

## HOW TO TRANSFORM XBRL DATA INTO USEFUL INFORMATION

**Donald Joyner, Norfolk State University Norfolk, USA [dtjoyner@nsu.edu](mailto:dtjoyner@nsu.edu)**

When the SEC mandated the usage of XBRL in 2009, its objectives included making information easier to analyze, increasing the overall usefulness of disclosures, and improved ability to make comparisons between companies, periods, and industries (SEC 2009). To that end, the SEC adopted a new reporting standard called eXtensible Business Reporting Language (XBRL). Since 2009, the SEC has required publicly traded companies to file their financial statements in XBRL format. XBRL was instituted to address the issues that cause stock mispricing. Unfortunately, the sheer volume of XBRL information released by the SEC in its raw format, make it difficult for the average investor to utilize without the aid of expensive programs.

If one were to examine an XBRL data set on the SEC website (<https://www.sec.gov/dera/data/financial-statement-data-sets.html>) the complexity of such data is obvious. In its raw form from the SEC website, isn't useful unless the data is manipulated. All necessary information is included in the information one can download from the SEC, but it isn't usable without some method of converting the raw data into a usable form. XBRL is described as "a machine-interpretable format that provides ways of exploiting and integrating data in an easy and comprehensive way" (<https://www.xbrl.org>). Wenger et al. (2014) note that the full potential of XBRL can only be realized if it is used in conjunction with software that "can locate and use specific items of financial information without human help."

The manner in which information is disseminated plays a crucial role in the way stock markets behave. According to Yuan, et al. (2014) most stock market analysts do not believe in the efficient market hypothesis. Fama (1970) defines the efficient market hypothesis as "a market in which prices provide accurate signals for resource allocation, that is, a market in which firms can make production-investment decisions, and investors can choose among the securities that represent ownership of firms' activities under the assumption that security prices at any time 'fully reflect' all available information." If the stock market actually behaved in accordance with the efficient market hypothesis, there would be no way to produce any predictions that the market hasn't already taken into account. If the stock price is always correct and always will be correct, there would be no risk in the market. However, that is clearly not the case.

Every day, investors make and lose money based on trading activities. Consequently, as observed by Tjung, et al. (2012), the efficient market hypothesis "has been vigorously challenged by behavioral finance since the 1980's." From a practical standpoint, the stock market will almost certainly never operate in a manner consistent with the efficient market hypothesis. In order for stocks to always be perfectly priced, every bit of consequential information would have to be known at all times. It's highly unlikely this level of information will ever be achieved. For one, companies have a legitimate reason to limit the amount of information they provide to the public. They have no obligation to reveal their strategy and internal estimates because it would benefit competitors. Stock prices rarely reflect the true value of the company they represent. Typically, a stock will either be undervalued or overvalued. The main reason for the difference between market price and actual value can be attributed to information asymmetry. Information asymmetry occurs because investors buy and sell stocks with imperfect information. Information asymmetry can be reduced by providing investors with more timely and more standardized information. XBRL is designed to address and reduce information asymmetry as it provides faster and more efficient access to financial information.

XBRL is based on a comprehensive set of rules which specify how to classify different financial items. XBRL preparation is based on Generally Accepted Accounting Principles (GAAP) taxonomies which specify how different data should be tagged and classified. Elam, et al. (2014) define XBRL taxonomies as a "universal set of schemas." They are the product of a cooperative effort amongst the SEC, Financial Accounting Standards Board (FASB), and other regulatory bodies. In terms of world-wide implications, International Financial Reporting Standards are also adopting XBRL technology. Since XBRL is intended to provide users with timely information based upon a single, commonly shared format, it is well-suited to reduce information asymmetry.

Hodge and Pronk (2006) identify the internet as "an increasingly important source of financial information for investors." They also identify the steps people apply to new information. The first step is information acquisition. It

is the manner in which people gather information. The second step is information evaluation. This entails how one interprets the information, specifically how one uses the information to draw conclusions about a company. The third step is information assimilation. This addresses the manner in which individuals combine current and previous information to reach an overall conclusion about a company's condition, performance, and prospects. Investors buy or sell a stock based on their overall conclusions.

A study by Kim, et al. (2012) examined the impact of XBRL on the financial information environment. The results of their study indicate that XBRL creates "an increase in information efficiency, a decrease in event return volatility, and reduction of change in stock returns volatility." In addition to that, they also found that XBRL "mitigates information risk in the market" even in instances where there is uncertainty in the information environment. Kim, et al. (2012) examined 428 firms (with a combined 1,536 10-K and 10-Q filings) that filed financial information after the SEC mandated use of XBRL. The results of their study showed that XBRL is having the impact the SEC desired. They showed that XBRL disclosure "decreases information risk and information asymmetry in both general and uncertain information environments."

According to Reformat and Yager (2015) information in XBRL form "does not unburden the users from looking at the data, analyzing it, and eventually drawing conclusions indicating the financial statuses of corporations." Since XBRL is an XML-based language, there are already a number of programs and tools which can be used to analyze it. Geiger, et al. (2014) discusses the ability of XBRL to tag all financial statement concepts. This includes financial statements, footnotes, and schedules. Footnotes and supplemental schedules to the financial statements are more difficult to analyze than the data associated with financial statements. They conclude "when corporate disclosures (including financial statements, footnotes, and management discussion, analysis, and forecasts) are equally available and analyzable for all interested market participants" greater market efficiency is the result.

This study focuses only on the numerical values contained within XBRL data.

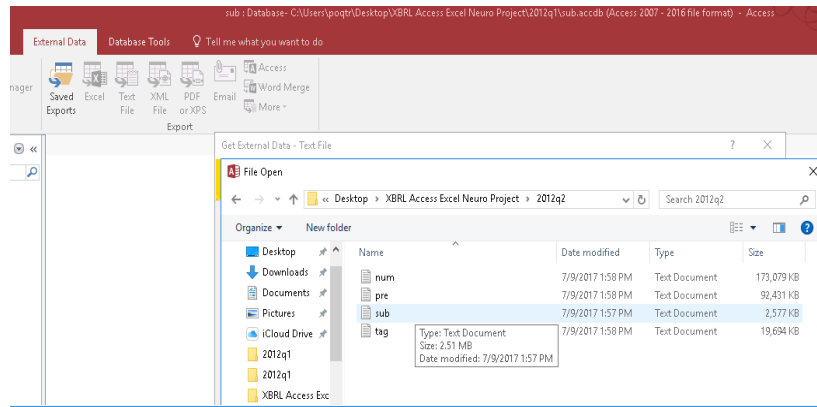
The SEC releases XBRL Data Sets quarterly. Each data set is comprised of four files: SUB, TAG, NUM, and PRE. Each file contains information pertaining to all companies that have filed XBRL data with the SEC for the specified time frame. The information is presented in the form it was presented by each filing company to the SEC, so it is possible this information may contain errors or inconsistencies. Each company that submits XBRL data to the SEC is assigned a unique identifying number.

The SUB file contains basic company information such as Company Names, Central Index Key #s, SIC #s, the period being reported on, Ticker Symbol, and date the data was submitted to the SEC. The TAG file contains all tags contained in each submission. A 'tag' is the name a company assigns to a particular line item in their financial statements (ex. revenue, cost of goods sold, depreciation, etc.). All tags must be labeled using a common taxonomy. Each company may call similar line items different names. The usage of a common taxonomy is intended to correct differences in terminology and make information between companies easier to compare. The NUM file displays the numerical values assigned to each tag. The PRE file contains the text of each line item as well as the order in which each line item appears in each financial statement. For purposes of this study only the SUB and NUM files are used.

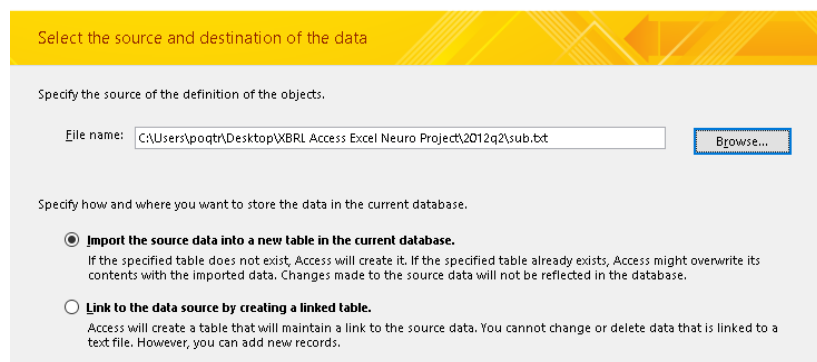
To transform a Quarterly XBRL filing into usable data, both Microsoft Access and Microsoft Excel need to be used. Excel is limited to only 1,048,576 rows whereas Access isn't. Some files contain more than the maximum number of rows Excel can handle, so it is easier to manipulate some of the data before importing it into Excel. The following is a step-by-step guide for transforming XBRL data into a useful Excel spreadsheet.

First, download the quarterly data you wish to analyze from the SEC website: <https://www.sec.gov/dera/data/financial-statement-data-sets.html>. Each quarter is in a zip file containing the 4 files discussed above. Once these files are unzipped, open Microsoft Access. Select External Data and choose 'From File' and choose 'Text File.'

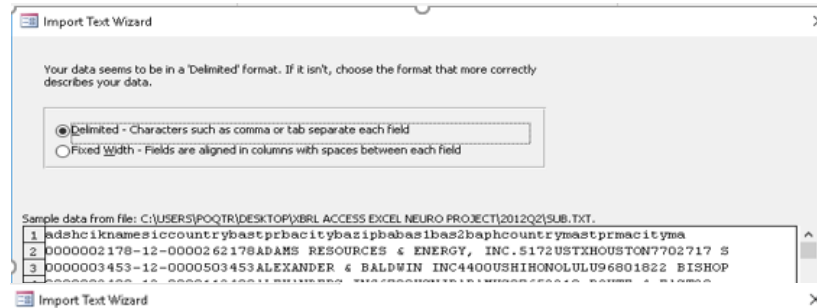
Select the *Sub* and *Num* files from the appropriate Year and Quarter



Select 'Import the source data into a new table in the current database.'



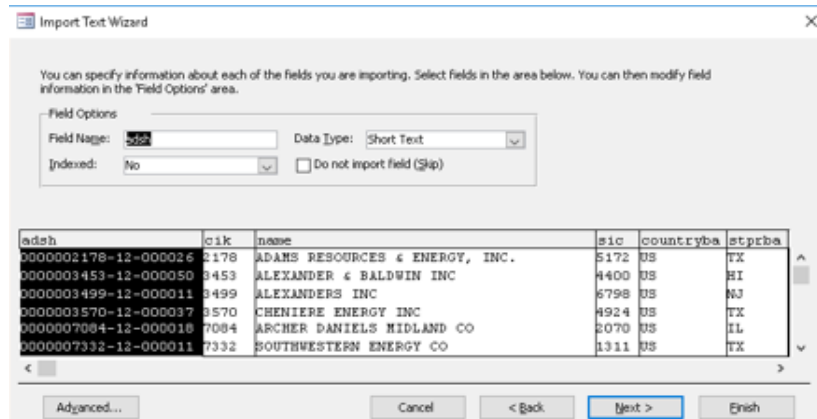
Select 'Delimited' and click 'Next.'



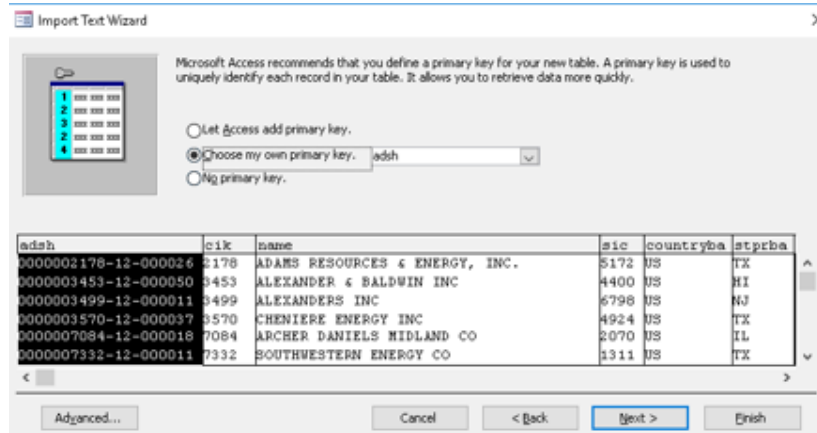
Check 'First Row Contains Field Names' and click 'Next'



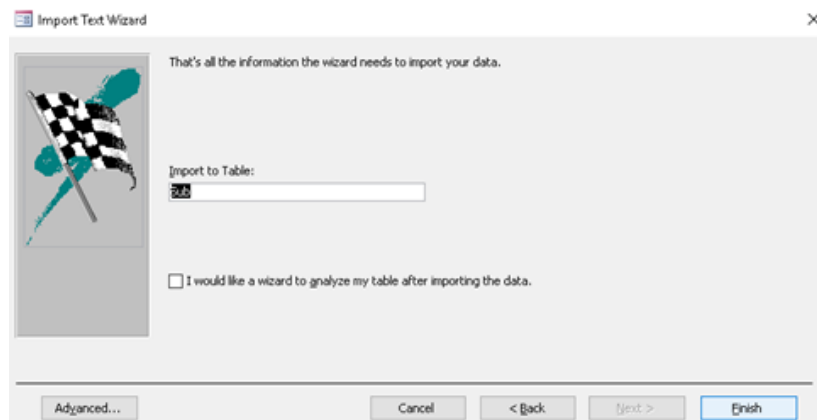
Make sure Field Name for *adsh* is 'Short Text' and click 'Next'



Make sure *adsh* is selected as the 'Primary Key' and select 'Next'



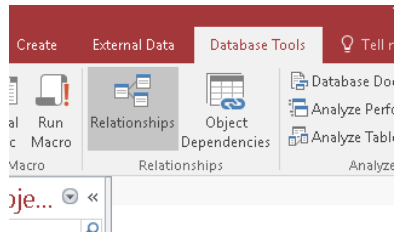
Import to Table: *sub* or *num* depending on the file being imported. Click 'Finish'



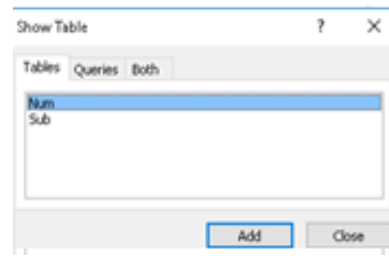
Delete '*num\_ImportErrors*'



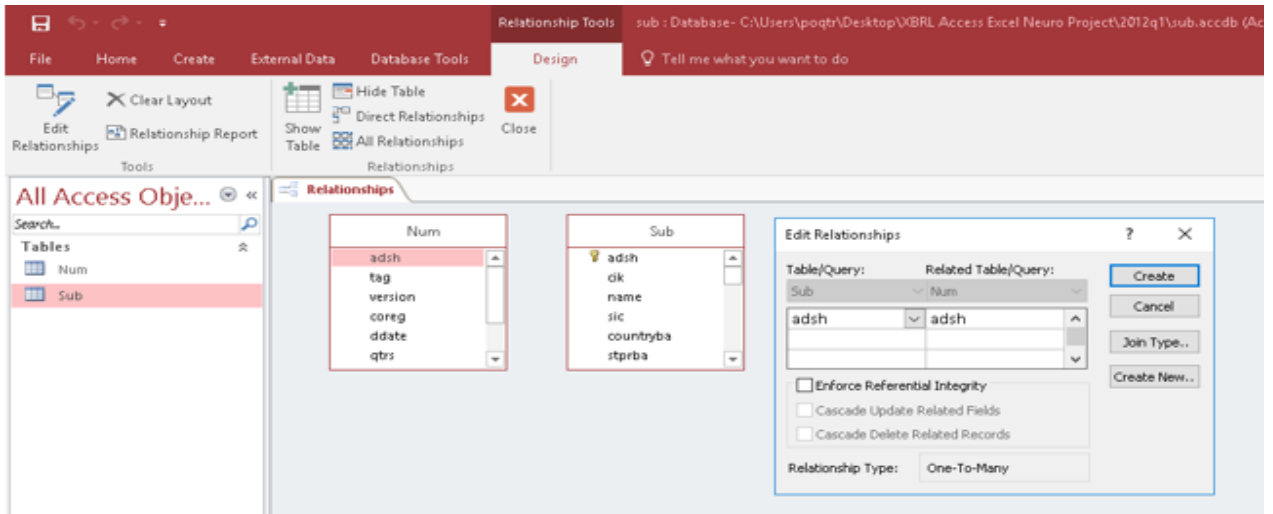
Select Database Tools and click 'Relationships'



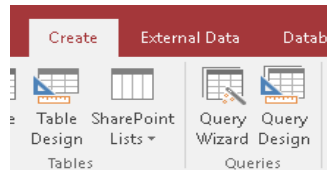
Then Add BOTH *num* and *sub*



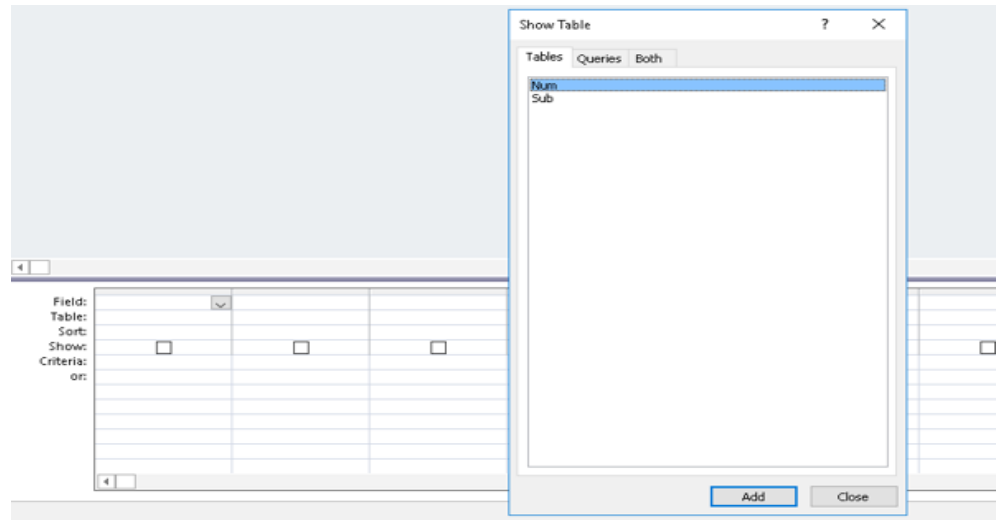
Create a link between the table based on the *adsh* column. Click 'create'



Create a new Query using 'Query Design'



Add both tables and click 'Add'



From the *sub* table add these items in the following order: adsh, name, cik, sic, countryinc, form, fye, period, fy, and instance

Field:	adsh	name	cik	sic	countryinc	form	fye	period	fy	Instance
Table:	Sub	Sub	Sub	Sub	Sub	Sub	Sub	Sub	Sub	Sub
Sort:										
Show:	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Criteria:										

From the *num* table add these items in the following order: version, qtrs, tag, uom, and value

version	qtrs	tag	uom	value
Num	Num	Num	Num	Num
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

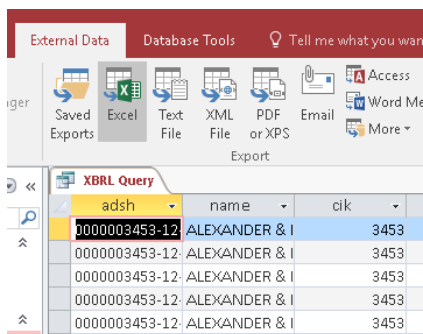
Save as XBRL Query

Add the following to the Query: countryinc, form, fye, period, fy, qtrs, tag, uom

countryinc	form	fye	period	fy	qtrs	tag	uom
Sub	Sub	Sub	Sub	Sub	Num	Num	Num
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
"US"	"10-Q"	331 Or 630 Or 930 Or	20120331	2012	Q Or 1		"USD" Or "shares"

NOTE: period has to be changed each quarter, example: 201XXXXX  
 201X1231 = For 1<sup>st</sup> Quarter (the year is one year before the fiscal year, ex. 20111231 for 2012)  
 201X0331 = For 2<sup>nd</sup> Quarter  
 201X630 = For 3<sup>rd</sup> Quarter  
 201X0930 = For 4<sup>th</sup> Quarter

Save as XBRL Database and Query. This query will work for each quarterly XBRL release. The only things that need to be adjusted are the Fiscal Year and Period. Open this file and adjust for each period. Save the Query as an Excel workbook. Use the name '201X QX Query' for each set of data. It should also be noted that the query eliminates many entries and reduces the number of rows contained in the data to far less than maximum number of rows Excel can handle.



Beforehand, an Excel spreadsheet called Data Worksheet.xlsx was created in order to manipulate the data contained in the query. The next step is to open this spreadsheet and import the Query data into it. The main problem posed by raw XBRL format is that it is in a row format. This worksheet is comprised of 6 sheets: XBRL\_Query, Data Staging, # of companies, @vlookup, Results, and 201XQX. The XBRL\_Query sheet is where the query data for the period is imported into the worksheet. The Query sheet contains one row for each tag used by each company. If a company has entries for Cash and Equivalents, Net Inventory, Accounts Receivable, and Property, Plant, and Equipment, etc. there will be a row for each entry. The data for a single company can be over 100 rows. The chief aim of the Data Worksheet is to create a final sheet that takes all company data and puts it into a single row with multiple columns.

Once the Query information is imported into the worksheet, the data is imported into the Data Staging sheet. The Data Staging sheet takes the information contained in the Query sheet and combines it into two columns. The first column combines the company's adsh number, the company's name, and the tag used for a particular row. The second column displays the numerical value associated with that row.

M	N
Combined ID	value
0000007084-12-000010ARCHER DANIELS MIDLAND COAccountsPayableCurrent	2581000000.0000
0000007084-12-000010ARCHER DANIELS MIDLAND COAccountsPayableCurrent	4136000000.0000
0000007084-12-000010ARCHER DANIELS MIDLAND COAccountsReceivableNetCurrent	4808000000.0000
0000007084-12-000010ARCHER DANIELS MIDLAND COAccountsReceivableNetCurrent	5093000000.0000

Data Staging Sheet

The # of companies sheet takes the information from the Query sheet and calculates how many individual companies are represented. It also displays key information associated with each company.

A	B	C	D	E	F	G
#	adsh	name	tag	cik	sic	ticker
1	0000007084	ARCHER DANIELS MIDLAND CO	AccountsPayableCurrent	7084	2070	adm
2	0000008670	AUTOMATIC DATA PROCESSING INC	AccountsPayableCurrent	8670	7374	adp
3	0000021510	COHERENT INC	AccountsPayableCurrent	21510	3826	cohr
4	0000038777	FRANKLIN RESOURCES INC	AccountsPayableCurrent	38777	6282	ben
5	0000047307	HICKOK INC	AccountsPayableCurrent	47307	3823	hicka

# of Companies Sheet

The @vlookup sheet then combines every possible tag with the adsh and name for each company.

A	B	C	D	E	F
adsh	name	InventoryNet	AssetsCurrent	AccountsReceivableNetCurrent	
0000007084-12-000010	ARCHER DANIELS MIDLAND CO	0000007084-12-000010ARCHER DANIELS MIDLAND COInventoryNet	0000007084-12-000010ARCHER DANIELS MIDLAND COAssetsCurrent	0000007084-12-000010ARCHER DANIELS MIDLAND COAccountsReceivableNetCurrent	
0000008670-12-000002	AUTOMATIC DATA PROCESSING INC	0000008670-12-000002AUTOMATIC DATA PROCESSING INCInventoryNet	0000008670-12-000002AUTOMATIC DATA PROCESSING INCAssetsCurrent	0000008670-12-000002AUTOMATIC DATA PROCESSING INCAccountsReceivableNetCurrent	
0000021510-12-000004	COHERENT INC	0000021510-12-000004COHERENT INCInventoryNet	0000021510-12-000004COHERENT INCAssetsCurrent	0000021510-12-000004COHERENT INCAccountsReceivableNetCurrent	
0000038777-12-000007	FRANKLIN RESOURCES INC	0000038777-12-000007FRANKLIN RESOURCES INCInventoryNet	0000038777-12-000007FRANKLIN RESOURCES INCAssetsCurrent	0000038777-12-000007FRANKLIN RESOURCES INCAccountsReceivableNetCurrent	
0000047307-12-000007	HICKOK INC	0000047307-12-000007HICKOK INCInventoryNet	0000047307-12-000007HICKOK INCAssetsCurrent	0000047307-12-000007HICKOK INCAccountsReceivableNetCurrent	
0000055772-12-000005	KIMBALL INTERNATIONAL INC	0000055772-12-000005KIMBALL INTERNATIONAL INCInventoryNet	0000055772-12-000005KIMBALL INTERNATIONAL INCAssetsCurrent	0000055772-12-000005KIMBALL INTERNATIONAL INCAccountsReceivableNetCurrent	
0000063296-12-000008	MATTHEWS INTERNATIONAL CORP	0000063296-12-000008MATTHEWS INTERNATIONAL CORPInventoryNet	0000063296-12-000008MATTHEWS INTERNATIONAL CORPAssetsCurrent	0000063296-12-000008MATTHEWS INTERNATIONAL CORPAccountsReceivableNetCurrent	

@vlookup Sheet

The @vlookup sheet is the first step into converting the data into column format. The Results sheet takes every value computed in the @vlookup sheet and displays its associated number from the Data Staging Sheet. If there is a number associated with the value in the @vlookup sheet, it will be displayed in the Results sheet.

A	B	C	D	E	F	G	H	I	J	K	L	M
adsh	name	tag	cik	sic	ticker	CashAndCashEquivalentsAtCarryingValue	InventoryNet	AssetsCurrent	AccountsReceivableNetCurrent	PropertyPlantAndEquip	OtherAssetsNoncurrent	
0000007084-12-000010	ARCHER DANIELS MIDLAND CO	AccountsPayableCurrent	7084	2070	adm	#N/A	2750400000.0000	4808000000.0000	9500000000.0000	6810000000.0000	42193000000.0000	
0000008670-12-000002	AUTOMATIC DATA PROCESSING INC	AccountsPayableCurrent	8670	7374	adp	#N/A	28583500000.0000	13648000000.0000	7162000000.0000	9226000000.0000	34239300000.0000	
0000021510-12-000004	COHERENT INC	AccountsPayableCurrent	21510	3826	cohr	1523950000.0000	5806460000.0000	1410370000.0000	1045040000.0000	641920000.0000	8432660000.0000	
0000038777-12-000007	FRANKLIN RESOURCES INC	AccountsPayableAndAccruedLiabilitiesCurrent	38777	6282	ben	#N/A	76167680000.0000	#N/A	5897480000.0000	1312780000.0000	137758480000.0000	
0000047307-12-000007	HICKOK INC	AccountsPayableCurrent	47307	3823	hicka	20508510.0000	2959318.0000	482878.0000	468893.0000	#N/A	3469061.0000	

Results Sheet

The results sheet displays calculated values. Therefore, sorting and data analysis wouldn't work as it could change the formulas. The last sheet 201XQX is a copy of the Results sheet, but everything is pasted as a number instead of a formula. The Results sheet can be used to apply statistical analyses. It can also be used to group companies together and compare companies directly.

While a time consuming and tedious task, each new quarterly release of XBRL data can be subjected to the processes described in this study. Microsoft Access and Excel are far cheaper than commercially available software. It only takes about an hour for the combined Access and Excel functions to compute the results sheet. If necessary, this process could be refined to provide more specific information.

The SEC adopted the XBRL format to facilitate the acquisition and analysis of financial data. Depending on the needs of the user, raw XBRL data can be interpreted and formatted in such a way as to provide a meaningful format. More and more users may be inclined to develop their own systems for processing XBRL data as it is a far cheaper alternative. Also, this study uses only 2 of the 4 files associated with quarterly XBRL data releases. There is no reason this method could not be modified to produce information related to notes to the financial statements and other non-numerical disclosures.

**References:**

Elam, R., Wenger, M. R., & Williams, K. L. (2012). XBRL Tagging of Financial Statement Data Using XMLSpy: The Small Company Case. *Issues in Accounting Education*, 27(3), 761-781. doi:10.2308/iace-50162

Fama, E. (1991) Efficient capital markets: II. *Journal of Finance*, 46(5), pp. 1575-1617.

Geiger, M. A., North, D. S., & Selby, D. D. (2014). Releasing Information In Xbrl: Does It Improve Information Asymmetry For Early U. S. Adopters? *Academy of Accounting & Financial Studies Journal*, 18(4), 66-83.

Hodge, F., & Pronk, M. (2006). The Impact of Expertise and Investment Familiarity on Investors' Use of Online Financial Report Information. *Journal Of Accounting, Auditing & Finance*, 21(3), 267-292.

Kim, J. W., Jee-Hae, L., & Won Gyun, N. (2012). The Effect of First Wave Mandatory XBRL Reporting across the Financial Information Environment. *Journal of Information Systems*, 26(1), 127-153.

Reformat, M. Z., & Yager, R. R. (2015). Soft Computing Techniques for Querying XBRL Data. *Intelligent Systems in Accounting, Finance & Management*, 22(3), 179-199.

Securities and Exchange Commission (SEC). 2009. Interactive Data to Improve Financial Reporting. Release Nos. 33-9002; 34-59324; 39-2461; IC-28609; File No. S7-11-08. Available at: <https://www.sec.gov/rules/final/2009/33-9002.pdf>

Tjung, L. C., Ojong, K., & Tseng, K. C. (2012). Comparison study on neural network and ordinary least squares model to stocks' prices forecasting. *Academy of Information & Management Sciences Journal*, 15(1), 1-35.

Wenger, M. R., R. Elam, and K. L. Williams. 2013. A tour of five XBRL tools: Products that help make tagged data work for you and your clients. *Journal of Accountancy*. 215(4):48-55.

Yuan, Y., Peasnell, K., Lubberink, M., & Hunt III, H. G. (2014). Determinants of Analysts' Target P/E Multiples. *Journal of Investing*, 23(3), 35-42.