

Office Binary (doc, xls, ppt) Translator to Open XML

(xls to xlsx Mapping v0.3; DiaLOGIKa/sare/wk 27 March 2009)

This is a very initial mapping table from the binary Excel format (xls) to SpreadsheetML -- the Open XML markup language for Excel documents (xlsx).

We will use this mapping table in the future to implement the converter from binary to Open XML Excel documents.

The following mapping information is currently available:

- Document Parts
- Workbook Part
- SharedStrings Part
- Worksheet Part
- External Link Part
- Styles

The following columns describe the mapping:

Stream:	<i>The binary stream</i>
Struct:	<i>The binary structure</i>
Field:	<i>The mapped field of the binary structure</i>
XML File:	<i>The corresponding XML file</i>
XML Element:	<i>The corresponding XML element</i>
XML Attribute:	<i>The corresponding XML attribute</i>
Comment:	<i>Interesting details</i>

Workbook Part					
BiffRecord	Field	XML element	XML attribute	Comment	Implemented
BOUNDSHEET	rgch	/workbook/sheets/sheet	name	This is a byte array which contains the name of a worksheet	✓
		/workbook/sheets/sheet	sheetId	Automatic generated sheet id	✓
		/workbook/sheets/sheet	r:Id	Automatic generated sheet id and an additional string	✓
SUPBOOK		/externalReferences/externalReference/		For each external link file there is a entry in the externalReferences element	✓
SUPBOOK		/externalReferences/externalReference/	r:Id	Automatic generated external link reference id	✓
NAME / Lbl		/workbook/definedNames/definedName/		Defined name structure	✓
	rgce + cce	/workbook/definedNames/definedName/		Formula for this defined name, the cce value defines the lenght of the parsed formula in the rgce field	✓
	Name + cch	/workbook/definedNames/definedName/	name	The cch field defines the lenght of the defined name string	✓
	itab	/workbook/definedNames/definedName/	localSheetId	This is the reference to a worksheet	✓
	chKey	/workbook/definedNames/definedName/	shortcutKey	Shortcut to this name	
	fGrp	/workbook/definedNames/definedName/	functionGroupId	ID from a functionGroup	
NAMECMT	comment	/workbook/definedNames/definedName/	comment	Comment to a defined name	

Worksheet Part					
BiffRecord	Field	XML element	XML attribute	Comment	Implemented
NUMBER, RK, MULRK, BLANK, MULBLANK, LABELSST	rw	/worksheet/sheetData/row	r	Indicates where the following data row is located	✓
NUMBER, RK, MULRK, BLANK, MULBLANK, LABELSST	Different fields: NUMBER, RK, BLANK and LABELSST use the col value. The MULBLANK and MULRK structs use the colFirst value + an integer value that indicates the position of the array.	/worksheet/sheetData/row/c	r	Indicates where the following data is located. This value has to be a mix of the column as letter and the row as a number value.	✓
		/worksheet/sheetData/row/c	t	This value indicates if the following data is a string or a number. If it is a number, the attribute "t" is missing else the attribute t is there and its value is "s".	✓
	ixfe	/worksheet/sheetData/row/c	s	Reference to the style structure	✓
LABELSST	isst	/worksheet/sheetData/row/c/v		The data. If the value is a string then it is only the reference to the shared string table.	✓
MULRK	rgrkrec -> array of rk records	/worksheet/sheetData/row/c/v		"	✓
MULBLANK, BLANK	no value	/worksheet/sheetData/row/c/v		"	✓
RK	rk	/worksheet/sheetData/row/c/v		"	✓
NUMBER	num	/worksheet/sheetData/row/c/v		"	✓
MERGECELL	cmcs	/worksheet/mergeCells	count		✓
MERGECELL	rwFirst, rwLast, colFirst, colLast	/worksheet/mergeCells/mergeCell	ref	Creates a reference string that has two values, the position of the top left corner from the mergecell and the bottom right position	✓
FORMULA	rw, col	/worksheet/sheetData/row/c	r	Indicates where the following data is located. This value has to be a mix of the column as letter and the row as a number value.	✓
	cce, rgce	/worksheet/sheetData/row/c/f		This field is used to store the whole formula. This formula is stored in the postfix notation and has to be translated into the infix notation.	✓
	fShrFmla	/worksheet/sheetData/row/c/f	t="shared"	If this formula is a shared formula this flag is set	✓
	val	/worksheet/sheetData/row/c/v		The calculated value from a formula, this can be a number, a string, an error value or a boolean.	✓
		/worksheet/sheetData/row/c/f	si	Shared group - automatically generated number	✓
ARRAY	rwFirst, rwLast, colFirst, colLast	/worksheet/sheetData/row/c	r	Indicates where the following data is located. This value has to be a mix of the column as letter and the row as a number value.	✓
	cce, rgce	/worksheet/sheetData/row/c/f		This field is used to store the whole formula. This formula is stored in the postfix notation and has to be translated into the infix notation.	✓
		/worksheet/sheetData/row/c/f	t="array"	Indicates that the formula is an array formula	✓
		/worksheet/sheetData/row/c/f	ref=" "	The reference value is the same value where the formula is located	✓
SHRFMLA	ref	/worksheet/sheetData/row/c/f	ref	Reference where this shared formula is positioned	✓
STRING		/worksheet/sheetData/row/c/v		This record specifies the string value of a formula.	✓
ROW	fCollapsed	/worksheet/sheetData/row	collapsed	If the row is collapsed this is true.	✓
	fGhostDirty	/worksheet/sheetData/row	customFormat		✓
	fUnsynced	/worksheet/sheetData/row	customHeight		✓

Worksheet Part					
BiffRecord	Field	XML element	XML attribute	Comment	Implemented
	fDyZero	/worksheet/sheetData/row	hidden	Hiddenflag	✓
	miyRw	/worksheet/sheetData/row	ht	Hight of this row	✓
	iOutLevel	/worksheet/sheetData/row	outlineLevel	Specifies the outline level	✓
	fPhonetic	/worksheet/sheetData/row	ph	phonetic data flag	✓
	ixfe_val	/worksheet/sheetData/row	s	A reference to the style data element	✓
	colMic, colMac	/worksheet/sheetData/row	spans	The area in which are values	✓
	fExDes	/worksheet/sheetData/row	thickBot	If the row above has a thick bottom border this value is true	✓
	fExAsc	/worksheet/sheetData/row	thickTop	If the row below has a thick top border this value is true	✓
COLINFO		/worksheet/cols/col		Defines some column information	✓
	colFirst	/worksheet/cols/col	min	First cell in this column	✓
	colLast	/worksheet/cols/col	max	Last cell in this column	✓
	coldx	/worksheet/cols/col	width	The widt from this column	✓
	fUserSet	/worksheet/cols/col	customWidth	User set the widht!	✓
	fHidden	/worksheet/cols/col	hidden	If the column is hidden this value is 1	✓
	fBestFit	/worksheet/cols/col	bestFit	bestFit flag	✓
	fPhonetic	/worksheet/cols/col	phonetic	phonetic data flag	✓
	iOutLevel	/worksheet/cols/col	outlineLevel	outlineLevel value	✓
	fCollapsed	/worksheet/cols/col	collapsed	collapsed flag	✓
	ixfe	/worksheet/cols/col	style	ID to the style record	✓
DEFCOLWIDTH	cchdefColWidth	/worksheet/sheetFormatPr	defaultColWidth	Default column width	✓
DEFROWHEIGHT	miyRW	/worksheet/sheetFormatPr	defaultRowHeight	Default row hight	✓
	fDyZero	/worksheet/sheetFormatPr	zeroHeight	Flag is set if the row height is 0	✓
	fUnsynced	/worksheet/sheetFormatPr	customHeight	If the height is set manually	✓
	fExAsc	/worksheet/sheetFormatPr	thickTop	If the row above has a thick bottom border this value is true	✓
	fExDsc	/worksheet/sheetFormatPr	thickBottom	If the row below has a thick top border this value is true	✓
SETUP		/worksheet/pageSetup/		This structure stores the page setup information	✓
	iPaperSize	/worksheet/pageSetup/	paperSize		✓
	iScale	/worksheet/pageSetup/	scale		✓
	fNoColor	/worksheet/pageSetup/	blackAndWhite		✓
	iCopies	/worksheet/pageSetup/	copies		✓
	fDraft	/worksheet/pageSetup/	draft		✓
	iPageStart	/worksheet/pageSetup/	firstPageNumber		✓
	iFitHeight	/worksheet/pageSetup/	fitToHeight		✓
	iFitWidth	/worksheet/pageSetup/	fitToWidth		✓
	iRes	/worksheet/pageSetup/	horizontalDpi		✓
	fUsePage	/worksheet/pageSetup/	useFirstPageNumb		✓
	iVRes	/worksheet/pageSetup/	verticalDpi		✓
	iError	/worksheet/pageSetup/	errors		✓
	fNoOrient, fLandscape	/worksheet/pageSetup/	orientation	Page orientation, none, landscape or portrait	✓
	fNotes, fEndNotes	/worksheet/pageSetup/	cellComments		✓
	fLefttoRight	/worksheet/pageSetup/	pageOrder		✓
SETUP	Setup.numHdr	/worksheet/pageMargins/	header	Size of the header	✓
	Setup.numFtr	/worksheet/pageMargins/	footer	Size of the footer	✓
LEFTMARGIN		/worksheet/pageMargins/	left		✓
RIGHTMARGIN		/worksheet/pageMargins/	right		✓

Worksheet Part					
BiffRecord	Field	XML element	XML attribute	Comment	Implemented
TOP		/worksheet/pageMargins/	top		✓
BOTTOM		/worksheet/pageMargins/	bottom		✓

Styles Part					
BiffRecord	Field	XML element	XML attribute	Comment	Implemented
		\styleSheet\numFmts\	count	Number of format elements	✓
FORMAT	ifmt	\styleSheet\numFmts\numFmt\	numFmtId	The ID from this format string	✓
	stFormat	\styleSheet\numFmts\numFmt\	formatCode	The string from this format	✓
		\styleSheet\fonts\	count	Number of font elements	✓
FONT		\styleSheet\fonts\font\		Font element	✓
	dyHeight	\styleSheet\fonts\font\sz\	val	Font size	✓
	fItalic	\styleSheet\fonts\font\i\		Is italic	✓
	fStrikeOut	\styleSheet\fonts\font\strike\	val		✓
	fOutline	\styleSheet\fonts\font\outline\	val		✓
	fShadow	\styleSheet\fonts\font\shadow\	val		✓
	fCondense	\styleSheet\fonts\font\condense\	val		✓
	icv	\styleSheet\fonts\font\color\	auto, rgb, indexed	The color is always the rgb value!	✓
	bls	\styleSheet\fonts\font\b\	val		✓
	sss	\styleSheet\fonts\font\vertAlign\	val		✓
	uls	\styleSheet\fonts\font\u\	val		✓
	bFamily	\styleSheet\fonts\font\family\	val		✓
	bCharSet	\styleSheet\fonts\font\charset\	val		✓
	fontName	\styleSheet\fonts\font\name\	val		✓
STYLE	rgch	\styleSheet\cellStyles\cellStyle\	name	Name of the style.	✓
	ixfe	\styleSheet\cellStyles\cellStyle\	xfld	ID to a xf element	✓
	istyBuiltin	\styleSheet\cellStyles\cellStyle\	builtinId		✓
		\styleSheet\fills\	count		✓
XF	fls	\styleSheet\fills\fill\patternFill\	patternType		✓
	icvFore	\styleSheet\fills\fill\patternFill\fgColor\	rgb	rgb color value if the icvFore is not auto	✓
		\styleSheet\fills\fill\patternFill\fgColor\	auto="1"	Auto color if the icvFore is auto	✓
	icvBack	\styleSheet\fills\fill\patternFill\bgColor\	rgb	rgb color value if the icvBack is not auto	✓
		\styleSheet\fills\fill\patternFill\bgColor\	auto="1"	Auto color if the icvBack is auto	✓
		\styleSheet\borders\	count	Number of border elements	✓
XF	grbitDiag	\styleSheet\borders\border\	diagonalDown="1"	If grbit == 1	✓
	grbitDiag	\styleSheet\borders\border\	diagonalUp="1"	If grbit == 2, if grbit == 3 -> booth values	✓
		\styleSheet\borders\border\left\			✓
	dgLeft	\styleSheet\borders\border\left\	style	Style from this borderpart	✓
	icvLeft	\styleSheet\borders\border\left\color\	rgb or auto		✓
		\styleSheet\borders\border\right\			✓
	dgRight	\styleSheet\borders\border\right\	style	Style from this borderpart	✓
	icvRight	\styleSheet\borders\border\right\color\	rgb or auto		✓
		\styleSheet\borders\border\top\			✓
	dgTop	\styleSheet\borders\border\top\	style	Style from this borderpart	✓
	icvTop	\styleSheet\borders\border\top\color\	rgb or auto		✓
		\styleSheet\borders\border\bottom\			✓
	dgBottom	\styleSheet\borders\border\bottom\	style	Style from this borderpart	✓
	icvBottom	\styleSheet\borders\border\bottom\color\	rgb or auto		✓
		\styleSheet\borders\border\diagonal\			✓

Styles Part					
BiffRecord	Field	XML element	XML attribute	Comment	Implemented
	dgDiag	\styleSheet\borders\border\diagonal\	style	Style from this borderpart	✓
	icvDiag	\styleSheet\borders\border\diagonal\color\	rgb or auto		✓
		\styleSheet\cellStyleXfs\	count	number of cellStyleXfs	✓
XF	ifmt	\styleSheet\cellStyleXfs\xf\	numFmtId	Format ID	✓
	ifnt	\styleSheet\cellStyleXfs\xf\	fontId	Font ID	✓
		\styleSheet\cellStyleXfs\xf\	fillId	Autogenerated	✓
		\styleSheet\cellStyleXfs\xf\	borderId	Autogenerated	✓
	fWrap	\styleSheet\cellStyleXfs\xf\alignment\	wrapText="1"		✓
	alc	\styleSheet\cellStyleXfs\xf\alignment\	horizontal		✓
	alcV	\styleSheet\cellStyleXfs\xf\alignment\	vertical		✓
	fJustLast	\styleSheet\cellStyleXfs\xf\alignment\	justifyLastLine="1"		✓
	fShrinkToFit	\styleSheet\cellStyleXfs\xf\alignment\	shrinkToFit="1"		✓
	trot	\styleSheet\cellStyleXfs\xf\alignment\	textRotation		✓
	cIndent	\styleSheet\cellStyleXfs\xf\alignment\	indent		✓
	iReadOrder	\styleSheet\cellStyleXfs\xf\alignment\	readingOrder		✓
		\styleSheet\cellXfs\	count	Number of cellXfs	✓
XF	ifmt	\styleSheet\cellXfs\xf\	numFmtId	Format ID	✓
	ifnt	\styleSheet\cellXfs\xf\	fontId	Font ID	✓
		\styleSheet\cellXfs\xf\	fillId	Autogenerated	✓
		\styleSheet\cellXfs\xf\	borderId	Autogenerated	✓
	ixfParent	\styleSheet\cellXfs\xf\	xfId		✓
	ifmt	\styleSheet\cellXfs\xf\	applyNumberFormat="1"	if Format != 0	✓
		\styleSheet\cellXfs\xf\	applyBorder="1"	if Border != 0	✓
		\styleSheet\cellXfs\xf\	applyFill="1"	if Fill != 0	✓
	ifnt	\styleSheet\cellXfs\xf\	applyFont="1"	if Font != 0	✓
	fWrap	\styleSheet\cellXfs\xf\alignment\	wrapText="1"		✓
	alc	\styleSheet\cellXfs\xf\alignment\	horizontal		✓
	alcV	\styleSheet\cellXfs\xf\alignment\	vertical		✓
	fJustLast	\styleSheet\cellXfs\xf\alignment\	justifyLastLine="1"		✓
	fShrinkToFit	\styleSheet\cellXfs\xf\alignment\	shrinkToFit="1"		✓
	trot	\styleSheet\cellXfs\xf\alignment\	textRotation		✓
	cIndent	\styleSheet\cellXfs\xf\alignment\	indent		✓
	iReadOrder	\styleSheet\cellXfs\xf\alignment\	readingOrder		✓

Styles Part					
BiffRecord	Field	XML element	XML attribute	Comment	Implemented