

inoxision ARCHIVEsuite

Datumsummschlüsselungen

Stand: 26. August 2013

1 Datumsformate

Das Datum ist in der Regel eine der wichtigsten Informationen auf einem eingescannten Beleg. Voraussetzung für eine archivumfassende Recherche ist jedoch ein einheitliches Datumsformat. Da es keine exakte, (inter)national verbindliche Norm gibt, nach welcher ein Datum erstellt werden muss, sind auf Belegen viele verschiedene Varianten anzutreffen.

Tabelle 1: Datumsformate nach verschiedenen Normen

Norm	Format	Beispiel
DIN 5008	dd.mm.yyyy	16.04.2012
DIN 5008	dd. [Monatsname] yyyy	16. April 2012
ISO 8601	yyymmdd	20120416
ISO 8601	yyyy/mm/dd	2012/04/16
ÖNORM, ISO, DIN	yyyy-mm-dd	2012-04-16
ANSI INCITS 30-1997	mm/dd/yyyy	04/16/2012

Die laut aller Normen zu bevorzugende Variante „yyyy-mm-dd“ hat sich aufgrund der allgemeinen Schreibgewohnheiten im deutschen Sprachraum nicht durchgesetzt, üblicherweise wird weiterhin das Datum mit dem Format dd.mm.yyyy bevorzugt. In dieser Dokumentation wird beschrieben, auf welchen Wegen man die in der Tabelle aufgezeigten Formate und ihre Abwandlungen in das gängige Format wandelt, um so ein einheitliches Datumsformat im Archiv zu schaffen.

2 Einzelschritte

Im Folgenden werden alle Einzelschritte dargestellt, welche bei einer Datumsumschlüsselung notwendig sind. Abschließend wird anhand eines Beispiels dargestellt, wie anhand einer Umschlüsselungsliste eine komplexe Umwandlung realisiert werden kann.

Alle Umschlüsselungen werden im Seitenbereichs-Reiter einer zur Bearbeitung geöffneten Maske erstellt. Klicken Sie in der Zeile „Umschlüsselung“ auf „Neu“.

2.1 Datum per Regulärem Ausdruck auslesen

Der einfachster Fall: das Datum liegt schon in der Form dd.mm.yyyy vor und muss schlicht aus einem Textblock ausgelesen werden. Geben Sie eine Bezeichnung ein und wählen als Kategorie „Textbearbeitung“ und als Art „Regulärer Ausdruck“. Die Vorlage für diesen Ausdruck finden Sie in der Dropdown-Liste in der Vorlagen-Zeile. Wählen Sie „Deutsches Datum auflösen“ und klicken auf Daten übernehmen. Der Reguläre Ausdruck wird nun mit einem Beispieltext angezeigt. Wenn Sie auf „Ausdruck testen“ klicken, sehen Sie als Ergebnis nur das Datum. Abschließend klicken Sie auf „Speichern/Verlassen“, die Umschlüsselung wurde erstellt und muss nun nur noch einem auszulesenden Bereich (bzw. einer mehrteiligen Umschlüsselung) auf der Seite zugewiesen werden (s. Abb. 1).

Abbildung 1: Datum per Regulärem Ausdruck auslesen

2.2 Datumsberechnung: bestimmte Formate mit mathematischer Operation wandeln

Bestimmte Formate lassen sich mit der Kategorie „Mathematische Operation“ und der Art „Datumsberechnung“ in einem Schritt umwandeln. Um die Konvertierung zu testen, geben Sie Ihr Quelldatum bei „Zeichenfolge“ ein, weitere Einstellungen müssen nicht vorgenommen werden. Als Ergebnis erhalten Sie dann das gängige deutsche Datum. Abschließend gehen Sie vor wie unter 2.1 beschrieben. Dieser Schritt funktioniert mit den Formaten: yyyyymmdd, yyyy/mm/dd, yyyy-mm-dd (s. Abb. 2)

Abbildung 2: Datumsberechnung: bestimmte Formate mit mathematischer Operation wandeln

2.3 Leerzeichen/Sonderzeichen ersetzen bzw. entfernen

Als vorbereitender Schritt ist es bei manchen Formaten notwendig, Leerzeichen aus dem Datum zu entfernen bzw. Sonderzeichen zu ersetzen. Dies erreicht man mit der Kategorie „Textbearbeitung“, Art „Ersetzen“. Stellen Sie die Vergleichsart auf „Einzelne Zeichen“. In der darunter liegenden Tabelle geben Sie die entsprechenden Zeichen an. Beim Ersetzen von Sonderzeichen muss eine bestimmte Schreibweise verwendet werden: &H[ASCII Hexadezimal-Wert], für ein Leerzeichen dementsprechend: &H20¹. Abschließend gehen Sie vor wie unter 2.1 beschrieben.

Beim Entfernen von Leerzeichen lassen Sie im „Nach“-Tabellenfeld den Wert „(NULL)“ einfach stehen (s. Abb. 3).

Allgemein										
Umschlüsselung										
Bezeichnung	Leerzeichen/Sonderzeichen ersetzen bzw. entfernen									
Kategorie	Textbearbeitung									
Art	Ersetzen									
Suchen und Ersetzen										
Vergleichsart	Einzelne Zeichen									
Standardwert (Typ)	Verwende Quelldaten									
Groß-/Kleinschreibung	<input checked="" type="checkbox"/>									
Wertedatei	<input type="checkbox"/>									
Werte	<table border="1"> <thead> <tr> <th></th> <th>Von</th> <th>Nach</th> </tr> </thead> <tbody> <tr> <td>...</td> <td>&H20</td> <td>(NULL)</td> </tr> <tr> <td>*</td> <td></td> <td></td> </tr> </tbody> </table>		Von	Nach	...	&H20	(NULL)	*		
	Von	Nach								
...	&H20	(NULL)								
*										

Abbildung 3: Leerzeichen/Sonderzeichen ersetzen bzw. entfernen

¹Wikipedia: ASCII-Tabelle

2.4 Ersten zehn Tageswerte des Datums mit „0“ ergänzen

Auf manchen Belegen finden sich Formate, welche teilweise nur eine Stelle für den Tageswert verwenden (Bsp: 7. November 2012). Hier müssen die Tage 1 - 9 um eine „0“ ergänzt werden. In diesem Fall kommt wieder die Art „Ersetzen“ in der Kategorie „Textbearbeitung“ zum Einsatz. Man wandelt von „\b1[.]\"b“ nach „01.“ für den ersten Tag des Monats, von „\b2[.]\"b“ nach „02.“ für den zweiten Tag usw. bis zum neunten Tag (s. Abb. 4). Abschließend gehen Sie vor wie unter 2.1 beschrieben.

Werte	Von	Nach
►	\b1[.]\"b	01.
	\b2[.]\"b	02.
	\b3[.]\"b	03.
	\b4[.]\"b	04.
	\b5[.]\"b	05.
	\b6[.]\"b	06.
	\b7[.]\"b	07.
	\b8[.]\"b	08.
	\b9[.]\"b	09.

Abbildung 4: Ersten zehn Tageswerte des Datums mit „0“ ergänzen

Falls Sie diese Umschlüsselung öfters verwenden, speichern Sie die Tabelle einfach als CSV-Datei ab und wählen diese als „Wertedatei“ aus. In der ersten Zeile tragen Sie die Spaltenüberschriften, getrennt durch ein Semikolon, ein. Im Anschluß folgt für jeden Wert eine neue Zeile. Alter und neuer Wert werden ebenfalls durch Semikolon getrennt (s. Lst. 1).

```
OLDVALUE;NEWVALUE
\b1[.]\"b;01.
\b2[.]\"b;02.
....
\b9[.]\"b;09.
```

Listing 1: Beispiel CSV-Tabelle

Diese Vorgehensweise sollte immer dann verwendet werden, wenn eine exakte Übereinstimmung benötigt wird. Eine Wandlung von „1.“ nach „01.“ würde an dieser Stelle nicht funktionieren, da die Kette „1.“ auch im Monat vorkommen kann: „5.11.2012“ würde ohne die speziellen Begrenzer mit „\b“ nach „05.101.2012“ gewandelt werden.

Die obige Situation kann man auch mit „Capture Groups“ lösen (genaue Beschreibung siehe [2.7](#)). Man benötigt dann nur eine einzelne Zeile und wandelt von „\b(?<tag>\d)[.]\b“ nach „0\$tag.“.

2.5 Ausgeschriebene Monatsnamen in Ziffernwerte wandeln

Auch bei ausgeschriebenen Monatsnamen behelfen Sie sich mit der „Ersetzen“-Art. Wählen Sie die Vergleichsart „Einzelne Zeichen“ und tragen Sie die Werte wie in Abb. 5 ein.

Allgemein

Umschlüsselung

Bezeichnung: Ausgeschriebene Monatsnamen in Ziffernwerte wandeln

Kategorie: Textbearbeitung

Art: Ersetzen

Suchen und Ersetzen

Vergleichsart: Einzelne Zeichen

Standardwert (Typ): Verwende Quelldaten

Groß-/Kleinschreibung: ☒

Wertedatei: ☐

Werte

	Von	Nach
►	Januar	01.
	Februar	02.
	März	03.
	April	04.
	Mai	05.
	Juni	06.
	Juli	07.
	August	08.
	September	09.
	Oktober	10.
	November	11.
	Dezember	12.
*		

Abbildung 5: Ausgeschriebene Monatsnamen in Ziffernwerte wandeln

Je nach Vorgehensweise bzw. Reihenfolge bei der kompletten Datumsumwandlung muss noch ein Punkt in der „Nach“-Spalte ergänzt bzw. entfernt werden (falls z.B. noch ein Punkt als Trennzeichen zwischen Tag und Monat notwendig ist). Ebenso wie in Kapitel [2.4](#) beschrieben, können Sie die Tabelle für mehrmaligen Gebrauch in einer CSV-Datei abspeichern. Abschließend gehen Sie vor wie unter [2.1](#) beschrieben.

2.6 Jahreszahl ergänzen

Falls ein Datumsformat mit nur zwei Ziffern für das Jahr vorliegt (dd.mm.yy), kann dies wie unter Kapitel [2.2](#) beschrieben mit Hilfe einer „Mathematischen Operation“ gelöst werden. Jahreswerte von 00 bis 29 werden

mit „20“ ergänzt, Werte von 30 bis 99 mit „19“ (s. Abb. 6).

Abbildung 6: Jahreszahl ergänzen

2.7 Amerikanisches Format

Eine Besonderheit stellt das us-amerikanische Datumsformat nach ANSI INCITS 30-1997 dar, welches Tages- und Monatswerte vertauscht und wie folgt anzutreffen ist: mm/dd/yyyy Vor jeder weiteren Bearbeitung muss deshalb die Reihenfolge der gängigen deutschen Form hergestellt werden. Verwenden Sie bitte wieder die Umschlüsselung „Textbearbeitung“/„Ersetzen“. Reguläre Ausdrücke erlauben es, einzelne Bestandteile einer Zeichenfolge als Gruppe zu definieren (sogenannte Capture Groups) und dann diese Gruppe umzustellen. Um z.B. die ersten beiden Ziffern des Datumswertes zu gruppieren verwendet man folgenden Regulären Ausdruck: (?<mm>\d{2})

Die Bezeichnung des Blockes ist frei wählbar, damit die Umwandlung übersichtlich gestaltet ist, wird hier „mm“ für die beiden Ziffern des Monatswertes des amerikanischen Datumsformats verwendet.

Der Ausdruck für das gesamte Datum wäre entsprechend: (?<mm>\d{2})/(?<dd>\d{2})/(?<yyyy>\d{4})

Um die Gruppen zu tauschen, spricht man diese mit dem Dollarzeichen und Ihren Gruppennamen an und bringt sie in die gewünschte Reihenfolge: \${dd}/\${mm}/\${yyyy}

Die korrekten Einstellungen sind unter der Abbildung 7 zu sehen.

Nachdem man die Reihenfolge geändert hat, kann das Datum weiter bearbeitet werden, z.B. „/“ mit Punkten ersetzt werden.

3 Beispiel: „Rezept“ für eine mehrteilige Umschlüsselung

Mit Hilfe der oben aufgeführten Einzelschritte kann man für nahezu alle Datumsformate eine Umschlüsselung in das gewünschte Zielformat konstruieren. Ausgehend vom Quellformat wird mit Hilfe mehrerer abgestimmter

Allgemein							
Umschlüsselung							
Bezeichnung	Amerikanisches Format						
Kategorie	Textbearbeitung						
Art	Ersetzen						
Suchen und Ersetzen							
Vergleichsart	Regulärer Ausdruck						
Standardwert (Typ)	Verwende Quelldaten						
Groß-/Kleinschreibung	<input checked="" type="checkbox"/>						
Wertedatei	<input type="checkbox"/>						
Werte	<table border="1"> <thead> <tr> <th>Von</th> <th>Nach</th> </tr> </thead> <tbody> <tr> <td>[?<mm>\d{2})/([?<dd>\d{2})/([?<yyyy>\d{4})</td> <td>\${dd}/\${mm}/\${yyyy}</td> </tr> <tr> <td>*</td> <td></td> </tr> </tbody> </table>	Von	Nach	[?<mm>\d{2})/([?<dd>\d{2})/([?<yyyy>\d{4})	\${dd}/\${mm}/\${yyyy}	*	
Von	Nach						
[?<mm>\d{2})/([?<dd>\d{2})/([?<yyyy>\d{4})	\${dd}/\${mm}/\${yyyy}						
*							

Abbildung 7: Amerikanisches Format

Einzelschritte als „Zutaten“ Schritt für Schritt zum Zielformat hingearbeitet.

Auf unserem Beispielbeleg findet sich folgendes Ausgangsformat: „1. November 2012“, das ausgelesene Datum soll jedoch folgendermaßen archiviert werden: „01.11.2012“

3.1 Ausgeschriebenes Datum auslesen

Der erste Schritt ist immer gleich: der zu verarbeitende Datumsstring muss aus dem markierten Maskenbereich komplett extrahiert werden: (siehe DokuARCHIVEgate.pdf Kapitel 3.1.4 unter Textbearbeitung/Regulärer Ausdruck) z.B. alle Zeichen nach einem bestimmten Ausdruck. In unserem Beispiel extrahieren wir das ausgeschriebene Datum mit folgendem Ausdruck:

```
\d{1,2}\. \s?(Januar|Februar|März|April|Mai|Juni|Juli|August|September|Oktober|November|Dezember)\s?(19|20)\d{2}
```

(s. Abb. 8)

Allgemein	
Umschlüsselung	
Bezeichnung	1. Ausgeschriebenes Datum auslesen
Kategorie	Textbearbeitung
Art	Regulärer Ausdruck
Regulärer Ausdruck	
<div> Lokal Öffentlich </div>	
Vorlagen	<div> </div>
<div> Daten übernehmen Vorlage löschen </div>	
Beispieltext	<div> </div> 11/440438 01. März 2012 645610403 DE1234567890
Ausdruck	\d{1,2}\. \s?(Januar Februar März April Mai Juni Juli August September Oktober November Dezember)\s?(19 20)\d{2}

Abbildung 8: Schritt 1

3.2 Leerzeichen entfernen, Monatsnamen in Zahlen wandeln

Als nächstes werden die Leerzeichen zwischen den Bestandteilen des Datums entfernt und die ausgeschriebenen Monatsnamen mit den entsprechenden Zahlenwerten ergänzt (s. Abb. 9).

	Von	Nach
►	&H20	
	Januar	01.
	Februar	02.
	März	03.
	April	04.
	Mai	05.
	Juni	06.
	Juli	07.
	August	08.
	September	09.
	Oktober	10.
	November	11.
	Dezember	12.

Abbildung 9: Schritt 2

3.3 Ersten zehn Tage mit 0 ergänzen

Wie im Beispiel-Quelldatum zu sehen, werden ausgeschriebene Daten sehr oft mit nur einer einzigen Ziffer für den Tag geschrieben. Um die „0“ zu ergänzen führt man wie unter 2.4 beschrieben eine Umschlüsselung durch (s. Abb. 10).

Kategorie	Textbearbeitung										
Art	Ersetzen										
Suchen und Ersetzen											
Vergleichsart	Regulärer Ausdruck										
Standardwert (Typ)	Verwende Quelldaten										
Groß-/Kleinschreibung	<input checked="" type="checkbox"/>										
Wertedatei	<input type="checkbox"/>										
Werte	<table border="1"> <tr> <th></th> <th>Von</th> <th>Nach</th> </tr> <tr> <td>►</td> <td>\b(?<tag>\d)[.]\b</td> <td>0\${tag}.</td> </tr> <tr> <td>*</td> <td></td> <td></td> </tr> </table>		Von	Nach	►	\b(?<tag>\d)[.]\b	0\${tag}.	*			
	Von	Nach									
►	\b(?<tag>\d)[.]\b	0\${tag}.									
*											

Abbildung 10: Schritt 3

3.4 Abschluss: Alle 3 Schritte zusammengefasst

Abschließend fügen Sie die drei Einzelschritte wie im ARCHIVEgate-Handbuch unter Kapitel 3.1.4 beschrieben zur einer Aufgabenliste zusammen. Wählen Sie hierzu bei einer neuen Umschlüsselung die Kategorie „Umschlüsselung“ und die Art „Aufgabenliste erstellen“. Danach können Sie aus den bisher erstellten Einzelschlüsselungen eine Liste zusammenstellen, in welcher das Datum Schritt für Schritt umgewandelt wird (s. Abb. 11).

The screenshot shows a software window with a tab labeled 'Allgemein'. Below the tab is a section titled 'Umschlüsselung' containing three fields: 'Bezeichnung' with the value 'Alle 3 Schritte zusammengefasst: Datum umwandeln', 'Kategorie' with 'Umschlüsselung', and 'Art' with 'Aufgabenliste erstellen'. Below this is a section titled 'Aufgabenliste mit Umschlüsselungen anlegen'. It features a list of three tasks, each with a checkbox, a description, and a hexadecimal code. Above the list are navigation arrows and a count '3'.

Aufgabenliste mit Umschlüsselungen anlegen	
<input type="checkbox"/>	1. Ausgeschriebenes Datum auslesen / #6E6C
<input type="checkbox"/>	2. Leerzeichen entfernen, Monatsnamen in Zahlen wandeln / #7D36
<input type="checkbox"/>	3. Ersten zehn Tage mit 0 ergänzen / #2EA4

Abbildung 11: Aufgabenliste

Rechts neben der Aufgabenliste befindet sich ein Eingabefeld, in welches man einen Beispieltext eintragen kann um die Funktionalität der Gesamtumschlüsselung zu testen.

Die obige Beispielmaske inkl. Testbeleg kann hier heruntergeladen werden: [Beispielmaske](#)