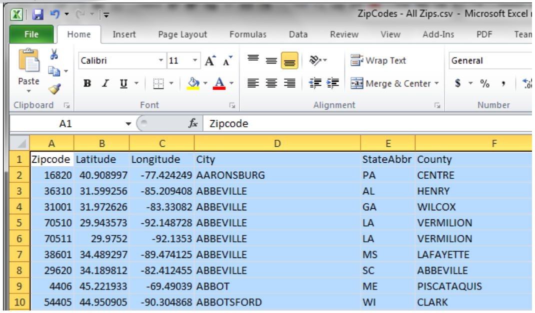
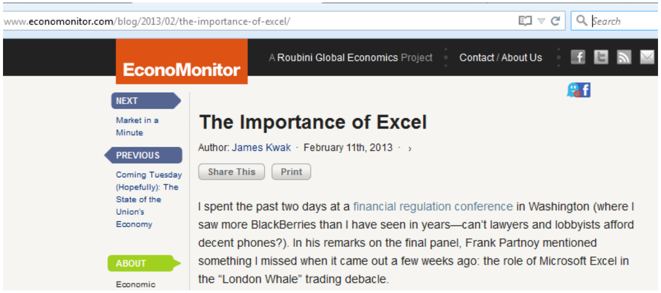
Mining the Depths of Excel

Case Study in Objects and Arrays



	A	B	C	D	E	F
1	Zipcode	Latitude	Longitude	City	StateAbbr	County
2	16820	40.908997	-77.424249	AARONSBURG	PA	CENTRE
3	36310	31.599256	-85.209408	ABBEVILLE	AL	HENRY
4	31001	31.972626	-83.33082	ABBEVILLE	GA	WILCOX
5	70510	29.943573	-92.148728	ABBEVILLE	LA	VERMILION
6	70511	29.9752	-92.1353	ABBEVILLE	LA	VERMILION
7	38601	34.489297	-89.474125	ABBEVILLE	MS	LAFAYETTE
8	29620	34.189812	-82.412455	ABBEVILLE	SC	ABBEVILLE
9	4406	45.221933	-69.49039	ABBOT	ME	PISCATAQUIS
10	54405	44.950905	-90.304868	ABBOTSFORD	WI	CLARK

Why Excel?


www.economonitor.com/blog/2013/02/the-importance-of-excel/

EconoMonitor
A Roubini Global Economics Project
Contact / About Us
NEXT
Market in a Minute
PREVIOUS
Coming Tuesday (Hopefully): The State of the Union's Economy
ABOUT
Economic
Share This
Print

The Importance of Excel

Author: James Kwak · February 11th, 2013

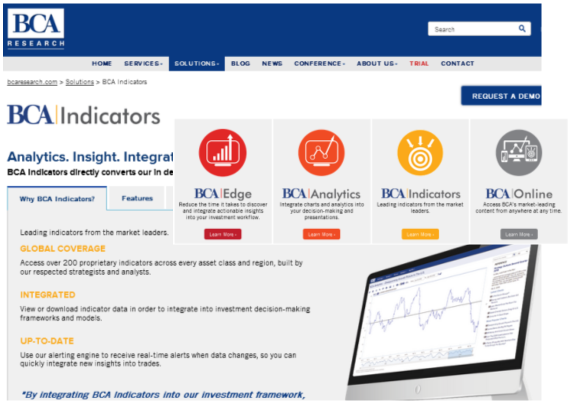
I spent the past two days at a financial regulation conference in Washington (where I saw more BlackBerries than I have seen in years—can't lawyers and lobbyists afford decent phones?). In his remarks on the final panel, Frank Partnoy mentioned something I missed when it came out a few weeks ago: the role of Microsoft Excel in the "London Whale" trading debacle.

James Kwak:
(prominent financial blogger?)
"...Microsoft Excel is one of the greatest, most powerful, most important software applications of all time..."

"...Excel is everywhere you look in the business world—especially in areas where people are adding up numbers a lot, ... I have a probably untestable hypothesis that, were you to come up with some measure of units of software output, ***Excel would be the most-used program in the business world****..."*

(http://www.economonitor.com/blog/2013/02/the-importance-of-excel/)

Excel at BCA Research...



bcaresearch.com > Solutions > BCA Indicators

BCA | Indicators

Analytics. Insight. Integrat

BCA Indicators directly converts our in de

BCA | Edge
Reduce the time it takes to discover and integrate actionable insights into your investment workflow.

BCA | Analytics
Integrate charts and analytics into your decision-making and presentations.

BCA | Indicators
Leading indicators from the market leaders.

BCA | Online
Access BCA's market-leading content from anywhere at any time.

Leading indicators from the market leaders.

GLOBAL COVERAGE

Access over 200 proprietary indicators across every asset class and region, built by our respected strategists and analysts.

INTEGRATED

View or download indicator data in order to integrate into investment decision-making frameworks and models.

UP-TO-DATE

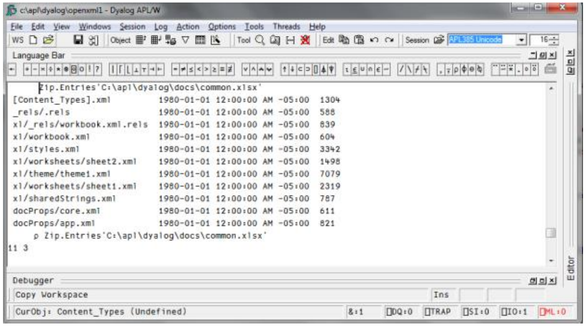
Use our alerting engine to receive real-time alerts when data changes, so you can quickly integrate new insights into trades.

"By integrating BCA Indicators into our investment framework,

- Lists - keeping track of..., eg. user profiles, data retrieval codes, publication files, etc.
- Data sources - (Bloomberg, ThomsonReuters...) make data available as .xlsx or .csv
- Interfaces for data collection - downloads and analytical tools are driven by Excel Add-ins
- Charts - if no other way to produce
- Statistics - if no better way to calculate
- Reports - presentation of analyses; lists of things that need attention, etc.
- Process control - "table-driven" tasks - determine what to do based on worksheet contents

In short, Excel is extremely important, and EVERYWHERE

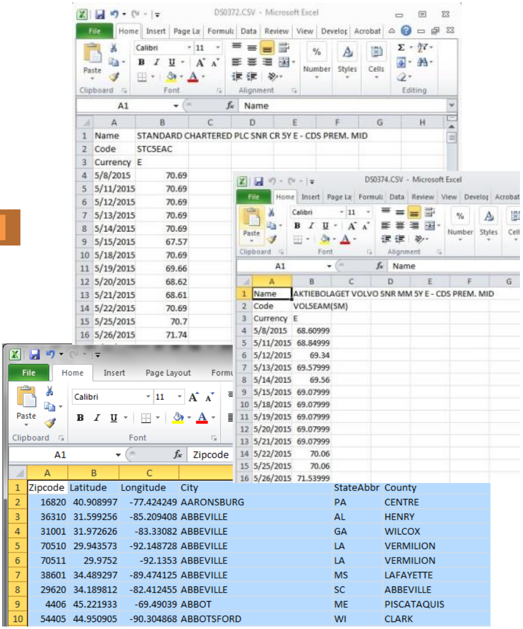
Integration with APL


c:\apl\dyalog\openxml1 - Dyalog APL/W
File Edit View Windows Session Log Action Options Tools Threads Help
Language Bar
Debugger
Editor

```
      Zip.Entries'C:\apl\dyalog\docs\common.xlsx'
[Content_Types].xml          1980-01-01 12:00:00 AM -05:00  1304
_rels/.rels                  1980-01-01 12:00:00 AM -05:00  588
xl/_rels/workbook.xml.rels   1980-01-01 12:00:00 AM -05:00  839
xl/workbook.xml              1980-01-01 12:00:00 AM -05:00  604
xl/styles.xml                1980-01-01 12:00:00 AM -05:00  3342
xl/worksheets/sheet2.xml     1980-01-01 12:00:00 AM -05:00  1498
xl/theme/theme1.xml          1980-01-01 12:00:00 AM -05:00  7079
xl/worksheets/sheet1.xml     1980-01-01 12:00:00 AM -05:00  2319
xl/sharedStrings.xml         1980-01-01 12:00:00 AM -05:00  787
docProps/core.xml            1980-01-01 12:00:00 AM -05:00  611
docProps/app.xml             1980-01-01 12:00:00 AM -05:00  821
      ⍴ Zip.Entries'C:\apl\dyalog\docs\common.xlsx'
11 3
```

XLSX
CSV
XLS

1000's


DS0372.CSV - Microsoft Excel
DS0374.CSV - Microsoft Excel
File Home Insert Page Layout Formulas Data Review View Developer Acrobat
Clipboard Font Alignment Number Styles Cells Editing

A1: Name

	A	B
1	Name	STANDARD CHARTERED PLC SNR CR 5Y E - CDS PREM. MID
2	Code	STC5EAC
3	Currency	E
4	5/8/2015	70.69
5	5/11/2015	70.69
6	5/12/2015	70.69
7	5/13/2015	70.69
8	5/14/2015	70.69
9	5/15/2015	67.57
10	5/18/2015	70.69
11	5/19/2015	69.66
12	5/20/2015	68.62
13	5/21/2015	68.61
14	5/22/2015	70.69
15	5/25/2015	70.7
16	5/26/2015	71.74

A1: Name

	A	B
1	Name	AKTIEBOLAGET VOLVO SNR MM 5Y E - CDS PREM. MID
2	Code	VOL5EAM(SM)
3	Currency	E
4	5/8/2015	68.60999
5	5/11/2015	68.84999
6	5/12/2015	69.34
7	5/13/2015	69.57999
8	5/14/2015	69.56
9	5/15/2015	69.07999
10	5/18/2015	69.07999
11	5/19/2015	69.07999
12	5/20/2015	69.07999
13	5/21/2015	69.07999
14	5/22/2015	70.06
15	5/25/2015	70.06
16	5/26/2015	71.53999

A1: Zipcode

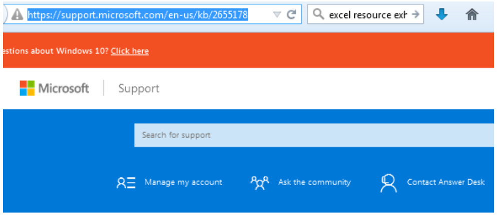
	A	B	C			
1	Zipcode	Latitude	Longitude	City	StateAbbr	County
2	16820	40.908997	-77.424249	AARONSBURG	PA	CENTRE
3	36310	31.599256	-85.209408	ABBEVILLE	AL	HENRY
4	31001	31.972626	-83.33082	ABBEVILLE	GA	WILCOX
5	70510	29.943573	-92.148728	ABBEVILLE	LA	VERMILION
6	70511	29.9752	-92.1353	ABBEVILLE	LA	VERMILION
7	38601	34.489297	-89.474125	ABBEVILLE	MS	LAFAYETTE
8	29620	34.189812	-82.412455	ABBEVILLE	SC	ABBEVILLE
9	4406	45.221933	-69.49039	ABBOT	ME	PISCATAQUIS
10	54405	44.950905	-90.304868	ABBOTSFORD	WI	CLARK

NEED for SPEED?

Automation?

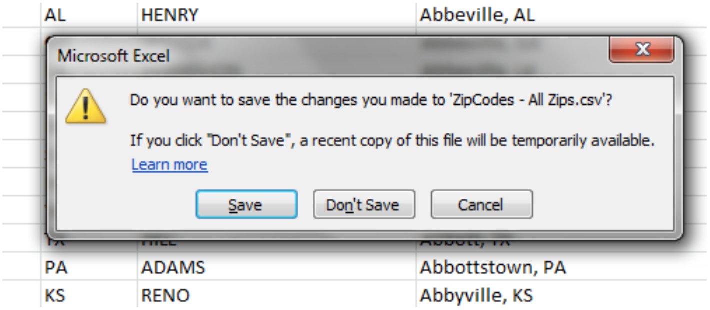
Glitches & Gotchas

- **Installation / Incompatibility issues - APL/Excel**
- **Resource exhaustion - Excel**
- **Automation failures, eg. pop-ups**


https://support.microsoft.com/en-us/kb/2655178

estions about Windows 10? Click here

Excel cannot complete this task with available resources error, Excel 2010


Microsoft Excel

Do you want to save the changes you made to 'ZipCodes - All Zips.csv'?

If you click "Don't Save", a recent copy of this file will be temporarily available.

Excel and XML

Office Open XML (OOXML)
en.wikipedia.org/wiki/Office_Open_XML

Office Open XML

From Wikipedia, the free encyclopedia

Not to be confused with Open Office XML

- **around 2000 - MS began to develop xml-based format for Office documents**
- **2000-2006 - standardization process (ECMA/ISO/IEC)**
- **MS Office 2007 - MS adopts the format as default**
- ***.xlsx, not *.xls**

Office Open XML (also informally known as OOXML or OpenXML) is a <u>zipped, XML-based</u> file format developed by Microsoft for representing spreadsheets, charts, presentations and word processing documents. The format was initially standardised by Ecma (as ECMA-376) and, in later versions, by ISO and IEC (as ISO/IEC 29500). Starting with Microsoft Office 2007...

Support for Office Open XML (non-MS)

https://en.wikipedia.org/wiki/List_of_software_that_supports_Office_Open_XML
and
en.wikipedia.org/wiki/Comparison_of_Office_Open_XML_software#Spreadsheet_documents

- **Many direct products**
- **other (C#/VB) programming components**
- **Office Open XML standard (ECMA-376 and ISO/IEC 29500:2008)**

List of software that supports Office Open XML

From Wikipedia, the free encyclopedia

This is an overview of software support for the Office Open XML

Office Open XML
- Office Open XML file formats
- Open Packaging Conventions
- Open Specification Promise
- Office Open XML software
- Comparison of Office Open XML software

Why not COM? (OLE, ActiveX...)

(http://ramblingcookiemonster.github.io/PSExcel-Intro/)

"Chances are you have worked with Excel through COM....this isn’t supported if you want to use it in an automated solution... "

...from Microsoft: (https://support.microsoft.com/en-us/kb/257757)
All current versions of Microsoft Office were designed, tested, and configured to run as end-user products on a client workstation. They assume an interactive desktop and user profile. They do not provide the level of reentrancy or security that is necessary to meet the needs of server-side components that are designed to run unattended.

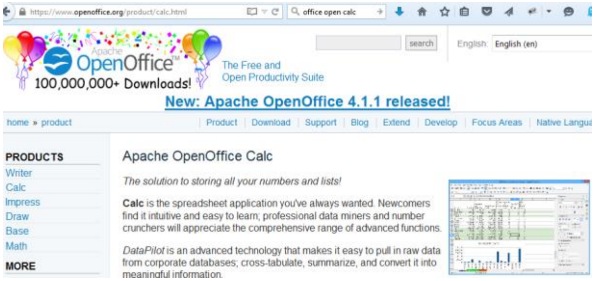
"Microsoft does not currently recommend, and does not support, Automation of Microsoft Office applications from any unattended, non-interactive client application or component (including ASP, ASP.NET, DCOM, and NT Services), because Office may exhibit unstable behavior and/or deadlock when Office is run in this environment."

So #1 = Automation; Plus - above-mentioned issues...

Options beyond COM/OLE

Application?, or API?

- Who does the UI?
- Is there a UI?, etc.
- (there's also ODBC?; ADO.Net?)



https://www.openoffice.org/product/calc.html

Apache OpenOffice™ 100,000,000+ Downloads! The Free and Open Productivity Suite

New: Apache OpenOffice 4.1.1 released!

home » product | Product | Download | Support | Blog | Extend | Develop | Focus Areas | Native Langua

PRODUCTS: Writer, Calc, Impress, Draw, Base, Math, MORE

Apache OpenOffice Calc

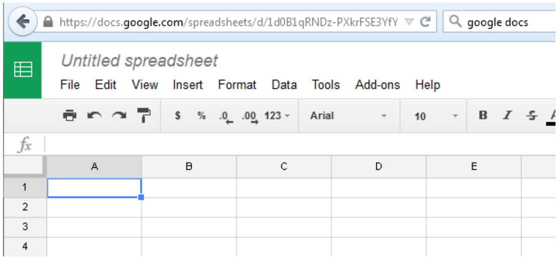
The solution to storing all your numbers and lists!

Calc is the spreadsheet application you've always wanted. Newcomers find it intuitive and easy to learn; professional data miners and number crunchers will appreciate the comprehensive range of advanced functions.

DataPilot is an advanced technology that makes it easy to pull in raw data from corporate databases; cross-tabulate, summarize, and convert it into meaningful information.

Applications

- www.openoffice.org - "Calc", formerly by Sun Microsystems, now Apache
- Google Docs; Android Apps - work but are features limited?
- (MS-OneDrive = Excel Online?)



https://docs.google.com/spreadsheets/d/1d0B1qRNDz-PXkrFSE3YfY
google docs
Untitled spreadsheet
File Edit View Insert Format Data Tools Add-ons Help
$ % .0 .00 123
Arial
10
B I
fx
A B C D E
1
2
3
4

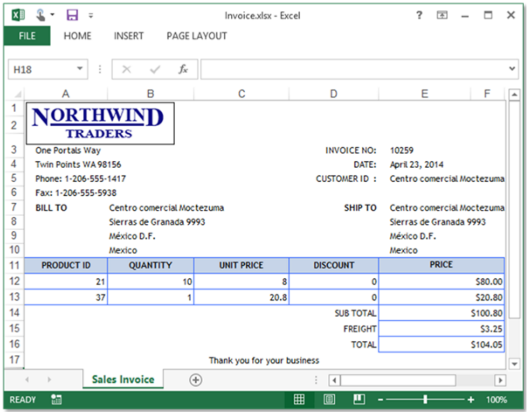
API / Utilities

Syncfusion XlsIO

Introduction to Essential XlsIO

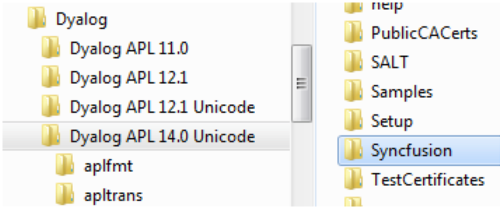
Essential XlsIO is a 100% native .NET library that generates fully functional **Microsoft Excel Spreadsheets** in **native Excel** format without depending on Microsoft Excel. **Essential XlsIO** is a perfect solution for those who need to read and write Microsoft Excel files from code. It does not require MS Excel to be installed in the report generation machine or server.

Essential XlsIO library can be used in any .NET environment including C#, VB.NET, and managed C++. It is a Non-UI component that can be used in **ASP.NET**, **Windows Forms**, **WPF**, **Silverlight**, **ASP.NET MVC**, **WinRT**, **Windows Phone 8** applications, without any change in the API.



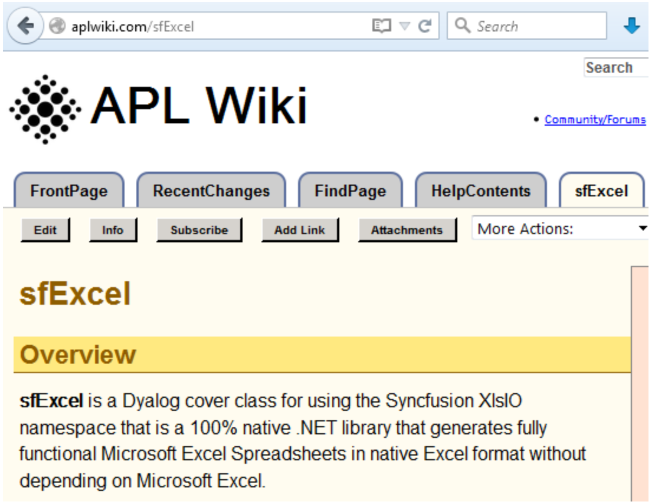
NORTHWIND TRADERS

	A	B	C	D	E	F
3	One Portals Way			INVOICE NO:	10259	
4	Twin Points WA 98156			DATE:	April 23, 2014	
5	Phone: 1-206-555-1417			CUSTOMER ID :	Centro comercial Moctezuma	
6	Fax: 1-206-555-5938					
7	BILL TO	Centro comercial Moctezuma		SHIP TO	Centro comercial Moctezuma	
8		Sierras de Granada 9993			Sierras de Granada 9993	
9		México D.F.			México D.F.	
10		Mexico			Mexico	
11	PRODUCT ID	QUANTITY	UNIT PRICE	DISCOUNT	PRICE	
12	21	10	8	0	$80.00	
13	37	1	20.8	0	$20.80	
14				SUB TOTAL	$100.80	
15				FREIGHT	$3.25	
16				TOTAL	$104.05	
17			Thank you for your business			



sfExcel

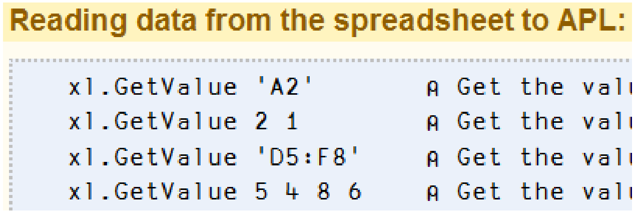
- excellent toolkit for Dyalog APL (no need to study XlsIO)
- Namespace
- speed issues if (1000's < 1↑⍴MAT)
- "DataTable" feature for speedups
- some formatting options
- data/text only (not "inclusions")
- Pierre Gilbert



sfExcel

Overview

sfExcel is a Dyalog cover class for using the Syncfusion XlsIO namespace that is a 100% native .NET library that generates fully functional Microsoft Excel Spreadsheets in native Excel format without depending on Microsoft Excel.



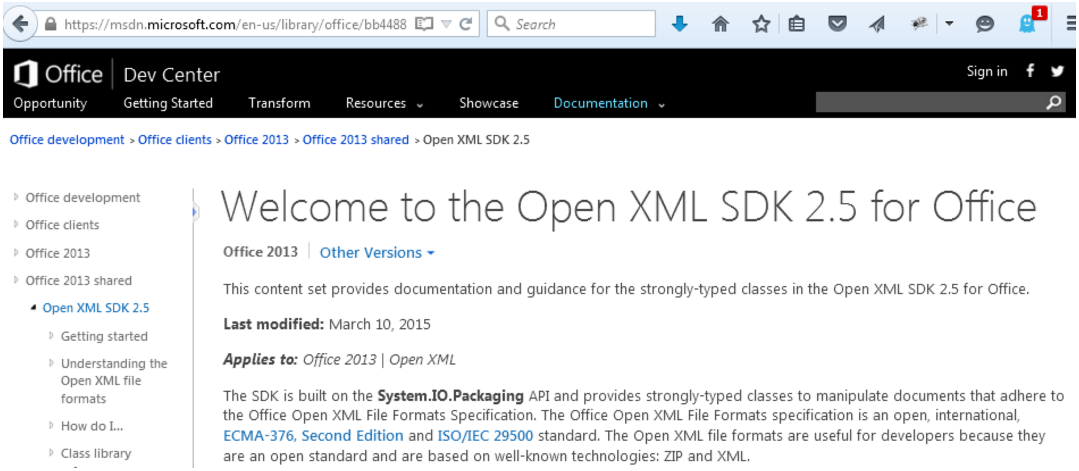
Reading data from the spreadsheet to APL:

```
    xl.GetValue 'A2'          ⍝ Get the valu
    xl.GetValue 2 1           ⍝ Get the valu
    xl.GetValue 'D5:F8'       ⍝ Get the valu
    xl.GetValue 5 4 8 6       ⍝ Get the valu
```

But what does the C# / VB / VisualStudio crowd do?

MS Open XML SDK

https://msdn.microsoft.com/en-us/library/office/bb448854.aspx



Welcome to the Open XML SDK 2.5 for Office

Office 2013

This content set provides documentation and guidance for the strongly-typed classes in the Open XML SDK 2.5 for Office.

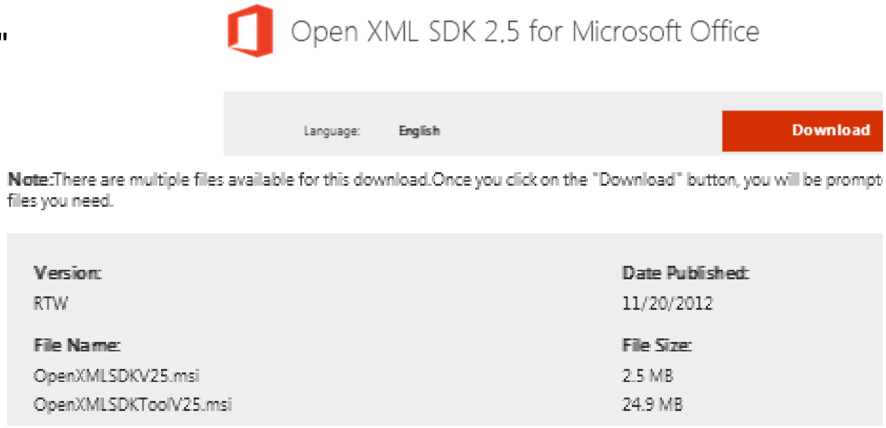
Last modified: March 10, 2015

Applies to: *Office 2013 | Open XML*

The SDK is built on the **System.IO.Packaging** API and provides strongly-typed classes to manipulate documents that adhere to the Office Open XML File Formats Specification. The Office Open XML File Formats specification is an open, international, ECMA-376, Second Edition and ISO/IEC 29500 standard. The Open XML file formats are useful for developers because they are an open standard and are based on well-known technologies: ZIP and XML.

Open XML SDK - Download

API and "Tool"


Open XML SDK 2.5 for Microsoft Office

Language: English

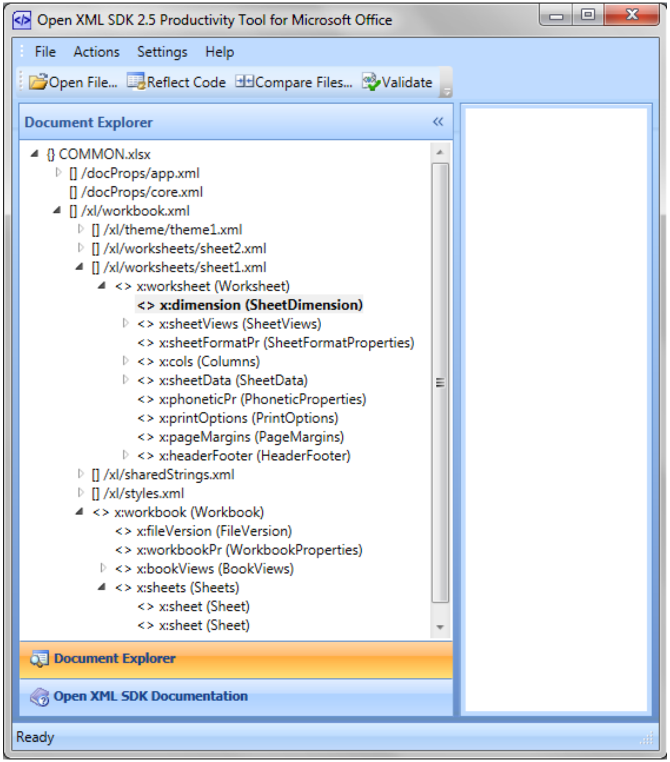
Note: There are multiple files available for this download.Once you click on the "Download" button, you will be prompt
files you need.

Version:	**Date Published:**
RTW	11/20/2012
File Name:	**File Size:**
OpenXMLSDKV25.msi	2.5 MB
OpenXMLSDKToolV25.msi	24.9 MB

Open XML SDK - Tool

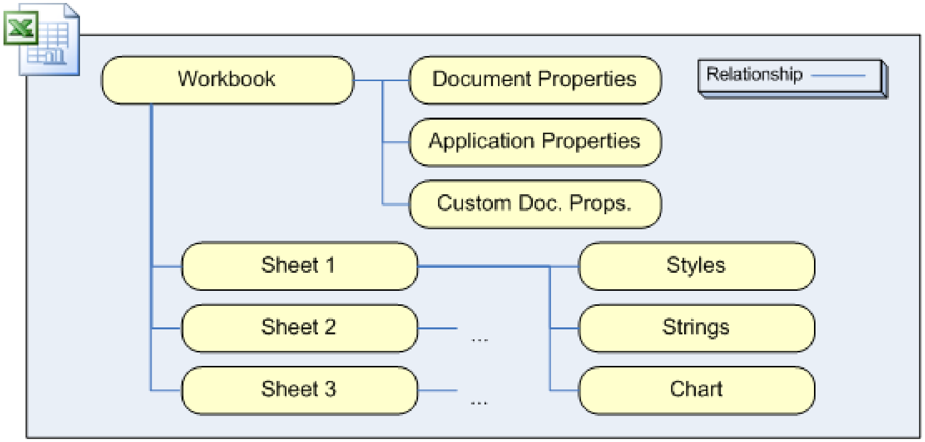
- **drill-down into XML structure**
- **structure, not content?**

file:///C:/Program Files (x86)/Open XML SDK/V2.5/tool/HelpPage.htm
*"...The **tool** integrates the following functionalities: Automatically generate Open XML SDK code based on document content. User will be able to directly copy, compile and run the code to re-generate the same document or specific parts of the documents..."*
(? not tried yet)



```
{} COMMON.xlsx
  [] /docProps/app.xml
  [] /docProps/core.xml
  [] /xl/workbook.xml
    [] /xl/theme/theme1.xml
    [] /xl/worksheets/sheet2.xml
    [] /xl/worksheets/sheet1.xml
      <> x:worksheet (Worksheet)
        <> x:dimension (SheetDimension)
        <> x:sheetViews (SheetViews)
        <> x:sheetFormatPr (SheetFormatProperties)
        <> x:cols (Columns)
        <> x:sheetData (SheetData)
        <> x:phoneticPr (PhoneticProperties)
        <> x:printOptions (PrintOptions)
        <> x:pageMargins (PageMargins)
        <> x:headerFooter (HeaderFooter)
    [] /xl/sharedStrings.xml
    [] /xl/styles.xml
    <> x:workbook (Workbook)
      <> x:fileVersion (FileVersion)
      <> x:workbookPr (WorkbookProperties)
      <> x:bookViews (BookViews)
      <> x:sheets (Sheets)
        <> x:sheet (Sheet)
        <> x:sheet (Sheet)
```

Open XML - Structure

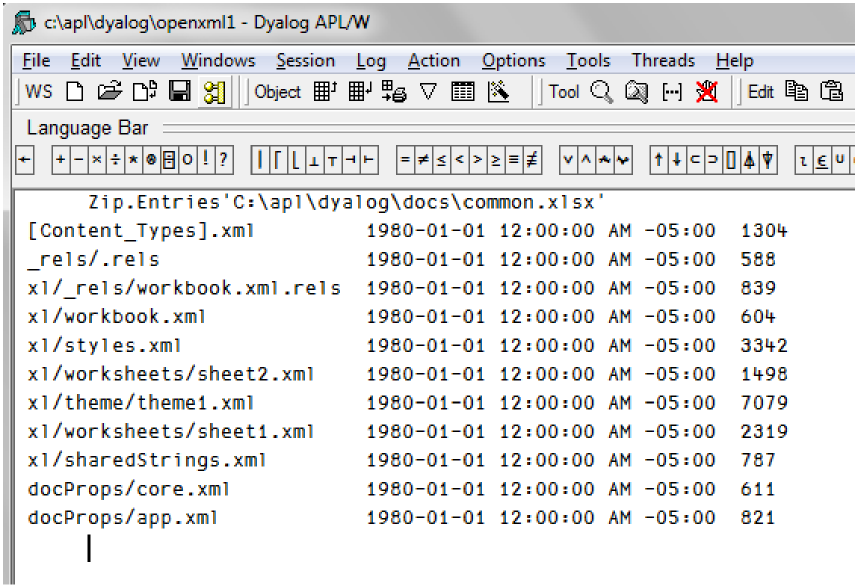


from: http://www.developerfusion.com/article/6170/read-and-write-open-xml-files-ms-office-2007/
- for a good explanation of XML internal structures and specifics, tag definitions, etc.

See also: "Open XML Explained e-book"
http://openxmldeveloper.org/blog/b/openxmldeveloper/archive/2007/08/13/1970.aspx

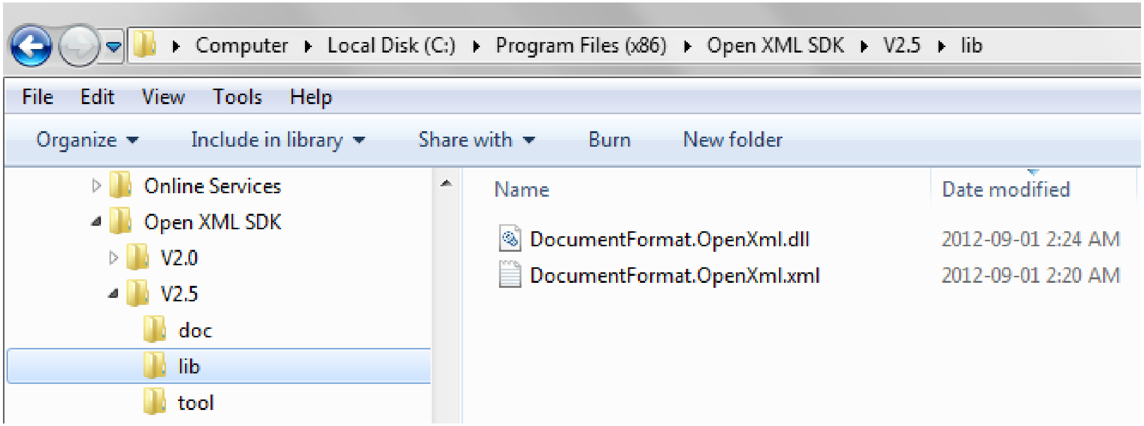
Open XML - Structure (from APL)

:USING System.IO.Compression (zip.dyalog by DanB)


c:\apl\dyalog\openxml1 - Dyalog APL/W
File Edit View Windows Session Log Action Options Tools Threads Help
WS Object Tool Edit
Language Bar

```
      Zip.Entries'C:\apl\dyalog\docs\common.xlsx'
[Content_Types].xml          1980-01-01 12:00:00 AM -05:00  1304
_rels/.rels                  1980-01-01 12:00:00 AM -05:00  588
xl/_rels/workbook.xml.rels   1980-01-01 12:00:00 AM -05:00  839
xl/workbook.xml              1980-01-01 12:00:00 AM -05:00  604
xl/styles.xml                1980-01-01 12:00:00 AM -05:00  3342
xl/worksheets/sheet2.xml     1980-01-01 12:00:00 AM -05:00  1498
xl/theme/theme1.xml          1980-01-01 12:00:00 AM -05:00  7079
xl/worksheets/sheet1.xml     1980-01-01 12:00:00 AM -05:00  2319
xl/sharedStrings.xml         1980-01-01 12:00:00 AM -05:00  787
docProps/core.xml            1980-01-01 12:00:00 AM -05:00  611
docProps/app.xml             1980-01-01 12:00:00 AM -05:00  821
```

Open XML SDK - .Net Library


Computer ▸ Local Disk (C:) ▸ Program Files (x86) ▸ Open XML SDK ▸ V2.5 ▸ lib
File Edit View Tools Help
Organize ▾ Include in library ▾ Share with ▾ Burn New folder
Online Services
Open XML SDK
V2.0
V2.5
doc
lib
tool

Name	Date modified
DocumentFormat.OpenXml.dll	2012-09-01 2:24 AM
DocumentFormat.OpenXml.xml	2012-09-01 2:20 AM

```
⎕USING,←⊂',C:\Program Files\Open XMLSDK\V2.5\lib\DocumentFormat.OpenXml.dll'
```

Open XML SDK (or similar) - Product Integration

numerous products built with this, or in similar fashion to it by making use of the Open Office Xml format, some free, some not, some out-of-date, eg.

- **EPPLus - epplus.codeplex.com - uses System.IO.Packaging namespace**
- **github.com/dfinke/ImportExcel - uses EPPLus.dll**
- **GemBoxSoftware - www.gemboxsoftware.com/support/articles/read-write-excel-spreadsheet-net**
- **EASYXLS - www.easyxls.com/**
- **Simple OOXML - http://simpleooxml.codeplex.com/ (2010)**
- **ExcelPackage - excelpackage.codeplex.com - (last update 2007?)**

- **MatLab - http://www.mathworks.com/help/matlab/ref/xlsread.html**
- **R - XLConnect - http://blog.datacamp.com/r-tutorial-read-excel-into-r/**

- **popular Blog: http://openxmldeveloper.org/**

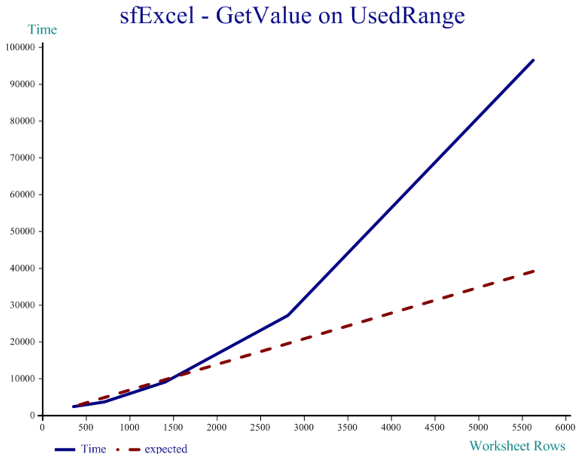
Welcome to the new

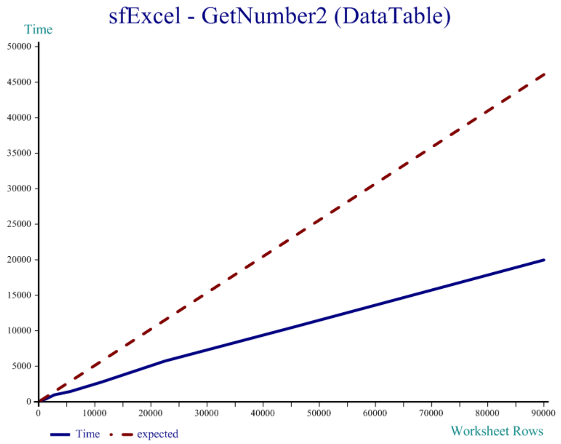
OpenXMLDeveloper.org

If you're interested in learning more about Open XML Formats, you're in the right place. OpenXmlDeveloper.org is your source for Open XML Formats news and information, a place to get your questions answered and share your insights as well. It's all free, and open to everyone. This site is regularly updated by a variety of contributors, so check back soon to see what's new!

Issues (XlsIO and Open XML SDK)

1. Speed - for sfExcel (XlsIO) - using a "DataTable" may help, BUT... what about datatype?

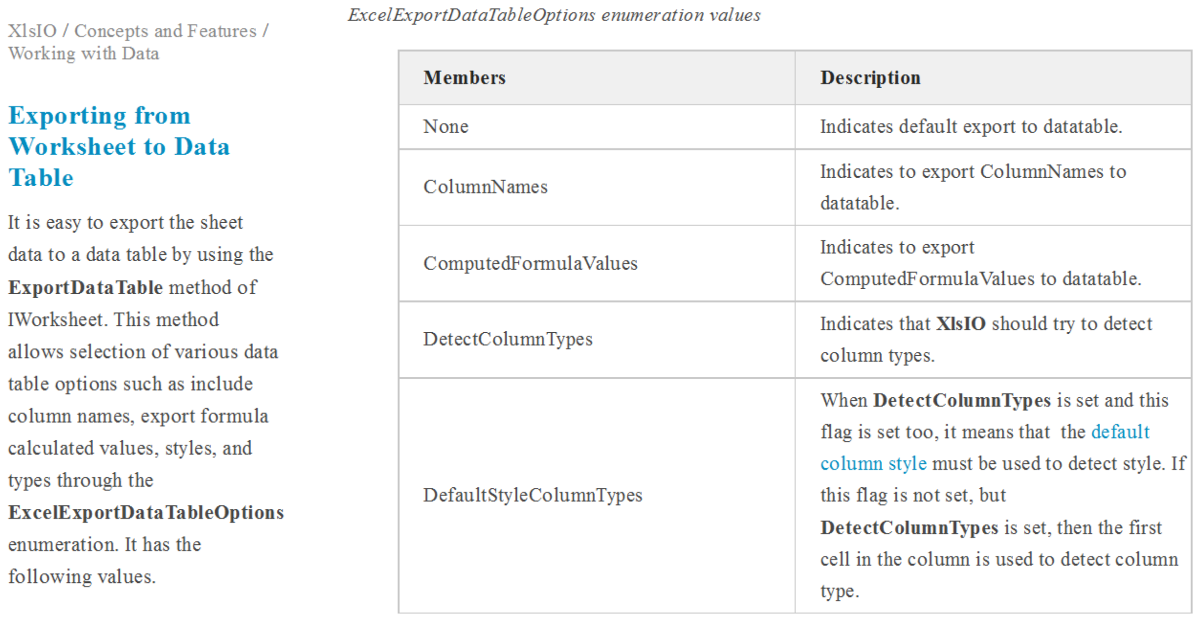




DataSet & DataTable (...represents one table of in-memory relational data):
https://msdn.microsoft.com/en-us/library/t31h6yhs%28v=vs.110%29.aspx

XlsIO - DataTable usage

http://docs.syncfusion.com/winrt/xlsio/working-with-data#exporting-from-worksheet-to-data-table



XlsIO / Concepts and Features / Working with Data

Exporting from Worksheet to Data Table

It is easy to export the sheet data to a data table by using the **ExportDataTable** method of IWorksheet. This method allows selection of various data table options such as include column names, export formula calculated values, styles, and types through the **ExcelExportDataTableOptions** enumeration. It has the following values.

ExcelExportDataTableOptions enumeration values

Members	Description
None	Indicates default export to datatable.
ColumnNames	Indicates to export ColumnNames to datatable.
ComputedFormulaValues	Indicates to export ComputedFormulaValues to datatable.
DetectColumnTypes	Indicates that **XlsIO** should try to detect column types.
DefaultStyleColumnTypes	When **DetectColumnTypes** is set and this flag is set too, it means that the default column style must be used to detect style. If this flag is not set, but **DetectColumnTypes** is set, then the first cell in the column is used to detect column type.

DataTable I-Beams - 2010⌶, 2011⌶

Dyalog Technical Note #1: (circa 2010)

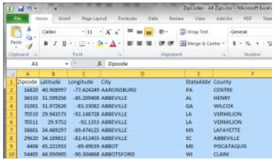
`SetDT←2010⌶`	**Set DataTable contents**
`GetDT←2011⌶`	**Get DataTable contents**

"Optimisations to Support the Microsoft.Net System.Data.DataTable Class"

System.Data.DataTable is a central object in the ADO.NET library in the Microsoft.Net Framework. It is used by DataSet and DataView objects as a container for data which has been extracted from, or will be written to, an ADO.NET data source.

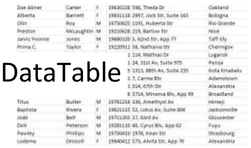
"...because we need to "loop" on each row – doing a noticeable amount of work each time. The new I-Beam does all the looping in compiled code..."

- Rapid conversion of source data to APL arrays
- compromises on datatype, easily overcome in APL
- sfExcel fns: GetNumber2, GetText2, SetText2...
- dt.Select(query) = select rows "on the way in" (to APL)?



	A	B	C	D	E	F
1	Zipcode	Latitude	Longitude	City	StateAbbr	County
2	16820	40.908997	-77.424249	AARONSBURG	PA	CENTRE
3	36310	31.599256	-85.209408	ABBEVILLE	AL	HENRY
4	31001	31.972626	-83.33082	ABBEVILLE	GA	WILCOX
5	70510	29.943573	-92.148728	ABBEVILLE	LA	VERMILION
6	70511	29.9752	-92.1353	ABBEVILLE	LA	VERMILION
7	38601	34.489297	-89.474125	ABBEVILLE	MS	LAFAYETTE
8	29620	34.189812	-82.412455	ABBEVILLE	SC	ABBEVILLE
9	4406	45.221933	-69.49039	ABBOT	ME	PISCATAQUIS
10	54405	44.950905	-90.304868	ABBOTSFORD	WI	CLARK

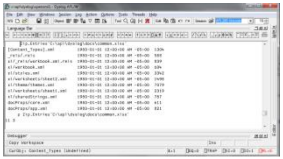
Excel



Zoe Abner	Carter	F	19630228	598, Theda Dr	Oakland
Alberta	Barnett	F	19831118	2997, Jock Str, Suite 163	Bologna
Olin	Roy	M	19750625	1191, Huberta Str	Rio Grande
Preston	McLaughlin	M	19210628	219, Barlow Str	Nice
Jarvis Yvonne	Jones	M	19680520	5, 62nd Str, App 77	Taff-Ely
Prima C.	Taylor	F	19220911	58, Nathania Str	Chernigov
			1	114, Mathias Dr	Lugansk
			1	24, 31st Av, Suite 975	Penza
			5	1313, 88th Av, Suite 235	Kota Kinabalu
			1	7, Carma Blv	Adamstown
			1	514, 67th Str	Alexandria
			6	3714, Minerva Blv, App 99	Broadland
Titus	Butler	M	19761216	336, Amethyst Av	Himeji
Baptista	Rivera	F	19621115	52, Lotus Av, Suite 804	Jacksonville
Joab	Bell	M	19711203	37, 63rd Av	Gloucester
Dirk	Peterson	M	19281110	40, Cyrus Blv, App 62	Fuyu
Pawley	Phillips	M	19750410	1978, Kean Str	Strasbourg
Lodema	Driscoll	F	19940412	573, Alvita Str, App 70	Alexandria

DataTable

APL



```
      Zip.Entries 'C:\apl\dyalog\docs\common.xlsx'
[Content_Types].xml          1980-01-01 12:00:00 AM -05:00  1304
_rels/.rels                  1980-01-01 12:00:00 AM -05:00  588
xl/_rels/workbook.xml.rels   1980-01-01 12:00:00 AM -05:00  839
xl/workbook.xml              1980-01-01 12:00:00 AM -05:00  604
xl/styles.xml                1980-01-01 12:00:00 AM -05:00  3342
xl/worksheets/sheet2.xml     1980-01-01 12:00:00 AM -05:00  1498
xl/theme/theme1.xml          1980-01-01 12:00:00 AM -05:00  7079
xl/worksheets/sheet1.xml     1980-01-01 12:00:00 AM -05:00  2319
xl/sharedStrings.xml         1980-01-01 12:00:00 AM -05:00  787
docProps/core.xml            1980-01-01 12:00:00 AM -05:00  611
docProps/app.xml             1980-01-01 12:00:00 AM -05:00  821
      ⍴ Zip.Entries 'C:\apl\dyalog\docs\common.xlsx'
11 3
```

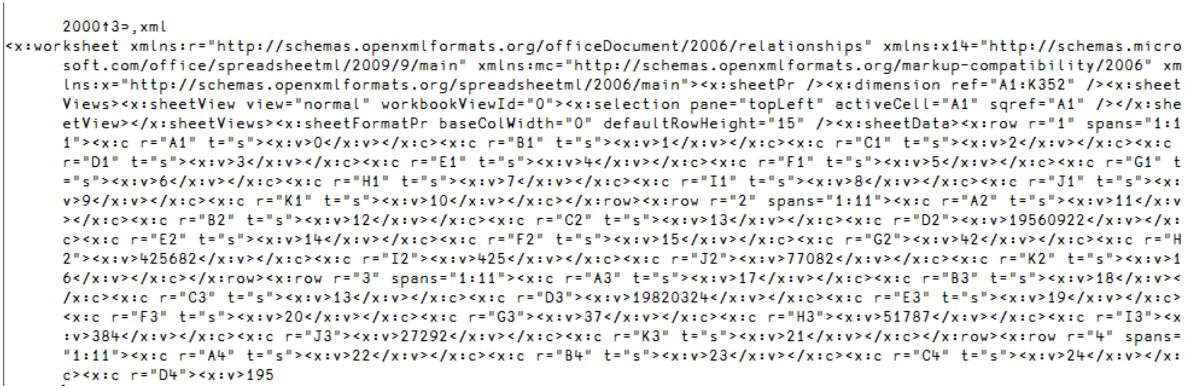
Open XML SDK - Issues

- **not many, really, other than complexity, since it offer lots of options**
- **(*guessing...*) similar performance issues as above for large tables** (*when used in "classic OO" manner*)
- **finding online examples is easy (C#), but adapting such code is challenging (for me)**
- **for really large files, blogs indicate workarounds using "Streaming" methods (?)**
- **is it really necessary? ie. if we are just pulling out entire worksheet xml vectors...**
- **may be helpful on inclusions (charts, etc.) - to be investigated; or just sorting out what the workbook contains if unknown?**
- **work in progress**

Using Open XML with APL - *what is the best strategy?*

since we "have" the data...

```
      ⍴xml← {Excel/OpenXML worksheet object}...Worksheet.OuterXml
159478
      2000↑xml
```



```
      2000↑3⊃,xml
<x:worksheet xmlns:r="http://schemas.openxmlformats.org/officeDocument/2006/relationships" xmlns:x14="http://schemas.micro
      soft.com/office/spreadsheetml/2009/9/main" xmlns:mc="http://schemas.openxmlformats.org/markup-compatibility/2006" xm
      lns:x="http://schemas.openxmlformats.org/spreadsheetml/2006/main"><x:sheetPr /><x:dimension ref="A1:K352" /><x:sheet
      Views><x:sheetView view="normal" workbookViewId="0"><x:selection pane="topLeft" activeCell="A1" sqref="A1" /></x:she
      etView></x:sheetViews><x:sheetFormatPr baseColWidth="0" defaultRowHeight="15" /><x:sheetData><x:row r="1" spans="1:1
      1"><x:c r="A1" t="s"><x:v>0</x:v></x:c><x:c r="B1" t="s"><x:v>1</x:v></x:c><x:c r="C1" t="s"><x:v>2</x:v></x:c><x:c
      r="D1" t="s"><x:v>3</x:v></x:c><x:c r="E1" t="s"><x:v>4</x:v></x:c><x:c r="F1" t="s"><x:v>5</x:v></x:c><x:c r="G1" t
      ="s"><x:v>6</x:v></x:c><x:c r="H1" t="s"><x:v>7</x:v></x:c><x:c r="I1" t="s"><x:v>8</x:v></x:c><x:c r="J1" t="s"><x:
      v>9</x:v></x:c><x:c r="K1" t="s"><x:v>10</x:v></x:c></x:row><x:row r="2" spans="1:11"><x:c r="A2" t="s"><x:v>11</x:v
      ></x:c><x:c r="B2" t="s"><x:v>12</x:v></x:c><x:c r="C2" t="s"><x:v>13</x:v></x:c><x:c r="D2"><x:v>19560922</x:v></x:
      c><x:c r="E2" t="s"><x:v>14</x:v></x:c><x:c r="F2" t="s"><x:v>15</x:v></x:c><x:c r="G2"><x:v>42</x:v></x:c><x:c r="H
      2"><x:v>425682</x:v></x:c><x:c r="I2"><x:v>425</x:v></x:c><x:c r="J2"><x:v>77082</x:v></x:c><x:c r="K2" t="s"><x:v>1
      6</x:v></x:c></x:row><x:row r="3" spans="1:11"><x:c r="A3" t="s"><x:v>17</x:v></x:c><x:c r="B3" t="s"><x:v>18</x:v><
      /x:c><x:c r="C3" t="s"><x:v>13</x:v></x:c><x:c r="D3"><x:v>19820324</x:v></x:c><x:c r="E3" t="s"><x:v>19</x:v></x:c>
      <x:c r="F3" t="s"><x:v>20</x:v></x:c><x:c r="G3"><x:v>37</x:v></x:c><x:c r="H3"><x:v>51787</x:v></x:c><x:c r="I3"><x
      :v>384</x:v></x:c><x:c r="J3"><x:v>27292</x:v></x:c><x:c r="K3" t="s"><x:v>21</x:v></x:c></x:row><x:row r="4" spans=
      "1:11"><x:c r="A4" t="s"><x:v>22</x:v></x:c><x:c r="B4" t="s"><x:v>23</x:v></x:c><x:c r="C4" t="s"><x:v>24</x:v></x:
      c><x:c r="D4"><x:v>195
```

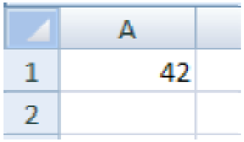
(thanks DanB...)

Open XML → APL

extract items from the "XML vector"... insert "shared strings"...

The *row* element defines a new row. Normally the first row in the *sheetData* is the first row in the visible sheet. There are optimizations that can alter this behavior, which is discussed later on in the chapter. Inside the row you create new cells using the *c* element.
Values for cells can be provided by storing a *v* element inside the cell. Usually the *v* element will contain the current value of the worksheet cell. To store a numeric value in the spreadsheet, all you have to do is to include its value in the *v* element.

```
<worksheet xmlns="http://.../spreadsheetml/2006/main" >
  <sheetData>
    <row>
      <c>
        <v>42</v>
      </c>
    </row>
  </sheetData>
</worksheet>
```

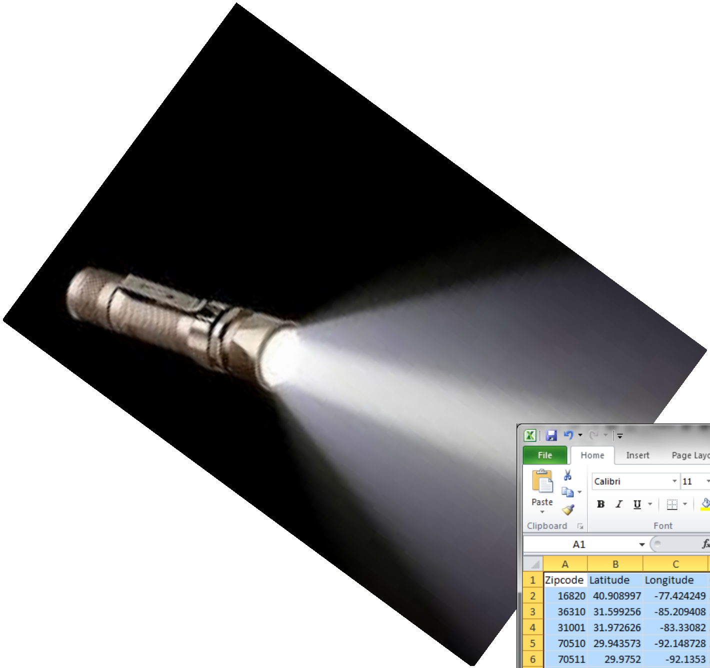


	A
1	42
2	

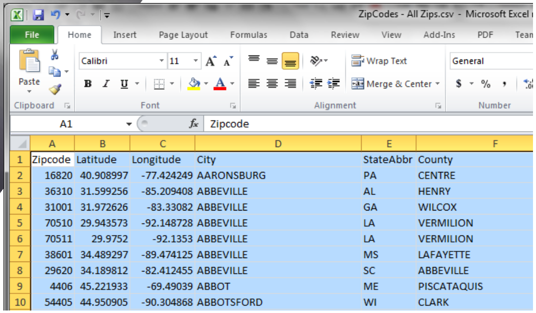
(from: "Open XML Explained": http://openxmldeveloper.org/..../Open-XML-Explained.pdf)

Open XML → APL - issues and further work

- ⎕XML vs. direct string manipulations? regular expressions?
- &amp; etc. ie. xml escapes
- missing cell references, special-case dates, N/As, etc.
- inclusions (charts, shapes, tables, etc.)
- selection on (named?) ranges, worksheets, etc.
- clever coding for DataTable and datatype issues (several options)
- APL write to Excel - more to be investigated



Thank you!



	A	B	C	D	E	F
1	Zipcode	Latitude	Longitude	City	StateAbbr	County
2	16820	40.908997	-77.424249	AARONSBURG	PA	CENTRE
3	36310	31.599256	-85.209408	ABBEVILLE	AL	HENRY
4	31001	31.972626	-83.33082	ABBEVILLE	GA	WILCOX
5	70510	29.943573	-92.148728	ABBEVILLE	LA	VERMILION
6	70511	29.9752	-92.1353	ABBEVILLE	LA	VERMILION
7	38601	34.489297	-89.474125	ABBEVILLE	MS	LAFAYETTE
8	29620	34.189812	-82.412455	ABBEVILLE	SC	ABBEVILLE
9	4406	45.221933	-69.49039	ABBOT	ME	PISCATAQUIS
10	54405	44.950905	-90.304868	ABBOTSFORD	WI	CLARK