

# Financial Statement Data Sets

## Contents

- [1 Overview](#)
- [2 Scope](#)
- [3 Organization](#)
- [4 File Formats](#)
- [5 Table Definitions](#)
  - [5.1 SUB \(Submissions\)](#)
  - [5.2 TAG \(Tags\)](#)
  - [5.3 NUM \(Numbers\)](#)
  - [5.4 PRE \(Presentation of Statements\)](#)

[Figure 1. Data relationships](#)

[Figure 2. Fields in the SUB data set](#)

[Figure 3. Fields in the TAG data set](#)

[Figure 4. Fields in the NUM data set](#)

[Figure 5. Fields in the PRE data set](#)

## 1 Overview

The following data sets provide information extracted from XBRL submissions filed with the Commission in a flattened data format to assist users in more easily consuming the data for analysis. The data is sourced from selected information found in the XBRL tagged financial statements submitted by filers to the Commission. These data sets currently include quarterly and annual numeric data rendered by the Commission in the primary financial statements submitted by filers. Certain additional fields (e.g. Standard Industrial Classification (SIC)) used in the Commission's EDGAR system are also included to help in supporting the use of the data. The information has been taken directly from submissions created by each registrant, and the data is "as filed" by the registrant. The information will be updated quarterly. Data contained in documents filed after the last business day of the quarter will be included in the next quarterly posting.

**DISCLAIMER:** The Financial Statement Data Sets contain information derived from structured data filed with the Commission by individual registrants as well as Commission-generated filing identifiers. Because the data sets are derived from information provided by individual registrants, we cannot guarantee the accuracy of the data sets. In addition, it is possible inaccuracies or other errors were introduced into the data sets during the process of extracting the data and compiling the data sets. Finally, the data sets do not reflect all available information, including certain metadata associated with Commission filings. The data sets are intended to assist the public in analyzing data contained in Commission filings; however, they are not a substitute for such filings. Investors should review the full Commission filings before making any investment decision.

The data extracted from the XBRL submissions is organized into four data sets containing information about submissions, numbers, taxonomy tags, and presentation. Each data set consists of rows and columns and is provided as a tab-delimited TXT format file. The data sets are as follows:

- **SUB** – Submission data set; this includes one record for each XBRL submission with amounts rendered by the Commission in the primary financial statements. The set includes fields of information pertinent to the submission and the filing entity. Information is extracted from the SEC’s EDGAR system and the filings submitted to the SEC by registrants.
- **NUM** – Number data set; this includes one row for each distinct amount appearing on the primary financial statements rendered by the Commission from each submission included in the SUB data set.
- **TAG** – Tag data set; includes defining information about each numerical tag. Information includes tag descriptions (documentation labels), taxonomy version information and other tag attributes.
- **PRE** – Presentation data set; this provides information about how the tags and numbers were presented in the primary financial statements as rendered by the Commission.

## 2 Scope

The scope of the data in the financial statement data sets consists of:

- Numeric data on the primary financial statements as rendered by the Commission (Balance Sheet, Income Statement, Cash Flows, Changes in Equity, and Comprehensive Income) and page footnotes on those statements;
- From XBRL submissions which include financial statements rendered by the Commission (e.g., 10-K, 10-Q, 20-F, 40-F);
- Submitted from 4/15/2009 through the “Data Cutoff Date” inclusive (there is a file named 2009q1.zip on the SEC website that contains data sets with column headings only and no rows, merely so that all years prior to this year will consist of four zip files).

All numeric data is “as filed.”

## 3 Organization

Note that this data set represents quarterly and annual uncorrected and “as filed” EDGAR document submissions containing multiple reporting periods (including amendments of prior submissions). Data in this submitted form may contain redundancies, inconsistencies, and discrepancies relative to other publication formats. There are four data sets.

1. SUB identifies all the EDGAR submissions with amounts rendered by the Commission on the primary financial statements in the data set, with each row having the unique (primary) key **adsh**, a 20 character EDGAR Accession Number with dashes in positions 11 and 14.
2. TAG is a data set of all numerical tags used in the submissions, both standard and custom. A unique key of each row is a combination of these fields:
  - 1) **tag** – tag used by the filer
  - 2) **version** – if a standard tag, the taxonomy of origin, otherwise equal to adsh.
3. NUM is a data set of all numeric XBRL facts presented on the primary financial statements as rendered by the Commission. A unique key of each row is a combination of the following fields:
  - 1) **adsh**- EDGAR accession number

- 2) **tag** – tag used by the filer
  - 3) **version** – if a standard tag, the taxonomy of origin, otherwise equal to adsh.
  - 4) **ddate** - period end date
  - 5) **qtrs** - duration in number of quarters
  - 6) **uom** - unit of measure
  - 7) **segments** – XBRL tags used to represent axis and member reporting
  - 8) **coreg** - coregistrant of the parent company registrant (if applicable)
4. PRE is a data set that provides the text assigned by the filer to each line item in the primary financial statements, the order in which the line item appeared, and the tag assigned to it. A unique key of each row is a combination of the following fields:
- 1) **adsh** – EDGAR accession number
  - 2) **report** – sequential number of report within the statements
  - 3) **line** – sequential number of line within a report.

The relationship of the data sets is as shown in Figure 1. The Accession Number (**adsh**) found in the NUM data set can be used to retrieve information about the submission in SUB. Each row of data in NUM was tagged by the filer using a tag. Information about the tag used can be found in TAG. Each row of data in NUM appears on one or more lines of reports detailed in PRE.

**Figure 1. Data relationships**

Dataset	Columns referencing other datasets	Referenced dataset	Referenced columns
<b>NUM</b>	adsh	SUB	adsh
	tag, version	TAG	tag, version
<b>PRE</b>	adsh	SUB	adsh
	tag, version	TAG	tag, version
	adsh, tag, version	NUM	adsh, tag, version

Note: The SEC website folder <http://www.sec.gov/Archives/edgar/data/{cik}/{accession}/> will always contain all the files for a given submission, where {accession} is the **adsh** with the '-' characters removed.

## 4 File Formats

Each of the four data sets is provided in a single encoding, as follows:

Tab Delimited Value (.txt): utf-8, tab-delimited, \n- terminated lines, with the first line containing the column names in lowercase.

## 5 Table Definitions

The columns in the figures below (figures 2 – 5) provide the following information: field name, description, source (SUB file only), data format, maximum field size, an indication of whether or not the field may be NULL (yes or no), and key.

The Source column in the SUB file has two possible values:

- EDGAR indicates that the source of the data is the filer's EDGAR submission header.
- XBRL indicates that the source of the data is the filer's XBRL submission.

The Key column indicates whether the field is part of a unique index on the data. There are two possible values for this column:

- "\*" – Indicates the field is part of a unique key for the row.
- Empty (nothing in column) – the column is a function of all or some of a unique key.

### 5.1 SUB (Submissions)

The submissions data set contains summary information about an entire EDGAR submission. Some fields were sourced directly from EDGAR submission information, while other columns of data were sourced from the XBRL submission. Note: EDGAR derived fields represent the most recent EDGAR assignment as of a given filing's submission date and do not necessarily represent the most current assignments.

**Figure 2. Fields in the SUB data set**

Field Name	Field Description	Source	Format	Max Size	May be NULL	Key
<b>adsh</b>	Accession Number. The 20-character string formed from the 18-digit number assigned by the SEC to each EDGAR submission.	EDGAR	ALPHANUMERIC (nnnnnnnnnnnn-nn-nnnnnn)	20	No	*
<b>cik</b>	Central Index Key (CIK). Ten digit number assigned by the SEC to each registrant that submits filings.	EDGAR	NUMERIC	10	No	
<b>name</b>	Name of registrant. This corresponds to the name of the legal entity as recorded in EDGAR as of the filing date.	EDGAR	ALPHANUMERIC	150	No	
<b>sic</b>	Standard Industrial Classification (SIC). Four digit code assigned by the SEC as of the filing date, indicating the registrant's type of business.	EDGAR	NUMERIC	4	Yes	
<b>countryba</b>	The ISO 3166-1 country of the registrant's business address.	EDGAR	ALPHANUMERIC	2	Yes	
<b>stprba</b>	The state or province of the registrant's business address, if field countryba is US or CA.	EDGAR	ALPHANUMERIC	2	Yes	

Field Name	Field Description	Source	Format	Max Size	May be NULL	Key
<b>cityba</b>	The city of the registrant's business address.	EDGAR	ALPHANUMERIC	30	Yes	
<b>zipba</b>	The zip code of the registrant's business address.	EDGAR	ALPHANUMERIC	10	Yes	
<b>bas1</b>	The first line of the street of the registrant's business address.	EDGAR	ALPHANUMERIC	40	Yes	
<b>bas2</b>	The second line of the street of the registrant's business address.	EDGAR	ALPHANUMERIC	40	Yes	
<b>baph</b>	The phone number of the registrant's business address.	EDGAR	ALPHANUMERIC	20	Yes	
<b>countryma</b>	The ISO 3166-1 country of the registrant's mailing address.	EDGAR	ALPHANUMERIC	2	Yes	
<b>stprma</b>	The state or province of the registrant's mailing address, if field countryma is US or CA.	EDGAR	ALPHANUMERIC	2	Yes	
<b>cityma</b>	The city of the registrant's mailing address.	EDGAR	ALPHANUMERIC	30	Yes	
<b>zipma</b>	The zip code of the registrant's mailing address.	EDGAR	ALPHANUMERIC	10	Yes	
<b>mas1</b>	The first line of the street of the registrant's mailing address.	EDGAR	ALPHANUMERIC	40	Yes	
<b>mas2</b>	The second line of the street of the registrant's mailing address.	EDGAR	ALPHANUMERIC	40	Yes	
<b>countryinc</b>	The ISO 3166-1 country of incorporation for the registrant.	EDGAR	ALPHANUMERIC	3	Yes	
<b>stprinc</b>	The state or province of incorporation for the registrant, if countryinc is US or CA.	EDGAR	ALPHANUMERIC	2	Yes	
<b>ein</b>	Employee Identification Number, 9 digit identification number assigned by the Internal Revenue Service to business entities operating in the United States.	EDGAR	NUMERIC	10	Yes	
<b>former</b>	Most recent former name of the registrant, if any.	EDGAR	ALPHANUMERIC	150	Yes	
<b>changed</b>	Date of change from the former name, if any.	EDGAR	ALPHANUMERIC	8	Yes	

Field Name	Field Description	Source	Format	Max Size	May be NULL	Key
<b>afs</b>	Filer status with the SEC at the time of submission: 1-LAF=Large Accelerated, 2-ACC=Accelerated, 3-SRA=Smaller Reporting Accelerated, 4-NON=Non-Accelerated, 5-SML=Smaller Reporting Filer, NULL=not assigned.	XBRL	ALPHANUMERIC	5	Yes	
<b>wksi</b>	Well Known Seasoned Issuer (WKSI). An issuer that meets specific SEC requirements at some point during a 60-day period preceding the date the issuer satisfies its obligation to update its shelf registration statement.	XBRL	BOOLEAN (1 if true and 0 if false)	1	No	
<b>fye</b>	Fiscal Year End Date, rounded to nearest month-end.	XBRL	ALPHANUMERIC (mmdd)	4	Yes	
<b>form</b>	The submission type of the registrant's filing.	EDGAR	ALPHANUMERIC	10	No	
<b>period</b>	Balance Sheet Date, rounded to nearest month-end.	XBRL	DATE (yyyymmdd)	8	No	
<b>fy</b>	Fiscal Year Focus (as defined in the EDGAR XBRL Guide Ch. 3.1.8).	XBRL	YEAR (yyyy)	4	Yes	
<b>fp</b>	Fiscal Period Focus (as defined in the EDGAR XBRL Guide Ch. 3.1.8) within Fiscal Year.	XBRL	ALPHANUMERIC (FY, Q1, Q2, Q3, Q4)	2	Yes	
<b>filed</b>	The date of the registrant's filing with the Commission.	EDGAR	DATE (yyyymmdd)	8	No	
<b>accepted</b>	The acceptance date and time of the registrant's filing with the Commission.	EDGAR	DATETIME (yyyy-mm-dd hh:mm:ss)	19	No	
<b>prevrpt</b>	Previous Report –TRUE indicates that the submission information was subsequently amended.	EDGAR	BOOLEAN (1 if true and 0 if false)	1	No	
<b>detail</b>	TRUE indicates that the XBRL submission contains quantitative disclosures within the footnotes and schedules at the required detail level (e.g., each amount).	XBRL	BOOLEAN (1 if true and 0 if false)	1	No	
<b>instance</b>	The name of the submitted XBRL Instance Document. The name often begins with the company ticker symbol.	EDGAR	ALPHANUMERIC (e.g. abcd-yyyymmdd.xml)	40	No	

Field Name	Field Description	Source	Format	Max Size	May be NULL	Key
<b>nciks</b>	Number of Central Index Keys (CIK) of registrants (i.e., business units) included in the consolidating entity's submitted filing.	EDGAR	NUMERIC	4	No	
<b>aciks</b>	Additional CIKs of co-registrants included in a consolidating entity's EDGAR submission, separated by spaces. If there are no other co-registrants (i.e., nciks=1), the value of aciks is NULL. For a very small number of filers, the entire list of co-registrants is too long to fit in the field. Where this is the case, users should refer to the complete submission file for all CIK information.	EDGAR	ALPHANUMERIC (space delimited)	120	Yes	

Note: To access the complete submission files for a given filing, please see the [SEC EDGAR website](http://www.sec.gov/Archives/edgar/data/{cik}/{accession}/). The SEC website folder <http://www.sec.gov/Archives/edgar/data/{cik}/{accession}/> will always contain all the files for a given submission. To assemble the folder address to any filing referenced in the SUB data set, simply substitute {cik} with the **cik** field and replace {accession} with the **adsh** field (after removing the dash character). The following sample SQL Query provides an example of how to generate a list of addresses for filings contained in the SUB data set:

- ```
select name,form,period, 'http://www.sec.gov/Archives/edgar/data/' +
ltrim(str(cik,10))+ '/' + replace(adsh, '-', '') + '/' + instance as url from SUB sub order by
period desc, name
```

## 5.2 TAG (Tags)

The TAG data set contains the standard taxonomy tags and the custom taxonomy tags defined in the submissions. The source is the "as filed" XBRL filer submissions. The standard tags are derived from taxonomies in <https://www.sec.gov/data-research/standard-taxonomies>.

**Figure 3. Fields in the TAG data set**

| Field Name     | Field Description                                                                                                                          | Field Type                         | Max Size | May be NULL | Key |
|----------------|--------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------|----------|-------------|-----|
| <b>tag</b>     | The unique identifier (name) for a tag in a specific taxonomy release.                                                                     | ALPHANUMERIC                       | 256      | No          | *   |
| <b>version</b> | For a standard tag, an identifier for the taxonomy; otherwise the accession number where the tag was defined.                              | ALPHANUMERIC                       | 20       | No          | *   |
| <b>custom</b>  | 1 if tag is custom (version=adsh), 0 if it is standard. <i>Note: This flag is technically redundant with the version and adsh columns.</i> | BOOLEAN (1 if true and 0 if false) | 1        | No          |     |

| Field Name      | Field Description                                                                                                                                                                         | Field Type                         | Max Size | May be NULL | Key |
|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------|----------|-------------|-----|
| <b>abstract</b> | 1 if the tag is not used to represent a numeric fact.                                                                                                                                     | BOOLEAN (1 if true and 0 if false) | 1        | No          |     |
| <b>datatype</b> | If abstract=1, then NULL, otherwise the data type (e.g., monetary) for the tag.                                                                                                           | ALPHANUMERIC                       | 20       | Yes         |     |
| <b>iord</b>     | If abstract=1, then NULL; otherwise, "I" if the value is a point-in time, or "D" if the value is a duration.                                                                              | ALPHANUMERIC                       | 1        | No          |     |
| <b>crdr</b>     | If datatype = monetary, then the tag's natural accounting balance (debit or credit); if not defined, then NULL.                                                                           | ALPHANUMERIC ("C" or "D")          | 1        | Yes         |     |
| <b>tlabel</b>   | If a standard tag, then the label text provided by the taxonomy, otherwise the text provided by the filer. A tag which had neither would have a NULL value here.                          | ALPHANUMERIC                       | 512      | Yes         |     |
| <b>doc</b>      | The detailed definition for the tag. If a standard tag, then the text provided by the taxonomy, otherwise the text assigned by the filer. Some tags have neither, and this field is NULL. | ALPHANUMERIC                       |          | Yes         |     |

### 5.3 NUM (Numbers)

The NUM data set contains numeric data, one row per data point as rendered by the Commission on the primary financial statements. The source for the table is the "as filed" XBRL filer submissions.

**Figure 4. Fields in the NUM data set**

| Field Name      | Field Description                                                                                                                                  | Field Type (format) | Max Size | May be NULL | Key |
|-----------------|----------------------------------------------------------------------------------------------------------------------------------------------------|---------------------|----------|-------------|-----|
| <b>adsh</b>     | Accession Number. The 20-character string formed from the 18-digit number assigned by the SEC to each EDGAR submission.                            | ALPHANUMERIC        | 20       | No          | *   |
| <b>tag</b>      | The unique identifier (name) for a tag in a specific taxonomy release.                                                                             | ALPHANUMERIC        | 256      | No          | *   |
| <b>version</b>  | For a standard tag, an identifier for the taxonomy; otherwise the accession number where the tag was defined.                                      | ALPHANUMERIC        | 20       | No          | *   |
| <b>ddate</b>    | The end date for the data value, rounded to the nearest month end.                                                                                 | DATE (yyyymmdd)     | 8        | No          | *   |
| <b>qtrs</b>     | The count of the number of quarters represented by the data value, rounded to the nearest whole number. "0" indicates it is a point-in-time value. | NUMERIC             | 8        | No          | *   |
| <b>uom</b>      | The unit of measure for the value.                                                                                                                 | ALPHANUMERIC        | 20       | No          | *   |
| <b>segments</b> | Tags used to represent axis and member reporting.                                                                                                  | ALPHANUMERIC        | 1024     | Yes         | *   |
| <b>coreg</b>    | If specified, indicates a specific co-registrant, the parent company, or other entity (e.g., guarantor). NULL indicates the consolidated entity.   | ALPHANUMERIC        | 256      | Yes         | *   |



| Field Name      | Field Description                                                                                                                                                             | Field Type (format) | Max Size | May be NULL | Key |
|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------|----------|-------------|-----|
| <b>value</b>    | The value. This is not scaled, it is as found in the Interactive Data file, but is limited to four digits to the right of the decimal point.                                  | NUMERIC(28,4)       | 16       | Yes         |     |
| <b>footnote</b> | The text of any superscripted footnotes on the value, as shown on the statement page, truncated to 512 characters, or if there is no footnote, then this field will be blank. | ALPHANUMERIC        | 512      | Yes         |     |

## 5.4 PRE (Presentation of Statements)

The PRE data set contains one row for each line of the financial statements tagged by the filer. The source for the data set is the “as filed” XBRL filer submissions. Note that there may be more than one row per entry in NUM because the same tag can appear in more than one statement (the tag NetIncome, for example can appear in both the Income Statement and Cash Flows in a single financial statement, and the tag Cash may appear in both the Balance Sheet and Cash Flows).

**Figure 5. Fields in the PRE data set**

| Field Name    | Field Description                                                                                                                                                                                      | Field Type (format)                                                                                                                                                            | Max Size | May be NULL | Key |
|---------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------|-------------|-----|
| <b>adsh</b>   | Accession Number. The 20-character string formed from the 18-digit number assigned by the SEC to each EDGAR submission.                                                                                | ALPHANUMERIC                                                                                                                                                                   | 20       | No          | *   |
| <b>report</b> | Represents the report grouping. This field corresponds to the statement (stmt) field, which indicates the type of statement. The numeric value refers to the “R file” as posted on the EDGAR Web site. | NUMERIC                                                                                                                                                                        | 6        | No          | *   |
| <b>line</b>   | Represents the tag’s presentation line order for a given report. Together with the statement and report field, presentation location, order and grouping can be derived.                               | NUMERIC                                                                                                                                                                        | 6        | No          | *   |
| <b>stmt</b>   | The financial statement location to which the value of the “report field pertains.                                                                                                                     | ALPHANUMERIC (BS = Balance Sheet, IS = Income Statement, CF = Cash Flow, EQ = Equity, CI = Comprehensive Income, SI = Schedule of Investments, UN = Unclassifiable Statement). | 2        | No          |     |
| <b>inpth</b>  | Value was presented “parenthetically” instead of in columns within the financial statements. For example: Receivables ( <i>net of allowance for bad debts of \$200 in 2012</i> ) \$700.                | BOOLEAN (1 if true and 0 if false)                                                                                                                                             | 1        | No          |     |
| <b>rfile</b>  | The type of interactive data file rendered on the EDGAR web site, H = .htm file, X = .xml file.                                                                                                        | ALPHANUMERIC                                                                                                                                                                   | 1        | No          |     |

| Field Name      | Field Description                                                       | Field Type (format)                | Max Size | May be NULL | Key |
|-----------------|-------------------------------------------------------------------------|------------------------------------|----------|-------------|-----|
| <b>tag</b>      | The tag chosen by the filer for this line item.                         | ALPHANUMERIC                       | 256      | No          |     |
| <b>version</b>  | The taxonomy identifier if the tag is a standard tag, otherwise adsh.   | ALPHANUMERIC                       | 20       | No          |     |
| <b>plabel</b>   | The text presented on the line item, also known as a "preferred" label. | ALPHANUMERIC                       | 512      | No          |     |
| <b>negating</b> | Flag to indicate whether the plabel is negating.                        | BOOLEAN (1 if true and 0 if false) | 1        | No          |     |