

Wartung und Pflege

Kurzbezeichnung	CR Umstellung Zeichensatz
Zeichen	2010-37
Status-Modellierung	abgenommen
Status-Test	erledigt
Umsetzung	G



OSCI® ist eine registrierte Marke
der Freien Hansestadt Bremen

Umstellung Zeichensatz auf Unicode

Problemstellung

Der DS Meld sieht ab dem 1. November 2011 die Verwendung von xoev:string.latin als Zeichensatz vor.

Lösung

Einführung von xoev:string.lating im Text der Spezifikation und in den Schema-Dateien.

Die mit der Einführung des neuen Datentyps verbundenen Probleme hinsichtlich von Suchalgorithmen sind nicht durch die Spezifikation zu lösen.

Antragsdetails

Antragsteller: OSCI Leitstelle

Erfasst am: 25.02.2009

Bezug: Spezifikation 1.4 – Kapitel 1

Analyse des Änderungsantrags

In den Registern des Personenstandswesens sind Namen von Personen und weitere Angaben buchstabengetreu ohne Änderung oder Übersetzung zu speichern. Lediglich die Transliteration ist zugelassen, sofern die zu Grunde liegende Urkunde in anderen Schriftzeichen als lateinischen geschrieben ist. Somit müssen im Personenstandswesen Namen auch diakritische Zeichen unterstützt werden. Die technische Vorgabe zur Registerführung im Meldewesen (DSMeld) schreibt hingegen einen sehr eingeschränkten Zeichensatz ohne diakritische Zeichen vor. Für die bei den Ausländerbehörden der Länder geführten Register gibt es keine Vorgaben zum Umgang mit diakritischen Zeichen. Für das Ausländerzentralregister ist eine Speicherung ohne diakritische Zeichen vorgeschrieben. Somit werden Namen, die in ihrer originären Darstellung diakritische Zeichen enthalten, in den Registern der Innenverwaltung unterschiedlich repräsentiert. Um Interoperabilitätsprobleme zukünftig zu vermeiden, ist die Festlegung eines einheitlichen Zeichensatzes für die elektronisch geführten Register der Innenverwaltung erforderlich. Es besteht Konsens, dass hierfür ISO/IEC 10646:2003 (Unicode) in der UTF-8 Codierung genutzt werden soll. Der Begriff der lateinischen Zeichen innerhalb Unicode bedarf jedoch einer Präzisierung.

Innerhalb des Meldewesens besteht die grundsätzliche Bereitschaft, die Vorgabe der Registerführung auf Unicode umzustellen. Da dies jedoch mit erheblichem technischen Umstellungsaufwand verbunden ist, sind den Verfahrensherstellern und Betreibern entsprechende Fristen zur Vorbereitung einzuräumen. Die DSMeld Arbeitsgruppe hat sich auf ihrer Sitzung Ende April 2008 einstimmig für eine Umstellung der Registerführung auf Unicode ausgesprochen. Sie hält als frühesten Termin den 1. Mai 2011 für realistisch.

Lösungsvorschlag im Änderungsantrag

1. Für die Führung der Register und für Datenübermittlungen im Bereich der Innenverwaltung soll ein einheitlicher Zeichensatz vorgegeben werden.
2. Die Vorgabe von Unicode bzw. ISO/IEC 10646:2003 ist durch eine abschließende Aufzählung aller zulässiger Zeichen aus Unicode zu präzisieren, um die rechtliche Vorgabe zur Registerführung mit lateinischen Schriftzeichenauch technisch umzusetzen zu können. Diese abschließende Aufzählung der zulässigen lateinischen Zeichen des Unicode ist als Anlage des Standards OSCI-XMeld zu veröffentlichen.
3. Im Zusammenhang mit der geänderten Vorgabe des Zeichensatzes für die Registerführung im Meldewesen sind auch rechtlich-organisatorische Vorgaben zum Übergang auf Unicode entwickeln. So ist zu gewährleisten, dass Personen, die bisher nach dem eingeschränkten Zeichensatz des DSMeld in Melderegistern gespeichert sind, auch dann sicher identifiziert werden, wenn deren Namen zukünftig unter Nutzung der dann zulässigen Zeichen erfasst und gesucht werden.
4. In Abstimmung mit den registerführenden Stellen sowie den Verfahrensherstellern im Meldewesen ist ein geeigneter Zeitpunkt für eine Umstellung des für die Registerführung vorgeschriebenen Zeichensatzes festzulegen.

Bewertung

Bewertungskriterien										Aufwandsschätzung	
Gesetzliche Vorgabe	Fehler	Eindeutigkeit	XÖV-Konformität	Erleichterung MB	Erleichterung AB	XMeld-Prozesse	Wartbarkeit	Fachlicher Aspekt	Detail	Modellierung	0,17
										Test	0,17
										Hersteller	3
										Betroffene Dokumente	
										Spezifikation	✓
										HAs	
										Verwandte CRs	CR 2010-36
1	0	0	1	0	0	0	0	0	0		

Bewertet durch: EG W&P

Bewertet am: EG09-03

Bearbeitung

Wartend auf:

Betrifft	Aktivität	Status Mod	Status Test
Spezifikation	Alle Vorkommen von xs:string ersetzen durch xoev-dt:String.Latin	2010-09-27	5.10.2010
PG Standard	Bedarf nach Vorgaben für Suche mit diakritischen Zeichen kundtun (ToDo 5 aus EG10-07)	2010-09-14	
Testsuite	Testfall 02.002.002.001 erweitert	---	5.10.2010
Testsuite	Erweiterung auf weitere Testfälle	---	16.11.2010
Referenznachrichten	Alle Nachrichten zu 02.002.002.001 / 02.002.010.002 / 02.002.003.001 prüfen	---	9.2.2011
Spezifikation	string durch String.Latin ersetzen	2010-11-08	9.2.2011

Bearbeitet durch: W&P Mod

Bearbeitet am: 2010-09-07

Der Typ xoev-dt:String.Latin wird zukünftig anstelle von xs:string XMeld-weit eingesetzt. Dies entspricht der Bekanntmachung des BMI.

Für die Umsetzung in den Fachverfahren bedarf es dringend Vorgaben für die Suche (innerhalb der Meldebehörde und fachübergreifend (z. B. BfJ)) mit diakritischen Zeichen. Dieser Aspekt wird von Leitstelle in der PG Standard eingebracht.

Bearbeitet durch: W&P Mod

Bearbeitet am: 2010-09-07

Der Bedarf nach Vorgaben für die Suche mit diakritischen Zeichen ist der PG Standard bereits bekannt und wird dort bearbeitet werden. Insofern ist ToDo 5 aus EG10-07 hinfällig.

Bearbeitet durch: AG Test

Bearbeitet am: 2010-10-05

Spezifikation / Schemata

Überprüfung der XSD-Dateien hinsichtlich der Verwendung von xs:string. xs:string wird nur bei Restriction verwendet. Eine Prüfung des Ersetzens von string durch String.Latin wird verzichtet.

Die Spezifikation vom 4.10.2010 verwendet String.Latin und verweist in Anhang H auf den Basisdatentyp im xrepository. In der Einleitung werden in OSCI-XMeld-Schemata die Attribute des Basisnachrichtentyps mit dem falschen (string) Datentype dargestellt. Empfehlung: type Angaben streichen. Die Darstellung im Schema ist ok
Zu 12.4 Datentypen muss die „Regelungen zu leeren Pflicht-Elementen“ ergänzt werden. In D1 Datentypen für Schlüsseltabellen muss der Text und die Angaben zum Type ebenfalls überarbeitet werden. Der Absatz „Wir gehen davon aus ...“ sollte gelöscht werden.

Testsuite

Suchalgorithmen sind noch nicht bekannt. Für die Darstellung von diakritischen Zeichen wird ein Testfall aus dem Bereich Zuzug entsprechend angereichert. Testfall 02.002.002.001 wird erweitert.

Eine Erweiterung der Verwendung von diakritischen Zeichen wird überlegt. Infrage kommen z. B. der/ein Geburtsort oder Straßenangaben in Anschriften. Zunächst soll abgewartet werden, wie der Generator mit diakritischen Zeichen umgeht

Bearbeitet durch: Ullrich Bartels

Bearbeitet am: 2010-11-08

Die am 05.10. angeregten Änderungen wurden in die Spezifikation eingearbeitet.

Bearbeitet durch: AG Test

Bearbeitet am: 2011-11-16

Änderungen geprüft.

Testfälle weiter bearbeitet:

- 02.002.010.002 durch Geburtsort mit diakritischen Zeichen angereichert.
- 02.002.003.001 durch Anschrift mit diakritischen Zeichen angereichert.

Status CR-Testsuite

➔ "Testfälle bearbeitet"

Bearbeitet durch: AG Test

Bearbeitet am: 2011-02-09

Prüfung der Referenznachrichten

Es sollte ein Zeichen verwendet werden, das aus mehr als einem Byte besteht und damit nicht auf ein Zeichen des ISO-Zeichensatzes abbildbar ist (1ED7)

02.002.002.001 Testfall enthält Zeichen 1ED7 (Frönke) - in der Referenznachricht korrekt umgesetzt

02.002.010.002 Geburtsort Umå - in der Referenznachricht korrekt umgesetzt

02.002.003.001 Strasse Credé - in der Referenznachricht korrekt umgesetzt

Zeichen sind in Nachrichten an alle Empfänger von XMeld (DSRV, BZST, XMELDIT)

Kombinierte Zeichen z. B. 006e+0302 (8 Byte) sind nicht in den Testfällen enthalten; unklar ist wie diese Zeichen in den Testfällen dargestellt werden könnten.

Status CR erledigt