Data Retention beim ISP

Daten-Ermittlung, -Speicherung und -Beauskunftung

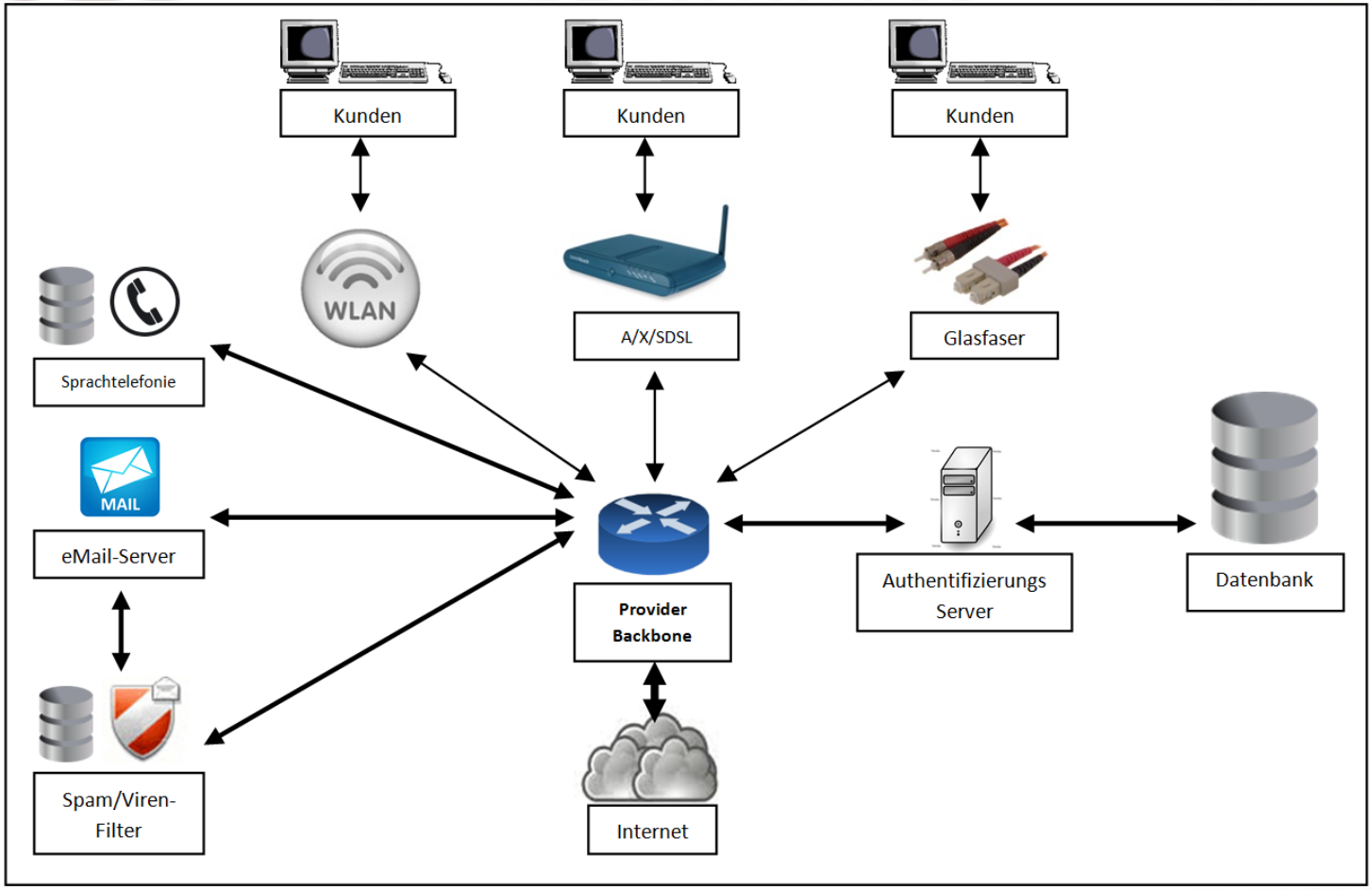
Vortragender:

Dipl.-Ing. Stephan Saalberg

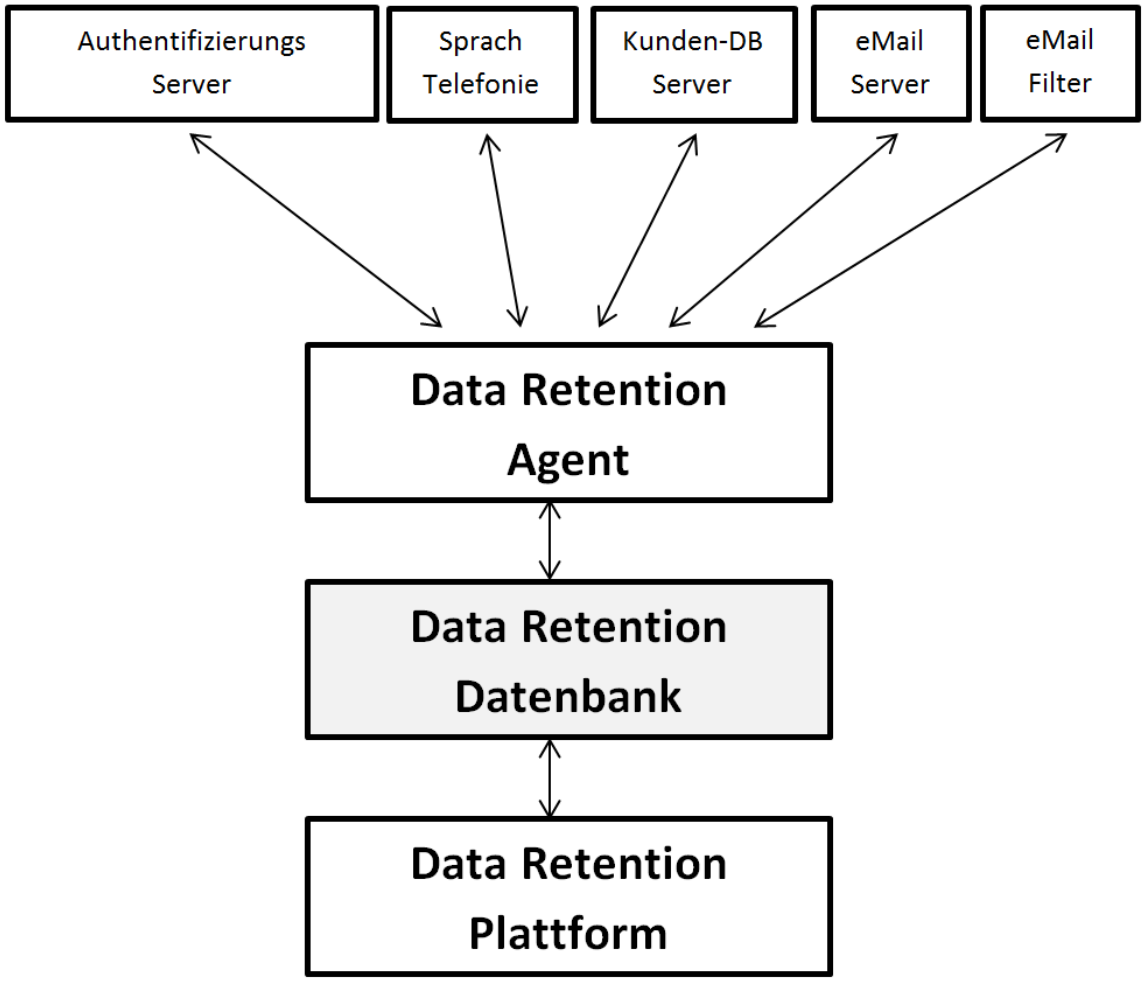
Data Retention: Überblick und Motivation

- **Was bedeutet die Vorratsdaten-Speicherung für die Provider?**
  - Unterschiedliche Infrastrukturen
  - Datenermittlung z.T. problematisch (?)
  - Zusätzliches Personal und Hardware
  - Zusätzliche Kosten und Erstattung?
  - Umsetzungsfristen ab der DSVO ausreichend lang?

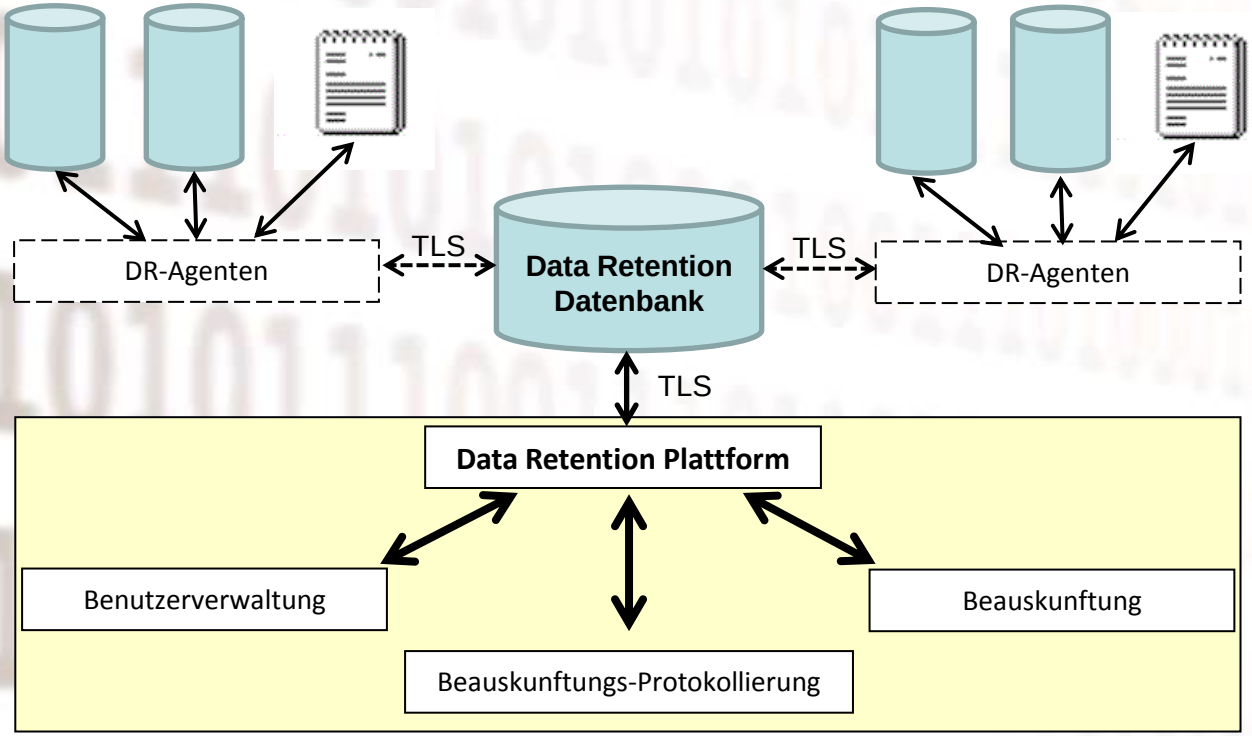
Realisierungskonzept [1]
Die Providerlandschaft



Realisierungskonzept [2]
DR-Agenten-Sicht



Realisierungskonzept [3]
DR-Plattform- / DR-Jobs-Sicht



DR-Jobs (automatisch): Löschung von Daten außerhalb der Aufbewahrungszeit
Überführung von Betriebsdaten zu Vorratsdaten
Kennzeichnung wenn in BD- und VD-Datenbank vorhanden

TLS … Transport Layer Security, ehemals bekannt als SSL (Secure Socket Layer)

Schnittstelle gemäß TKG § 94 (4) - EP 020 [1]

- EP 020 – Ausgabe 3: Aktualisierung Ausgabe 2 (Ende Juni 2011)
  ➔ Übernommen in DSVO Anhang

- Zusammenfassung:
  - Kurzskizzierung der DLS
    - Beschreibung der Protokollierung
    - Übermittlung von Zusatzinformationen zur Beauskunftung
  - Beschreibung der zu erhebenden Datenarten (nächste Folie)
  - Definitionen und Richtlinien als Basis für einheitliche Umsetzungen
  - Beschreibung der Ausgaben inkl. Beispiele

Schnittstelle gemäß TKG § 94 (4) - EP 020 [2]

- Überblick Datenarten:
  - **Internetzugangsdienste (bestehend)**
    z.B. Zuweisung und Entzug von Public IP's (keine NAT-Adressen)
    Gesetzliche Grundlage: § 102a (2) Z 1 - 4 TKG
  - **Öffentliche Telefoniedienste (bestehend)**
    z.B. Telefonnummern und Gesprächszeiten der Teilnehmer
    Gesetzliche Grundlage: § 102a (3) Z 1 – 6
  - **Erstanmeldung (bestehend)**
    z.B. Erstinbetriebnahme von (Prepaid-) Handys bzw. Daten-Modems
    Gesetzliche Grundlage: § 102a (3) Z 6 c
  - **eMail-Verkehrsdaten (neu)**
    z.B. WER hat mit WEM kommuniziert? U.a. Relevante eMail- und IP-Adressen
    Gesetzliche Grundlage: § 102a (4) Z 1 - 4
  - **An-/Abmeldungen am eMail-Server (neu)**
    z.B. mittels Webmail-Oberfläche oder eMail-Clients inkl. IP-Adressen
    Gesetzliche Grundlage: § 102a (4) Z 5

➔ Inklusive die Identitäten der Teilnehmer bei allen Datenarten, sofern ermittelbar!
➔ Ermittlung ausschließlich beim Provider! Keine Deep-Packet-Inspection!

Schnittstelle gemäß TKG § 94 (4) - EP 020 [3]

- Beispiel: Datenformat

Das Datenformat für die Abfrage von Vorratsdaten zu öffentlichen Telefondiensten wird wie folgt festgelegt:

Feldname	Beschreibung	Syntax
Referenz		siehe Kapitel 3.1.3
IndikatorArt	NR, MSIS, ZIEL, IMSI, IMEI, CELL	siehe Kapitel 3.1.4
Indikator	Festnetznummer, MSISDN, Zielrufnummer, IMSI, IMEI, Cell-Id	
IndikatorMSISDN		siehe Kapitel 3.1.6
IndikatorIMSI		
IndikatorIMEI		
IndikatorVorname		siehe Kapitel 3.1.15
IndikatorFamilienname		
IndikatorAdresse		
BetreiberId	diese Information bezieht sich auf den Indikator und ist nur für Mobilfunkbetreiber relevant	siehe Kapitel 3.1.12
CellId	die CellId ist Netzbetreiber-spezifisch	
GeoKoordinaten	das sind die geografischen Koordinaten des Senderstandortes	siehe Kapitel 3.1.11
Zeit	Datum und Uhrzeit nach ISO 8601	siehe Kapitel 3.1.9
Dauer	in Sekunden	Zahl
Ruftyp	Telefonie (T), SMS (S) oder MMS (M)	siehe Kapitel 3.1.7
Richtung	aktiv (A) oder passiv (P)	siehe Kapitel 3.1.8
PartnerMSISDN		siehe Kapitel 3.1.6
PartnerIMSI	IMSI und IMEI werden nur angegeben, wenn sich der Partner im eigenen (Mobilfunk-) Netz befindet.	
PartnerIMEI		
PartnerVorname	Die Stammdaten können nur ermittelt werden, wenn sich der Partner im eigenen Netz befindet.	siehe Kapitel 3.1.15
PartnerFamilienname		
PartnerAdresse		
Anrufumleitung	gibt an, ob es sich um eine Anrufumleitung handelt (JA) oder enthält die Zielrufnummer der Anrufumleitung	siehe Kapitel 3.1.6

Schnittstelle [1]
DR-Agenten / DR-Plattform

- Aufgabenverteilung: DR-Agenten / DR-Plattform
- Gemeinsame Basis: Datenbank (z.B. PostgreSQL)
- Gemeinsames Datenmodell:
  - auf Basis des DSVO Anhangs (EP 020)
  - mit definierten Zugriffsrechten für die DR-Agenten, und –Plattform und Jobs
  - Daten-Formate und -Inhalte übergebener Daten
  - Verschiedene weitere Rahmenbedingungen und Konventionen (Speicherung IP-Adressen, #, n.a., etc.)
  - Ev. Aufteilung in 2 Datenbanken:
    - Betriebsdatenbank und Vorratsdatenbank

Ab jetzt folgt die Datenmodellierung …

Schnittstelle [2]
DR-Agenten / DR-Plattform

Mögliches Datenbank-Tabellen-Modell: ZUGRIFFSRECHTE

	Gruppenrolle Agent				Gruppenrolle Plattform				Gruppenrolle Jobs			
Funktion / Tabellen-Name	SELECT	INSERT	UPDATE	DELETE	SELECT	INSERT	UPDATE	DELETE	SELECT	INSERT	UPDATE	DELETE
IPAdressen		X			X						X	X
TelefonieAktivitaeten		X			X						X	X
Erstaktivierung		X			X						X	X
eMailVerkehrsdaten		X			X						X	X
eMailServerAnAbmeldung		X			X						X	X
Login					X	X	X	X				
DRAgent_log	X	X			(X)							X
Deletion_log					X					X		X
DRPlattform_log					X	X						X
BeauskunftungsProtokoll					X	X						X

Schnittstelle [3]
DR-Agenten / DR-Plattform

Verwendete Datentypen: Beispiel PostgreSQL

Name	Speicher Größe	Beschreibung
bigint	8 Bytes	ganze Zahl mit großer Reichweite -9223372036854775808 bis 9223372036854775807
character varying(n)	4 Bytes	beliebige Zeichenketten, n definiert die max. Länge max. 1 GB lang
inet	12 Bytes	IP-Hosts und –Netzwerke für IPv4 und IPv6 Adressen
timestamp with time zone	8 Bytes	Datum und Zeit 4713 v.u.Z. bis 1465001 u.Z. bis zu 1 Mikrosekunde / 14 Stellen

Schnittstelle [5]
DR-Agenten / DR-Plattform

Tabelle TelefonieAktivitaeten: PostgreSQL Syntax

Feldbezeichnung	Datentyp
id	BIGINT
indikatorart	CHARACTER VARYING(10)
indikator	CHARACTER VARYING(50)
indikatormsisdn	CHARACTER VARYING(40)
indikatorimsi	CHARACTER VARYING(40)
indikatorimei	CHARACTER VARYING(40)
indikatorvorname	CHARACTER VARYING(100)
indikatorfamilienname	CHARACTER VARYING(100)
indikatorakademischergrad	CHARACTER VARYING(100)
indikatoradresse	CHARACTER VARYING(150)
betreiberid	CHARACTER VARYING(50)
cellid	CHARACTER VARYING(50)
geokoordinaten	CHARACTER VARYING(50)
zeit	TIMESTAMPTZ
dauer	BIGINT

Feldbezeichnung	Datentyp
ruftyp	CHARACTER VARYING(1)
richtung	CHARACTER VARYING(1)
partnerindikatorart	CHARACTER VARYING(10)
partnerindikator	CHARACTER VARYING(50)
partnermsisdn	CHARACTER VARYING(40)
partnerimsi	CHARACTER VARYING(40)
partnerimei	CHARACTER VARYING(40)
partnervorname	CHARACTER VARYING(100)
partnerfamilienname	CHARACTER VARYING(100)
partnerakademischergrad	CHARACTER VARYING(100)
partneradresse	CHARACTER VARYING(150)
anrufumleitung	CHARACTER VARYING(40)
betriebsdatum	TIMESTAMPTZ
vorratssdatum	TIMESTAMPTZ

Schnittstelle [7]
DR-Agenten / DR-Plattform

Tabelle eMailVerkehrsdaten: PostgreSQL Syntax

Feldbezeichnung	Datentyp
id	BIGINT
senderidentikatorart	CHARACTER VARYING(10)
senderidentikator	CHARACTER VARYING(50)
sendervorname	CHARACTER VARYING(100)
senderfamilienname	CHARACTER VARYING(100)
senderakademischergrad	CHARACTER VARYING(100)
senderadresse	CHARACTER VARYING(150)
empfaengeridentikatorart	CHARACTER VARYING(10)
empfaengeridentikator	CHARACTER VARYING(50)
empfaengervorname	CHARACTER VARYING(100)
empfaengerfamilienname	CHARACTER VARYING(100)
empfaengerakademischergrad	CHARACTER VARYING(100)
empfaengeradresse	CHARACTER VARYING(150)
zeit	TIMESTAMPTZ
gesendetabsender	CHARACTER VARYING(320)
gesendetabsenderip_adresse	INET
gesendetempfaenger	CHARACTER VARYING(320)
empfangabsender	CHARACTER VARYING(320)
empfangabsenderip_adresse	INET
empfangziel	CHARACTER VARYING(320)
betriebsdatum	TIMESTAMPTZ
vorratsdatum	TIMESTAMPTZ

Beispiel: Speicherbedarf
2,5 KB pro eMail
ca. 10.000 eMails/Tag
➔ ca. 25 MB/Tag
➔ ca. 9125 MB/Jahr

Schnittstelle [9]
DR-Agenten / DR-Plattform

Tabelle Logon: PostgreSQL Syntax

Feldbezeichnung	Datentyp
id	BIGINT
benutzername	CHARACTER VARYING(100)
passwort	CHARACTER VARYING(100)
salz	CHARACTER VARYING(20)
erstellt	TIMESTAMPTZ
aktualisiert	TIMESTAMPTZ

Hinweis: Passwort z.B. vor/bei Speicherung verschlüsseln (z.B. SHA1) und verschlüsseltes Abbild + ev. verwendetes “Salz“ speichern
➔ Höhere Sicherheit!!!
erstellt und aktualisiert sind automatische Zeitstempel der Datenbank

Schnittstelle [11]
DR-Agenten / DR-Plattform

Tabelle BeauskunftungsProtokoll (Vorschlag): PostgreSQL Syntax

Feldbezeichnung	Datentyp
id	BIGINT
erstellt	TIMESTAMPTZ
transaktionsnummer	CHARACTER VARYING(50)
anforderung	TIMESTAMPTZ
einlangung	TIMESTAMPTZ
betreiber	CHARACTER VARYING(50)
beauskunftungvon	TIMESTAMPTZ
beauskunftungbis	TIMESTAMPTZ
auskunftserteilung	TIMESTAMPTZ
referenz_anordnung	CHARACTER VARYING(30)
aktenzahl	CHARACTER VARYING(30)
datenart	CHARACTER VARYING(50)
identikatorart	CHARACTER VARYING(30)
identikator	CHARACTER VARYING(100)
beauskunftung_betriebsdaten	CHARACTER VARYING(500)
beauskunftung_vorratsdaten	CHARACTER VARYING(500)
speicherdauer_betriebsdaten	CHARACTER VARYING(100)
speicherdauer_vorratsdaten	CHARACTER VARYING(100)
targetteilnehmer	CHARACTER VARYING(500)
targetteilnehmerkennungart	CHARACTER VARYING(10)
targetteilnehmerkennung	CHARACTER VARYING(50)
benutzer1	CHARACTER VARYING(50)
benutzer2	CHARACTER VARYING(50)

➔ Oder Stammdaten des Teilnehmers als einzelne Tabellen-Felder

Schnittstelle [12]
DR-Agenten / DR-Plattform

- DR-Jobs:
  - Überführen von Daten aus Betriebsdatenbank in Vorratsdatenbank
  - Aktualisierung von zusätzlichen Feldern, sofern Betriebsdaten bereits in Vorratsdatenbank gespeichert wurden
  - Entfernung von Daten (auch Protokoll-Daten) außerhalb der Speicherfristen

- Hinweise zu bestehendem Tabellen-Vorschlag
  - (+) Vollständige Tabellen und Datensätze
  - (+) leicht abfragbar ohne Tabellen-Joins
  - (+) weniger Fehleranfällig
  - (-) nicht in 3. bzw. BC-Normalform

- Alternativen???
  - Aufbau einer Tabellen-Struktur in 3. bzw. BC-Normalform
  - Mehr Tabellen und Fremdschlüsselattribute erforderlich
    - Komplexer: Tabellen-Joins bei Abfrage erforderlich (!)
    - Versionierung bei Stamm-Daten erforderlich (gültig von/bis)

Mögliche Sicherheitskonzepte und Überwachung

- Sicherheit auf Datenbank-Ebene
  - Konfiguration von TLS, Login- und Gruppen-Rollen
- Sicherheit auf Netzwerk-Ebene
  - Konfiguration von Firewalls, VPN, …
- Sicherheit auf DR-Anwendungsebene
  - Anmeldung mittels 4-Augen Prinzip, SHA256
- Physikalische und organisatorische Sicherheit
  - Server in Rack versperrt, Zutrittskontrolle, …
- Überwachung mittels geeigneter Monitoring Tools
  - Nagios, The Dude, DR-Agent, etc.
    - Benachrichtigung bei Ausfall bzw. nicht Erreichbarkeit (SMS, eMail)

Anfrage/Beauskunftungs-Vorgang beim Provider

1) Beauskunftungsanfrage trifft ein

- Prüfung (Annahme/Ablehnung) und Weiterleitung an zuständiges DR-Personal

2) Erfassung der Anfrage (manuell)

- Anmeldung an DR-Plattform
- Konfiguration der Beauskunftungs-Anfrage in der DR-Plattform

3) Ausführung der Beauskunftungs-Anfrage ([halb]automatisch)

- Ermittlung angefragte Daten aus DR-Datenbank
- Erstellung Ausgabe- und SHA1-Hash-Dateien (4-Augen-Prinzip)
- Erstellung der Statistik-Protokollierung für die DLS
- Protokollierung Anfrage- und Beauskunftungsergebnis (intern!)

4) Übermittlung der Ergebnis-Dateien (manuell/automatisiert)

- Per HTTPS-Upload und Web-Browser (HTLM5-fähig) an DLS (gezippt)
- Per Soap-Schnittstelle an DLS
- Zusatzinformationen als Klartext in DLS erfassbar

Beauskunftungs-Protokollierung

- Grundsätzlich DLS-Aufgabe, aber Provider muss auch Daten liefern
  - Zur Kontrolle des Datenschutzes und zur Gewährleistung der Datensicherheit an die Datenschutzkommission und den Datenschutzrat
  - Zur Berichterstattung an die Europäische Kommission und an den Nationalrat bzw. an den Bundesminister für Justiz, etc.
- Inhalt:
  Keine beauskunfteten Daten selbst, nur WER hat WANN, WORÜBER angefragt und über WEN wurde WIEVIEL (Anzahl der Zeilen) beauskunftet
- Protokollierung über Beauskunftungs-Protokoll-Anfragen erfolgen allerdings nur intern bzw. zur Prüfung
- Siehe Vorschlag Tabelle BeauskunftungsProtokoll und DLS-Spezifikation

Aufbau Beauskunftungs-Ausgabe-Dateien [1]

Ausgabe-/BeauskunftungsProtokoll-Dateien

- Dateiname (Ausgabe-Dateien): Referenz_<NR>.csv
- Dateiname (BeauskunftungsProtokoll-Dateien): ???
- CSV-Dateiformat mit Feld-Separator: Komma (Hexadezimal 2C)
- Zeichensatz: UTF-8 (RFC 3629)
- Zeilenende: CR LF (Hexadezimal 0D 0A)
- Datenfelder (Header + Daten) durch doppelte Anführungszeichen (Hexadezimal 22) begrenzen, außer # (Hexadezimal 23) und n.a.
- Erste Zeile: Header
- Ab zweiter Zeile: Daten

Prüfsummen-Dateien:

- Dateiname: Referenz_<NR>.csv.sha1
- Enthält SHA1-Hash (Zeichenkette) zur Prüfung für Empfänger

Aufbau Beauskunftungs Ausgabe-Dateien [2]

Beispiel: eMailVerkehrsdaten-Abfrage

Anforderung: Auskunft gemäß § 76a (2) Z 4 StPO nach der E-Mail Adresse **max@example.com**
➔ Beispiel für empfangene E-Mail

Ausgabe-Datei Beispiel:
Dateiname: **200003.csv**

"Referenz","IndikatorArt","Indikator","TeilnehmerkennungArt",
"Teilnehmerkennung","Zeit","GesendetAbsender","GesendetAbsenderIP_Adresse",
"GesendetEmpfaenger","EmpfangAbsender","EmpfangZiel","EmpfangIP_Adresse" CRLF
"200003","MAIL","max@example.com","KENN","mustermannmax1234",
"2010-01-12T21:23:12+01",#,#,#,"mona@example.com","max@example.com",
"192.0.2.10" CRLF

… nicht relevante Daten, da hier ein empfangenes eMail beauskunftet wurde, kein gesendetes eMail

SHA1-Hash Beispiel:
Datei: **200003.csv.sha1**

SHA1-Hash: "68ac906495480a3404beee4874ed853a037a7a8f"

Entwicklungs-Tipp: Hibernate

- Web: http://www.hibernate.org/
- Portierte Version für .net: nHibernate Web: http://nhforge.org
- Hauptaufgabe: Object-Relational Mapping (ORM)
  - Speicherung von gewöhnlichen Objekten mit Attributen und Methoden in Datenbanken
  - Umkehrung: Erzeugung von Objekten aus gespeicherten Datenbank-Datensätzen
- Datenbankunabhängigkeit
- Mechanismen zur Kompatibilität mit mehrerer Datenbanken
  - Oracle, Microsoft SQL Server, PostGreSQL, MySQL, DB2, SAP DB, etc.
- Abfrage-Statements werden nicht explizit in SQL programmiert, sondern von Hibernate in Abhängigkeit vom verwendeten SQL-Dialekt der Datenbank generiert

Probleme/Fragen vor/bei der Implementierung

- Implementierung selbst oder durch externe IT-Dienstleistung? Kosten!?
- Datentypen für IP-Adressen-Speicherung: IPv4 und IPv6
- Mögliche Problem Datentypen-Mapping bei (n)Hibernate: IPAdressen vs. INET
- Definition der Schnittstelle zwischen DR-Agenten und DR-Plattform
  - Vollständige Tabellen oder normalisiertes Datenmodell (3. NF bzw. BCNF)
    - Ev. Versionierung von Stammdaten!?
  - Verteilung der Aufgaben von DR-Agenten und DR-Plattform: Absprache und Definition im Team!!!
- Datentypen und Speicherung zusätzlicher Daten wie Geokoordinaten, Cell-ID, usw. obwohl beim Provider (derzeit) nicht vorhanden?
  - Empfehlung: Andenken und Umsetzen, soweit als möglich!
- Uhrzeit-Synchronisation bei ALLEN beteiligten Systemen um Unterschiede bei gelesenen/geschriebenen Zeitstempeln (UTC) vorzubeugen!
- Weitere Probleme bzw. Denkansätze: z.B.
  - Integration/Auslesen von "gewachsenen" IT-Landschaften
  - Ermittlung der Verkehrsdaten aus den bestehenden Provider-Backbone-Systemen
  - eMail-Adressen Mapping zu Kunden
  - Ausfälle der eigenen IT → Probleme bei der Ermittlung von Verkehrsdaten!
  - Protokollierung
  - Konsistenz-Prüfungen
  - Backup's
  - Sicherstellung der Löschung

Kostenersatz für die Provider [1]
Investitionskostenverordnung

Gesamte Rechtsvorschrift für Investitionskostenverordnung, Fassung vom 28.11.2011
http://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=20005982
(neue Fassung wird folgen)

Bemessungsgrundlage für Personal- und Sachaufwendungen der IKVO
1. Anschaffungskosten
2. Einrichtungskosten
3. Netzanpassungskosten
4. Lizenzkosten

Geltendmachung:
- Bei Bundesministerin für Verkehr, Technologie und Transport
- 2-Fache Ausfertigung
- Binnen 3 Monaten
- Alle Kostenbestandteile einzeln auflisten

Kostenbestimmung:
- Bundesministerin für Justiz entscheidet über die Höhe der zu ersetzenden Kosten
- Kostenersatz für den Betreiber bis zu 80% der Bemessungsgrundlage
- im Topf sind 17 Millionen EUR (exkl. UST) vorgesehen
- Erstattung erfolgt ca. 14 Tage nach Bescheid

Weitere erforderliche Maßnahmen beim Betreiber

- **Definition einer Speicher-Policy**
  - WELCHE Daten sind wie lange Betriebsdaten?
  - WANN Daten zu Vorratsdaten?
  - SCHRIFTLICH definieren und dokumentieren!
  - TKG-DSVO §5 Abs 5: Die tatsächliche Speicherdauer von Betriebsdaten sowie allfällige diesbezügliche interne Richtlinien sind gegenüber der Datenschutzkommission in Falle einer Prüfung zu beauskunften (Näheres im Vortrag von Hr. Wolfger)
- **Definition einer Security-Policy**
  - WELCHE Sicherheitsmaßnahmen gibt es beim Betreiber allgemein (und bei der Vorratsdatenspeicherung)
  - SCHRIFTLICH definieren und dokumentieren!
- **Definition einer Backup-Policy**
  - WANN (Zeitintervall) und WIE werden die Backups durchgeführt?
  - WIE erfolgt ein Recovery?
  - SCHRIFTLICH definieren und dokumentieren!
- **Definition von Data Retention Personal**
  - WER zählt zum Data Retention Personal? Ermächtigung von Personen!
  - WELCHE Aufgaben/Befugnisse haben diese Personen?
  - SCHRIFTLICH definieren und dokumentieren!
- **Was könnte noch wichtig sein?** (Liste ev. nicht vollständig!!!)
  - Mehr dazu im nächsten Vortrag …

Danke für die Aufmerksamkeit!

Fragen, Diskussion, nächster Vortrag…

Vortragender: Dipl.-Ing. Stephan Saalberg
eMail: stephan@saalberg.at