User Guide: Document Distributor – Transcripts V1.0.2

Last Edited: July 2017

Table of Contents

Notice

Any SQL code provided in this document is intended for illustrative purposes only. The functions and queries herein are provided in SQL formatted for MSSQL. It may to be necessary adjust the query language to suit a different SQL DBMS. Sample queries should be viewed as possible examples of data calculations and are subject to change. The queries provided should be used as a guide to understand how the underlying data items can be used but are not guaranteed to represent the same methodology as the FactSet Workstation or other industry calculations. Queries cannot be guaranteed to be written for performance and/or efficiency.

The information contained in this document is provided "as is" and all representations, warranties, terms and conditions, oral or written, express or implied (by common law, statute or otherwise), in relation to the information are hereby excluded and disclaimed to the fullest extent permitted by law. In particular, FactSet, its affiliates, and its suppliers disclaim implied warranties of merchantability and fitness for a particular purpose and make no warranty of accuracy, completeness or reliability of the information.

Updates

Effective Date	Version	Change Type	Description	Affected Files
18-JUL-2017	1.0.2	Document Update	Edit DocRetrieval queries Edit maintaining the local repository	
12-MAR-2017	1.0.1	Document Update	Edit DocRetrieval queries	
03-MAR-2017	1.0.0	Document Update	Document created	

Overview

This user guide reviews FactSet's Document Distributor product for Transcripts, which covers events for over 49,000 companies (about 39,000 active companies). This product is comprised of three components: a Document Archive, FactSet Events Standard Datafeed, and OnDemand's DocRetrieval Service.

The Document Archives contains XML versions of all FactSet transcripts since 1999.

The FactSet Events Standard DataFeed provides metadata for over 1.7 million company corporate events, such as conference call date and time, phone number and password, type of conference call, and company investor relations contact information.

FactSet's OnDemand DocRetrieval provides robust searching capabilities and intraday access to XML transcripts.

Packages

FactSet's *Document Distributor – Transcripts* product consists of five main packages:

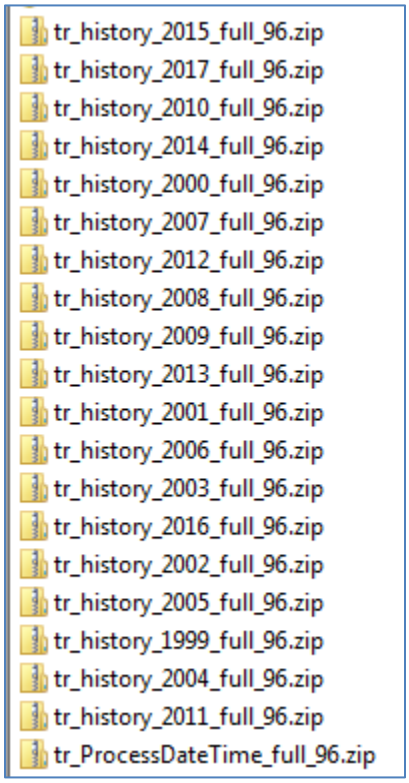
- *Document Archives – Transcripts*
  - Includes the most recent Raw and Corrected versions of all transcripts in the FactSet universe in XML format.
- *Standard DataFeed - Future Events with Full History Package*
  - Includes metadata associated with all future and historical events
- *Standard DataFeed – Transcript Versions*
  - Includes metadata, such as the upload time, associated with current and prior versions of transcripts
- *OnDemand – DocRetreival*
  - This service gives subscribers the ability to download new and revised transcript files on an ongoing basis.
- *Standard DataFeed - Symbology - Historical Mappings - Equity*
  - Includes historical CUSIPs and SEDOLs to permid mappings for equity securities.

Document Archives – Transcripts

Clients subscribing to the Document Distributor – Transcripts product are entitled to download the complete history of transcripts in the FactSet universe from the Document Archives. This download provides one zip file for each year that transcripts are available based on the Event Date and is updated **daily at approximately 12:00 pm EST**. These zip files have the following naming convention:

tr_history_<YYYY>_full_<sequence>.zip

The <YYYY> string indicates the year of the Event for every transcript contained in the zip. The <sequence> string will be a number, unique to each complete set of transcript produced by FactSet (i.e. each version of the Document Archives cut by FactSet). To ensure the download was successful, it is important that all files downloaded contain the same sequence in their name. Below is an example of what the download will look like, once complete, in your local directory:



The transcript xml files, within their respective zip files, will be in the following format: ***<event date>-<report id>-<transcript type>.xml.*** Only the most recent Raw and Corrected versions are provided for each transcript. For details on the contents of these XML files, please see the "Transcript XML Data Dictionary" section of this document (below).

Aside from transcript zip files, a zip file called "tr_ProcessDateTime_full_<sequence>.zip" will also be available. This zip contains a single text file that bears the same name as its zip ("tr_ProcessDateTime_full_<sequence>.txt"). The text file will contain a date and timestamp (in UTC) indicating when the version of Document Archives that was downloaded was produced. Once the Document Archives are completely downloaded, subscribers should query OnDemand to retreive all transcripts produced by FactSet after their version of the Document Archives was generated. The timestamp within the "tr_ProcessDateTime_full_<sequence>.txt" file is reduced by 10 minutes as a precaution to ensure there is some overlap between the transcript archive and the OnDemand query results (i.e. to ensure there is no gap in coverage of transcripts). Subscribers should continue querying OnDemand on an ongoing basis to ensure their local repository of transcripts is complete and current. More information on querying OnDemand can be found in the "OnDemand – DocRetreival" section of this document.

The Document Archives can be downloaded with the DataFeed Loader using the normal "FDSLoader.exe" command. Once the Loader has completed downloading the Document Archives, zip files containing the XML transcripts can be found in the directory specified during the DataFeed Loader's setup. If FactSet adds a transcript with a never before seen Event Year, a new zip file will be added to the Document Archives. This will occur, for example, on or near the first day of each calendar year. It is important to note that, when this happens, the DataFeed Loader will recognize a new zip file has been added and will automatically re-download the Document Archives to the local directory specified in during the Loader' setup on its next

scheduled run. Previously downloaded versions (sequences) of the Document Archives ***will not*** be automatically removed from your local directory.

For more information on the Loader setup process and commands (including rebuild commands), please see the User Guide here.

Standard DataFeed Packages

The "Standard DataFeed - Future Events with Full History" package provides events and report metadata for all events in the FactSet universe. The feed provides full files weekly and delta files hourly, to the extent historical events have changed or new events have been added. Subscribers can use this package to:

- Identify recently released or recently revised transcripts that must be fetched from the OnDemand DocRetreival service (described below).
- Identify specific events and transcripts of interest (i.e. year, event type, company, etc). Once events and transcripts of interest are identified, they can be retrieved from your local download of the Document Archives or fetched from the OnDemand DocRetreival service for further analysis.
- Easily connect transcripts with other FactSet content sets, such as Estimates and People.

The tables available in the "Standard DataFeed - Future Events with Full" package, along with their descriptions, are provided below:

Table Name	Table Description
CE_CONFERENCES	This table contains detailed information on conferences including their locations and times.
CE_CONFERENCES_COVERAGE	This table links conferences to their hosts.
CE_COVERAGE	This table contains Investor Relation information for all companies covered within the Events DataFeed.
CE_EVENTS	This table contains detailed information on all events types collected by FactSet. This table includes the Event Date and Event Type.
CE_EVENTS_COVERAGE	This table links each entity to their respective event(s).
CE_PARTICIPANTS	This table contains detailed information on individual participants of event, including their company and job titles.
CE_REPORTS	This table contains information on the transcription progress of an event.
CE_REPORT_SLIDES	This table contains the URL links for slides or other presentation material that were presented during the conference call.
CE_TRANSCRIPT_VERSIONS	This table contains metadata associated with previously uploaded versions and current versions of transcripts. This table includes the "upload_datetime" field, which indicates when the transcription of each call was completed by FactSet.
CE_SEC_ENTITY	This table contains company to equity security mappings for all companies covered within the Events DataFeed.

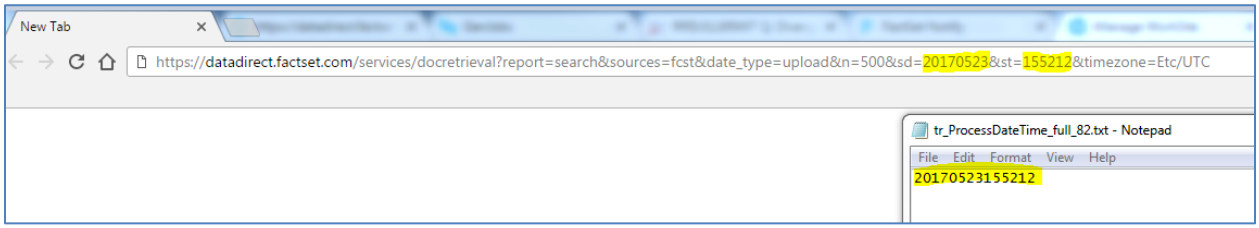
For a complete description of the "Standard DataFeed - Future Events with Full History" package, please see the User Guide on page 15222 of FactSet's Online Assistant (OA). You must log in using your factset.net ID as your username. If you do not have a username, please contact your FactSet Account Representative.

OnDemand – DocRetreival

After your historical transcript backfill, the OnDemand DocRetreival service is used to maintain your local repository. This service can be queried, using multiple parameters, to identify and download new and updated transcripts. Below are a few examples of common DocRetrieval requests that can be made:

https://datadirect.factset.com/services/docretrieval?report=search&sources=fcst&date_type=upload&n=500&sd=20170531&st=155212&timezone=Etc/UTC

Transcripts FactSet completed transcribing since the 31st of May 2017 at 15:21:12UTC, limited to 500 results. This may include newer versions of transcripts for events that have already been transcribed. **Note** – This is the preferred request for your first download of transcripts from OnDemand after your initial download of the Document Archives. To ensure no gaps exist between your version of the Document Archives and the first transcript downloaded via OnDemand, enter the date and time sent in the tr_ProcessDateTime_full_<sequence>.txt file (see the "Document Archives – Transcripts" section of this document) into this URL. The results will list the most recent versions of each distinct transcript added by FactSet since your downloaded version of the Document Archives was produced:



https://datadirect.factset.com/services/docretrieval?report=search&sources=fcst&report_ids=1954376,1946003,1956312&sd=-20

This looks for specific transcripts via the unique ReportIDs (report_ids=). It is advisable to include a date range in order to speed up the request. In this case, a start date of -20 days has been used to optimise the search window (sd=-20).

https://datadirect.factset.com/services/docretrieval?report=search&sources=fcst&sd=20170508&ed=20170508&event_ids=5761198

This looks for all reports tied to a specific event: the acquisition of Kate Spade & Company by Coach (event_ids= 5761198). Each company has its own report for this event with unique Report IDs. However, they share the same Event ID as there was only one physical call that took place.

https://datadirect.factset.com/services/docretrieval?report=search&sources=fcst&ids=IBM-US,MSFT-US,CSCO-US&sd=-100

Transcripts of IBM, Microsoft and Cisco's (ids=IMB-US,MSFT-US,CSCO-US) events that occurred in the last 100 days (sd=100).

Note – The suggested file name provided by OnDemand will include an Event Date string. It is important to remember that, because and Event's date can change, the suggested file name for a transcript could differ from the transcript's file name in your local transcript document repository. If you wish to replace old transcripts with new transcripts fetched from OnDemand, determine whether the transcript exists locally based on the Report ID and Transcript Type ("C" or "T") strings of the file name.

For a full list of parameters, please see the "FactSet OnDemand - DocRetrieval Service" User Guide on page 16925 of FactSet's **Online Assistant (OA)**. You must log in using your factset.net ID as your username. If you do not have a username, please contact your FactSet Account Representative.

Using the Standard DataFeed to Query OnDemand

For subscribers who are interested in a particular category or subset of transcripts, the Standard DataFeed package can be leveraged to identify transcripts of interest. Subscribers can generate a list of Report IDs from SQL to be used as a search parameter in the OnDemand – DocRetrieval Service. Below are common queries performed on the Standard DataFeed to identify specific types of transcripts:

1) Transcripts from a particular company:

```
select distinct r.report_id
from evt_v1.ce_reports r
where transcript_completed is not null
and r.factset_entity_id = <FACTSET ENTITY ID>
```

2) Transcripts from companies in a particular country:

```
select distinct r.report_id
from evt_v1.ce_reports r
left join sym_v1.sym_entity e on r.factset_entity_id=e.factset_entity_id
where transcript_completed is not null
and e.iso_country=<ISO_COUNTRY>
```

3) Connection the Events feed to the Standard DataFeed - Peoples feed. View the Names and bios of all participants on a call (**NOTE** – This query requires you to have access to the StandardDataFeed – Peoples feed):

```
select
c.entity_proper_name,b.bio
from sdf_coombs_dftester.evt_v1.ce_participants a
left join sdfdemo.ppl_v1.ppl_bio b on a.factset_person_id=b.factset_person_id
left join sdfdemo.sym_v1.sym_entity c on a.factset_person_id=c.factset_entity_id
where a.report_id=<REPORT ID>
```

Similarly, the Standard DataFeed can be used as a master record to ensure your local repository is up to date with new or updated transcripts. Users can check for updates based on upload or event date then use the associated Report ID to retrieve the transcript from OnDemand DocRetrieval. For example:

1) Transcripts newly transcribed or updated today:

```
select distinct report_id
from evt_v1.ce_transcript_versions tv
where convert(date,upload_datetime)=convert(date,getdate())
```

2) Transcripts with Event Date/Time from a specific period of time:

```
select distinct r.report_id
from evt_v1.ce_reports r
join evt_v1.ce_events e  on r.event_id = e.event_id
where transcript_completed is not null
and convert(date, event_datetime_utc) between <START DATE> and <END DATE>
```

Lastly, the recommended file naming convention for a given report or list of reports can be generated using the metadata provided in the Standard DataFeed. This is the same naming convention used for transcript XMLs found in the Document Archives. Below is an example:

```
select
replace(CAST(CAST(event_datetime_utc as date) as varchar),'-','')+'-
'+CAST(r.report_id as varchar)+'-'+replace(r.transcript_completed,'R','T')+'.xml'
,e.event_type
from evt_v1.ce_reports r
join evt_v1.ce_events e  on r.event_id = e.event_id
where r.report_id=<REPORT ID>
```

Recommended for Maintaining Your Local Repository

Below are the recommended steps to maintain a repository of the transcripts available from FactSet. The steps assume that you want every transcript available:

1. Use the DataFeed Loader to download the full XML Transcript Documents Archive
2. Use DocRetrieval to search for every transcript with upload_datetime after the time provided in the tr_ProcessDateTime_full_#.zip file
    a. Note the time in tr_ProcessDateTime_full_#.zip file is in UTC whereas DocRetrieval by default is in New York time. To remain consistent, it is advised to add the argument, timezone=Etc/UTC to your DocRetrieval request.
3. Store the most recent upload_datetime of all transcripts from the results.
4. Each subsequent DocRetrieval request should search for every transcript with upload_datetime after the most recent known upload_datetime found in Step 3.
5. Repeat steps 3 and 4 indefinitely.

The Events Stanadard DataFeed can be used to perform a reconciliation between your local repository and available transcripts. The following steps outline this process:

1. Within the Events Standard DataFeed, run the following query to get a list of expected transcripts:

```
select distinct report_id from evt_v1.ce_reports where transcript_completed is not null
```

2. Compare this list of IDs generated from the above query with the list of transcripts available in your local repository.
3. A comma-separated list of Report IDs for the missing transcripts (identified in the above comparison) should be included in the following OnDemand command and fetched from the DocRetrieval Service:

https://datadirect.factset.com/services/docretrieval?report=search&sources=fcst&report_ids=**[LIST_REPORT_IDS]**

FactSet recommends this reconciliation process be performed on a regular basis to ensure historical transcripts that are outside of FactSet's the normal coverage are not missing from your local repository. For more information on these transcripts, please see the "Embargoed Transcripts" section of the Content Methodology Guide on Page 15222 of Online Assistant (OA).

Transcript XML Files

Transcripts are provided in XML format. Two versions of a transcript could exist: the Raw version (identified with "T" in the XML name) and a Corrected version (identified with "C" in the XML name). It is possible to receive multiple corrected versions if there are multiple rounds of editing performed. Prior to 2008, it is common that only one version (Raw or Corrected) for a given transcript is provided. However, nearly all transcripts after 2008 will have both a Raw and Corrected version.

The XML contains two major sections: Meta and Body. The Meta section will display basic event information, such as title and date, as well as detailed participant information. The Participants section will contain the Speaker ID associated with each individual and will be used to define the speaker in the Body. Participant information was added to our transcripts after May 2011. Therefore, earlier transcripts will typically lack participant details. Participant ID metadata may be missing or incomplete for Raw transcripts, even after May 2011. If a participant cannot be identified, "Unverified Participant" is displayed in the participants section of the XML.

The Body section is the transcript itself. The Body of the transcript is broken into sections, each prefaced by a header that includes the section name. Each section will contain an attribute defining the speaker, transcribed content broken out by paragraphs, and the type of text where possible (defining a question or answer during the Q&A section).

Within these transcripts, special characters are encoded. Therefore, it is advisable to decode when parsing. A full list of characters can be found here: http://www.ascii.cl/htmlcodes.htm

Below is an example of a typical transcript XML file:

```
<?xml version="1.0" encoding="us-ascii"?>
<transcript id="1912991" product="RawTranscript" xmlns="http://www.factset.com/callstreet/xmllayout/v0.1">
  <meta>
    <title>Q4 2016 Earnings Call</title>
    <date>2017-02-28</date>
    <companies>
      <company>48184</company>
    </companies>
    <participants>
      <participant id="0" type="operator">Operator</participant>
      <participant id="1" type="corprep" affiliation="Delek US Holdings, Inc." affiliation_entity="063VZF-E" title="V
      <participant id="2" type="corprep" affiliation="Delek US Holdings, Inc." affiliation_entity="00F0S1-E" title="E
      <participant id="3" type="corprep" affiliation="Delek US Holdings, Inc." affiliation_entity="00F0S1-E" title="V
      <participant id="4" type="corprep" affiliation="Chairman, President &amp;amp; Chief Executive Officer">Ezra Uzi Yem
    </participants>
  </meta>
  <body>
    <section name="MANAGEMENT DISCUSSION SECTION">
      <speaker id="0">
        <plist>
          <p>Good morning. My name is Tiffany, and I will be your conference operator today. At this time, I would li
          <p>Keith Johnson, you may begin your conference.</p>
        </plist>
      </speaker>
      <speaker id="1">
        <plist>
```

Note: Empty tags are provided in the following format: **<TagName/>**

XML Tags – Data Dictionary

Below are the descriptions for the tags found within the transcript XML files:

TAG	Data Type	VALUES	DESCRIPTION
Transcript Product	Varchar	CorrectedTranscript	XML product
Transcript ID	Integer	[0-9]{1-10}	Unique identifier for the transcript. This is the Report ID from the metadata file
Title	Integer + Text	No Limitations	Title of the event
Date	Date	YYYY-MM-DDTHH:MM:SSZ	Time of the event
Company Entity	Integer + Text	[0-9A-Z]{6}-E	Unique identity for the company
Company	Integer	[0-9]{1-5}	Unique identifier for the company. This is the Coverage ID from the metadata file
Participant ID	Integer	Ordinal system, 0+	Transcript specific ID
Participant Type	Varchar	Operator, corprep, other	Operator, Corprep = Hosting company participant, Other = Analysts
Participant Title	Text	No Limitations	Professional title
Participant Entity	Integer + Text	[0-9A-Z]{6}-E	Unique identifier for the participant. This is the Factset Person ID from the metadata file
Participant Affiliation Entity	Integer + Text	[0-9A-Z]{6}-E	Unique identifier for the participant's company. This is the Factset Entity ID from the metadata file
Participant Affiliation	Integer + Text	No Limitations	Company name
Section Name	Varchar	MANAGEMENT DISCUSSION SECTION, QUESTION AND ANSWER SECTION	Headers

Global Client Support

For assistance, please contact your local FactSet Consultant or Salesperson or e-mail feed_support@factset.com.

North and South America

United States and Canada	1.877.FACTSET
Brazil	0800.8917850
Mexico	001.888.542.9899

Europe

United Kingdom	0800.169.5954
Belgium	0800.94108
France	0800.484.414
Germany	0800.200.0320
Ireland, Republic of	1800.409.937
Italy	800.510.858
Netherlands	0800.228.8024
Norway	800.30365
Spain	900.811.921
Sweden	0200.110.263
Switzerland	0800.881.720
Europe/Middle Eastern countries not listed above	44.(0)20.7374.4445

Africa/Asia/Australia

Australia	1800.33.28.33
Hong Kong Consulting Services (for Hong Kong, China, India, Malaysia, Singapore, Sri Lanka, and Taiwan clients)	852.3011.4888
Japan Consulting Services (Japan/Korea)	0120.779.465 (Within Japan) 81.3.6268.5200 (Outside Japan)
Korea	080.411.0880
South Africa	0800.166.509
United Arab Emirates	800.0444.0014