



The Global Standard for Business Research

wrds

WHARTON RESEARCH DATA SERVICES

# SEC Filings Data on WRDS

WRDS Research

May, 2020

# SEC filings are great resources for research

One-stop research platform on SEC filing

Familiarize yourself with the SEC Analytics Suite

Learn how to access information

Discover how the SEC Analytics Suite can expedite & enhance your research

The screenshot shows the WRDS SEC Analytics Suite website. The header includes navigation tabs for HOME, TOOLS, SUPPORT, E-LEARNING, COMMUNITY, NEWS, ABOUT, and MyWRDS, along with a search bar labeled 'Search WRDS'. The main content area is titled 'WRDS SEC Analytics Suite (Admin Edit - alert)'. It features a 'Select a Data Set:' dropdown menu and a list of available datasets including 'Filings Index Data', 'Filings on WRDS', 'SEC Filings Queries', 'Filings Parser', 'List of Filings Exhibits', 'List of 8K Items', 'SEC Linking Tables', 'Historical Company Names', 'K-CUSIP Link Table', and 'GVKEY-CIK Link Table'. Below this is a search bar and an 'Advanced Search' button. The main text describes the suite's capabilities, such as simplifying research by allowing users to quickly and easily develop tailored datasets from all SEC filings, parsing millions of regulatory reports based on custom criteria. It also lists benefits like daily updates, comprehensive historical linking, and flexible data extraction options. A 'Highlights' section lists features such as access to over 12 million aggregated SEC filings index records, search capabilities for various filing types, and support for multiple data output formats. The 'Features' section details mass extraction, links to original HTML, easy-to-use web queries, and compatibility with various data analysis tools like SAS, STATA, Excel, MATLAB, HTML, and SPSS.

# SEC Filings on WRDS

WRDS SEC Analytics Suite Data offerings have expanded substantially in recent years

1

WRDS SEC Analytics Suite: Filings and Metadata

2

WRDS SEC Analytics Suite: Web Queries

3

Textual Analytics and Datasets: Bag of Words/  
Readability/Sentiment

4

Datasets from Parsed XML Forms: 13F, Insiders, etc

# Why use Regulatory Filings

## **Regulatory filings are a trove of financial and accounting data**

There are over 400 different types of forms available on EDGAR – and expect more to come.

## **Go beyond what's available in Compustat**

Filings with fundamental or accounting data contain way more information than the 3 main Accounting Tables and their footnotes.

## **SEC data extraction has never been easier**

Since 2009 U.S. companies and foreign issuers must file in XBRL, a spreadsheet-like XML format for businesses.



U.S. Securities and Exchange Commission  
[www.sec.gov](http://www.sec.gov)

# WRDS SEC Analytics Suite

Centralized storage & parsing of SEC filing contents

**19.8 million+** records of electronic filings with the SEC since 1994, as well as the text, html, and pdf filings available on wrds server.

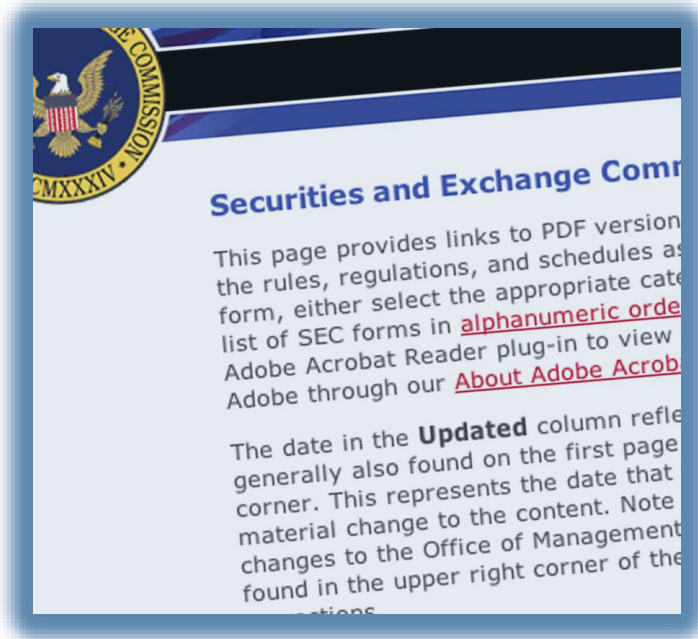
**Fast Solr search over 4 million filings** for all 10-K, 10-Q, 8-K, IPO Prospectuses, Proxy filings, and SEC Correspondences since 1994

## Derived Datasets:

- over 3.4 million 8-K events/items
- 75+ million filing exhibits for all filings
- Readability and Sentiment measures for all filings
- *Bag of Words*: word frequency distributions for all filings
- **pre-parsed data** including confirmed period of report, time of filings, historical state of incorporation + more

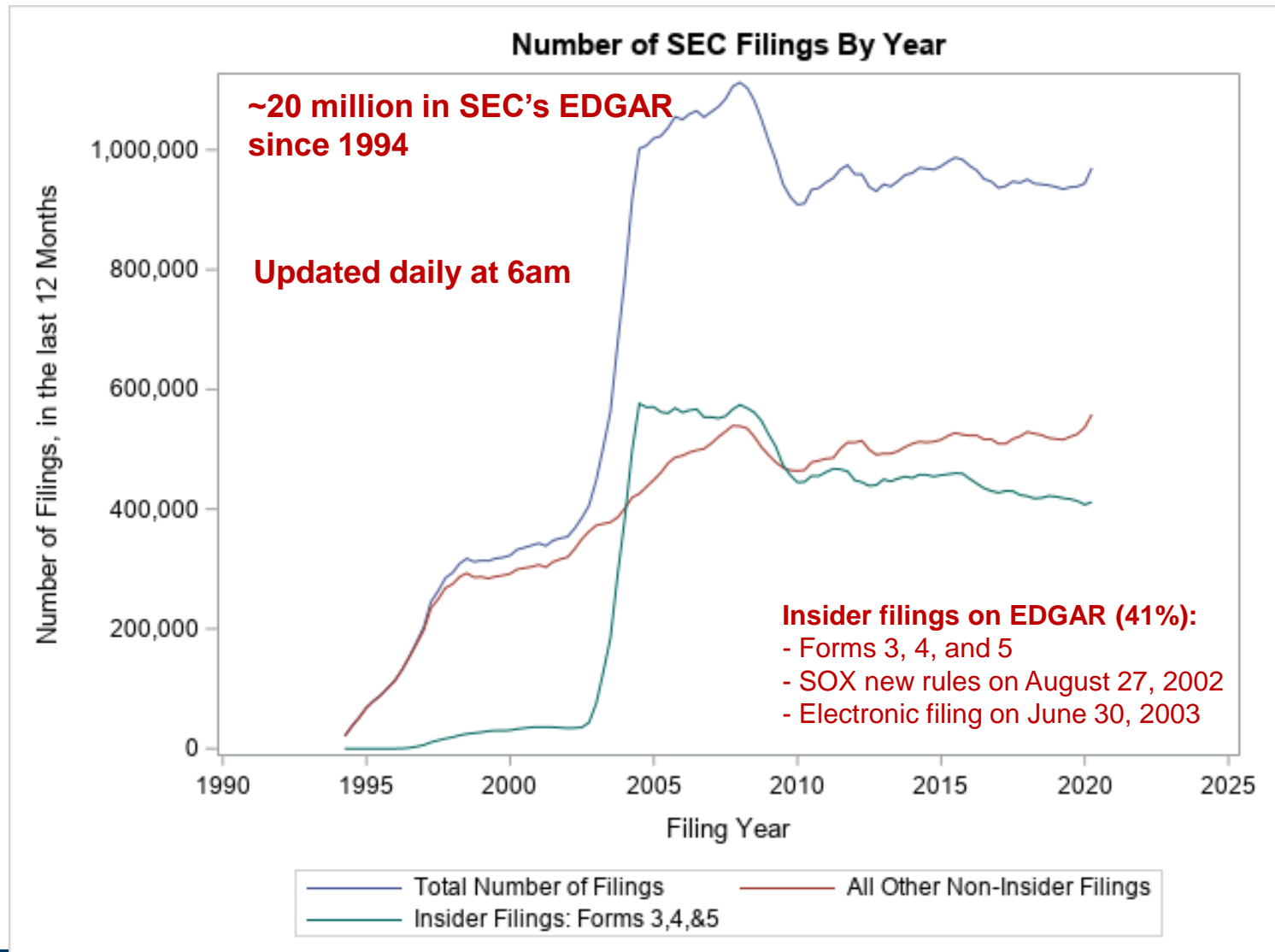
Historical **GVKEY**, **CUSIP** and **CIK** link tables

**Additional XML-based data: Insiders, 13F, + more**



# Records of all electronic filings on EDGAR

## SEC filings continue to grow every year



# SEC Filings on WRDS

**1**

**WRDS SEC: Filings and Metadata**

# SEC Filings Index Data on WRDS

## Easy access to the latest SEC filings

- The SEC Analytics Suite contains the **records of all electronic filings** with SEC since 1994
- **Over 19.8 million filings since 1994, as of June 2020**
- Filings are **updated daily at 6 a.m.**; access the previous day's filing records for all companies
- Identify **who** filed **what** and **when** + link to physical filing location
- **Monitor** new filings and reporting requirements
- After the Sarbanes-Oxley Act of 2002, electronic filings by insiders **increased**
  - **Nearly 41% of all filings** are insider filings



# All Filings Records: Identify Who filed What and When

## **WRDS\_FORMS** and **WRDS\_FORMS\_REG** datasets

Compustat Global Company Key - Header	SEC Central Index Key	Filing Date	SEC Form	Company Name	Reference Name of Complete Report Filing	Complete Report File Size	Conformed Period of Report / Fiscal Period End	SEC Acceptanc Time	Standard Industrial Classification	State of Incorporati	State of Business Address
012141	0000789019	20100128	8-K	MICROSOFT CORP	edgar/data/789019/0001193125-10-015589.txt	202923	28JAN2010	160836	7372	WA	WA
012141	0000789019	20100129	SC 13G	MICROSOFT CORP	edgar/data/789019/0001086364-10-008414.txt	15384	.	161709	7372	WA	WA
012141	0000789019	20100208	CORRESP	MICROSOFT CORP	edgar/data/789019/0001193125-10-024738.txt	24474	.	200844	7372	WA	WA
012141	0000789019	20100216	CORRESP	MICROSOFT CORP	edgar/data/789019/0001193125-10-032435.txt	26037	.	170915	7372	WA	WA
012141	0000789019	20100216	SC 13G/A	MICROSOFT CORP	edgar/data/789019/0001104659-10-007434.txt	60047	.	171557	7372	WA	WA
012141	0000789019	20100302	UPLOAD	MICROSOFT CORP	edgar/data/789019/0000000000-10-011854.txt	19265	.	165914	7372	WA	WA
012141	0000789019	20100422	10-Q	MICROSOFT CORP	edgar/data/789019/0001193125-10-090116.txt	4757325	31MAR2010	161852	7372	WA	WA
012141	0000789019	20100422	8-K	MICROSOFT CORP	edgar/data/789019/0001193125-10-090109.txt	194512	22APR2010	161630	7372	WA	WA
012141	0000789019	20100525	8-K	MICROSOFT CORP	edgar/data/789019/0001193125-10-128133.txt	26907	25MAY2010	145021	7372	WA	WA
012141	0000789019	20100608	8-K	MICROSOFT CORP	edgar/data/789019/0001193125-10-135038.txt	41900	08JUN2010	172821	7372	WA	WA
012141	0000789019	20100609	8-K	MICROSOFT CORP	edgar/data/789019/0001193125-10-136086.txt	39009	08JUN2010	164221	7372	WA	WA
012141	0000789019	20100618	8-K	MICROSOFT CORP	edgar/data/789019/0001193125-10-142376.txt	533578	14JUN2010	172501	7372	WA	WA
012141	0000789019	20100722	8-K	MICROSOFT CORP	edgar/data/789019/0001193125-10-163839.txt	339672	22JUL2010	161932	7372	WA	WA
012141	0000789019	20100730	10-K	MICROSOFT CORP	edgar/data/789019/0001193125-10-171791.txt	20098498	30JUN2010	113038	7372	WA	WA
012141	0000789019	20100913	8-K	MICROSOFT CORP	edgar/data/789019/0001193125-10-209027.txt	13834	09SEP2010	170149	7372	WA	WA
012141	0000789019	20100927	8-K	MICROSOFT CORP	edgar/data/789019/0001193125-10-216531.txt	304558	22SEP2010	192422	7372	WA	WA
012141	0000789019	20100930	DEF 14A	MICROSOFT CORP	edgar/data/789019/0001193125-10-221150.txt	1695884	16NOV2010	164529	7372	WA	WA
012141	0000789019	20101001	DEFA14A	MICROSOFT CORP	edgar/data/789019/0001193125-10-222170.txt	1128446	.	164835	7372	WA	WA
012141	0000789019	20101004	DEFA14A	MICROSOFT CORP	edgar/data/789019/0001193125-10-222844.txt	22659	.	142554	7372	WA	WA
012141	0000789019	20101022	DEFA14A	MICROSOFT CORP	edgar/data/789019/0001193125-10-234650.txt	22739	.	164534	7372	WA	WA
012141	0000789019	20101028	10-Q	MICROSOFT CORP	edgar/data/789019/0001193125-10-239825.txt	11723855	30SEP2010	163615	7372	WA	WA
012141	0000789019	20101028	8-K	MICROSOFT CORP	edgar/data/789019/0001193125-10-239800.txt	158303	28OCT2010	162849	7372	WA	WA
012141	0000789019	20101109	SC 13G/A	MICROSOFT CORP	edgar/data/789019/0001086364-10-009924.txt	12977	.	153412	7372	WA	WA
012141	0000789019	20101119	8-K	MICROSOFT CORP	edgar/data/789019/0001193125-10-265243.txt	45454	16NOV2010	161359	7372	WA	WA
012141	0000789019	20110127	10-Q	MICROSOFT CORP	edgar/data/789019/0001193125-11-015947.txt	13808138	31DEC2010	160033	7372	WA	WA
012141	0000789019	20110127	8-K	MICROSOFT CORP	edgar/data/789019/0001193125-11-015944.txt	243072	27JAN2011	155759	7372	WA	WA
012141	0000789019	20110208	8-K	MICROSOFT CORP	edgar/data/789019/0001193125-11-026916.txt	283575	03FEB2011	123530	7372	WA	WA

Example of the available and ready-to-use parsed content

# SEC Filings on WRDS

## Explore the different types of SEC filings

- **Filings archive updated daily.** Accessible by SAS, R or Python, and stored in `/wrds/sec/warchives/`
  - **WRDS\_FORMS** dataset contains the information to access these filings
  - **WRDS\_FORMS\_REG** contains additional registrant entities information
  - **WRDS FILE NAME (or WRDSFNAME)**, in *WRDS\_FORMS* provides reference to the filings on WRDS server  
**FSIZE>0** is a condition to be used when determining available filings
- **All filings are cleaned**, and stored in `/wrds/sec/wrds_clean_filings/`
- **SAS datasets** in `/wrds/sec/sasdata/` with parsed contents: e.g. **WRDS\_FORMS** and **WRDS\_FORMS\_REG** datasets
  - **Filing size**, fiscal year end
  - **Date and Time Report** of SEC Acceptance (Available after May 2002)
  - **Confirmed Period of Report** including Fiscal Period End for 10-K and 10-Q, Event Date for 8-K, and Meeting Date for proxy filings
  - Historical state of **incorporation** and **headquarters**
  - Historical as-reported **SIC code**
  - **+ many others**



# WRDS Cleaned Text Filings

- **All filings on EDGAR are downloaded** , and stored in `/wrds/sec/warchives/`
- **All filings are cleaned**, and stored in `/wrds/sec/wrds_clean_filings/`
- **Daily Process** to download SEC Index Files
  - Compares daily index with full index to ensure completeness
  - Uses the Index Files to create a list of added filings
  - Downloads the full text of the individual filings to `/wrds/sec/warchives/` as WRDSFNAME
  - Parse header and clean body of document: update WRDS\_FORMS & WRDS\_FORMS\_REG
  - Remove presentation tags, convert PDF files to text using OCR, and convert tables to text
  - Cleaned filings are stored in `/wrds/sec/wrds_clean_filings/`
- **Auditing and Redundancy Checks**
  - Compares the complete index files to the list of processed filings every quarter to ensure that we have all the filings
  - Calculates the number of registrants to ensure that all data is collected
  - Any files that are unavailable from the SEC are stored in the `missing_filings` dataset for reference. 11

# Prepared Contents of all SEC Filings

## WRDS\_FORMS

Variable	Description
<b>fdate</b>	Filing Date
<b>cik</b>	SEC Central Index Key
<b>form</b>	Form Type
<b>coname</b>	Company Name
<b>wrdsfname</b>	Reference Name of Complete Report Filing
<b>fsize</b>	File Size
<b>dccount</b>	Public Document Count
<b>fname</b>	Reference Name of Complete Report Filing
<b>rdate</b>	Conformed Period of Report
<b>secadate</b>	SEC Acceptance Date
<b>secatime</b>	SEC Acceptance Time
<b>secpdate</b>	Filing Publication Date
<b>accession</b>	Accession Number
<b>regcount</b>	Total Number of Reporting Registrants

## WRDS\_FORMS\_REG

Variable	Description
<b>fdate</b>	Filing Date
<b>accession</b>	Accession Number
<b>regseq</b>	Reporting Registrant Sequence Number
<b>regrole</b>	Reporting Registrant Role
<b>regcik</b>	Registrant Central Index Key
<b>regfile_no</b>	Registrant SEC File Number
<b>regconame</b>	Registrant Company Name
<b>regfye</b>	Registrant Fiscal Year End
<b>regsic</b>	Registrant Standard Industrial Classification
<b>regstreet_hdq</b>	Street of Registrant Business Address
<b>regcity_hdq</b>	City of Registrant Business Address
<b>regstate_hdq</b>	State of Registrant Business Address
<b>regzip_hdq</b>	Zip Code of Registrant Business Address
<b>regstate_inc</b>	Registrant State of Incorporation
<b>regphone</b>	Phone Number of Registrant Business Address
<b>regfconame</b>	Former Registrant Company Name
<b>regfchangedate</b>	Date of Registrant Name Change

# Ex 1: Registrants Info, Carl Icahn 13D Filings

WRDS\_FORMS: at the text filing level where FNAME is primary identifier

WRDS\_FORMS\_REG: Registrant info where ACCESSION is main identifier. Merge it back with WRDS\_FORMS using ACCESSION

Filing Date	Accession Number	Reporting Registrant Sequence Number	Reporting Registrant Role	Registrant Central Index Key	Registrant SEC File Number	Registrant Company Name	Registrant Fiscal Year End	Registrant Standard Industrial Classification	City of Registrant Business Address	Zip Code of Registrant Business Address	Registrant State of Incorporation
20160129	0000928464-16-000138	1	SUBJECT COMPANY	0000108772	005-33994	XEROX CORP	1231	7389	NORWALK	06856	NY
20160129	0000928464-16-000138	2	FILED BY	0000921669		ICAHN CARL C	1231	0000	NEW YORK	10153	NY
20160208	0000928464-16-000142	1	SUBJECT COMPANY	0000077449	005-18769	PEP BOYS MANNY MOE & JACK	0202	5531	PHILADELPHIA	19132	PA
20160208	0000928464-16-000142	2	FILED BY	0000921669		ICAHN CARL C	1231	0000	NEW YORK	10153	NY
20160219	0000928464-16-000145	1	SUBJECT COMPANY	0000701811	005-38367	MENTOR GRAPHICS CORP	0131	7373	WILSONVILLE	97070-7777	OR
20160219	0000928464-16-000145	2	FILED BY	0000921669		ICAHN CARL C	1231	0000	NEW YORK	10153	NY
20160224	0000928464-16-000147	1	SUBJECT COMPANY	0001344596	005-81342	American Railcar Industries, Inc.	1231	3743	ST. CHARLES	63301	ND
20160224	0000928464-16-000147	2	FILED BY	0000921669		ICAHN CARL C	1231	0000	NEW YORK	10153	NY
20160229	0000928464-16-000149	1	SUBJECT COMPANY	0001419581	005-83475	Federal-Mogul Holdings Corp	1231	3714	SOUTHFIELD	48034	DE
20160229	0000928464-16-000149	2	FILED BY	0000921669		ICAHN CARL C	1231	0000	NEW YORK	10153	NY
20160229	0000928464-16-000153	1	SUBJECT COMPANY	0000039899	005-19132	TEGNA INC	1228	4833	MCLEAN	22107-0910	DE
20160229	0000928464-16-000153	2	FILED BY	0000921669		ICAHN CARL C	1231	0000	NEW YORK	10153	NY
20160301	0000928464-16-000157	1	SUBJECT COMPANY	0000813762	005-38593	ICAHN ENTERPRISES L.P.	1231	2911	NEW YORK	10153	DE
20160301	0000928464-16-000157	2	FILED BY	0000921669		ICAHN CARL C	1231	0000	NEW YORK	10153	NY
20160304	0000928464-16-000160	1	SUBJECT COMPANY	0000859737	005-41074	HOLOGIC INC	0924	3844	MARLBOROUGH	01752	DE
20160304	0000928464-16-000160	2	FILED BY	0000921669		ICAHN CARL C	1231	0000	NEW YORK	10153	NY
20160304	0000928464-16-000162	1	SUBJECT COMPANY	0000859737	005-41074	HOLOGIC INC	0924	3844	MARLBOROUGH	01752	DE
20160304	0000928464-16-000162	2	FILED BY	0000921669		ICAHN CARL C	1231	0000	NEW YORK	10153	NY
20160304	0000928464-16-000164	1	SUBJECT COMPANY	0000701811	005-38367	MENTOR GRAPHICS CORP	0131	7373	WILSONVILLE	97070-7777	OR
20160304	0000928464-16-000164	2	FILED BY	0000921669		ICAHN CARL C	1231	0000	NEW YORK	10153	NY
20160308	0000928464-16-000166	1	SUBJECT COMPANY	0000701811	005-38367	MENTOR GRAPHICS CORP	0131	7373	WILSONVILLE	97070-7777	OR
20160308	0000928464-16-000166	1	SUBJECT COMPANY	0000701811	005-38367	MENTOR GRAPHICS CORP	0131	7373	WILSONVILLE	97070-7777	OR
20160308	0000928464-16-000166	2	FILED BY	0000921669		ICAHN CARL C	1231	0000	NEW YORK	10153	NY
20160308	0000928464-16-000166	2	FILED BY	0000921669		ICAHN CARL C	1231	0000	NEW YORK	10153	NY
20160309	0000928464-16-000169	1	SUBJECT COMPANY	0000039899	005-19132	TEGNA INC	1228	4833	MCLEAN	22107-0910	DE
20160309	0000928464-16-000169	2	FILED BY	0000921669		ICAHN CARL C	1231	0000	NEW YORK	10153	NY

Registrants are identified in the REGROLE Variable

Activist vs. Subject company, or Reporting Owner vs. Issuer, etc.

Use it to identify relationships between filer and company

# Registrant Info: Collected from Filing Headers

## SUBJECT COMPANY:

### COMPANY DATA:

COMPANY CONFORMED NAME:	PEP BOYS MANNY MOE & JACK
CENTRAL INDEX KEY:	0000077449
STANDARD INDUSTRIAL CLASSIFICATION:	RETAIL-AUTO & HOME SUPPLY STORES [5531]
IRS NUMBER:	230962915
STATE OF INCORPORATION:	PA
FISCAL YEAR END:	0202

### FILING VALUES:

FORM TYPE:	SC 13D/A
SEC ACT:	1934 Act
SEC FILE NUMBER:	005-18769
FILM NUMBER:	161393533

### BUSINESS ADDRESS:

STREET 1:	3111 W ALLEGHENY AVE
CITY:	PHILADELPHIA
STATE:	PA
ZIP:	19132
BUSINESS PHONE:	2152299000

## FILED BY:

### COMPANY DATA:

COMPANY CONFORMED NAME:	ICAHN CARL C
CENTRAL INDEX KEY:	0000921669
STANDARD INDUSTRIAL CLASSIFICATION:	UNKNOWN SIC - 0000 [0000]
IRS NUMBER:	000000000
STATE OF INCORPORATION:	NY
FISCAL YEAR END:	1231

### FILING VALUES:

FORM TYPE:	SC 13D/A
------------	----------

### BUSINESS ADDRESS:

STREET 1:	C/O ICAHN ASSOCIATES HOLDING LLC
STREET 2:	767 FIFTH AVE., SUITE 4700
CITY:	NEW YORK
STATE:	NY
ZIP:	10153
BUSINESS PHONE:	212-702-4300

## REGROLE:

FILER  
REPORTING OWNER  
SUBJECT COMPANY  
FILED BY  
FILED FOR  
ISSUER  
SERIAL COMPANY

Source: <https://www.sec.gov/Archives/edgar/data/77449/0000928464-16-000142.txt>

# SEC Filings on WRDS

**2**

**WRDS SEC Web Queries & Data**

# Web-based access to SEC filings

The screenshot displays the WRDS SEC Analytics Suite - Filings Parser interface. At the top, there is a navigation menu with links for HOME, TOOLS, SUPPORT, E-LEARNING, FORUMS, NEWS, ABOUT, and MyWRDS. A search bar is located in the top right corner. The main content area is titled "WRDS SEC Analytics Suite - Filings Parser" and includes a link to "Detailed Documentation". Below this, there is a section for "Step 1: What date range do you want to use?" with a date range selector set to "Jan 2008 to Dec 2010". A second section, "Step 2: How would you like to search this dataset?", is also visible. A red arrow points from the text "queries" to the "WRDS SEC Analytics Suite" menu item. Another red arrow points from the text "Detailed Documentation" to the corresponding link. Red boxes highlight the "Detailed Documentation" link and the "Step 2" section.

queries →

→ Detailed Documentation

- Easy-to-use web queries and similar to any other WRDS queries
- Flexible output format and Live html links to actual filings
- Parser query with various input and line extract options



# Web-based access to SEC filings

## WRDS SEC Filings Index Data

SEC Filings Index

SEC Filings on WRDS

## WRDS SEC Filings Queries

Readability and Sentiment

SEC Filings Search

List of Filings Exhibits

List of 8K Items

## WRDS SEC Text Analysis

» WRDS Bag of Words

## WRDS SEC Linking Tables

Historical Company Names

CIK-CUSIP Link Table

GVKEY-CIK Link Table

1. Complete Index Data: Records of ALL electronic filings on EDGAR (~20 million)
2. Archive of downloaded filings on WRDS server (19.8 million + additional information (filing time, FPE, incorp, ...))
3. Readability and Sentiment data
4. Search SEC Filings using solr syntax
5. Get the list of Filings Exhibits
6. Extract or Filter by 8K Items
7. Extract word counts using Bag of Words
8. Linking tables

# Example: Microsoft Corp recent 10-K

<http://www.sec.gov/Archives/edgar/data/789019/0001193125-10-171791-index.htm>

[TV](#) [V](#) [EDGAR Overview / R...](#) [EDGAR](#) [WRDSN](#) [WRDS2](#) [P](#) [R](#) [U@Penn](#) [Lippincott](#) [E-Resources](#) [My Accounts](#) [Blogs](#) [S](#)

**Form 10-K - Annual report [Section 13 and 15(d), not S-K Item 405]**

<b>Filing Date</b> 2010-07-30	<b>Period of Report</b> 2010-06-30
<b>Accepted</b> 2010-07-30 11:30:38	<b>Filing Date Changed</b> 2010-07-30
<b>Documents</b> 16	
<a href="#">Interactive Data</a>	

**19+ million Filing with  
75+ million Exhibits  
since 1994**

**Document Format Files**

Seq	Description	Document	Type	Size
1	ANNUAL REPORT ON FORM 10-K	d10k.htm	10-K	1668256
2	2009 OFFICERS' INDEMNIFICATION TRUST AGREEMENT	dex1012.htm	EX-10.12	85831
3	AMENDED AND RESTATED 2003 INDEMNIFICATION TRUST AGREEMENT	dex1013.htm	EX-10.13	81483
4	COMPUTATION OF RATIO OF EARNINGS TO FIXED CHARGES	dex12.htm	EX-12	32211
5	SUBSIDIARIES OF REGISTRANT	dex21.htm	EX-21	4613
6	CONSENT OF INDEPENDENT REGISTERED PUBLIC ACCOUNTING FIRM	dex231.htm	EX-23.1	2139
7	CERTIFICATIONS OF CHIEF EXECUTIVE OFFICER PURSUANT TO SECTION 302	dex311.htm	EX-31.1	6105
8	CERTIFICATIONS OF CHIEF FINANCIAL OFFICER PURSUANT TO SECTION 302	dex312.htm	EX-31.2	6096
9	CERTIFICATIONS OF CHIEF EXECUTIVE OFFICER PURSUANT TO SECTION 906	dex321.htm	EX-32.1	3265
10	CERTIFICATIONS OF CHIEF FINANCIAL OFFICER PURSUANT TO SECTION 906	dex322.htm	EX-32.2	3262
	Complete submission text file	0001193125-10-171791.txt		20098498

**Data Files**

Seq	Description	Document	Type	Size
11	XBRL INSTANCE DOCUMENT	msft-20100630.xml	EX-101.INS	3150912
12	XBRL TAXONOMY EXTENSION SCHEMA	msft-20100630.xsd	EX-101.SCH	128451
13	XBRL TAXONOMY EXTENSION CALCULATION LINKBASE	msft-20100630_cal.xml	EX-101.CAL	143706
14	XBRL TAXONOMY EXTENSION DEFINITION LINKBASE	msft-20100630_def.xml	EX-101.DEF	687511
15	XBRL TAXONOMY EXTENSION LABEL LINKBASE	msft-20100630_lab.xml	EX-101.LAB	643036
16	XBRL TAXONOMY EXTENSION PRESENTATION LINKBASE	msft-20100630_pre.xml	EX-101.PRE	739125

# Example: Valeant Pharma's 8-K

Form 8-K - Current report

**Filing Date**

2010-10-01

**Accepted**

2010-10-01 17:03:16

**Documents**

15

**Period of Report**

2010-09-27

**Filing Date Changed**

2010-10-01

**Items**

- Item 1.01: Entry into a Material Definitive Agreement
- Item 1.02: Termination of a Material Definitive Agreement
- Item 2.01: Completion of Acquisition or Disposition of Assets
- Item 2.03: Creation of a Direct Financial Obligation or an Obligation under an Off-Balance Sheet Arrangement of a Registrant
- Item 3.03: Material Modifications to Rights of Security Holders
- Item 5.02: Departure of Directors or Certain Officers; Election of Directors; Appointment of Certain Officers; Compensatory Arrangements of Certain Officers
- Item 5.03: Amendments to Articles of Incorporation or Bylaws; Change in Fiscal Year
- Item 5.07: Submission of Matters to a Vote of Security Holders
- Item 9.01: Financial Statements and Exhibits

**Time of Filing or SEC Acceptance Time**

**3.4+ million Corporate Events for 1.7+ million 8-Ks hat triggered 8-K filings since 1994**

**New 8-K Item starting in March 2010**

Document Format Files

Seq	Description	Document	Type	Size
1	FORM 8-K	y86806e8vk.htm	8-K	104731
2	EX-3.1	y86806ex3w1.htm	EX-3.1	10818
3	EX-4.1	y86806ex4w1.htm	EX-4.1	653099
4	EX-4.2	y86806ex4w2.htm	EX-4.2	36748
5	EX-4.3	y86806ex4w3.txt	EX-4.3	336777
6	EX-10.1	y86806ex10w1.htm	EX-10.1	809643
7	EX-10.2	y86806ex10w2.htm	EX-10.2	11258
8	EX-99.1	y86806ex99w1.htm	EX-99.1	14348
9	EX-99.2	y86806ex99w2.htm	EX-99.2	8494
10	GRAPHIC	y86806y8680605.gif	GRAPHIC	672
11	GRAPHIC	y86806y8680606.gif	GRAPHIC	1990
12	GRAPHIC	y86806y8680607.gif	GRAPHIC	1784
13	GRAPHIC	y86806y8680608.gif	GRAPHIC	766
14	GRAPHIC	y86806y8680609.gif	GRAPHIC	5879
15	GRAPHIC	y86806y8680610.gif	GRAPHIC	5317
	Complete submission text file	0000950123-10-090858.txt		2011916

# SEC Filings Search

- Web query that uses Apache Lucene and Solr to provide full-text search of all 10Ks, 10Qs, 8Ks, Proxy and Registration Statements, 40-F Annual Reports, Uploads and SEC correspondence filings

Search through the contents of 3,710 Proxy and Registration Statements, 4 correspondence.

Enter your search here

+ More Options (date range, identifiers, etc.)

- Search Syntax

The following example uses all of the

"shareholder rights plan" +adopt

"shareholder rights plan" is a ref within the quotes exactly.

+adopt makes the term adopt mand

On the other hand, -repeal makes t

Additional syntax features can be found in Section 4 of the "Overview of WRDS SEC Analytics Suite" document.

Search
Code Lookup

+ More Options (date range, identifiers, etc.)

+ Search Syntax

← Prev
Viewing 1 to 25 of 40,659 documents
Next →
Download

Document	Company Name	Form Type	Fiscal Year
1669021/0001669021-16-000007	IATSE	CORRESP	
IRS Number: 131575490 CIK: 0001669021 SIC: Former Name: Incorp. State: Business Addr.: TOLUCA LAKE CA 91602		File Number: Accession No.: 0001669021-16-000007 Filing Date: 25-Apr-2016 Period End Date: Acceptance Date: 26-Apr-2016 13:03:42	
Text matches within document: "-really-hot-quick-score-analysis-2015s- real-governance-trends/ - Common Wisdom: <b>Poison pill</b> adoptions"			
<p>+ View 7 more matches</p>			
Document	Company Name	Form Type	Fiscal Year
1009881/0001009881-97-000003	PIER 1 IMPORTS INC/DE	PREN14A	1997

# SEC Filings Search

- Query allows versatile searches
  - *Simple search*: `-compensation` searches for all filings that do not contain the word 'compensation'.
  - *Phrase search*: `"executive compensation"` returns filings with that exact phrase in them.
  - *Vicinity search*: `"performance compensation"~8` returns hits for "Management Performance Compensation Plan", "Performance Based Executive Compensation Plan", "Performance Based incentive Compensation Plan" but also "performance-based vesting criteria determined by the Compensation Committee", "performance metrics for executive compensation", etc.
  - *Compound search*: A compound search is two or more of the above search items, either joined with a Boolean 'AND' or 'NOT' operator, or with each search item prepended with a '+' or '-'. 'AND' or '+' return filings that contain all search terms, whereas 'NOT' or '-' return filings without the following term. If you do not specify an operator, the search will return filings that contain any of the search terms, which is generally not useful.
- See Lucene Solr Syntax help for additional information:  
[https://lucene.apache.org/core/2\\_9\\_4/queryparsersyntax.html](https://lucene.apache.org/core/2_9_4/queryparsersyntax.html)

# CIK Link Tables

- CIK link tables are **datasets that map CIK to all historical** company legal names, CUSIP numbers, and other identification information
  - **WCIKLINK\_NAMES** lists of all company names for a given CIK
  - **WCIKLINK\_CUSIP** maps a CIK to all CUSIPs that appear in a company's filings
  - **WCIKLINK\_GVKEY** maps between GVKEY and 'Historical' CIKs
- **Helps retain historical records** for companies that are undergoing restructuring and who are more likely to change their CIK filing number
  - **Essential tool** for when you want to track all historical filings for public companies
  - Researchers use GVKEY-CIK historical maps to **avoid selection and survivorship bias concerns**

# Example: K-Mart Historical GVKEY-CIK Map

Compustat Global Company Key - GVKEY	Company Name in Compustat - Header	First Date of Compustat Data	Last Date of Compustat Data	CIK Number	GVKEY-CIK Link Source	Company Name in SEC Filings - Header	First Filing Date	Last Filing Date	Number of 10-K Filings	Number of 10-Q Filings	Total Number of Filings	Link Flag: 0-No Link, 1-No Filings, 2-Name Mismatch, 3-Valid Link
004034	DOMTAR INC	31DEC1950	31DEC2006	0000801644	CUSIP	DOMTAR INC /FI	13FEB2006	14FEB2007	0	0	4	3
004034	DOMTAR INC	31DEC1950	31DEC2006	0000819817	COMPEN	DOMTAR INC /CANADA	13FEB1995	14FEB2008	0	0	20	3
004034	DOMTAR INC	31DEC1950	31DEC2006	0001381531	CUSIP	DOMTAR CORP	12DEC2006	25FEB2011	4	13	116	3
006307	SEARS HOLDINGS CORP	31DEC1950	31JAN2010	0000056824	CUSIP	KMART CORP	04FEB1994	29AUG2003	12	34	177	2
006307	SEARS HOLDINGS CORP	31DEC1950	31JAN2010	0001229206	FDATE	KMART HOLDING CORP	15MAY2003	21MAR2005	3	10	63	3
006307	SEARS HOLDINGS CORP	31DEC1950	31JAN2010	0001310067	COMPEN	SEARS HOLDINGS CORP	24MAR2005	25FEB2011	6	18	139	3
018284	1ST BANCORP IND	30JUN1993	30JUN1998	0000840458	COMPEN	FIRST BANCORP /IN/	28DEC1995	13NOV1998	4	9	21	3
018284	1ST BANCORP IND	30JUN1993	30JUN1998	0000915327	COMPEN							1
018284	1ST BANCORP IND	30JUN1993	30JUN1998	0001074543	FDATE	FIRST BANCORP OF INDIANA INC	17MAY1999	14MAY2008	9	30	109	2

# SEC Filings on WRDS

**3**

Textual Analytics: Bag of Words/Sentiment



# Readability and Sentiment

- Surge of interest in text analysis
  - a need to make it easier for researchers to process, manipulate, and analyze the text content of SEC filings
  - Cleaned set of text files for every SEC filing
    - Including OCRing image and pdf files for “UPLOAD” and “CORRESP” filings among others
    - Stripping out html tables and exhibits to keep only material text within the filing: fine-tuning in progress
  - Baseline sentiment and readability scores
- Researchers can use the pre-computed scores to further academic research, and can also compute their own features based on the raw text or using the new “**Bag of Words**” dataset
  - Dataset containing series of variables relating to sentiment polarity and readability.
  - Many Readability Indices: Coleman-Liau, Gunning Fog, Flesch Reading Ease Indices, etc.
  - Sentiment based on “bag of words” methodology: Loughran and McDonald (2011) and on Harvard GI dictionary.
- Coverage: Every single filing on SEC’s EDGAR website since 1994

# Readability and Sentiment: List of measures

## Readability

Feature	Description
Character count	Total # of characters in document
Word count	Total # of words in document
Sentence count	Total # of sentences in document
Average Characters per Sentence	Average # of characters per sentence
Average Words per Sentence	Average # of words per sentence
Average Characters per Word	Average # of characters per word
Complex word count	Total # of 3 syllable or more words in document
Automated Readability Index	$4.71(\text{characters/words}) + 0.5(\text{words/sentences}) - 21.43$
Coleman-Liau Index	$0.0588(\text{avg characters}/100 \text{ words}) - 0.296(\text{avg sentences}/100 \text{ words}) - 15.8$
Gunning Fog Index	$0.4 ((\text{words/sentences})+100(\text{complex words/words}))$
Flesch Reading Ease	$206.835 - 1.015(\text{total words}/\text{total sentences}) - 84.6(\text{total syllables}/\text{total words})$
Flesch-Kincaid Grade Level	$0.39(\text{total words}/\text{total sentences}) + 11.8(\text{total syllables}/\text{total words}) - 15.59$
SMOG Index	$1.043 * \text{sqrt}(\text{complex words} * 30 / \text{sentences}) + 3.1291$
LIX	$\text{words}/(\text{sentences marked by periods, colons, or capital first letter}) + (\text{words over 6 letters} * 100)/\text{words}$

## Sentiment

Feature	Description
Harvard GI Negative count	Based on the Harvard General Enquirer negative word list
FinTerms_Positive count	L&M word list
FinTerms_Negative count	L&M word list
FinTerms_Uncertainty count	L&M word list
FinTerms_Litigious count	L&M word list
FinTerms_ModelStrong count	L&M word list
FinTerms_ModalWeak count	L&M word list

# WRDS SEC: Readability and Sentiment

	SEC Central Index Key	Filing Date	SEC Form	Company Name	Flesch Reading Ease Index	Flesch-Kincaid Readability Index	Paragraph count	Raw character count	Average number of words per sentence	Syllable count	Complex word count	Loughran-McDonald Negative word proportion	Loughran-M modal weak count	Loughran-McDonald modal strong count	Loughran-McDonald positive word proportion
537	0001415813	20090102	10-K	ASIARIM CORP	30.633518744	14.450631945	212	82356	21.994575045	22123	3110	0.0176044184	107	83	0.00923369
538	0001417561	20090102	SC 13D/A	OCH DANIEL	20.995668239	20.205861286	14	9259	39.75	2461	313	0.0105092967	7	1	0.0040420372
539	0001418970	20090102	497J	PHL VARIABLE ACCUMULATION ACCOUNT II	28.002258065	17.05483871	5	915	31	216	27	0.0105263158	0	0	0
540	0001421182	20090102	8-K	VERSO PAPER CORP.	21.206150371	17.411903632	11	7431	28.621621622	1960	290	0.0139037433	2	6	0.0106951872
541	0001422183	20090102	497AD	FS INVESTMENT CORP	27.865357078	14.786981908	221	97269	21.794303798	25537	3692	0.0130040012	73	91	0.006232687
542	0001422509	20090102	497J	ALLSTATE FINANCIAL INVESTMENT TRUST	33.049068966	14.413172414	2	913	23.2	206	30	0.0102040816	0	0	0
543	0001427300	20090102	10-D	CHL MORTGAGE PASS-THROUGH TRUST RESECURITIZATION 2008-2R	20.848743702	16.458049768	25	8528	24.581395349	2012	306	0.0204962244	3	0	0.0021574973
544	0001427301	20090102	10-D	CHL MORTGAGE PASS-THROUGH TRUST 2008-1	23.178409449	16.336472441	22	8488	25.4	1896	288	0.0212527964	3	0	0.0022371365
545	0001431328	20090102	CORRESP	NATIONAL RETAIL FUND III	43.724588336	12.72915952	17	8181	22.41509434	1971	261	0.0140449438	2	22	0.0028089888
546	0001431329	20090102	CORRESP	NATIONAL RETAIL FUND II	43.795800457	12.71922686	17	8180	22.41509434	1970	260	0.0140449438	2	22	0.0028089888
547	0001431835	20090102	UPLOAD	LIVERPOOLGROUP, INC.	50.190191569	10.698449983	3	3413	17.870967742	907	86	0.0121212121	10	1	0.0060606061
548	0001435260	20090102	SC 13D/A	MILL ROAD CAPITAL, L.P.	47.906619955	12.512967649	16	4874	23.892857143	1065	112	0.0080906149	0	1	0.0064724919
549	0001437425	20090102	UPLOAD	HIBERNIA HOMESTEAD BANCORP, INC.	44.554759969	11.415479246	8	6580	17.593220339	1772	203	0.0209863589	10	5	0.0052465897
550	0001437425	20090102	UPLOAD	HIBERNIA HOMESTEAD BANCORP, INC.	42.341634615	11.328365385	4	2697	16	729	84	0.0190735695	7	1	0.0108991826
551	0001438382	20090102	SC TO-T/A	ALKALOIDA CHEMICAL CO EXCLUSIVE GROUP LTD.	42.257405034	12.397814306	67	18589	20.257575758	4552	534	0.0226302306	10	8	0.0072587532
552	0001438945	20090102	UPLOAD	STRIKE AXE, INC	52.640228463	10.230683801	3	3504	17.363636364	925	83	0.0116731518	10	1	0.0058365759
553	0001439133	20090102	UPLOAD	SOLARTE HOTEL CORP	46.407198204	11.375469925	41	12487	18.472222222	3341	366	0.0102869518	12	15	0.0054141852
554	0001439133	20090102	UPLOAD	SOLARTE HOTEL CORP	48.56224365	10.783415578	12	10492	17.298969072	2791	314	0.0109536082	11	11	0.0096649485
555	0001439133	20090102	UPLOAD	SOLARTE HOTEL CORP	45.117335655	11.681208854	8	5617	18.978723404	1502	174	0.0085679315	9	5	0.0085679315
556	0001439133	20090102	UPLOAD	SOLARTE HOTEL CORP	40.496503244	12.31826793	10	4632	18.948717949	1285	157	0.0120300752	10	3	0.0090225564
557	0001439133	20090102	UPLOAD	SOLARTE HOTEL CORP	42.062696887	12.058841143	6	4378	18.783783784	1197	140	0.0128824477	7	5	0.0096618357
558	0001440420	20090102	487	VAN KAMPEN UNIT TRUSTS SERIES 828	29.117293845	15.36258049	545	323664	24.814168378	87152	11765	0.0175795053	591	472	0.0069567138
559	0001442228	20090102	487	MATRIX DEFINED TRUSTS 5	34.330890078	15.114209663	308	167758	26.741573034	44983	5605	0.0146626401	280	303	0.006278312
560	0001442865	20090102	487	CLAYMORE SECURITIES DEFINED PORTFOLIOS, SERIES 552	33.454558758	14.595142774	714	302343	24.160149973	79368	10265	0.0161189883	462	425	0.0074726959
561	0001445081	20090102	487	FT 1898	30.5553167	16.202726816	544	225873	29.003398471	59252	7396	0.01204397	218	307	0.0039827943
562	0001446926	20090102	S-11/A	1ST & 10 PROPERTIES, CORP.	33.802011616	14.189152815	267	112506	22.720990874	30893	4138	0.0128919017	130	208	0.0104708873
563	0001448230	20090102	SC 13D	PIRAMAL HEALTHCARE, INC.	28.094575616	16.109139873	47	19086	27.245098039	4963	657	0.0121046466	4	11	0.0085903944
564	0001448551	20090102	SC 13D/A	HBK SERVICES LLC	56.165965987	8.8243534549	59	10132	13.682242991	2367	250	0.0039525692	1	1	0.0118577075
565	0001448551	20090102	SC 13G	HBK SERVICES LLC	52.737250205	9.3018814749	71	12452	13.679389313	2970	336	0.006232687	1	2	0.0117728532
566	0001448759	20090102	SC 13D/A	HBK NEW YORK LLC	56.165965987	8.8243534549	59	10132	13.682242991	2367	250	0.0039525692	1	1	0.0118577075
567	0001448759	20090102	SC 13G	HBK NEW YORK LLC	52.737250205	9.3018814749	71	12452	13.679389313	2970	336	0.006232687	1	2	0.0117728532

# Bag of Words: On-Demand Word Distribution

- Exciting new product: Sentiment On-Demand
  - Dataset: Frequency distribution of all words in all filings since 1993
  - Objective: Users can load personal list / bag of words + search within subsections of filings  
→ Customized Analysis for Distancing / Sentiment / Deceptive / Uncertainty / Truthfulness / Forensic / Geographies / Products / Patents / Names etc.
  - Detailed [manual](#) on how the frequency counts are created
  - Access on [web](#) or server: `/wrds/wrdsapps/sasdata/bagofwords/`
- Web queries for comparison of filings using various similarity measures:
- Construct measures for changes in filings: 10Ks and 10Qs
  - Cosine Similarity =  $\frac{\sum w_i \times w_j}{\sqrt{\sum w_i^2} \times \sqrt{\sum w_j^2}}$ , where w is the # of word occurrences
  - Jaccard Similarity =  $\frac{\|W_i \cap W_j\|}{\|W_i \cup W_j\|}$
  - Minimal Edit Distance =  $\frac{|w_i - w_j|}{\max(\sum w_i, \sum w_j)}$
- Vectors of words: use as input Lasso/Ridge/MF/LDA applications: bankruptcy/forensic/linkages/themes etc.

# Advanced Access using WRDS Server

Take advantage of local storage of filings and index datasets with **PC-SAS** or **UNIX-SAS**

**Use Python, R, or SAS capabilities** to parse thousands of filings and build custom-tailored data sets in one step

**WRDS Research Macros** are standardized and well-documented SAS programs that can be modified and invoked in one line

**Effective, transparent and extensible SAS codes**, including:

- `LineParse`: Line-by-Line parser that preserves tabular format.
- `TextParse`: Parses out the match line & a pre-specified number of preceding characters.
- `ParaParse`: Extracts a paragraph with pre-specified number of lines around a string.



# SEC Filings on WRDS

**4**

**Derived Data Products**

# WRDS SEC: Derived Datasets

- Objective: “liberate numbers from textual reports” by capitalizing on XML and XBRL filings
- WRDS 13F Data:
  - Complete history from Jun 2013, including original filings & amendments
  - Confidential treatments flags + list of subadvisors + all reported holdings
- WRDS Insiders Data:
  - Complete Stock and Derivatives history from 2003 + original filings & amendments
  - Footnotes (e.g. collars, hedges/swaps, 10b-5, 14e-3 etc) + detailed filing contents
- Coming soon: more derived products and datasets (e.g. WRDS SEC Fundamentals for 10K and 10Q XBRL data and footnotes, Form D, etc.)

# WRDS SEC: Added Value

- *To level the playing field in Textual Analysis*
  - Make it easier/less costly to implement textual based research on SEC filings
  - Provide intuitive Tools/Macros/Webqueries that perform complex programming algorithms: *Bag of Words Platform*, *Readability/Sentiment*
- *Provide new data products*
  - SEC is upgrading tons of forms to include xml tags: *liberating numbers from filings*
  - Focus should be on forms that provide new data elements, relative to existing WRDS data: *WRDS SEC Fundamentals database*
- “Scale” is a differentiating element
- No Black Box: Simplicity + Transparency



# SEC Filings Data on WRDS

Thank you for attending this WRDS E-Learning session.

Research Applications, Macros and additional research content can be found in the [Research tab](#) on WRDS main page.

If you have any questions about the material covered in this session, please contact [wrds-support](#)



Wharton  
UNIVERSITY *of* PENNSYLVANIA

The Global Standard for Business Research

wrds