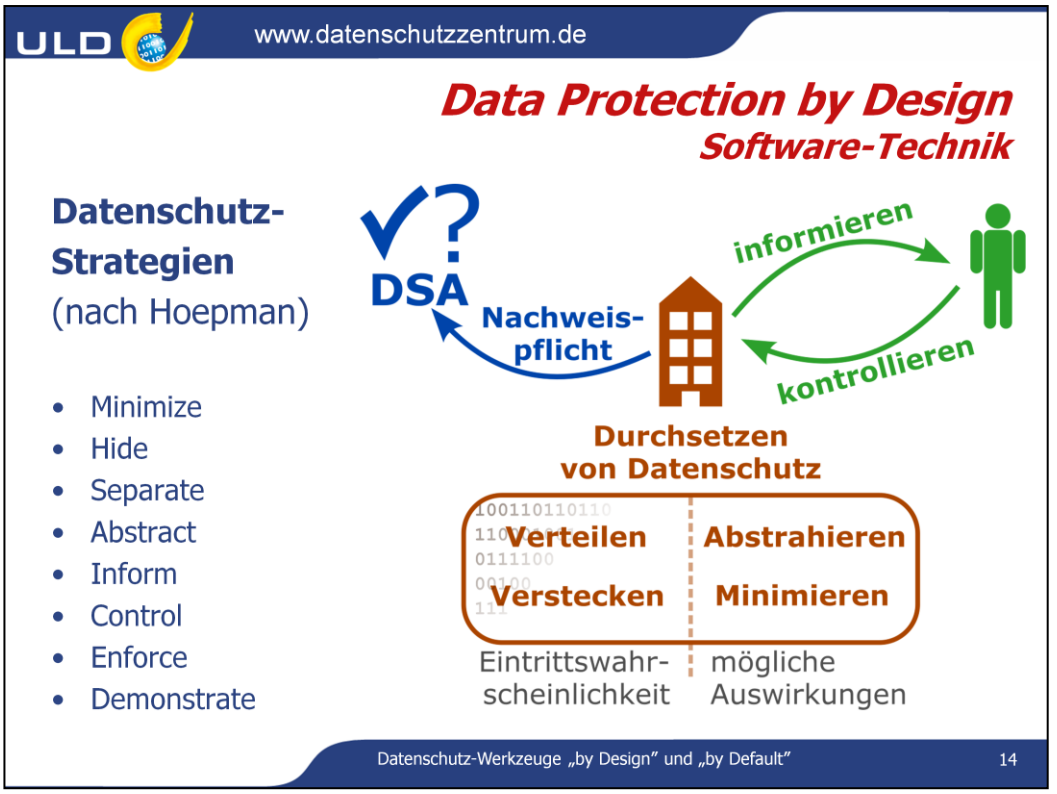
„Data Protection by Design" in der Software-Entwicklung

Eine klassische Sicht auf Software-Entwicklung ist die (zyklische) Aufteilung nach Phasen. Die folgenden Überlegungen können auch auf verschiedene Entwurfsstrategien (einfach: Wasserfall, V-Modell; komplexer: Agile Programming, DDD) übertragen werden.

„Data Protection by Design" als begleitenden Prozess der Software-Entwicklung wollen wir hier in drei Phasen unterteilen:

- Datenschutz-Strategien (privacy strategies)
- Datenschutz-Muster (privacy patterns)
- Konkrete Datenschutztechnologien (Privacy Enhancing Technologies)

Die Perspektive eines „Redesigns" darf dabei nicht aus den Augen verloren werden, denn an vielen Stellen kann es effektiver sein, vorhandene Software zu überarbeiten.



Datenschutz-Strategien

Jaap-Henk Hoepman hat acht Datenschutz-Strategien herausgearbeitet:

- Minimieren von Daten, z.B. durch Nichterheben oder Löschen
- Verstecken, z.B. durch Pseudonymisierung, Mischen
- Verteilen von Daten, physisch oder logisch um Verkettbarkeit zu erschweren
- Abstrahieren, z.B. Daten gruppieren oder generalisieren
- Informieren, d.h. Personen über Datenerhebung und –verarbeitung informieren
- Kontrolle, d.h. Recht auf Löschen bzw. Bearbeitung
- Durchsetzen der (rechtlichen) Anforderungen
- Nachweisen der Einhaltung von rechtlichen Vorgaben und der getroffenen Maßnahmen

Die Strategien *Verteilen* und *Verstecken* dienen dazu, die Eintrittswahrscheinlichkeit von Datenschutzverstößen zu verringern, die Strategien Abstrahieren und Minimieren zielen darauf ab, dass die möglichen Auswirkungen gering sind.

Quelle: http://www.cs.ru.nl/~jhh/publications/iwpe-privacy-strategies.pdf

Datenschutz-Muster (privacy patterns)

In der Software-Entwicklung werden immer wieder Entwurfsmuster (design pattern) benutzt, da sie als Lösungs-Schablonen für immer wieder auftauchende Entwicklungsprobleme genutzt werden können.

Bekannte Entwurfsmuster sind

- Model-View-Controller (Trennung von Daten, Ansicht und Verarbeitung)
- Observer (Weitergabe von Änderungen an einem Objekt)
- Factory (Erzeugen eines Objekts durch Aufrufen einer Methode)

Die Datenschutz-Muster folgen den Datenschutz-Strategien und bieten Lösungsansätze für typische Probleme.

Die Sammlung von Mustern auf www.privacypatterns.org ist bei weitem nicht vollständig, bietet aber einen guten ersten Überblick.

Ergänzend kann auch ein Blick auf Anti-Pattern helfen, um überhaupt Datenschutz-Risiken zu erkennen. Nick Doty und Mohit Gupta haben hierzu ein Paper geschrieben: https://cups.cs.cmu.edu/soups/2013/trustbusters2013/Privacy_Design_Patterns-Antipatterns_Doty.pdf

„Verschlüsselung auf Nutzer-Seite“ /
„Encryption with user-managed keys“

Problem: Das Speichern von personenbezogenen Daten soll bei Online-Diensten oft verschlüsselt erfolgen, um den Schaden im Fall eines Datenlecks gering zu halten.

Lösungsidee: Die Verschlüsselung erfolgt bereits auf Seite des Nutzers, sodass nicht das Problem entsteht, dass verschlüsselte Daten und Schlüssel am selben Ort zu finden sind.

Berücksichtigte Strategien: Verteilen, Verstecken

Beispiele: Online-Backups, die vor der Übertragung verschlüsselt werden.

https://privacypatterns.org/patterns/Encryption-user-managed-keys

„Meta-Daten löschen" / „Strip invisible Metadata"

Problem: Bei der Nutzung von digitalen Geräten fallen oft diverse Meta-Daten an, von denen die meisten unbemerkt verarbeitet werden und der Nutzer nur durch besondere Fachkenntnis von der Existenz erfährt.

Lösungsidee: Sofern die Meta-Daten nicht benötigt werden, werden sie gelöscht oder standardmäßig vor Dritten verborgen.

Berücksichtigte Strategien: Minimieren, Verstecken, Kontrolle

Beispiele:

EXIF-Daten von Bild-Dateien (z.B. Datum/Uhrzeit als ein Foto aufgenommen wurde) werden von Twitter beim Hochladen gelöscht und bei Flickr können sie verborgen werden, sodass Dritte sie nicht sehen können.

https://privacypatterns.org/patterns/Strip-invisible-metadata

„Datenschutz-Verletzungen melden“ / „Data Breach Notification Pattern“

Problem: Beispielsweise bei unbefugtem Zugriff auf Daten muss (innerhalb einer gewissen Frist) informiert werden – bspw. das beauftragende Unternehmen, die betroffenen Personen oder eine Aufsichtsbehörde (siehe Artikel 33 DS-GVO).

Lösungsidee: Ein Monitoring-System protokolliert Zugriffe auf Daten und sofern ein Verstoß nicht sofort verhindert aber (nachträglich) erkannt wird, meldet es die Verstöße an die vorab definierten Adressaten.

Berücksichtigte Strategien: Informieren, Nachweispflicht

https://privacypatterns.org/patterns/Data-breach-notification-pattern

Modellierung von Sigrid Gürgens (SIT/Fraunhofer)

„Erhebung von Nutzdaten einschränken“ /
„User data confinement pattern“

Problem: Daten, die bspw. aus Abrechnungsgründen fortwährend von einzelnen Nutzern an eine zentrale Stelle übertragen werden, erhöhen Datenschutz-Risiken.

Lösungsidee: Sofern möglich, sollte die Verarbeitung oder ein Aggregieren von Daten vor Ort passieren. Die Anzahl an Übertragungen wird verringert und auch die Informationen werden weniger aussagekräftig.

Berücksichtigte Strategien: Abstrahieren, Minimieren

Beispiel: Intelligente Wasserzähler: An Stelle der Verbrauchsdaten-Übertragung eines Kunden an eine zentrale Abrechnungsstelle können bereits im Ablesegerät alle Abrechnungen erstellt werden. Dadurch verlassen die sensiblen Daten (Gewohnheiten bei Aufstehen und Zubettgehen, Rückschlüsse auf verwendete Geräte, …) niemals die Wohnung.

https://privacypatterns.org/patterns/User-data-confinement-pattern

Datenschutz-Technologien (Privacy Enhancing Technologies)

Der Begriff „Privacy Enhancing Technologies“ (PET) ist weit verbreitet, weswegen an dieser Stelle „Privacy“ statt „Data Protection“ genutzt wird.

Als PET werden solche konkreten Werkzeuge und Services bezeichnen, die dabei helfen, Datenschutz zu unterstützen.

Zum einen durch einen Einsatz in Software bzw. Hardware, um den Schutz von personenbezogenen Daten zu verbessern.

Zum anderen aber auch zur Unterstützung des Selbstdatenschutzes, indem Nutzerinnen und Nutzern Werkzeuge zugänglich gemacht werden, mit denen sie mehr Kontrolle über ihre persönlichen Daten erhalten.

Datenschutz-Technologien – Beispiele

Beim Begriff PET werden unterschiedliche Ansätze verstanden

- Perspektive der Software-Entwicklung: Konkrete Technologien, die erprobt sind und eingesetzt werden können, z.B.
  - Verschlüsselung (Messenger, HTTPS)
  - Anonymisierung (kryptografische/mathematische Ansätze)
  - Berechnungen auf verschlüsselten Daten – insbesondere auf verschiedenen Rechnern
- Werkzeuge, die den Nutzerinnen und Nutzern die Möglichkeit bieten, mehr Souveränität über ihre Daten zu erlangen (Ad-Blocker, datenschutzfreundliche Suchmaschinen)
  siehe auch: http://cyberlaw.stanford.edu/wiki/index.php/PET

Weiterführende Literatur und Links:

„Readiness Analysis for the Adoption and Evolution of Privacy Enhancing Technologies"

"Privacy Enhancing Technologies: Evolution and State of the Art"

http://app-pets.org/home/

http://cyberlaw.stanford.edu/wiki/index.php/PET

Software-Technik - Diskussion

Data Protection by Default

In der DS-GVO wird gefordert: Datenschutz von Anfang an
(siehe Folie 5)

Ziel ist dabei:

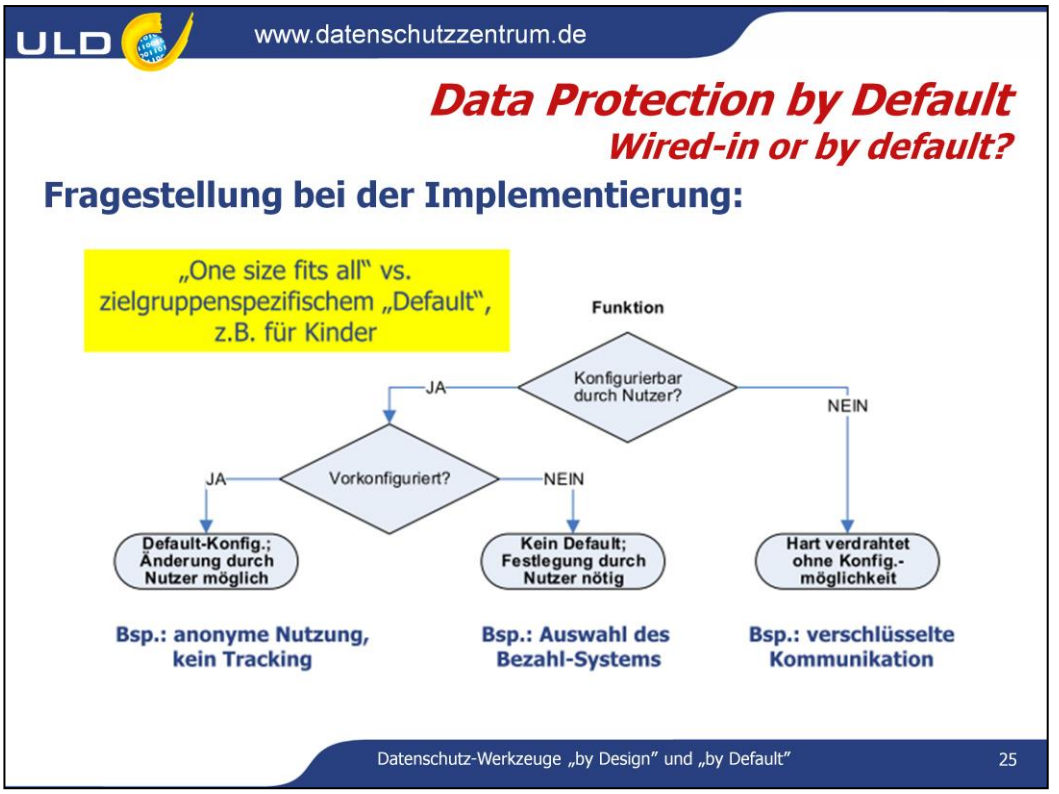
- Möglichst wenig Daten werden erhoben.
- Möglichst wenig Daten werden verarbeitet.
- Daten werden nur so lange gespeichert wie erlaubt / vorgeschrieben.
- Daten sollen grundsätzlich nicht Dritten zugänglich gemacht werden

„by Default" benötigten keinen Nutzer-Eingriff

Die Standard-Einstellung soll ohne Eingreifen „der Person" gelten – hier bezieht sich die DS-GVO auf das Individuum, dessen personenbezogene Daten verarbeitet werden.

Beim Veröffentlichen von Daten darf auch in Sozialen Netzwerken nicht automatisch davon ausgegangen werden, dass diese grundsätzlich für alle Personen bestimmt sind.

Beispielsweise bedeutet dies, dass standardmäßig nur die direkt verbundenen Kontakte („Freunde", „Follower", …) einen neuen Beitrag sehen können.



Was genau ist „by Default"?

Nicht in jedem Fall ist sofort klar, was als eine Standard-Einstellung („Default") betrachtet werden kann.

In der Grafik wird verdeutlicht, dass in unterschiedlichen Anwendungsfällen ein „Default" unterschiedlich eingeschätzt werden kann – oder es keinen Default gibt wie z.B. bei Bezahl-Systemen.

Was genau ist „by Default"? – Beispiel 1

Zwei HTTPS-Beispiele werden vorgestellt, die diese Problematik illustrieren: Soll ein Webserver nur HTTPS-Zugriffe zulassen oder diese nur bevorzugen?

Ggf. würden einzelne Browser/Systeme bei einer strengen HTTPS-Pflicht ausgeschlossen …

Markant ist dabei, dass der Einfluss von Dritten (hier: Suchmaschinen) auf die verbreitete Nutzung von HTTPS am größten eingeschätzt wird. Es gibt z.B. Suchmaschinen, die HTTPS-Seiten bevorzugen und dadurch einen Umstellungsdruck erzeugen.

Was genau ist „by Default"? – Beispiel 2

Mit dem Plugin HTTPS Everywhere gibt es eine Möglichkeit, den Browser – wenn möglich – zum Verwenden des verschlüsselten HTTPS-Datenverkehrs zu zwingen.

Eine Standard-Einstellung könnte daher sein, dass ein Browser mit diesem Plugin installiert werden soll.

Wo sind die Grenzen von „Data Protection by Default“?

Problematisch ist, dass einige Anwendungen (z.B. Soziale Netzwerke) mit einer konsequenten „by Default“-Regel nicht sinnvoll nutzbar sind. Gerade in diesem Beispiel ist schließlich das Erfahren von (veröffentlichten) Informationen Dritter von Interesse.

Eine strenge „by Default“-Regel, die zunächst alle (ungewollten und gewollten) Datenübertragungen ausschließt, allerdings mit einem Klick alle Datentransfers zugelassen werden können, kann aus Datenschutz-Sicht nicht die Lösung sein.

Wo sind die Grenzen von „Data Protection by Default“?

„by Default“ sollte daher im Einzelfall technisch so umgesetzt werden, dass mit unterschiedlichen Profilen bzw. situationsabhängig der „Default“ festgelegt wird – beispielsweise in Sozialen Netzwerken durch Gruppen von Kontakten, die einzeln ausgewählt werden können, bevor ein Beitrag veröffentlicht wird.

Beispielhaft dafür sind Standard-Konfigurationen von Firewalls, bei denen z.B. Internetzugriffe von typischen Browser- oder E-Mail-Anwendungen standardmäßig zugelassen werden.

Einsatzbeispiel: Zertifikatsprotokolle

Um die Authentizität eines (Web-)Servers nachzuweisen, kann sich dieser beim Verbindungsaufbau mit einem Zertifikat ausweisen. Solche Zertifikate können (aus unterschiedlichen Gründen) ablaufen bzw. verworfen werden, sodass eine Prüfung von Zertifikaten notwendig ist.

Früher wurden zu diesem Zweck Sperrlisten der nicht mehr gültigen Zertifikate angeboten, die z.B. von einem Browser heruntergeladen und später Zertifikate damit geprüft werden konnten. Der zeitliche Verzug wurde jedoch als Sicherheitsproblem gesehen.

Mit einem neuen Protokoll wurde festgelegt, dass ein einzelnes Zertifikat bei jedem Zugriff bei einer zentralen Stelle geprüft wird – wodurch jedoch diese zentralen Stellen die Online-Aktivitäten von einzelnen Personen hätten verfolgen können.

Die datenschutzfreundliche Lösung ist nun, dass die Gültigkeit eines Zertifikats vom Zertifizierten selbst nachgewiesen wird, indem er einen aktuellen signierten Gültigkeitsnachweis der Zertifizierungsstelle bereitstellt – so erfährt die Zertifizierungsstelle lediglich, dass ein Webserver kontaktiert wurde, nicht jedoch von wem.

Damit einher geht allerdings wieder ein (kleinerer) zeitlicher Verzug, da zwischen zwei Gültigkeitsnachweisen ein Missbrauch der Zertifikate theoretisch möglich wäre.

Diskussion zum Abschluss

Vielen Dank

ULD

Unabhängiges Landeszentrum für
Datenschutz Schleswig-Holstein