

Makros zum Export von SAS-Tabellen nach Excel / Access

Marcus Seiler
ArcelorMittal Bremen
Carl-Benz-Straße 30
Bremen
Marcus.Seiler@arcelormittal.com

Zusammenfassung

Das SAS-Makro %SAS2XML exportiert SAS-Tabellen in eine Excel-lesbare xml-Datei. Hierbei können einige Excel-Besonderheiten vorgegeben werden, wie z.B. Fensterfixierung, Autofilter, bedingte Formatierung, Schriftart/-größe und (alternierende) Zeilenfarben.

Schlüsselwörter: Makro, Export, Excel, Access, XML

1 Leitgedanke

Wie kann ich – mit SAS/BASE allein – Daten von SAS in die Microsoft-Office-Welt nach Excel zuverlässig exportieren?

2 Plädoyer für Excel

Natürlich kann Excel keine großen Datenmengen effizient bearbeiten und es eignet sich nicht gut zum Verknüpfen von Tabellen. Dennoch sollte man Excel nicht verteufeln. Die Erfahrung zeigt, dass man Excel gut zur Ansicht von Tabellen bis mittelgroßen Umfangs nutzen kann. Die unter Anwendern weit verbreiteten Such-Filter-Pivot-Kenntnisse decken die meisten Bedürfnisse. Daher verwundert es nicht, dass in der Praxis bei einer Auswertung mit einer einfachen Ergebnistabelle zumeist eine Exceldatei erfragt wird.

3 Historie

3.1 Makro Excel-importierbare Textdatei erzeugen

Die Forderung, SAS-Daten in Excel-Tabellen zu exportieren, trat mehrfach auf. Entsprechend wurde 2001 ein Makro programmiert, das Excel-importierbare Textdateien erzeugt. Die wiederkehrenden Probleme zu Trennzeichen, Dezimaltrenner, Umlaute und Sonderzeichen sollten einmalig zentral gelöst werden.

Störend bleibt, wenn z.B. Trennzeichen als normaler Text auftreten. Und es sind Quotings, also Maskierungen, erforderlich. Dass der Dezimaltrenner bei Excel von der Einstellung des einlesenden Betriebssystems abhängt, führt zu Fehlinterpretationen.

Beim Einlesen in Excel müssen Parameter gesetzt werden (z.B. Trennzeichen statt fester Breite), so dass ein vollautomatischer Ablauf nicht möglich ist.

3.2 Makro %EXCEL4 nutzt BIFF4

Die Idee ist, das BIFF (Binary Interchange File Format) von Excel direkt zu erzeugen. Für das alte BIFF4 aus Excel4 (1992) war eine relativ vollständige Kodierungsanleitung im Internet auffindbar. Für die rudimentäre Umsetzung (nur einfache Tabelle ohne Grafiken etc.) reicht das Format auch vollkommen aus. Bei einer leichten Verletzung der Spezifikation spielt Excel auch wunderbar mit: Excel4 kennt z.B. nur 16.384 Zeilen, neuere Excel-Versionen können dagegen BIFF4 mit 65.536 Zeilen problemlos lesen. Somit war 2002 das Makro %EXCEL4 geboren und der erste, echte Schritt in Richtung Excel ausgeführt. Die Excel4-Unzulänglichkeiten (BIFF4 überreizt) bereiteten Sorge: Wann kommt die Version, die hier nicht mehr ein Auge zukneift, sondern Formatverletzungen hart abmahnt und die Datei verweigert? Zwar funktioniert %EXCEL4 gut, aber es entspricht einer tickenden Zeitbombe. Andererseits ist der Aufwand zur Umsetzung von höheren BIFF-Versionen aufgrund der Komplexität wesentlich zu hoch. Das Dilemma rief nach einem neuen Weg.

3.3 Makro %SAS2XML verwendet Zwischenprodukt XML

XML (eXtensible Markup Language) wird in zahlreichen Varianten zum Datenaustausch eingesetzt. Excel definiert hier sein eigenes, spezielles XML-Schema. Dessen Struktur wurde quasi durch Reverse-Engineering ergründet: startend mit einer einzelligen Tabelle wurde diese immer weiter leicht variiert und die Auswirkungen beim XML-Export analysiert. So wurden peu-a-peu die notwendigen Excel-XML-Tags ergründet. Seit 2005 hat sich somit ein 62-parametriges Makro herauskristallisiert, das eine Vielzahl von Vorgaben ermöglicht (Zellenformate (Farbe, Fontart/-größe), alternierende Zeilenformate, bedingte Formatierung, Autofilter, Fensterfixierung, mehrere Tabellenblätter u.v.m.). Fast 200 automatische SAS-Programme nutzen bei ArcelorMittal Bremen das Makro %SAS2XML.

3.4 Konvertierung von XML nach XLS

XML-Dateien sind mit ihrem "sprechenden" Textinhalt bei etwaigen Fehlern natürlich wesentlich angenehmer zu diagnostizieren als die eher unbeliebten Binärdateien. Wenn der Erstellungsprozess aber verlässlich, fehlerfrei läuft, dann verlieren die um ca. Faktor 4 längeren XML-Dateien schnell ihre Daseinsberechtigung. Vielmehr sind echte, smarte XLS-Dateien gefragt. Zur Konvertierung von XML nach XLS kann man selbstverständlich Excel selbst nutzen. Die einfache Idee ist hierbei, dass eine Excel-VBA die XML-Datei öffnet und dann trivialerweise im XLS-Format speichert.

Unsere Lösungsmethode sieht vor, dass die XML-Datei und eine paarige, einzeilige Textdatei (quasi Ticket, beinhaltet nur den Zielort für die XLS-Datei) in ein spezielles Verzeichnis kopiert werden; dann reagiert ein Verzeichnisüberwacher und führt die erwähnte Konvertierung aus. Anfangs hatten wir als Verzeichnisüberwacher eine eigene

Excel-VBA. Diese war aber zu fehleranfällig (defekte XML-Datei, Netzwerkprobleme, Seiteneffekte durch andere Excel-Applikationen etc.) und wurde daher durch ein .NET-Programm ausgetauscht. Das .NET-Programm läuft stabil und sendet bei Laufzeit-Problemen Warnmails (Zielpfad defekt; bereits existierende Zieldatei nicht austauschbar wegen Anwenderblockierung, ...).

3.5 Makro %SAS2XML4ACCESS für Access

Nachdem die Excel-XML-Variante lief, folgte 2006 eine Access-XML-Variante. Aufgrund der vorhandenen Erfahrung und vergleichsweise wenigen 22 Parametern ist dieses Makro entsprechend schnell realisiert worden. Hier existieren mittlerweile ebenfalls über 20 aktive, automatische SAS-Programme in unserer Berichtswesen-Abteilung.

4 Das Makro %SAS2XML

4.1 Implementierung

Der Quellcode ist sowohl unter www.redscope.org als auch im deutschsprachigen SAS-Wiki (de.saswiki.org) zu finden.

Zwei systemabhängige Programmteile (Pfad für temporäre Datei und Kopiermechanismus) sind durch einen Kommentar mit dem Text "!!!SYSTEMABHÄNGIG!!!" markiert und somit leicht aufzufinden. Der aktuelle Programmcode ist lauffähig unter SAS für Linux Version 8.2.

4.2 Parameter-TOP4

Das Makro %SAS2XML funktioniert auch ohne Angabe von Parametern; dann wird die zuletzt erzeugte SAS-Tabelle (&SYSLAST) mit allen ihren Spalten und der Standardzellformatierung in eine XML-Datei im aktuellen Verzeichnis exportiert. Als adhoc-Aufgabe ist das in Ordnung. Im Programmcode empfiehlt es sich jedoch, mehr Wert auf Dokumentation und Lesbarkeit zu legen und somit die Makroparameter explizit anzugeben. Andererseits sind alle 62 Parameter eher zuviel und verwischen den Blick auf das Wesentliche. In der Praxis ist es sinnvoll, einen Kompromiss einzugehen und daher einen Minimalparametersatz immer anzugeben, auch wenn die Angaben redundant sind, und die restlichen Parameter auf ihren Defaultwerten zu belassen. Im Normalfall reichen insofern die folgenden vier Hauptparameter.

Tabelle 1: Hauptparameter

Parameter	Kurzbeschreibung
DATA	... beinhaltet eine Aufzählung der zu exportierenden SAS-Tabellen, getrennt durch das Pipe-Zeichen „ “. Pro SAS-Tabelle wird ein eigenes Tabellenblatt angelegt.
VAR	... spezifiziert die zu exportierenden Tabellenspalten durch Aufzählung von Variablennamenlisten, getrennt durch das Pipe-Zeichen „ “ korrespondierend zu DATA. Jede Liste wiederum ist eine durch Leerzeichen getrennte Aufzählung von Variablennamen. Ist eine Variablenliste leer, so werden alle Variablen aus der entsprechenden SAS-Tabelle genommen. Ist diese Aufzählung komplett leer, dann werden entsprechend alle DATA-Tabellen mit jeweils allen Variablen selektiert.
OUT	... beinhaltet in der Regel den Pfad inklusive Dateiname für die zu erzeugende XML-Datei. Sonderfall: Beginnt die Angabe allerdings mit einem Ausrufezeichen (!name), dann wird die nachfolgende Zeichenkette als logischer SAS-filename betrachtet; damit ist dann z.B. ein direktes Schreiben via FTP möglich. Ist die Angabe leer, dann wird ein XML-Dateiname aus dem ersten Tabellennamen in DATA generiert und das aktuelle Verzeichnis als Zielpfad verwendet.
KEY	... beinhaltet die Aufzählung von Anzahlen der Keyvariablen (werden in VAR jeweils als erstes aufgezählt) getrennt durch das Pipe-Zeichen „ “ korrespondierend zu DATA. Die Keyvariablen werden per Default farblich abgesetzt dargestellt und steuern die Fensterfixierung und ggf. alternierende Zeilenfarben. Dadurch wird das Manövrieren in der Excel-Datei erleichtert und eine einfache Orientierung erreicht.

4.3 Beispiel mittels sashelp

Der Makro-Aufruf

```
%SAS2XML (DATA=sashelp.class | sashelp.shoes,
          KEY =1 |,
          OUT =/mein/ausgabe/pfad/KSFE_beispiel.xml);
```

erzeugt beispielsweise

	A	B	C	D	E	F
1	Name	Sex	Age	Height	Weight	
2	Alfred	M	14	69	112,5	
3	Alice	F	13	56,5	84	
4	Barbara	F	13	65,3	98	
5	Carol	F	14	62,8	102,5	
6	Henry	M	14	63,5	102,5	
7	James	M	12	57,3	83	
8	Jane	F	12	59,8	84,5	

sashelp.class | sashelp.shoes | log

Abbildung 1: Excel-Beispiel

5 Ausblick

Ab Excel 12 (2007) sind Tabellenblätter mit 1.048.576 Zeilen und 16.384 Spalten möglich. Die aktuelle %SAS2XML-Version ist nur für Excel 2002 - also 65.536 Zeilen und 256 Spalten ausgelegt. Die Begrenzungen sind vorausschauend in Makrovariablen hinterlegt (MAXZEILEN, MAXSPALTEN) und daher simpel anzupassen. Damit müssen vermutlich nur noch die XML-Namensräume (xmlns-Verweise) umversioniert werden, während die wesentliche Tag-Struktur der Tabellenblätter hoffentlich stabil bzw. abwärtskompatibel bleibt. Daher sollte eine geplante Umstellung in diesem Jahr keine Probleme aufwerfen.