



Koordinierungsstelle
für IT-Standards



Freie
Hansestadt
Bremen

String.Latin und die Identifikation von Personen

Eine Standortbestimmung

Frank Steimke | Koordinierungsstelle für IT-Standards (KoSIT)
9. XÖV-Konferenz | 16. September 2016 | Bremen



Namen von Personen sollen in allen automatisiert geführten Registern identisch geführt werden

IT Verfahren müssen Namen so verarbeiten können, wie sie in Personenstandsregistern geführt werden

IT Verfahren müssen den Zeichensatz beherrschen, der für Personenstandsregister erforderlich ist

Alle lateinischen Zeichen



Wo stehen wir

- Die Innenverwaltung verfolgt das Ziel aktiv
 - String.Latin ist seit 2011 verbindlich
 - Für „Name“ ist die Datenstruktur des Personenstandswesens führend
- Der Planungsrat hat String.Latin beschlossen
 - Umstellungsfrist endet 2017
 - Begleitung der Umsetzung: nicht erforderlich
 - Repräsentation des Namens: Gegenstand der Standardisierungsagenda
- Viele offene Fragen
 - Sichere Identifizierung / Altdaten / Phonetik / Rechtsgrundlagen / technische Umsetzung / Transliteration / Neue Version / Schriftfont
 - Neue Version: Was / Wann / Wie / Übergangsregelungen / ...
 - Wir brauchen ein Fachgremium



Weiterer Ablauf

Zeit (ca.)	Thema	Referent
9:00 ... 9:15	Begrüßung, Einführung	Frank Steimke, KoSIT
9:15 ... 10:15	... bei der BA	Frau A. Lorenz, BA
	... Landeshauptstadt München	Frau Dr. M. Jurisch, MUC
	... in der Versicherungswirtschaft	Herr M. Heinemann, GdV
10:15...10:35	Fragen & Lösungsansätze ordnen und zusammenfassen	
11:00...11:15	EU-weite Situation	Herr B. Kappenberg
11:15...11:50	Gemeinsam Diskussion der identifizierten Themen	
11:50...12:00	Zusammenfassung, Abschluss	



Vielen Dank für Ihre Aufmerksamkeit!

Frank Steimke | [kosit \(at\) finanzen.bremen.de](mailto:kosit@finanzen.bremen.de) | www.xoev.de

Einführung von String Latin in der Bundesagentur für Arbeit

Rahmenbedingungen, Handlungsfelder, aktueller Stand und Notwendigkeit übergreifender Koordination



Inhaltsübersicht

❑ Rahmenbedingungen

- ❑ Vorgehen in der BA
- ❑ Systematische Betrachtung unserer IT-Architektur und der daraus abgeleiteten Handlungsfelder und Lösungsansätze
- ❑ Fazit

❑ Fachliche und technische Handlungsfelder als Gegenstand übergreifender Koordinierung

- ❑ Fachliche Konkretisierung des erwünschten Scopes
- ❑ Technische Unterstützung von String Latin (Fonts, abgestimmter Zeichensatz, usw.)
- ❑ Notwendigkeit eines bundesweiten Umstellungsplans und die Erarbeitung behördenübergreifender neuer Schnittstellenverträge

❑ Weiteres Vorgehen

- ❑ Aktueller Stand der Umstellung auf String Latin in der BA
- ❑ Überschneidung mit anderen Vorgaben/Projekten

Inhaltsübersicht

Rahmenbedingungen

- Vorgehen in der BA
- Systematische Betrachtung unserer IT-Architektur und der daraus abgeleiteten Handlungsfelder und Lösungsansätze
- Fazit

Fachliche und technische Handlungsfelder als Gegenstand übergreifender Koordinierung

- Fachliche Konkretisierung des erwünschten Scopes
- Technische Unterstützung von String Latin (Fonts, abgestimmter Zeichensatz, usw.)
- Notwendigkeit eines bundesweiten Umstellungsplans und die Erarbeitung behördenübergreifender neuer Schnittstellenverträge

Weiteres Vorgehen

- Aktueller Stand der Umstellung auf String Latin in der BA
- Überschneidung mit anderen Vorgaben/Projekten

Rahmenbedingungen/Vorgehen in der BA

Rechtliche Grundlagen

- ❑ Jede Person hat das Recht auf eine korrekte Schreibweise des Namens. Artikel 2 des Grundgesetzes schützt das allgemeine Persönlichkeitsrecht, das u. a. auch die korrekte Schreibweise des Namens umfasst.

Standardisierung durch den IT-Planungsrat

- ❑ Im April 2014 hat der IT-Planungsrat die verbindliche Anwendung des Interoperabilitätsstandards „Lateinische Zeichen in UNICODE“ als Mindeststandard beschlossen.
- ❑ Der Standard „Lateinische Zeichen in UNICODE“ wurde im Auftrag des IT-Planungsrats von der Koordinierungsstelle für IT-Standards (KoSIT) herausgegeben.
- ❑ Für IT-Verfahren, die dem bund-länderübergreifenden Datenaustausch oder dem Datenaustausch mit Bürgern und Wirtschaft dienen, wurden folgende Fristen für die Konformität festgelegt:
 - ❑ mit Beschlussfassung – für die IT-Verfahren, die neu aufgebaut oder in wesentlichem Umfang überarbeitet werden
 - ❑ drei Jahre nach Beschlussfassung – für andere IT-Verfahren (April 2017)

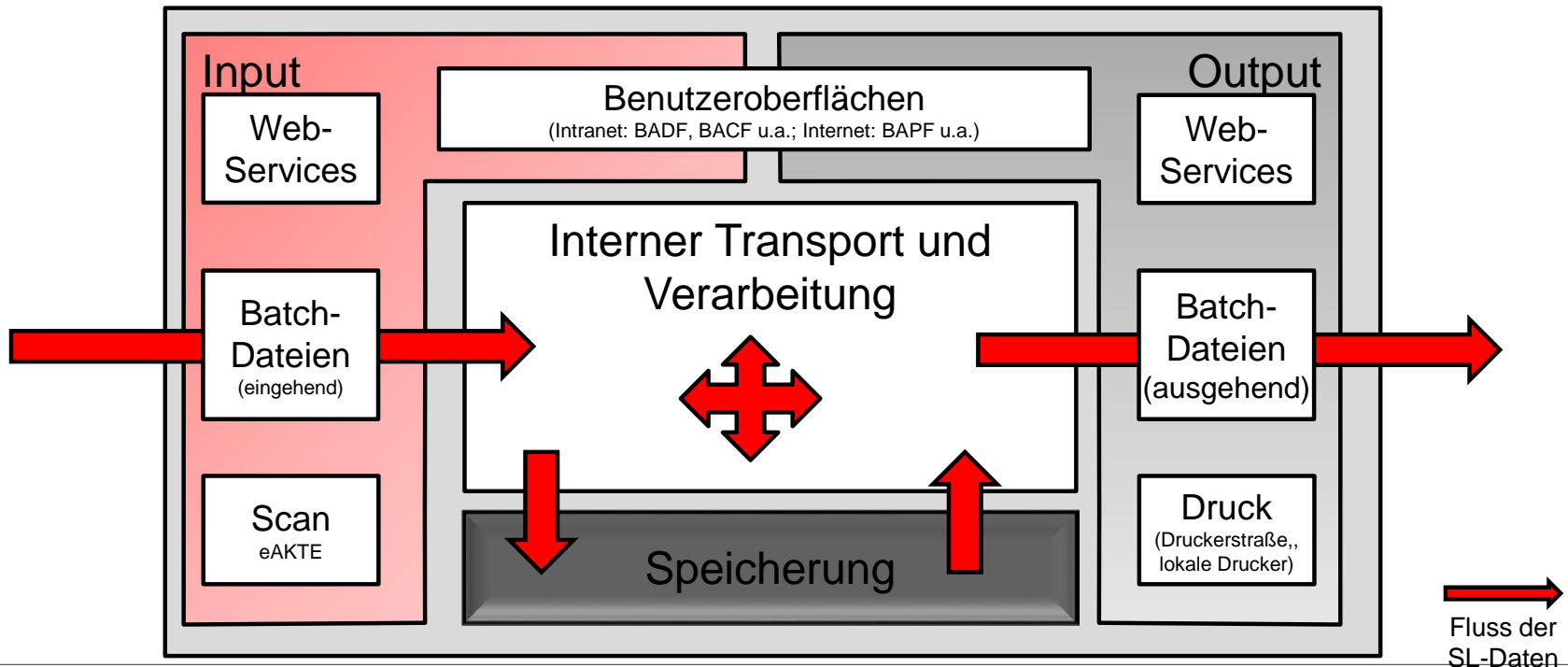
Beschlussfassung in der BA

- ❑ Der Beschluss des IT-Planungsrats ist für die BA rechtlich nicht bindend.
- ❑ Jedoch hat der Führungskreis der IT in der BA in seiner Sitzung am 29.05.2015 beschlossen, den Beschluss des IT-Planungsrats für die BA umzusetzen.
- ❑ Die Umsetzung erfolgt im Rahmen eines Projekts STRING-LATIN.

Systematische Betrachtung unsere IT-Architektur

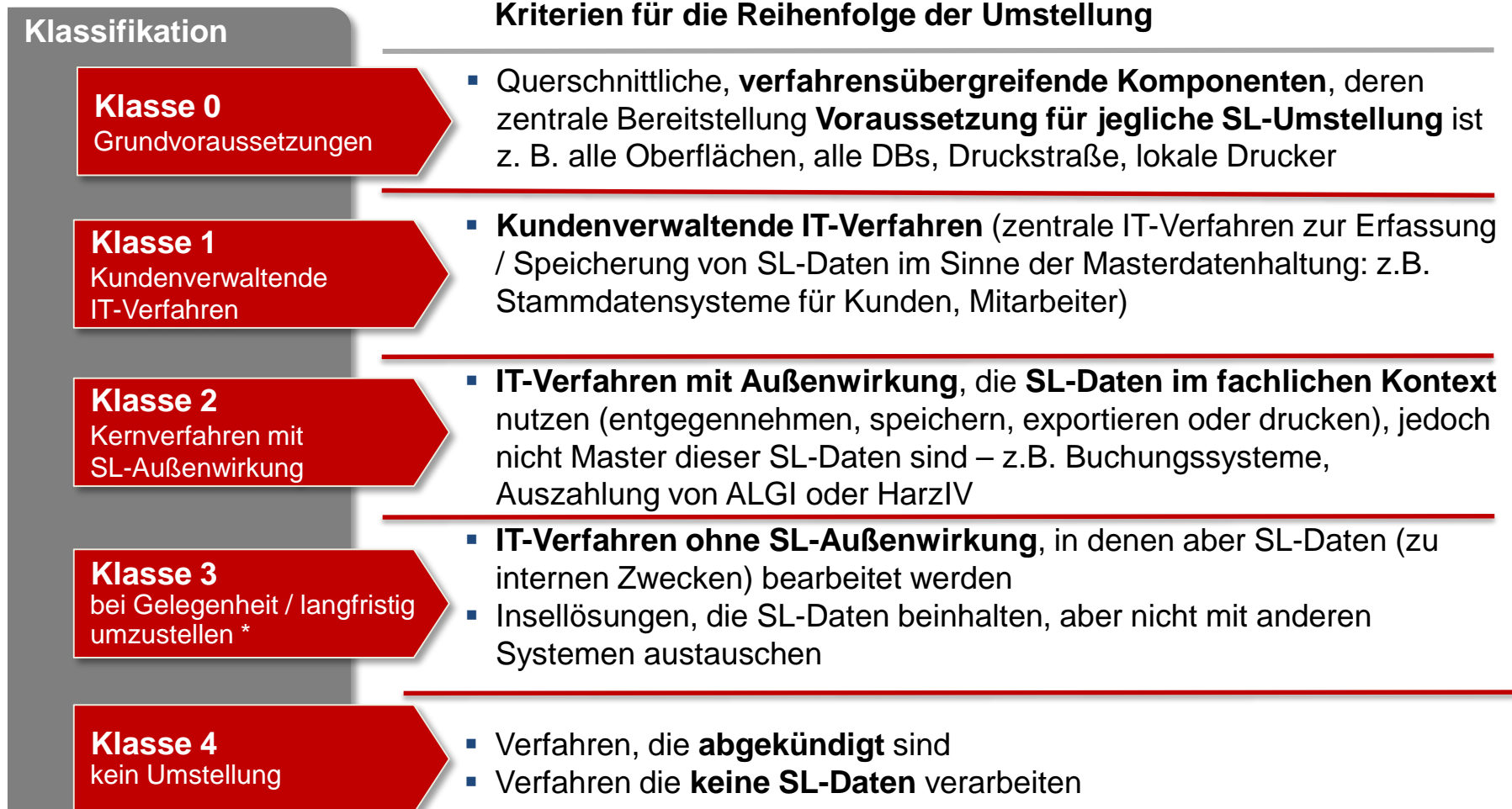
Zur Erreichung der SL-Konformität ist sicherzustellen, dass der Datenfluss der SL-Daten von der Erfassung (Input) über die Verarbeitung und Speicherung bis zum Druck bzw. zur Übermittlung nach außen (Output) durchgängig SL-konform erfolgt. Durchläuft ein Datenfluss nämlich eine Komponente, die SL nicht unterstützt, gehen SL-Zeichen verloren, und die SL-Konformität der Anwendung ist nicht mehr gegeben.

Aus der systematischen Betrachtung der BA-Anwendungslandschaft und der eingesetzten Technologiebausteine werden die SL-relevanten Bereiche für Ein-/Ausgabe, Transport, Verarbeitung und Speicherung der SL-Daten extrahiert. Die identifizierten Bereiche werden soweit untergliedert, bis sie mit konkreten Lösungsbausteinen unterstützt werden können.



Systematische Betrachtung unsere IT-Architektur und die daraus abgeleiteten Handlungsfelder und Lösungsansätze

- Die systematische Betrachtung der IT-Architektur ist bzgl. der identifizierten übergreifenden Komponenten auf 180 Verfahren anzuwenden.



Wichtigste offene Themen als Ergebnis genauerer Betrachtung

Die aktuelle fachliche Grundlage der PDP (herausgegeben vom IT-Planungsrat) ist veraltet (2012) und lässt viel Raum für Interpretationen:

- ❑ Der Beschluss des IT-Planungsrates fordert die Umstellung auf SL-Zeichen ohne konkrete Aussage zum Umfang und Art den betroffenen Daten → **betrifft den Umfang der Umstellung (dadurch Zeit und Kosten)**
- ❑ String.Latin ist kein ISO-Standard. Aus diesem Grund ist Kompatibilität der standardmäßig genutzten Schriftfonts mit String.Latin auf Dauer nicht gewährleistet. → **betrifft das grundsätzliche Vorgehen; ohne einen einheitlichen Schriftfont, kann das primäre Scope des Projektes nicht erreicht werden**
- ❑ Solange externe Schnittstellen-Partner nicht auch umstellen, geht an diesen Schnittstellen der Mehrwert der Umstellung verloren, da die BA an Liefervereinbarungen gebunden ist → **betrifft das Scope des Vorhabens**
- ❑ String Latin als deutschlandweites Standard tritt in Konkurrenz zu dem EU-Projekt EESSI, das aber als EU-Beschluss schon feststeht → **ohne Konsolidierung dieser Vorgaben besteht die Gefahr, dass die Anwendungslandschaft eines deutschen Sozialversicherungsträgers in Zukunft mit drei verschiedenen Zeichensätzen zurechtkommen muss: ISO8859-x, EESSI, String.Latin**

Inhaltsübersicht

□ Rahmenbedingungen

- Vorgehen in der BA
- Systematische Betrachtung unserer IT-Architektur und der daraus abgeleiteten Handlungsfelder und Lösungsansätze

□ Fazit

□ **Fachliche und technische Handlungsfelder als Gegenstand übergreifender Koordinierung**

- Fachliche Konkretisierung des erwünschten Scopes
- Technische Unterstützung von String Latin (Fonts, abgestimmter Zeichensatz, usw.)
- Notwendigkeit eines bundesweiten Umstellungsplans und die Erarbeitung behördenübergreifender neuer Schnittstellenverträge

□ Weiteres Vorgehen

- Aktueller Stand der Umstellung auf String Latin in der BA
- Überschneidung mit anderen Vorgaben/Projekten

Fachliche Konkretisierung des erwünschten Scopes

Der Beschluss des IT-Planungsrates fordert die Umstellung auf SL-Zeichen ohne konkrete Aussage zu den betroffenen Daten. In der Sekundärliteratur der KoSIT werden beispielhaft folgende Daten genannt: **Name und Vorname von natürlichen Personen, Firmennamen, Ortsnamen, Straßennamen, Länderbezeichnungen, Adresszusätze, Produkte, Markenbezeichnungen.**

Es sind grundsätzlich zwei sehr unterschiedliche Perspektiven bzgl. des Scopes möglich:

Maximale Auslegung des Beschlusses des IT-Planungsrates

Prinzipiell müssen alle Attribute eines Verfahrens auf SL-Zeichen umgestellt werden.

Minimale Auslegung anhand der verfügbaren Sekundärliteratur der KoSIT

Einschränkung auf die in der Sekundärliteratur der KoSIT genannten Attribute: Personennamen, Vornamen, Geburtsnamen, Firmennamen, Ortsnamen, Straßennamen, Länderbezeichnungen, Adresszusätze

Daher sind grundsätzlich folgende Lösungskombinationen denkbar:

8 Attribute / alle Verfahren	Alle Attribute / alle Verfahren
8 Attribute / betroffene Verfahren	Alle Attribute / betroffene Verfahren

Fazit: Die Bandbreite zwischen diesen Annahmen ist in den Kosten, Ihren Auswirkungen auf den Aufwand und das Risiko immens. Die Notwendigkeit, die Unterlagen des IT-Planungsrates zu interpretieren, birgt die Gefahr, dass die getroffenen Annahmen im Nachhinein mit einer eventuellen Konkretisierung der Vorgaben des IT-Planungsrates nicht übereinstimmen und aufwendige Nachbesserungen erforderlich werden.

Analyse des einzusetzenden Schriftfonts für die Ein- und Ausgabe und Konsequenzen

String.Latin ist kein ISO-Standard. Aus diesem Grund ist Kompatibilität der standardmäßig genutzten Schriftfonts mit String.Latin auf Dauer nicht gewährleistet.

Die KoSIT stellt derzeit offiziell den Schriftfont UnicodeDOC zur Unterstützung von SL-Zeichen bereit, aber:

- ❑ es gibt momentan keinerlei Weiterentwicklung oder technische Unterstützung für UnicodeDOC
- ❑ eine fachliche Erweiterung der SL-Zeichenmenge von 490 auf 690 Zeichen wird gerade von der KoSIT erarbeitet; allerdings ist nicht sichergestellt, dass UnicodeDOC diese zusätzlichen Zeichen aufnimmt
- ❑ UnicodeDOC mutet altmodisch an und ist für GUIs ungeeignet (Aussage SW-Ergonomie)
- ❑ Folgende Standard-Windows-Schriftarten unterstützen alle aktuellen SL-Zeichen:
 Arial (aktuell überwiegend bei der BA für Massendrucke eingesetzt), Times New Roman, Calibri, Helvetica, Segoe UI und Tahoma sowie die Corporate Design Schriftart der BA (CorpoS)
- ❑ Die zukünftige Unterstützung der geplanten 690 SL-Zeichen dürfte aber nicht gewährleistet werden.

Arial:

UnicodeDOC

Ausrüstungsgegenstände beantragen Sie mit dem Formular „Beschaffung von Ausrüstungsgegenständen (Kauf oder Verleih)“. Sie erhalten das Formular im Jobcenter oder über die Internetseite www.hamburg.de/vordrucke/. Dieses senden Sie an die zentrale Abrechnungsstelle des Bezirksamts Eimsbüttel.

Ausrüstungsgegenstände beantragen Sie mit dem Formular „Beschaffung von Ausrüstungsgegenständen (Kauf oder Verleih)“. Sie erhalten das Formular im Jobcenter oder über die Internetseite www.hamburg.de/vordrucke/.

Fazit: Auf Grund des zu erwartenden immensen Aufwands für die SL-Umstellung und der Unsicherheit bzgl. der Unterstützung zukünftiger SL-Erweiterungen eignet sich weder UnicodeDOC noch eine andere Schriftart für die SL-Umstellung.

Notwendigkeit eines bundesweiten Umstellungsplans und die Erarbeitung behördenübergreifender neuer Schnittstellenverträge

- ❑ Die bundesweite Umstellung auf String Latin bedarf einer übergreifender Abstimmungsreihenfolge mit entsprechender Priorisierung, um den vom IT-Planungsrat erwünschten SL-Datenfluss zu erreichen. Zur Erreichung der SL-Konformität ist sicherzustellen, dass der Datenfluss der SL-Daten von der Erfassung (Input) über die Verarbeitung und Speicherung bis zum Druck bzw. zur Übermittlung nach außen (Output) durchgängig SL-konform erfolgt. Durchläuft ein Datenfluss nämlich eine Komponente, die SL nicht unterstützt, gehen SL-Zeichen verloren, und die SL-Konformität der Anwendung ist nicht mehr gegeben.
- ❑ Formate externer Schnittstellen werden durch Verträge festgelegt und können nur in mühsamen Abstimmungen mit Vertragspartnern geändert werden. Geschieht dies nicht, so wird die Umstellung der BA-Systeme auf String.Latin von begrenztem Nutzen bleiben.
- ❑ Solange externe Schnittstellen-Partner nicht auch umstellen, geht an diesen Schnittstellen der Mehrwert der Umstellung verloren, da die BA an Liefervereinbarungen gebunden ist (z.B. Krankenversicherungen und Rentenversicherung prüfen aktuell Klagen gegen die Umstellung).
- ❑ Das Fehlen behördenübergreifender Vorgaben und Koordination kann dazu führen, dass Schnittstellen-Partner unter anderen, eigenen Annahmen umstellen, so dass Daten fachlich und technisch nicht korrekt ausgetauscht werden können.

Inhaltsübersicht

❑ Rahmenbedingungen

- ❑ Vorgehen in der BA
- ❑ Systematische Betrachtung unserer IT-Architektur und der daraus abgeleiteten Handlungsfelder und Lösungsansätze
- ❑ Fazit

❑ Fachliche und technische Handlungsfelder als Gegenstand übergreifender Koordinierung

- ❑ Fachliche Konkretisierung des erwünschten Scopes
- ❑ Technische Unterstützung von String Latin (Fonts, abgestimmter Zeichensatz, usw.)
- ❑ Notwendigkeit eines bundesweiten Umstellungsplans und die Erarbeitung behördenübergreifender neuer Schnittstellenverträge

❑ Weiteres Vorgehen

- ❑ Aktueller Stand der Umstellung auf String Latin in der BA
- ❑ Überschneidung mit anderen Vorgaben/Projekten

EESSI

Scope und Terminvorgabe

- ❑ EESSI steht für **E**uropean **E**xchange of **S**ocial **S**ecurity **I**nformation.
- ❑ EESSI soll es den Sozialversicherungsträgern („Competent Institutions“, kurz: „CI“) in der EU – 27 Staaten zzgl. Schweiz, Liechtenstein, Island und Norwegen erlauben, Sozialversicherungsdaten auf elektronischem Wege auszutauschen.
- ❑ Zu den Competent Institutions zählen in Deutschland neben der BA auch alle Renten- und Krankenversicherungsträger.
- ❑ EESSI setzt auf sog. Structured Electronic Documents (SED), die über ein Web-basiertes Protokoll (XML-Strukturen) zwischen den CIs transportiert werden. (Zurzeit werden diese Informationen in Form von Papier-Formularen per Fax ausgetauscht).
- ❑ Die EU wird hierfür im Juni 2017 den kompletten Backbone (Netz, Central Service Node und Access Software) zur Verfügung stellen.
- ❑ Die EU-Verordnung Nr. 987/2009 legt fest, dass alle CIs in allen Mitglieds-staaten bis Juni 2019 betriebsbereit an EESSI angebunden sein müssen.

String.Latin vs. EESSI-Zeichensatz

- ❑ Der von EESSI vorgegebene Zeichensatz ist nicht endgültig festgelegt.
- ❑ Die Bemühungen der KoSIT, String.Latin als europäischen Zeichensatz standardisieren zu lassen, waren bisher nicht erfolgreich, aber:
- ❑ Falls die EU einen umfangreicheren Zeichensatz als ISO8859-x fordern sollte, wäre es sinnvoll, String.Latin für EESSI zugrunde zu legen, denn:
 - ❑ String.Latin ist aus einer ausgiebigen sprachwissenschaftlichen Analyse hervorgegangen und somit eine fundierte Basis für die Bearbeitung von Sozialversicherungsdokumenten im europäischen Kontext.
 - ❑ String.Latin eignet sich als einheitlicher Zeichensatz sowohl für den innerdeutschen als auch für den europaweiten Datenaustausch.
 - ❑ Es besteht sonst die Gefahr, dass die Anwendungslandschaft eines deutschen Sozialversicherungsträgers in Zukunft mit drei verschiedenen Zeichensätzen zurechtkommen muss: ISO8859-x, EESSI, String.Latin.

Empfehlung: Die deutschen Sozialversicherungsträger und die KoSIT sollten bei der Technical Commission der EU darauf dringen, dass String.Latin der grundlegende Zeichensatz für EESSI wird.

Aktueller Stand von String Latin in der BA

- ❑ **In der Entscheidung zwischen** Qualifizierter Unterbrechung der SL-Umstellungsaktivitäten und der annahmebasierter Fortführung des Projektes hat die BA folgende Wahl getroffen:

Qualifizierte Unterbrechung der PDP-Aktivitäten, aktive Beteiligung an den bundesweiten Arbeitsgruppen mit anschließender Wiederaufnahme des Umstellungsvorhabens.

Dabei wurde das Risiko, der Notwendigkeit, die Unterlagen des IT-Planungsrates zu interpretieren und die Gefahr, dass die getroffenen Annahmen im Nachhinein mit einer eventuellen Konkretisierung der Vorgaben des IT-Planungsrates nicht übereinstimmen und aufwendige Nachbesserungen erforderlich werden, als zu hoch angesehen.

❑ **Aktuelle Aktivitäten**

- ❑ Es wird gerade die rein technische Umstellung der bestehenden Oracle-Datenbanken auf UTF-8 vorgenommen (ca. 420 DBs mit einer geplanten Lauzeit von 2 Jahren)
- ❑ Durch die DB-Umstellung auf UTF8 wird nur die technische Unicode-Fähigkeit für die Datenbank-Ebene hergestellt, die eigentliche Anwendung ändert sich dadurch nicht, d.h.: die höheren Anwendungsebenen arbeiten weiterhin mit den bislang verfügbaren 256 Zeichen, diese werden nur in der Datenbank anders codiert
- ❑ Durch die Umstellung erhalten die Datenbanken die Möglichkeit, zukünftig grundsätzlich **alle Unicodezeichen zu speichern und zu verarbeiten**

Backup

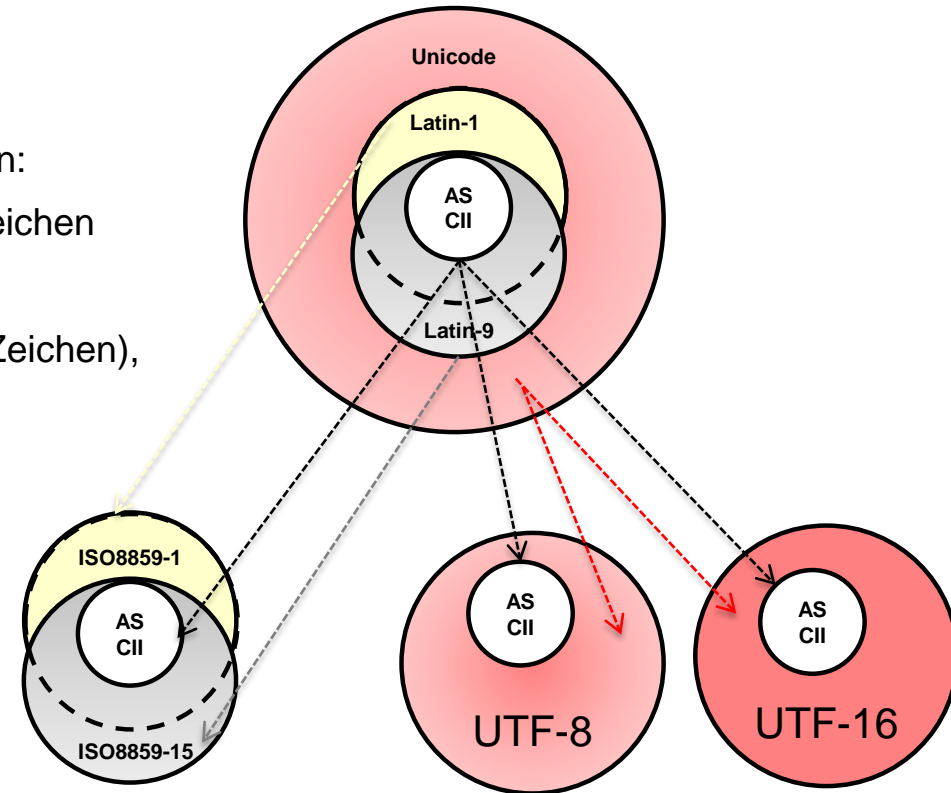
Grundlagen: Zeichensätze und Codierungen

Zeichensatz

- ❑ Menge von zur Verfügung stehenden Zeichen:
Buchstaben, Ziffern, Sonderzeichen, Steuerzeichen
- ❑ In der BA verwendete Zeichensätze:
ASCII (128 Zeichen), Latin-1, Latin-9 (je 256 Zeichen),
Unicode (110.000 Zeichen),
String.Latin (demnächst mit ca. 490 Zeichen)

Codierung

- ❑ Bildet jedes Zeichen eines Zeichensatzes
auf eine Bytefolge ab
- ❑ In der BA verwendete Codierungen:
ISO8859-1, ISO8859-15, UTF-8, UTF-16



- ❑ Weitere Zeichensätze oder Codierungen kommen laut den Recherchen des PDP-Teams String.Latin bei der BA nicht zur Anwendung



Umsetzung des Standards „Lateinische Zeichen in UNICODE“ bei der LHM

Bremen, 16.09.2016

Dr. Marlen Jurisch
Direktorium – Hauptabteilung III – IT-Strategie und IT-Steuerung / IT-Controlling
(STRAC) - Multiprojektmanagement (MPM)



1 Ausgangssituation

- **12 Referate & 5 Eigenbetriebe (EB) – über 800 IT-Verfahren**
 - Wie viele der 800 IT-Verfahren der LHM vom Standard betroffen sind ist heute noch nicht abschätzbar
 - Über die Schnittstellen existieren aktuell keine validierten / ganzheitlichen Informationen
 - Im schlimmsten Fall müssen die meisten IT-Verfahren angepasst werden
- Die LHM hat nur partielle Erfahrungen mit UNICODE-Umstellungen (bspw. Im Kontext von xMeld, xPersonenstand, etc.)
- Die **städtische IT hat verschiedenste Aufgaben** und die Umsetzung des Standards kann nur eine davon sein
 - Erste Aufwandsschätzungen im Jahr 2015 haben ergeben, dass eine UNICODE-Umstellung die IT-Ressourcen der LHM auf einen Zeitraum von 2 Jahren gebunden hätte
 - Bspw. für die Umstellung einer über 30 Jahre gewachsenen Natural-Anwendung auf UNICODE wurden ca. 5000 Personentage veranschlagt
- Die Umsetzungsplanung der UNICODE-Vorhaben muss stets unter **Berücksichtigung der Haushaltslage** erfolgen. Das heißt, unwirtschaftliche Umstellungen auf den Standard müssen vermieden werden



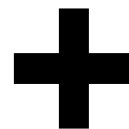
2 Vorgehen

Umsetzungszeitraum bei der LHM: 2016-2020 (ggf. 2020ff)

- ✓ Der Standard soll so **ressourcensparend**, wie möglich umgesetzt werden:
 - Wenn möglich immer in Verbindung mit laufenden / geplanten IT-Vorhaben oder IT-Projekten
 - Nur wenn es zwingend notwendig ist – z.B. wenn der Bund oder der Freistaat Bayern kurzfristig stichtagsbezogene Umstellungen vollziehen, auf welche die LHM reagieren muss oder aufgrund anderer fachlicher Gründe – werden Einzel-IT-Vorhaben für die Umsetzung des Standards aufgesetzt
- ✓ Es werden **keine Systeme mehrfach „angefasst“**
 - Bspw. werden nicht in 2017 für UC Schnittstellen an einem Fachverfahren umgestellt und in 2020 wird dieses dann durch ein neues IT-Verfahren ersetzt = Reduzierung von Doppelaufwänden

Stufe 1 – Analyse:

Analyse aller betroffenen IT-Verfahren
& deren Schnittstellen
(zentral koordiniert durch die
Stabsstelle Multiprojektmanagement)

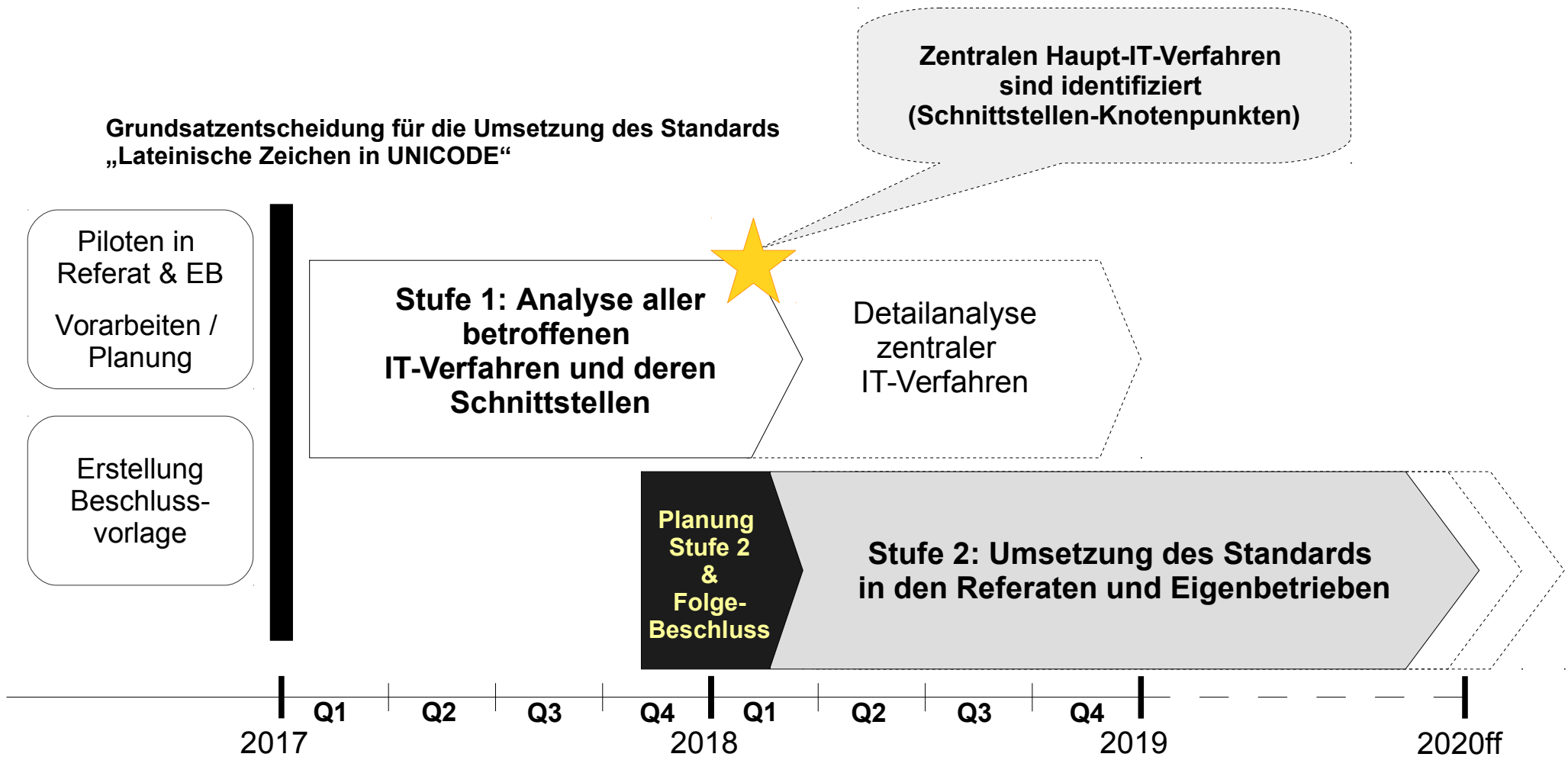


Stufe 2 – Umsetzungen:

Bedarfsorientierte Umsetzung in den
Referaten/EB
+ Unterstützung/Koordination durch
Multiprojektmanagement

3 Zeitplanung

Stufenweise Umstellung auf UNICODE bis 2020 (ggf. ff)





4 Analyse- und Priorisierungskriterien

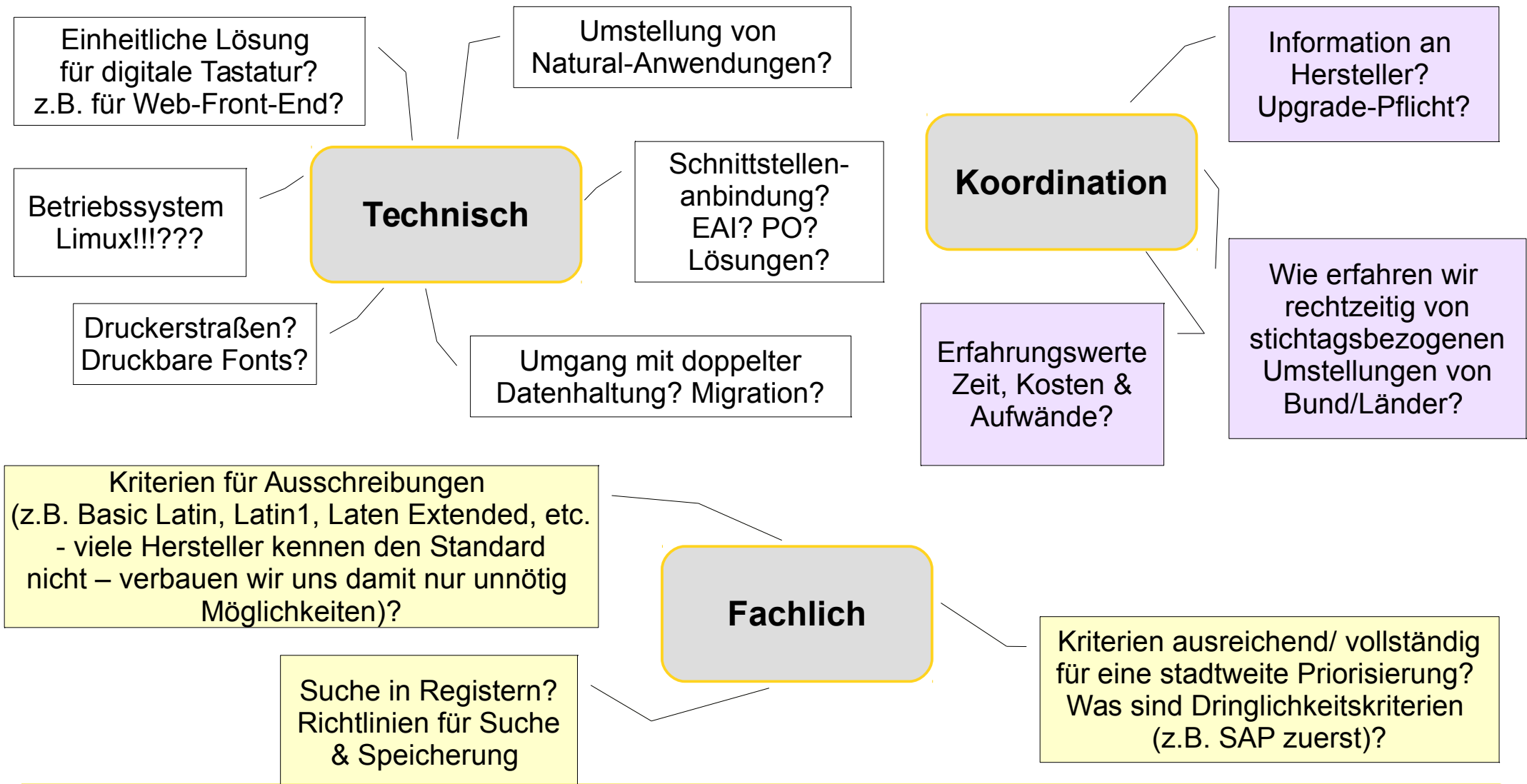
Fachliche und technische Kriterien

Entwicklung	<ul style="list-style-type: none">• Softwaretyp• Entwicklungsjahr	<ul style="list-style-type: none">• Entwickler
Technologie	<ul style="list-style-type: none">• OS Server• OS Client	<ul style="list-style-type: none">• Datenbank• Programmiersprache
Dokumentation	<ul style="list-style-type: none">• Schnittstellen• Datenmodell	<ul style="list-style-type: none">• Architektur• Quellcode
Schnittstellen	<ul style="list-style-type: none">• Intern: Referatsintern ausgehend/eingehend• Intern: zu/aus anderen Referaten	<ul style="list-style-type: none">• Extern: zu/aus Behörden• Extern: zu/aus Personen / Unternehmen a) gedruckt, b) digital und c) Online-Portal
UNICODE	<ul style="list-style-type: none">• UNICODE relevant• Aufwandsschätzung	<ul style="list-style-type: none">• Planung Betrieb / Ablösung• Planung Vorhaben zur Überarbeitung• Dringlichkeit
Kalkulationsgrundlage	<ul style="list-style-type: none">• Anzahl Masken• Anzahl Schnittstellen (intern & extern)	<ul style="list-style-type: none">• Anzahl Datenbankfelder• Anzahl Dokumentenvorlagen / Templates
Ansprechpartner	<ul style="list-style-type: none">• Referat• IT-Dienstleister	<ul style="list-style-type: none">• Hersteller



5 Ausgewählte Fragen in die Runde

Benötigte Rahmenbedingungen





Vielen Danke für Ihre Aufmerksamkeit!

Dr. Marlen Jurisch
Landeshauptstadt München
Direktorium STRAC MPM
Tel. 089-233-82390
marlen.jurisch@muenchen.de

Umstellung auf Lateinische Zeichen in Unicode in der öffentlichen Verwaltung

Darstellung der möglichen Betroffenheit am Beispiel der deutschen Lebensversicherer



Wir über uns

Gesamtverband der Deutschen Versicherungswirtschaft e.V.



Europabüro

51, rue Montoyer
B-1000 Brüssel
Tel.: 0032-2-2 82 47 30
Fax: 0032-2-2 82 47 39
E-Mail: bruessel@gdv.de

Verbandssitz Berlin

Wilhelmstraße 43 / 43G
10117 Berlin
Tel.: 030-2020 5000
Fax: 030-2020 6000
E-Mail: berlin@gdv.de

Büro Hamburg

Frankenstraße 18 A
20097 Hamburg
Tel.: 040-32107-5000
Fax: 040-32107-7030
E-Mail: hamburg@gdv.de

Quelle: gdv.de

16.09.2016

Interessenvertreter der deutschen Versicherer

Der Gesamtverband auf einen Blick



bieten durch **428 Millionen** **Versicherungsverträge** Bürgern und Unternehmen **Risikoschutz** und **Vorsorge**



generieren in der Erstversicherung insgesamt **194 Mrd. EUR** an **Beiträgen** und **zahlen** **157 Mrd. EUR** an **Leistungen**



führende Kapitalgeber für langfristige Investitionen mit Kapitalanlagen von **1.509 Mrd. EUR**



460

Mitgliedsunternehmen* des GDV



Versicherer sind Erwerbsquelle für **210 Tsd.** Beschäftigte. Insgesamt (inkl. Versicherungsvermittlung) arbeiten im Versicherungssektor **529 Tsd.** Menschen



repräsentieren über **95%** des **Erstversicherungsmarktes** in Deutschland

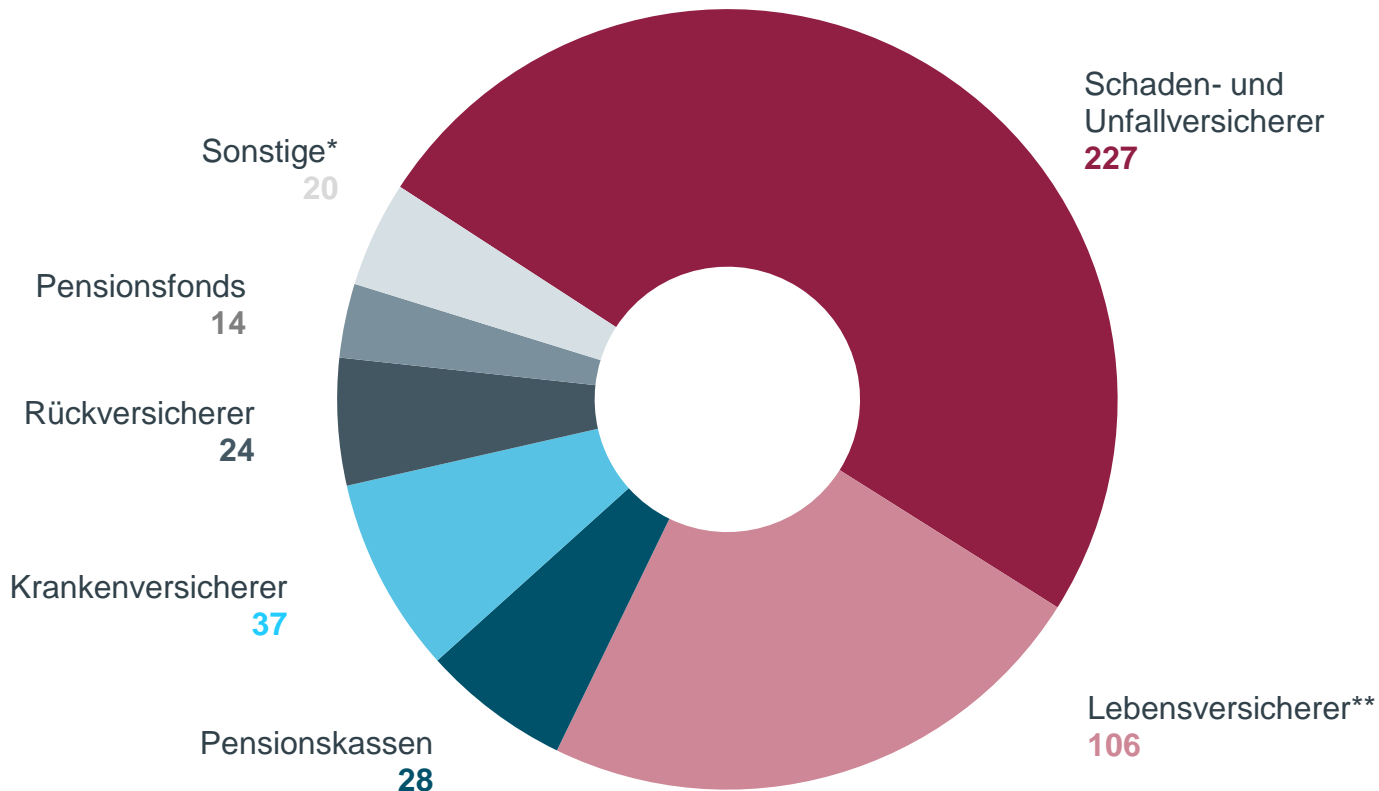
* Zahl gerundet, inkl. PKV-Verband

Quelle: gdv.de

16.09.2016

Anzahl Versicherungsunternehmen im GDV

nach Sparten



* funktionsausgegliederte, konzernangehörige Gesellschaften und Dienstleister (z. B. Vermögensverwaltungs-, Vertriebs- oder Assistancegesellschaften)

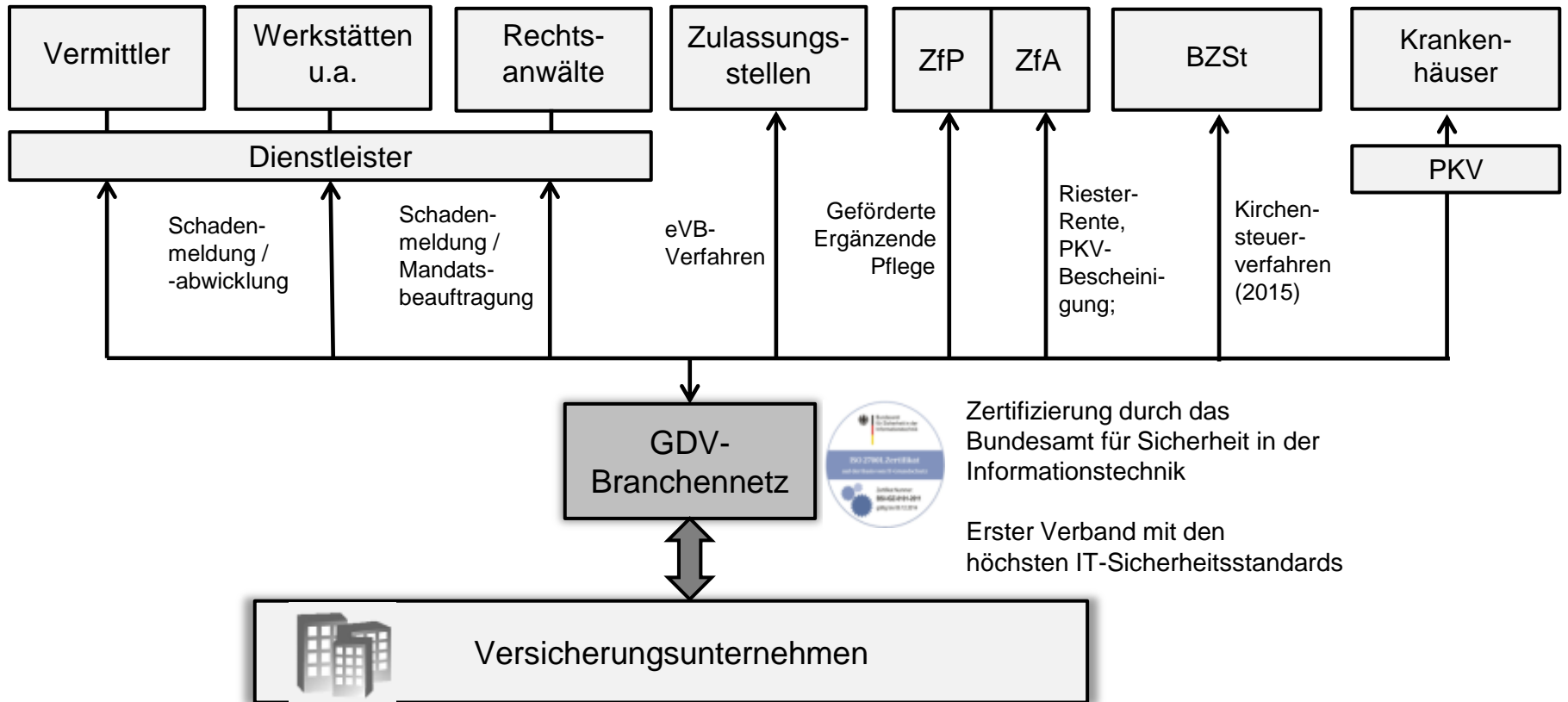
** inkl. Zweigniederlassungen ausländischer Versicherungsunternehmen

Quelle: GDV

16.09.2016

Versicherungsverband als Service-Dienstleister

GDV- Branchennetz (Auszug)



Nutzungszahlen / Transaktionen (2015): rd. 180 Mio. Transaktionen

Betroffenheit von E-Government-Verfahren

hier: Datenkommunikation der Lebensversicherungen mit Behörden

Verfahren	Art	zuständiges Amt / Ministerium	Schnittstellen	Zeichensatz
Riester-Rente (Zulagebeantragung etc.)	Melde-/Bescheidverfahren	ZfA (DRV Bund) / BMF	GRV, Landwirtschafts-, Familienkasse, Finanzämter	ISO 8859-1
Basis-Rente (Rürup)	Melde-/Bescheidverfahren	ZfA (DRV Bund) / BMF	Finanzämter	ISO 8859-15
Steuer-ID- Abfrage (MAV)	Abfrageverfahren	ZfA (DRV Bund) / BMF	Melderegister	ISO 8859-15
Rentenbezugsmitteilungen	Meldeverfahren	ZfA (DRV Bund) / BMF	Finanzämter	ISO 8859-15
Beiträge priv. Kranken-/Pflegevers.	Bescheinigungsverfahren	ZfA (DRV Bund) / BMF	Finanzämter	ISO-8859-15
Private Pflegezusatzversicherung	Melde-/Bescheidverfahren	ZfP (DRV Bund) / BMG	-	ISO 8859-1
Kirchensteuer (seit 2015)	Abfrageverfahren	BZSt / BMF	Finanzämter	Latin Unicode
Steuerdatenaustausch USA (FATCA, seit 2015)	Meldeverfahren	BZSt / BMF	IRS (US-Steuerbehörde)	Latin Unicode
Steuerdatenaustausch weltweit (CRS, ab 2017)	Meldeverfahren	BZSt / BMF	div. weltweite Knotenpunkte	ISO 8859-1
Zahlstellenmeldeverfahren	Meldeverfahren	GKV Spitzenverband / BMAS	Verband gesetzliche Krankenk.	ISO 8859-1

Meldeprozesse am Beispiel der Riester-Rente (ZfA)



Quelle: Anbieter-Kommunikationshandbuch der ZfA

Meldeprozesse am Beispiel der Riester-Rente (ZfA)

	Zentrale Zulagenstelle für Altersvermögen Kommunikationshandbuch Anbieter	Seite: 2-12 von 225 Version: 32.00.00 Stand: 20.07.2016
--	---	---

2.3 Datenaustausch im Rahmen der Zulageermittlung

2.3.1 Zulageantrag

Das Szenario beschreibt das Verfahren bei der Übermittlung eines Zulageantrages vom Anbieter an die ZfA.

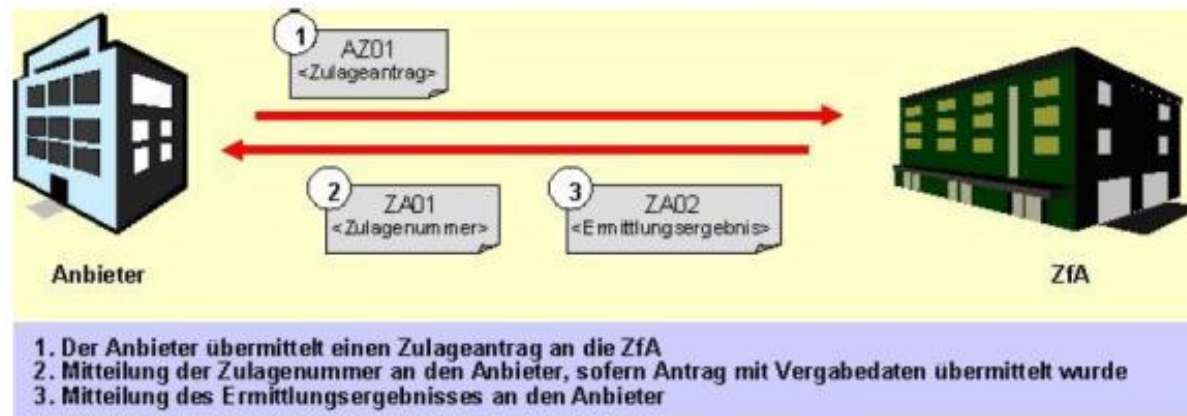


Abbildung 2: Szenario "Zulageantrag"

Meldeprozesse am Beispiel der Riester-Rente (ZfA)

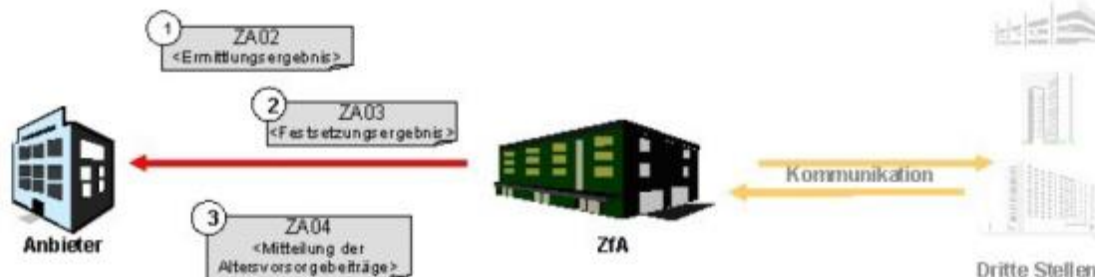
	Zentrale Zulagenstelle für Altersvermögen Kommunikationshandbuch Anbieter	Seite: 2-43 von 225 Version: 32.00.00 Stand: 20.07.2016
--	---	---

2.4 Meldungen aufgrund der Kommunikation mit Dritten Stellen

Im Rahmen Ihrer Aufgaben tritt die ZfA mit unterschiedlichen Stellen in Verbindung. Dies dient zum einen zur Überprüfung der ermittelten Zulage (§ 91 EStG); zum anderen wird die Kommunikation für die Konsolidierung sämtlicher Förderungen des Zulageberechtigten im Rahmen des Altersvermögensgesetzes vorgenommen (§ 10a Absatz 4 EStG).

2.4.1 Änderung des Ermittlungs- bzw. Festsetzungsergebnisses

Aufgrund der Angaben des Zulageantrages (AZ01-Meldung) überprüft die ZfA bei den betroffenen Stellen (z. B. DSRV, Familienkasse) die Nachhaltigkeit der übermittelten Daten. Werden durch die Überprüfungen Abweichungen zu den übermittelten Daten festgestellt, die zu einem veränderten Ermittlungs- oder Festsetzungsergebnis führen, wird die Zulage durch die ZfA entsprechend korrigiert. Der Anbieter erhält daraufhin erneut eine ZA02- bzw. ZA03-Meldung.



Datenfelder Riester-Zulageantrag

Auszug der ZfA-Datensatzbeschreibung, hier: Daten des Anbieters

Übermittlung der Daten des Zulageantrags nach § 89 Abs. 2 Satz 2 EStG und der Mitteilung des Antragstellers nach § 89 Abs. 2 Satz 3 i.V.m. § 89 Abs. 1 Satz 5 EStG sowie der Übermittlung der Daten des Zulageantrags nach § 89 Abs. 3 i.V.m. Abs. 2 Satz 2 EStG und der Mitteilung des Antragstellers nach § 89 Abs. 3 i.V.m. Abs. 2 Satz 3 und Abs. 1 Satz 5 EStG

Baustein 2 bis 4

Lg	Typ	Art	Name	Inhalt / Erläuterung
			ANLEGERDATEN Baustein 2	Dieses Element muss 1 mal enthalten sein
011	n	K	IDNR	Identifikationsnummer (§ 139 b AO)
012	an	m	ZUNR	Inhalt = VSNR / ZUNR in der Form: bbttmmjjAssp
011	n	m	AK-MITGLIEDNR	Mitgliedsnummer der Alterskasse
035	an	M	NNAME	Familienname
035	an	M	VNAME	Vorname
035	an	m	GNAME	Geburtsname
010	an	M	GEBDT	Geburtsdatum in der Form: jhjj-mm-tt
035	an	M	GORT	Geburtsort

ergänzend optionale Bausteine 3 und 9 (Personendaten Ehegatte und Kinder)



Auf dieser Grundlage wird bei erstmaliger Beantragung durch den Versicherer ein Zulagenkonto von der ZfA eröffnet.

Datenfelder Riester-Zulageantrag

Auszug der ZfA-Datensatzbeschreibung, hier: Daten der Zulagenstelle (ZfA)

Mitteilung der Zulagenummer (§ 90 Abs. 1 Satz 3 i.V.m. § 89 Abs. 1 Satz 4 EStG)

ANLEGERDATEN				Dieses Element muss 1 mal enthalten sein
Baustein 2				
012	an	M	ZUNR	Inhalt = VSNR / ZUNR in der Form: bbttmmjjAssp
035	an	M	NNAME	Familienname
035	an	M	VNAME	Vorname
035	an	m	GNAME	Geburtsname
020	an	K	TITEL	Titel
020	an	K	VWORT	Vorsatzwort
020	an	K	NAMZUS	Namenszusatz
035	an	M	GORT	Geburtsort
010	an	M	GEBDT	Geburtsdatum in der Form: jhjj-mm-tt



ZfA-Rückmeldung von Zulagekontonummer und Personendaten

Planungen am Beispiel der ZfA

UNICODE Latin String

- Anbieterinformationstage
- Berlin, 7. Juni 2016 bis 8. Juni 2016

Ilona Durmus / Carsten Dittrich



Planungen am Beispiel der ZfA

3. UNICODE Latin String - Vorgehen



- Der UNICODE Standard umfasst alle gebräuchlichen Alphabete und weist jedem Zeichen eine eindeutige Zahl zu
- Planungen für Umsetzungsstrategie wurden aufgenommen
- Für das sogenannte Encoding der Daten wird standardmäßig UTF8 verwendet
- Neben den technischen Umsetzungen sind auch Anpassungen der Rechtsgrundlagen (z.B. AltvDV, Datensätze) erforderlich

Planungen am Beispiel der ZfA

5. „Latin Unicode“ bei der ZfA Fragestellungen



Frage:

Werden nicht dem Latin Code entsprechend dargestellte Personennamen mit diakritischen Zeichen bei amtlichen Meldungen (z. B. Zulagenantrag, Rentenbezugsmitteilung) der Anbieter / Mitteilungspflichtigen zukünftig Datensatzabweisungen durch die ZfA verursachen?

Antwort:

Ja, langfristig ist davon auszugehen.
Alle nicht dem Datentyp „String.Latin“ (gemäß XÖV Standard „Lateinische Zeichen in Unicode“) entsprechenden Zeichen im Datensatz führen zur Abweisung.
Da eine Stichtagsregelung für alle Beteiligte unwahrscheinlich ist, müssen Klärungen für die Überleitungsphase erfolgen.

Hinweise aus Sicht der Versicherungswirtschaft (I.)

Status

- Einzelne Behörden haben scheinbar unterschiedliche Vorstellungen von der Verbindlichkeit des Beschlusses des IT-Planungsrates. Darüber hinaus besteht z. T. die Annahme, dass für die technische Umstellung auf String Latin zunächst gesetzliche Grundlagen geschaffen werden müssten.
- Die Mehrheit der mit der Versicherungswirtschaft elektronisch kommunizierenden Behörden haben bislang keine verbindlichen Details zur Vorgehensweise und Terminplanung kommuniziert. Es bestehen z. T. große Planungsunsicherheiten seitens der Versicherer, da die Betroffenheit bzw. der Budgetbedarf nicht abgeschätzt werden kann.
- Die Vorgaben des IT-Planungsrates zielen auf die öffentlichen Verwaltung. Etwaig entstehende Erfüllungsaufwände der Wirtschaft für Anpassungen des Encodings müssten u. E. in die Kostenbilanz zum Bürokratieabbau eingehen.

Hinweise aus Sicht der Versicherungswirtschaft (II.)

Best Case- Szenario

- Keine Abweisung elektronischer Datenmeldungen nach Zeichensatz-Umstellung in den Behörden, wenn in Datenfeldern für Personennamen einzelne Zeichen nicht dem Unicode String Latin entsprechen.
Die Identifikation von Personen erfolgt vorrangig mit Hilfe gesetzlich zulässiger Nummern (Sozialversicherungsnummern, Steuer-ID etc.). Ergänzende Abgleiche mit amtlichen Beständen zur eindeutigen Feststellung der Identität der gemeldete Personen werden mit entsprechenden Toleranzen vorgenommen.
- Keine indirekte Pflicht zur Speicherung von Personennamen gemäß Unicode String Latin durch Versicherer, wenn eine entsprechende Datenmeldung oder -rückmeldung zur Person amtlicherseits bereits erfolgt ist. Die weitere Verwendung der nicht dem Latin Code entsprechenden Personennamen kann in Datensätzen auch zukünftig erfolgen.
- Etwaig benötigte Konvertierungsprogramme für ausgehende und eingehende Datenlieferungen werden jeweils seitens der verfahrensführenden Behörde eingesetzt. Damit entfällt die Notwendigkeit, mehrfach höhere Investitionen bei den zahllosen Kommunikationspartnern tätigen zu müssen.

Hinweise aus Sicht der Versicherungswirtschaft (III.)

Empfehlungen im Worst Case- Szenario (Meldungen und Speicherung gemäß Latin Code)

- Koordiniertes Vorgehen durch Vorgabe eines engen terminlichen Umsetzungskorridors und Definition von Übergangsszenarien.
- Einbeziehung aller jeweils prozessbeteiligten Stellen bzw. deren Vertretungen bereits in der Konzeptionsphase und bei der Erstellung des Terminplans der Umsetzung.
- Die Erstellung einer „behördlichen Prozesslandkarte“ ergänzt um eine Betroffenheitsmatrix unter Berücksichtigung aller Wirtschaftszweige.
- Gewährleistung einer Vorlaufzeit von 12 Monaten (von der Veröffentlichung der Detailvorgaben bis zur technischen Produktivsetzung).

Wilhelmstraße 43 / 43 G, D-10117 Berlin
Postfach 08 02 64, D-10002 Berlin
Tel.: +49 30 2020-5000
Fax: +49 30 2020-6000

51, rue Montoyer
B-1000 Brüssel
Tel.: +32 2 28247-30
Fax: +32 2 28247-39

www.gdv.de  [@gdv_de](https://twitter.com/gdv_de)



String.Latin und die Identifikation von Personen

eine Standortbestimmung

9. XÖV-Konferenz
15./16. September 2016, Bremen
WORKSHOP 3

Bernd Kappenberg

bernd.kappenberg@gmx.de

Definitionen

- **Zeichensatz** (character set, character repertoire): Ein Vorrat an Zeichen (Buchstaben, Zahlen, Zeichensetzung, Diakritika...).
- Davon zu unterscheiden ist die **Kodierung**, bei der die Zeichen jeweils durch einen numerischen Wert codiert sind und sich daher für die digitale Speicherung und Übertragung von Daten eignen.
- **Schriftart** (typeface): Grafisches Design eines Zeichensatzes. Es ordnet die abstrakten numerischen Werte bestimmten Formen (Glyphen) und einer oder mehreren **Schriftgrößen** (fonts) zu.

Beispiel

ISO/IEC 10646 / Unicode: 8 bis 32 Bit

UTF-8 ist die sparsamste Kodierung für lateinische Zeichen.

(Es gibt auch UTF-16 / UCS-2, UTF-32 / UCS-4.)

ISO-8859-1

(Latin-1)

8 Bit



ISO-8859-15

(Latin-9)

8 Bit

ASCII:

7 Bit

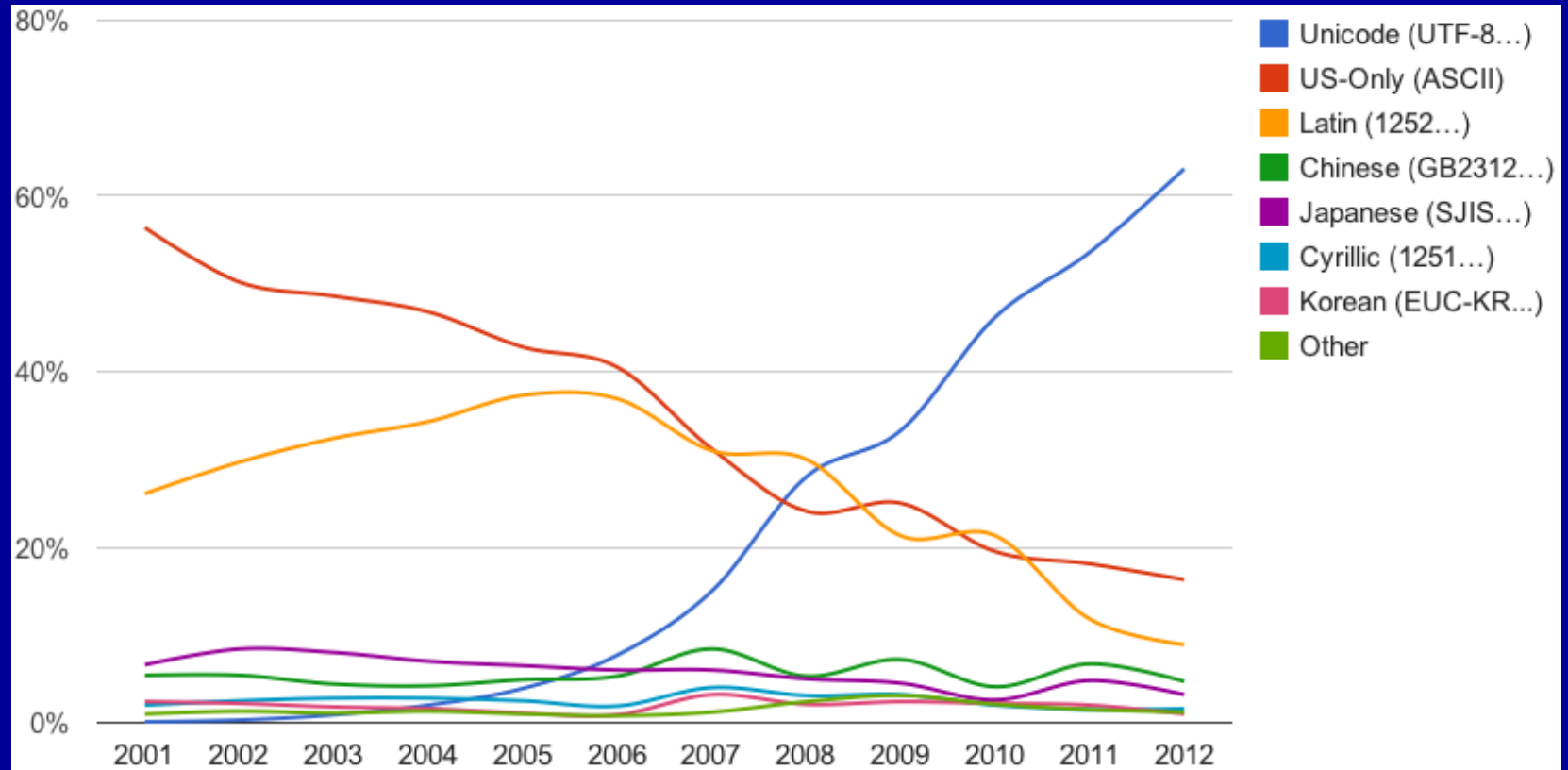
Kritischer Faktor: Zeichensatz

- Benötigt für String.Latin: 481 Zeichen (nächste Version \approx 600)
- Kodierung basiert auf Fernschreiber, nicht EDV!

Kodierung:

- | | | | |
|--------------------|--------|-----------------------|----------|
| • 1870 Baudot-Code | 5 Bit | 58 Zeichen* | zu wenig |
| • 1963 ASCII | 7 Bit | 128 Zeichen | zu wenig |
| • 1980er (diverse) | 8 Bit | 256 Zeichen | zu wenig |
| • 1991 Unicode | 16 Bit | 128,172 Zeichen (9.0) | OK? |

Unicode im Web



European Scripts	African Scripts	South Asian Scripts	Indo-European Scripts
Armenian	Adlam	Ahom	Balinese
Armenian Ligatures	Bamum	Bengali and Assamese	Batak
Caucasian Albanian	Bamum Supplement	Bhaiksuki	Buginese
Cypriot Syllabary	Bassa Vah	Brahmi	Buhaya
Cyrillic	Coptic	Chakma	Hanunó'o
Cyrillic Supplement	Coptic in Greek block	Devanagari	Javanese
Cyrillic Extended-A	Coptic Epact Numbers	Devanagari Extended	Rejang
Cyrillic Extended-B	Egyptian Hieroglyphs (1MB)	Grantha	Sundanese
Cyrillic Extended-C	Ethiopic	Gujarati	Sundanese Supplement
Elbasan	Ethiopic Supplement	Gurmukhi	Tagalog
Georgian	Ethiopic Extended	Kaithi	Tagbanwa
Georgian Supplement	Ethiopic Extended-A	Kannada	East Asian
Glagolitic	Mende Kikakui	Kharoshthi	Bopomofo
Glagolitic Supplement	Meroitic	Khojki	Bopomofo Extended
Gothic	Meroitic Cursive	Khudawadi	CJK
Greek	Meroitic Hieroglyphs	Lepcha	CJK Extension A
Greek Extended	N'Ko	Limbu	CJK Extension B
Ancient Greek Numbers	Osmanya	Mahajani	CJK Extension C
Latin	Tifinagh	Malayalam	CJK Extension D
Basic Latin (ASCII)	Vai	Meetei Mayek	CJK Extension E
Latin-1 Supplement	Middle Eastern Scripts	Meetei Mayek Extensions	(see also: CJK Extension F)
Latin Extended-A	Anatolian Hieroglyphs	Modi	CJK
Latin Extended-B	Arabic	Mro	CJK Extension G
Latin Extended-C	Arabic Supplement	Multani	CJK
Latin Extended-D	Arabic Extended-A	Newa	CJK Extension H
Latin Extended-E	Arabic Presentation Forms-A	Oi Chiki	CJK Extension I
Latin Extended Additional	Arabic Presentation Forms-B	Oriya (Odia)	Indic
Latin Ligatures	Aramaic, Imperial	Saurashtra	Hangeul
Fullwidth Latin Letters	Avestan	Sharada	Hangeul Supplement
IPA Extensions	Carian	Siddham	Hangeul Jamo
Phonetic Extensions	Cuneiform (1MB)	Sinhala	Hangeul Jamo Supplement
Phonetic Extensions Supplement	Cuneiform Numbers and Punctuation	Sinhala Archaic Numbers	Hangeul Jamo Supplement 2
Limousin		Sora Sambar	Hangeul Jamo Supplement 3

Kritischer Faktor: Zeichensatz

- Benötigt für String.Latin: 481 Zeichen (nächste Version \approx 600)
- Kodierung basiert auf Fernschreiber, nicht EDV!

Kodierung:

- | | | | |
|--------------------|-------|-------------|----------|
| • 1870 Baudot-Code | 5 Bit | 58 Zeichen* | zu wenig |
| • 1963 ASCII | 7 Bit | 128 Zeichen | zu wenig |
| • 1980er (diverse) | 8 Bit | 256 Zeichen | zu wenig |

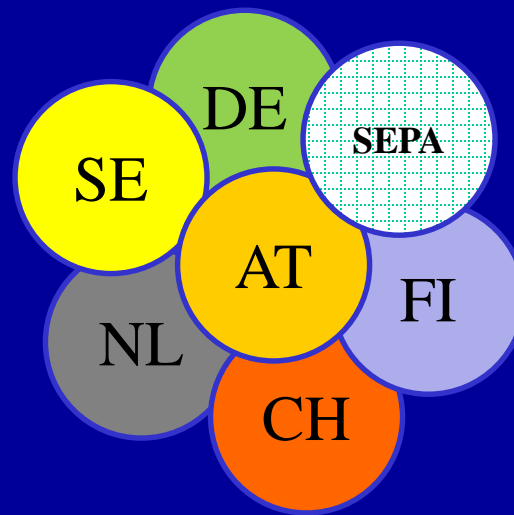
Lücke! = **Unicode**, aber reduziert auf lateinische Zeichen

- | | | | |
|----------------|--------|-----------------------|----------|
| • 1991 Unicode | 16 Bit | 128,172 Zeichen (9.0) | zu viel! |
|----------------|--------|-----------------------|----------|



Kritischer Faktor: Zeichensatz

- Die „Latein-Lücke“ wird bisher gefüllt mit nicht vollständig kompatiblen, meist nationalen Eigenkreationen (immer Unicode-basiert):



Kritischer Faktor: Zeichensatz

Status der Latein-Lückenfüller (1):

- NL (Beschluss Standardschreibweise Personendaten): rechtlich verpflichtend für die allgemeine Verwaltung (1993)
- SE (e-Namen): Richtlinie für die elektronische Verwaltung; bisher *nicht* bindend (2006)
- AT (Handbuch diakritische Zeichen): Richtlinie; *de facto* verpflichtend für Pässe und Personalausweise (Personenstandsgesetz; Meldegesetz nur durch Normverweis) (2006)
- DE (String.Latin): rechtlich verpflichtend für Pässe, Personalausweise, Meldedaten... (2012)

Kritischer Faktor: Zeichensatz

Status der Latein-Lückenfüller (2):

- FI: ...?
- CH (eCH-0011 Datenstandard Personendaten; eCH-0018 XML Best Practices): Unicode (2014) bei Polizei, Diplomaten, aber Zivilstand nur Latin-9, Migration: Latin-1
- UK: nur Aa-Zz (ICAO 9303)

Kritischer Faktor: Zeichensatz

Status der Latein-Lückenfüller (3):

- SEPA-Zeichensatz (alle Diakritika der Länder aus dem Euro-Zahlungsverkehrsraum): nur Vorschlag; bis jetzt ist SEPA (ISO 20022) zwar Unicode, aber *de facto* eingeschränkt auf Aa-Zz.
- Deutsche Kreditwirtschaft ist für Erweiterung!
- Bedenkenträger #1:
 - „Unbekannte Entwicklungskosten für interne Systeme“ (immer noch 7-bit ASCII & EBCDIC??)

12. JANUAR 2011

NEUER PERSONALAUSWEIS

Gute Zeichen, schlechte Zeichen

Von SEBASTIAN MICHAEL BRAUNS



Bringt das System zum Absturz: Der Ausweis von Bundesinnenminister Thomas de Maizière.

Foto: dpa

- Bedenkenträger #2:
 - „Transaktions-Screening (Geldwäsche, Terrorfinanzierung)?“ - kann einfach gelöst werden durch *interne* Repräsentation ohne Diakritika (z.B. gemäß ICAO 9303):



Viñas

Vinxxas

P<ESPVINXXAS<VINXXAS<<MARIA<MERCEDES<<<-

Kritischer Faktor: Zeichensatz

- Frage: Kann man daraus einen einheitlichen „EU-Zeichensatz“ auf Unicode-Basis machen?



Ist eine EU-Vorgabe möglich?

- Vorteil: in allen 28 Mitgliedstaaten einheitlich
- Zuständigkeit der Union gegeben?
 - AEU-Vertrag Art. 18: Verbot von Diskriminierung aus Gründen der Staatsangehörigkeit
 - Grundrechtecharta
 - Art. 4.2 (h), 170, 171: Transeuropäische Netze
- **DIN/CEN-Vorschlag:** “Define the summary repertoire for use in name writing in European public registers, especially in the light of current and potential future legal requirements.” (2011 abgelehnt)

ISA² (Interoperabilitätsstandards für die Administration)

- Vorschlag (Call for proposals): 26/06/2017 bis 28/08/2017
 - Andere EU-Länder mit ins Boot nehmen!
 - Muss mehr politisch sein als 2011! (Stichwort Flüchtlinge, EESSI)
 - Einheitlicher lateinischer Zeichensatz plus Suchalgorithmen, schönere Schriftart, automatische Transkription...
 - Evaluation (Work programme preparation): 09/2017 bis 03/2018
 - Annahme (Europäische Kommission): 03/2018
 - Durchführung (Work programme implementation): 04/2018 bis 12/2020
- http://ec.europa.eu/isa/isa2/index_en.htm

Vielen Dank für Ihre
Aufmerksamkeit



Setting Signs for Europe - Why Diacritics Matter for European Integration (ibidem 2015: ISBN 978-3-8382-0663-9)

www.sonderzeichen.com



Zeichen setzen für Europa - Der Gebrauch europäischer lateinischer Sonderzeichen in der deutschen Öffentlichkeit (ibidem 2012: ISBN 978-3-89821-749-1)



PS: Was ich gern von Ihnen wüsste:
Wie suchen Sie nach Namen?

- Suchalgorithmen
 - englisch: Soundex, Phonix, Metaphone
 - deutsch: Kölner Phonetik, PHONEM, Phonet
 - osteuropäisch: Daitch-Mokotoff
 - international: IPA, SAMPA
 - Hamming-Distanz, N-Gram, Damerau-Levenshtein...
 - Wildcard für Buchstaben mit Diakritika
 - Laufzeitverhalten?
- Identifikationsalgorithmus?