

March 29, 2016

# STRUCTURED DATA ANALYTICS



Mike Willis

Assistant Director

Office of Structured Disclosure

Division of Economic and Risk Analysis

# Disclaimer

The Securities and Exchange Commission, as a matter of policy, disclaims responsibility for any private publication or statement by any of its employees. Therefore, the views expressed today are my own, and do not necessarily reflect the views of the Commission or the other members of the staff of the Commission.



# Discussion Outline

- Where did we come from
- Where are we going
- How does it work
- Analytical Examples
  - Renderer
  - Corporate Issuer Risk Assessment
  - Inline viewer
  - Financial Statement Query Viewer
  - Text Analytics
  - Third Party Applications
- What we need to improve



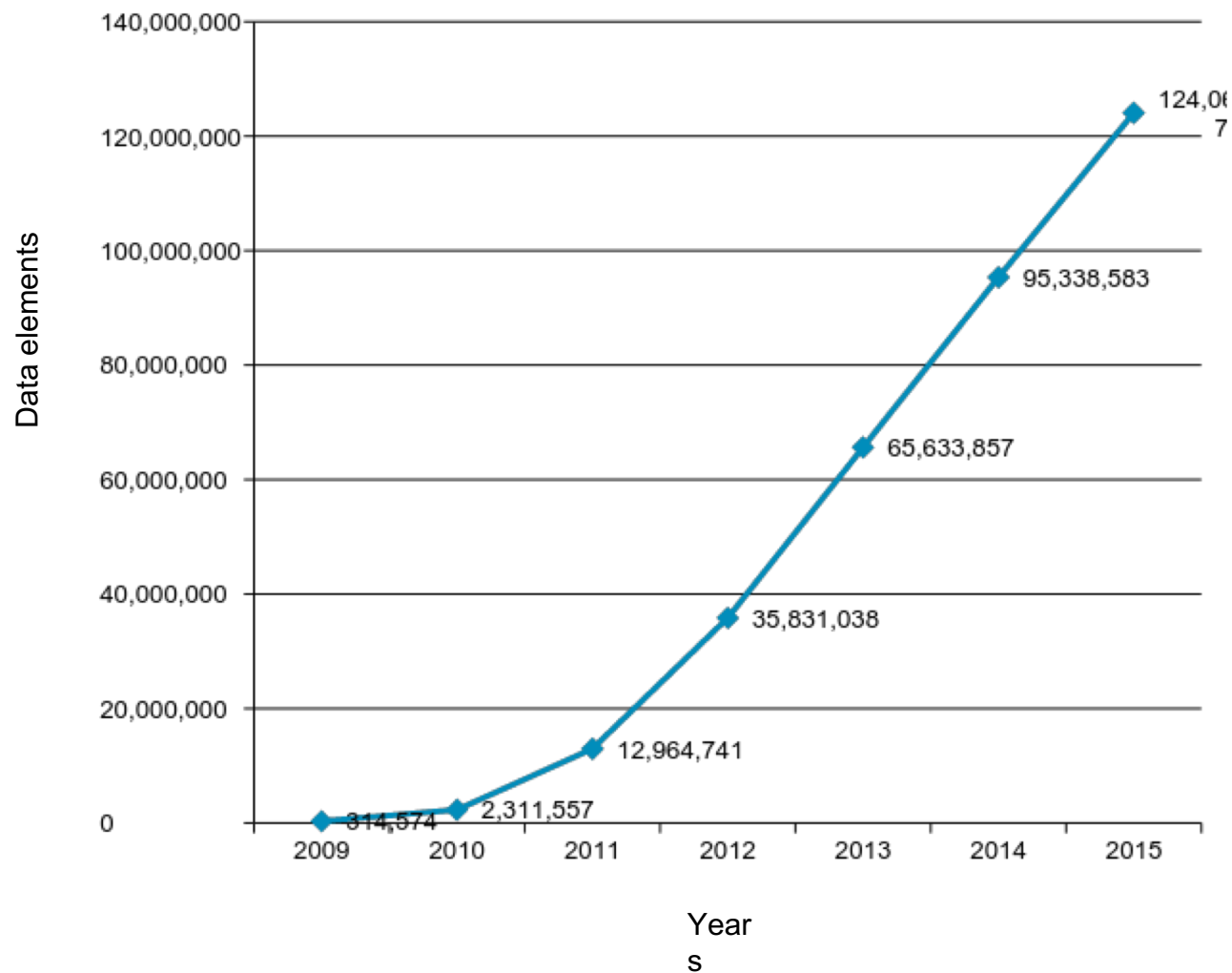
# SEC's Use of Structured Data Formats

- In 2008, the Commission initially adopted custom XML
  - Form D for Exempt Offerings of Securities
  - Forms 3, 4, and 5 for these insider ownership disclosures.
- In 2009, the first use of XBRL by the Commission for all public companies after a voluntary filing period that began in 2005.
  - Now also extends to Mutual Fund risk/return summaries, NRSROs, and SDRs
- In 2014, adopted rule using custom XML for Reg. AB governing the offering process, disclosure, and reporting for asset-backed securities (“ABS”).
- In 2015, proposed use of FpML and FIXML for SDRs to make SBS data available to the Commission
- The Commission’s structured disclosures reflect a range of structures:
  - Machine-readable, web-fillable PDF
  - Specialized knowledge of a particular software reporting language

# What is Structured Data

- Standardized Information – Format & Semantics
- Machine-readable
  - Organized so that the information can be automatically recognized and processed by software
  - Information can then be easily organized for viewing, manipulation, and analysis
  - Enables advanced search capabilities based on definitional information in the data tags
- How is it different from human-readable
- Not static, text-based information
- Users can discern patterns from large quantities of information
- Can it be both machine-readable and human-readable?

# Accelerating Volume of Structured Data



# Growth in Structured Forms & Filings

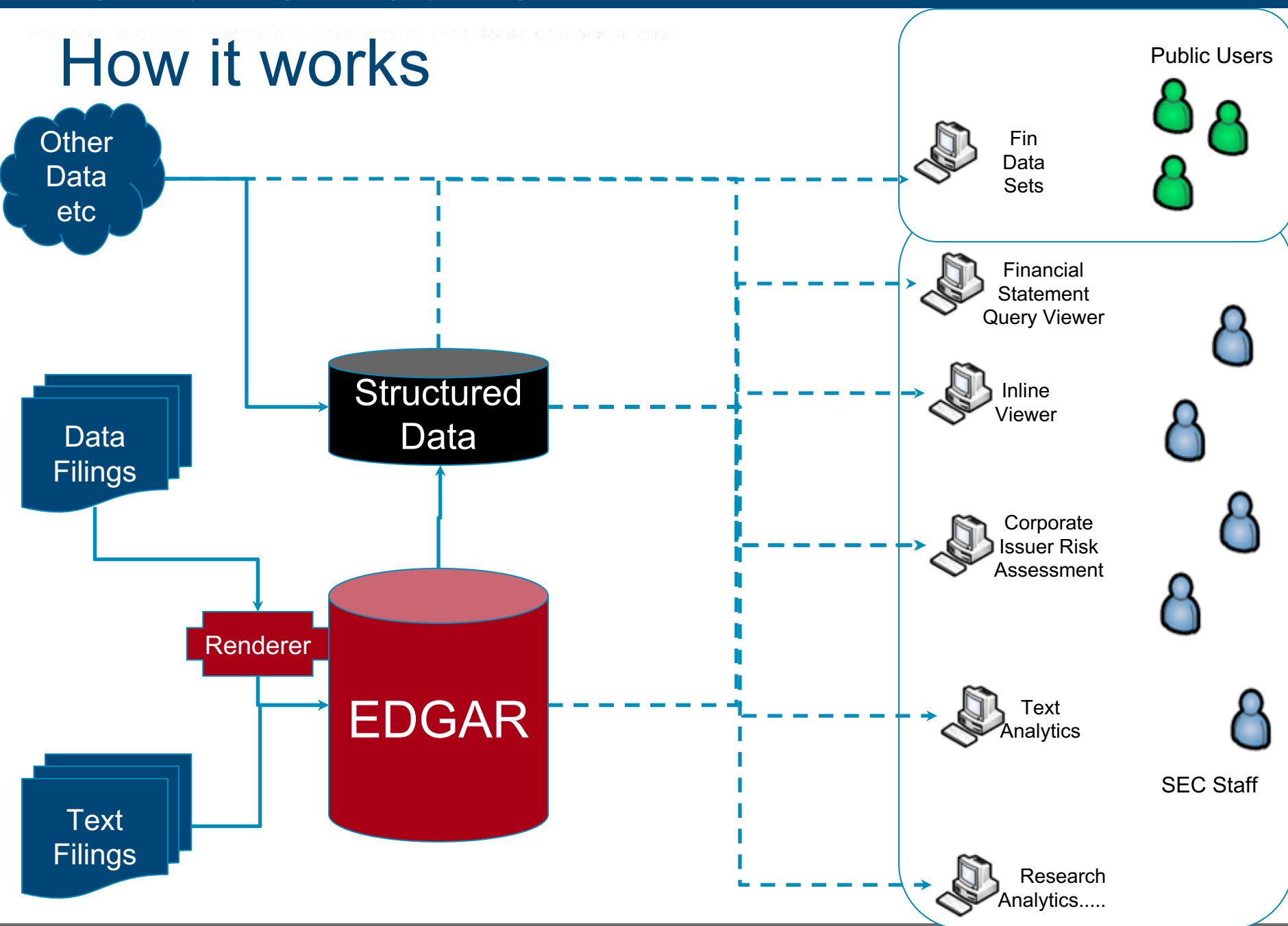
## Proposed rules:

- Disclosure of Payments by Resource Extraction Issuers
  - (Proposed December 11, 2015)
- Establishing the Form and Manner with which Security-Based Swap Data Repositories Must Make Security-Based Swap Data Available to the Commission
  - (Proposed December 11, 2015)
- Listing Standards for Recovery of Erroneously Awarded Compensation
  - (Proposed July 1, 2015)
- Investment Company Reporting Modernization
  - (Proposed May 20, 2015)
- Pay Versus Performance
  - (Proposed April 29, 2015)

## Final Rules:

- Security-Based Swap Data Repository Registration, Duties, and Core Principles
  - (Adopted February 11, 2015)
- Asset-Backed Securities Disclosure and Registration
  - (Adopted September 4, 2014)
- Nationally Recognized Statistical Rating Organizations
  - (Adopted August 27, 2014)
- Money Market Fund Reform; Amendments to Form PF
  - (Adopted July 23, 2014)
- Crowdfunding
  - (Adopted October 31, 2015)

# How it works



# Renderer

- ‘Renders’ from machine readable to human readable presentation
- Applies rules to assess compliance with data/technical requirements
- Same rules are available within the Previewer that should be used prior to filing
- Persistent assessment of all SD filings
- Provides warning and error messages
- Capacity for incremental application of additional rules

# Renderer

INCOME STATEMENTS - USD (\$) shares in Millions, \$ in Millions		12 Jun. 30, 2015
Revenue	\$ 93,580	
Cost of revenue	33,038	
Gross margin	60,542	
Research and development	12,046	
Sales and marketing	15,713	
General and administrative	4,611	
Impairment, integration, and restructuring	10,011	
Operating income	18,161	
Other income, net	346	
Income before income taxes	18,507	
Provision for income taxes	6,314	
Net income	\$ 12,193 <sup>[1]</sup>	
<b>Earnings per share:</b>		
Basic	\$ 1.49	
Diluted	\$ 1.48 <sup>[1]</sup>	
<b>Weighted average shares outstanding:</b>		
Basic	8,177	
Diluted	8,254	
Cash dividends declared per common share	\$ 1.24	

```

<us-gaap:PaymentsForRestructuring contextRef="eol_PE8528----1510-K0009_STD_365_20150630_0" unitRef="iso4217_USD"
decimals="-6" id="id_6568047_52AF2ACF-E9E6-4AE5-8446-74A2CB287309_5003_3">835000000</us-
gaap:PaymentsForRestructuring>
<us-gaap:OperatingIncomeLoss contextRef="eol_PE8528----1510-K0009_STD_365_20150630_0" unitRef="iso4217_USD"
decimals="-6" id="id_6568047_F7A26265-66FD-401E-B561-BB7F8F6F188F_1_7">18161000000</us-gaap:OperatingIncomeLoss>
<us-gaap:OtherComprehensiveIncomeLossAvailableForSaleSecuritiesAdjustmentNetOfTax contextRef="eol_PE8528----1510-
K0009_STD_365_20150630_0" unitRef="iso4217_USD" decimals="-6" id="id_6568047_26682282-9633-4E55-93C6-
8B0C08CCB2F7_1_4">-362000000</us-gaap:OtherComprehensiveIncomeLossAvailableForSaleSecuritiesAdjustmentNetOfTax>
<us-gaap:IncomeLossFromContinuingOperationsBeforeIncomeTaxes foreign contextRef="eol_PE8528----1510-
K0009_STD_365_20150630_0" unitRef="iso4217_USD" decimals="-6" id="id_6568047_C8982409-6AF9-418D-A766-
ABCD3EBA626_1_1">11144000000</us-gaap:IncomeLossFromContinuingOperationsBeforeIncomeTaxes foreign>
<us-gaap:IncreaseDecreaseInInventories contextRef="eol_PE8528----1510-K0009_STD_365_20150630_0" unitRef="iso4217_USD"
decimals="-6" id="id_6568047_2F147908-337E-4197-A624-CD7058360B39_1_13">272000000</us-
gaap:IncreaseDecreaseInInventories>
<us-gaap:IncomeTaxesPaidNet contextRef="eol_PE8528----1510-K0009_STD_365_20150630_0" unitRef="iso4217_USD" decimals="-6"
id="id_6568047_76B79994FB0C_1_2">1091000000</us-gaap:IncomeTaxesPaidNet>
<us-gaap:SalesRevenueNet contextRef="eol_PE8528----1510-K0009_STD_365_20150630_0" unitRef="iso4217_USD" decimals="-6"
id="id_6568047_F7A26265-66FD-401E-B561-BB7F8F6F188F_1_0">93580000000</us-gaap:SalesRevenueNet>
unitRef="iso4217_USD" decimals="-6" id="id_6568047_2F147908-337E-4197-A624-CD7058360B39_1_23">1500000000</us-
gaap:RepaymentsOfDebtMaturingInMoreThanThreeMonths>
<us-gaap:IncreaseDecreaseInAccountsReceivable contextRef="eol_PE8528----1510-K0009_STD_365_20150630_0"
unitRef="iso4217_USD" decimals="-6" id="id_6568047_2F147908-337E-4197-A624-CD7058360B39_1_12">-1456000000</us-
gaap:IncreaseDecreaseInAccountsReceivable>
<us-gaap:UnrecognizedTaxBenefitsIncreasesResultingFromCurrentPeriodTaxPositions contextRef="eol_PE8528----1510-
K0009_STD_365_20150630_0" unitRef="iso4217_USD" decimals="-6" id="id_6568047_B51837B0-B5C2-4416-A57C-
76B79994FB0C_1_2">1091000000</us-gaap:UnrecognizedTaxBenefitsIncreasesResultingFromCurrentPeriodTaxPositions>
<us-gaap:IncreaseDecreaseInOtherCurrentAssets contextRef="eol_PE8528----1510-K0009_STD_365_20150630_0"
unitRef="iso4217_USD" decimals="-6" id="id_6568047_2F147908-337E-4197-A624-CD7058360B39_1_14">-620000000</us-
gaap:IncreaseDecreaseInOtherCurrentAssets>
<us-gaap:RecognitionOfDeferredRevenue contextRef="eol_PE8528----1510-K0009_STD_365_20150630_0" unitRef="iso4217_USD"
decimals="-6" id="id_6568047_2F147908-337E-4197-A624-CD7058360B39_1_10">44920000000</us-
gaap:RecognitionOfDeferredRevenue>
<us-gaap:OtherComprehensiveIncomeLossDerivativesQualifyingAsHedgesNetOfTax contextRef="eol_PE8528----1510-
K0009_STD_365_20150630_0" unitRef="iso4217_USD" decimals="-6" id="id_6568047_26682282-9633-4E55-93C6-
8B0C08CCB2F7_1_2">559000000</us-gaap:OtherComprehensiveIncomeLossDerivativesQualifyingAsHedgesNetOfTax>
<us-gaap:InvestmentIncomeNet contextRef="eol_PE8528----1510-K0009_STD_365_20150630_0" unitRef="iso4217_USD"
decimals="-6" id="id_6568047_36F55639-5E8C-46C9-B2BF-CC88C41BC95E_1_0">766000000</us-gaap:InvestmentIncomeNet>
<us-gaap:NonoperatingIncomeExpense contextRef="eol_PE8528----1510-K0009_STD_365_20150630_0" unitRef="iso4217_USD"

```



# Multiple Presentations

INCOME STATEMENTS - USD (\$) shares in Millions, \$ in Millions		12
		Jun. 30, 2015
Revenue	\$ 93,580	

## Earnings Release FY15 Q4

Income Statements (in millions, except per share amounts)				
	Year Ended June 30,			
	2015	2014	2013	
Revenue	\$ 93,580	\$ 86,833	\$ 77,849	
Cost of revenue	55,038	27,078	20,385	
Gross margin	60,542	59,755	57,464	
Research and development	12,046	11,381	10,411	
Sales and marketing	15,713	15,811	15,276	
General and administrative	4,611	4,677	5,013	
Impairment, integration, and restructuring	10,011	127	0	
Operating income	18,161	27,759	26,764	
Other income, net	346	61	288	
Income before income taxes	18,507	27,820	27,052	
Provision for income taxes	6,314	5,746	5,189	
Net income	\$ 12,193	\$ 22,074	\$ 21,863	
Diluted	8,254			
Cash dividends declared per common share	\$ 1.24			

gaap:PaymentsForAcquireInvestments  
<us-gaap:PaymentsForRestructuring contextRef="eol\_PE8528----1510-K0009\_STD\_365\_20150630\_0" unitRef="iso4217\_USD" decimals="-6" id="id\_6568047\_52AF2ACF-E9E6-4AE5-8446-78A2CB287309\_5003\_3">835000000</us-



alifyingAsHedgesNetOfTax contextRef="eol\_PE8528----1510-SD" decimals="-6" id="id\_6568047\_26682282-9633-4E55-93C6-comprehensiveIncomeLossDerivativesQualifyingAsHedgesNetOfTax">28----1510-K0009\_STD\_365\_20150630\_0" unitRef="iso4217\_USD" 39-B2BF-CC88C41BC95E\_1\_0">766000000</us-gaap:InvestmentIncomeNet>\_PE8528----1510-K0009\_STD\_365\_20150630\_0" unitRef="iso4217\_USD"

# Corporate Issuer Risk Assessment (CIRA)

- **Analytical tool:** provides detailed information on various aspects of a company's business activities and financial reporting environment
- **Dashboard:** enables the user to search, compare and analyze a variety of information about companies through a single intuitive visual interface
- **Identify patterns:** Helps users assess the risks associated with financial reporting with more than 200 variables for thousands of SEC registrants across multiple years
- **Approach:** Based on database approaches used by academic financial accountants and large sample evidence documented in academic literature
- **Data sources:** Uses a variety of data including the SEC's own structured data

# Inline XBRL

What is Inline:

- Common Internet method of combining file formats
- Today the tagged data exhibit (XBRL) is separate from the actual Filing (HTML)
- Inline includes both HTML and XBRL (human readable and machine readable) together in the same file

Why Inline:

- Lower filing costs
- Improve Quality
- Advanced review capabilities
- Search, highlight and filter structured disclosures

# Inline Viewer Prototype (under Development)

- **Single document** – embed structure within actual filing disclosures rather than attaching a separate structured copy
- **Familiar view** – familiar financial statement browser view rather than a separate rendered view
- **Detailed review** – easily pinpoint and review the details of a specific structured disclosure (including summary information)
- **Navigate** – quickly navigate the structured disclosure sections
- **Why Inline Viewer** – ‘heads up’ display vs. ‘texting while driving’

Public Online Investors | Investor Access | Market Integrity | Investor Education | Capital Formation

# Inline Viewer Prototype Example

## (under Development)

	September 30, 2013 (unaudited)	December 31, 2012
	(in thousands, except share and per share data)	
<b>ASSETS</b>		
<b>CURRENT ASSETS</b>		
Cash and cash equivalents	\$ 320,870	\$ 403,095
Trade accounts receivable, less allowance for doubtful accounts (2013: \$20,770; 2012: \$10,180)	1,759,517	1,490,028
Merchandise inventories, net--at lower of cost or market	2,832,108	2,602,560
Prepaid expenses and other current assets	382,980	324,448
<b>TOTAL CURRENT ASSETS</b>	5,295,475	4,820,131
Goodwill and other intangible assets, less accumulated amortization	1,280,182	497,839
Deferred tax assets	257,571	279,463
Other assets	462,172	643,263
Property, plant and equipment, less allowance for depreciation	648,503	566,365
<b>TOTAL ASSETS</b>	\$ 7,943,903	\$ 6,807,061
<b>LIABILITIES AND EQUITY</b>		
<b>CURRENT LIABILITIES</b>		
Trade accounts payable	\$ 2,213,190	\$ 1,681,900
Current portion of debt	583,896	250,000
Income taxes payable	19,176	4,354
Dividends payable	82,982	76,641
Other current liabilities	554,850	474,743
<b>TOTAL CURRENT LIABILITIES</b>	3,454,094	2,487,638
Long-term debt	250,000	250,000

Attributes

Accounts Notes And Loans Receivable Net Current

Tag	us-gaap:AccountsNotesAndLoansReceivableNetCurrent
Fact	1,759,517,000
Period	As of 9/30/2013
Measure	USD
Scale	Thousands
Decimal	Thousands
Balance	Debit
Sign	Positive

# Inline Viewer Prototype Example

## (under Development)

**Search**

h XBRL Data **Revenue** GO

☐ Match  
☐ Case Search Definitions  
☐ Search Names

**Selection Detail**

Label: us-gaap:\_Revenues \*

Value: 58,437,000,000  
Calendar: 01/01/2009 – 12/13/2009  
Unit: USD  
Scale / Precision: 6/6  
Balance / Weight: credit / 1  
Type: monetaryItem

Definition: Amount of revenue recognized from goods sold, services rendered, insurance premiums ...

**Dimension Details**

**Labels**

\* Custom Save As...

Highlight Search About

Item 8

ITEM 8. FINANCIAL STATEMENTS AND SUPPLEMENTARY DATA

INCOME STATEMENTS

(In millions, except per share amounts)

Year Ended December 31,

	2009	2008	2007
Revenue	\$58,437	\$60,420	\$51,122
Operating expenses:			
Cost of revenue	12,155	11,598	10,693
Research and development	9,010	8,164	7,121
Sales and marketing	12,879	13,260	11,541
General and administrative	3,700	5,127	3,329
Employee severance	330	-	-
Total operating expenses	38,074	38,149	32,684
Operating income	20,363	22,271	18,438
Other income (expense)	(542)	1,543	1,663
Income before income taxes	19,821	23,814	20,101
Provision for income taxes	5,252	6,133	6,036
Net income	\$14,569	\$17,681	\$14,065
Earnings per share:			
Basic	\$ 1.63	\$ 1.90	\$ 1.44
Diluted	\$ 1.62	\$ 1.87	\$ 1.42
Weighted average shares outstanding:			
Basic	8,945	9,328	9,742
Diluted	8,996	9,470	9,886
Cash dividends declared per common share	\$ 0.52	\$ 0.44	\$ 0.40

See accompanying notes.

1 of 14 Highlight Matches

# Financial Statement Query Viewer (FSQV)

## FSQV Key features:

- **User interface:** Intuitive and easy to use web browser interface with filer profile
- **Filing Search:** Quickly search across all “as-reported” financial statement XBRL filings for all US GAAP filers since 2009
- **Item Search:** Search through all financial statement line items by a variety of criteria (e.g., keyword, filer size, date, CIK, ticker, industry etc.)
- **Review and compare:** Review a filer’s financial statements and compare footnote differences between periods
- **Save results:** Save all search results locally to Excel and Word to perform additional analysis



# Text Analytics Initiative

- Text analytics and sentiment analysis also use XBRL structured disclosures.
- Textual-analytical engines are more effective at analyzing structured disclosures when they know what they are looking at versus attempting to consume and analyze an entire unstructured report
- Useful in assessing:
  - Inconsistencies in narrative disclosures
  - Inconsistencies in narrative vs numeric disclosures
  - Narrative trending
  - Changes in risk profile based on sentiment
  - Others

# Text Analytics Initiative

## Natural language processing

- Using text analytic techniques to identify ‘themes’ in unstructured narrative disclosures.
- This is done by applying statistical methods that look at distributions of (root) words in a structured document in relation to those in other unstructured documents
- Modeling methods are used to identify topics (themes) in disclosure that, once ‘trained’ on a set of structured documents, can identify similar themes in other unstructured documents.
- For example, registrant disclosures associated with misconduct can be used as a basis for training to identify similar disclosures in other documents

DERA has structured (at a text block level) 200,000 SEC filings, from 1993 to 2014 and is developing Text Analytic capabilities using this structured data.

# Text Analytics – No Ambiguity

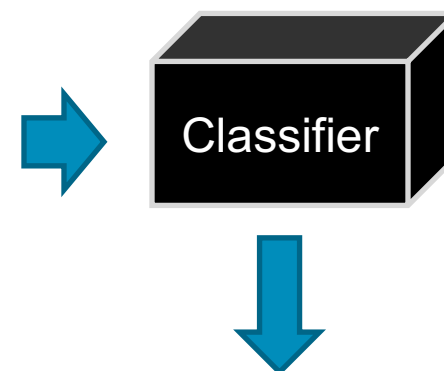
## 7. Discontinued Operations

In accordance with the provisions of ASC 360, *Property, Plant, and Equipment*, we have reclassified the revenues and expenses of properties and businesses sold or held for sale to “income (loss) from discontinued operations” and the related assets and liabilities to “assets related to discontinued operations” and “liabilities related to discontinued operations” for all periods presented in the accompanying consolidated financial statements. The net gains resulting from the sale of the properties below are included in “income (loss) from discontinued operations” on our consolidated statements of income.

On January 6, 2015, we completed the sale of 920 Holyoke Center, a 1.2 million square foot office building in Holyoke, Massachusetts, for \$238,000,000 in cash, which resulted in a net gain of \$52,200,000 that will be recognized in the first quarter of 2012.

On March 31, 2014, the receiver completed the disposition of Periphery Center in Alabama. In connection therewith, the property and related debt were removed from our consolidated balance sheet and we recognized a net gain of \$85,907,000 on the extinguishment of debt.

On January 12, 2011, we sold 110 Oregon Avenue and 127 15<sup>th</sup> Street in Washington, DC, for \$227,000,000 in cash, which resulted in a net gain of \$4,862,000.



Top 5 US GAAP Tags	P
DisposalGroupsIncludingDiscontinuedOperationsDisclosure	99 %
MergersAcquisitionsAndDispositionsDisclosures	0.08 %
OtherIncomeAndOtherExpenseDisclosure	0.0006 %
BusinessDescriptionAndAccountingPolicies	0.0004 %
LoansNotesTradeAndOtherReceivablesExcludingAllowanceForCreditLosses	0.0003 %

# Text Analytics – With Ambiguity

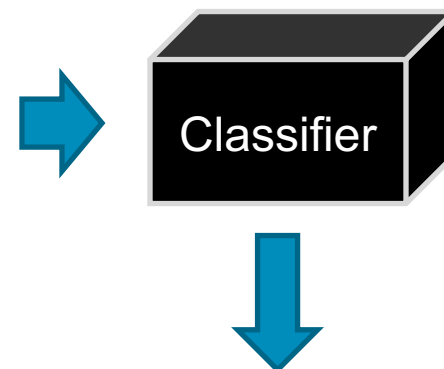
## NOTES TO CONSOLIDATED FINANCIAL STATEMENTS (CONTINUED)

### 11. Shareholders' Equity

The following table sets forth the details of our preferred shares of beneficial interest as of December 31, 2015 and 2014.

Preferred Shares	Balance as of December 31,		Shares Outstanding at December 31,		Per Share Liquidation Preference
	2015	2014	2015	2014	
6.75% Series H: auth. 4,600,000 shares	108,549	108,549	4,500,000	4,500,000	\$ 25.00
6.625% Series I: auth. 12,050,000 shares	262,379	262,379	10,800,000	10,800,000	\$ 25.00
6.875% Series J: auth. 9,850,000 shares	238,842	-	9,850,000	-	\$ 25.00
	<u>\$ 1,021,660</u>	<u>\$ 783,088</u>	<u>42,186,709</u>	<u>32,340,009</u>	

On April 20, 2011, we sold 7,000,000 6.875% Series J Cumulative Redeemable Preferred Shares at a price of \$25.00 per share, in an underwritten public offering pursuant to an effective registration statement. On April 21, 2011, the underwriters exercised their option to purchase an additional 1,050,000 shares to cover over-allotments. On May 5, 2011 and August 5, 2011 we sold additional 800,000 and 1,000,000 shares, respectively, at a price of \$25.00 per share. We retained aggregate net proceeds of \$238,842,000, after underwriters' discounts and issuance costs and contributed the net proceeds to the Operating Partnership in exchange for 9,850,000 Series J Preferred Units (with economic terms that mirror those of the Series J Preferred Shares).



US GAAP Tag	P
StockholdersEquityNoteDisclosure	68.0 %
PreferredStock	25.4 %
ShareholdersEquityAndShareBasedPayme nts	1.4 %
TreasuryStock	0.7 %
EquityMethodInvestmentsDisclosure	0.3 %

# Third Party Applications

- XBRL Subscription Service enables SEC staff to use interactive data in their regulatory work.
- Includes all public XBRL filings submitted to the SEC via the EDGAR since June 30, 2009.
- Accessible via the Internet via intuitive and user-friendly interface for accessing and analyzing the service's data via browser and an Excel application
- Robust analysis and reporting features including ratio analyses, trend analysis across time for a single company and across multiple user selected or industry-specific companies.
- Used across range of Divisions and Offices

Company Ticker	Filing Date	SIC	Retail-Catalog & Mail-Order Houses
Costco Wholesale Corporation	1/23/2016	5961	RETAIL-CATALOG & MAIL-ORDER HOUSES

Financial Data

Ratio Metrics

Topical Metrics

Company

Peers

Peers

# Common Data Quality Errors

- Company specific extensions
  - Elements
  - Axis
- Negative Values
- Scaling
- Incorrect values
- Incomplete filings
- Inconsistent dates
- Missing calculations
- No detail tagging



# SEC Data Quality Efforts

- New OSD Website periodic updates
- Dear CFO letters on calculations
- Rendering Engine Warning & Error messages
- Staff Observations: Axis Extensions, Negative values; Extensions; Scaling; others
- Publishing the data sets (listening to feedback)
- FASB US GAAP Taxonomy Simplification
- IFRS GAAP Taxonomy
- Inline XBRL Viewer
- Market education & Awareness
- Monitoring Data Quality Rules
- Others

# New OSD Website

News Updates w/  
frequent updates

Sign up for emails

Contact us

Information for:

- Data Users
- Preparers
- Technical users

Updated Content



U.S. SECURITIES AND  
EXCHANGE COMMISSION

[COMPANY FILINGS](#) | [MORE SEARCH OPTIONS](#)

[ABOUT](#) | [DIVISIONS](#) | [ENFORCEMENT](#) | [REGULATION](#) | [EDUCATION](#) | [FILINGS](#) | [NEWS](#)

Office of Structured Disclosure

Contact Us

What is Structured Data?

History and Rulemakings

Data Users >

**Preparers >**

Technical >

## Office of Structured Disclosure







**Sign up for E-mail Updates**

### OSD News

[XBRL Taxonomy Update](#)  
March 7, 2016

[SEC Interactive Data Test Suite Update 35-160213](#)  
Feb. 23, 2016

[Update to Staff Interpretations and FAQs Related to Interactive Data Disclosure](#)  
Jan. 26, 2016

[Draft 2015 FpML and FIXML Schemas for Security-Based Swap Data Repositories](#)  
Jan. 11, 2016

[Changes to the Securities and Exchange](#)



Category	2013	2014	2015
Large accelerated*	81%	78%	74%
Accelerated*	69%	65%	59%
Non accelerated*	68%	61%	52%
Smaller reporting company*	40%	38%	34%

The Office of Structured Disclosure (OSD), within the [Division of Economic and Risk Analysis](#), supports the SEC's efforts to make data accessible and easy to use. OSD works closely with other SEC Divisions and Offices to design data structuring approaches for required disclosures, and supports the SEC's data collections and data usage by designing taxonomies, validation rules, data quality assessments, and tools for conducting data analyses. OSD also works with investors, regulated entities, and the public to support the submission and use of structured data.

- [What Is Structured Data?](#)
- [History and Rulemakings](#)

Questions?