THOMSON REUTERS NEWS ANALYTICS

USER GUIDE

NEWS ANALYTICS VERSION 4.0.6

DATA MODEL VERSION 2.13

Document Version 1.0
Date of Issue: 08 June 2017

THOMSON REUTERS

CONTENTS

ABOUT THIS DOCUMENT

INTENDED READERSHIP

This document is intended for two groups of readers.

- Users of News Analytics (TRNA) content interested in the data model that is present in both Elektron feed and archive data
- Developers writing applications to consume this content, either from an Elektron feed or from archive files

IN THIS GUIDE

This document describes the structure of News Analytics data and how to consume it from Elektron. It also describes the MRN FTP site that provides reference and archive files.

FEEDBACK

If you have any comments on this document, please contact the Thomson Reuters News Analytics team on tf.TRNASupport@thomsonreuters.com.

CHAPTER 1 NEWS ANALYTICS OVERVIEW

Today's volume of information across traditional News, Internet News and Social Media is so vast that extracting meaning is an uphill struggle. Capturing, analyzing and producing actionable data from this volume is one of today's biggest challenges but also presents big opportunities for advances in financial analysis. Up-to-date knowledge and interpretation of the events that shape the financial landscape provide crucial tools for both human and machine traders to generate value. This enables algorithms to exploit the power of news to seize opportunities, capitalize on market inefficiencies and manage event risk.

Powered by a unique computational linguistic processing system, Thomson Reuters News Analytics (TRNA) provides real-time numerical insight into the events on multiple news sources, in a format that can be directly consumed by algorithmic trading systems.

The live and archived versions of the data share a common structure and format.

1.1 CORE NEWS ANALYTICS METADATA

Each item of news within the story body feed is scored individually for each asset that is mentioned, specifically companies and commodities. For each asset, the following categories of derived data are calculated:

- Relevance: A number of measures of how relevant the news item is to the asset
- Sentiment: Whether the news item talks about the asset in a positive, neutral or negative manner
- Novelty: A measure of similarity of this item to previously seen news items
- Volume: Counts of the number of recent items mentioning the asset
- Headline Classification: Specific analysis of the headline

The TRNA feed consists of the calculated News Analytics scores along with news item metadata derived from the input.

1.2 SUPPORT AND NOTIFICATIONS

For support issues, one may contact Thomson Reuters via the My Account site at https://my.thomsonreuters.com/ContactUsNew?CUType=Incident. Here are some useful facts for you to state when making support inquiries, especially when calling, in order to help Thomson Reuters route your issue to the appropriate team most efficiently:

- Name of the product: News Analytics
- Context: feed, archive, or deployed TRNA software
- If a feed, how it is delivered: Elektron Connect, Elektron Deployed, Elektron Managed Services, or legacy delivery method News Feed Direct

Clients are advised to subscribe to service alerts regarding the live and archive TRNA services. To subscribe, go the Service Alert-Edit Subscriptions page, here. Then check the box at Application -> Financial Information Applications -> Enterprise Information Products ->Thomson Reuters Machine Readable News -> Thomson Reuters News Analytics. Then click the Update button at bottom right.

For updates on releases of new versions of TRNA or of the company/commodity coverage lists and for related updates, Thomson Reuters strongly recommends that clients sign up for Product Client Notifications (PCNs). This will send notifications via email in advance of relevant releases and updates. To sign up for PCNs on News Analytics, go to the My Subscriptions page, here. For News Analytics, check the box at All Products -> Reuters Enterprise Information -> Thomson Reuters News Analytics. Toward the bottom, make sure that the box for "All Client impacts", or at least "For Action" under that, is checked. After setting your reminder period, click the submit → link in the lower or upper right to confirm your selection.

For general information on TRNA, see the My Account page for this product at http://my.thomsonreuters.com/products?productname=thomson+reuters+news+analytics.

CHAPTER 2 STORY CONSTRUCTION AND NEWS CODING

2.1 EVOLUTION OF A STORY

A story[1] is a related collection of published news items.

The first part of a story may be an alert. This is a brief item containing the most essential information relating to an emerging story. Sometimes several alerts for the same story are filed in a quick succession.

Alerts may be followed up by a headline and the first textual piece of the story, i.e., story body. This text together with its associated headline is called a “take”. Subsequent takes may also be filed to contain additional text as needed.

A story is also attached to a set of category codes. These are transmitted with the alert and headline(s), and are described in the next section.

Each alert or take of the same story contains two time stamps: (1) the story date and time, and (2) the take date and time. The story date and time is the time (in UTC) that the first alert or take for that story was filed and remains the same for all alerts and takes of that story. The take date and time describe when that a particular alert or take was filed.

All parts of a story also contain a common identifier, the Primary News Access Code (PNAC). Because PNACs are reused, it is insufficient to identify a story. One must use the PNAC and story date/time together.

Typically, only news agencies such as Reuters News will ever publish stories over multiple news items. News items from press wires and exchange wires tend to publish a story within a single news item and do not issue alerts.

2.2 NEWS CODING

Machine Readable News (MRN) news items sourced from Thomson Reuters, including Reuters News and third-party non-Internet news sources, carry category codes which describe the content of the story.

The codes described in sections 2.2.1 through 2.2.4 can be found in the News Analytics data as well as the Real-time News, a separate product containing the raw news content. A reference file containing a list of active codes can be found on Customer Zone or My Account here.

2.2.1. Product Codes

Product codes identify which desktop news product(s) the news item belongs to. They are typically tailored to specific audiences.

Examples: “M” for Money International News Service and “FB” for the French General News Service

Identification: audiences field, with values prefixed by “NP:”

2.2.2. Topic Codes

Topic codes describe the news item‘s subject matter. These can cover asset classes, geographies, events, industries/sectors, and other types.

Examples: “CDV” for credit default swaps, “ENVS” for environmental services sector, and “DIV” for dividends

Identification: subjects field, prefixed by “N2:”. Note that in the data, adjacent to the N2 there will be alternative representations of those codes. For example, geographic codes begin with “G:”.

2.2.3. Named Item Codes

Named item codes, also known as recurring report codes, identify news items that follow a pattern. Items sharing the same code cover periodic updates to the same subject matter and often have very similar headlines.

[1] The use of “story” here should not be confused with the story body feed, known as Real-time News, which uses the MRN_STORY RIC. The latter is called “story” because it includes what is commonly referred to as “story body text”.

Examples: ".L" for news on UK stocks and "MEAL/DEL" for CBOT Soymeal Deliveries

Identification: instancesOf field, with values prefixed by "NI:"

2.2.4. Attributions

The attribution denotes the organization that published the news item.

Examples: "RTRS" for Reuters News and "BSW" for Business Wire

Identification: provider field, with values prefixed by "NS:"

2.2.5. RICs

This denotes a RIC, or Reuters Instrument Code, that is tagged to the news item. The most commonly used types of RICs in news typically represent companies, economic indicators, or foreign exchange pairs.

Examples: "IBM.N" for IBM, "EURGBP=" for the Euro/British Pound exchange rate, and "USNAHB=ECI" for the NAHB Housing Market Index

Identification:

- subjects field, with values prefixed by "R:"
- assetCodes field, among identifiers for the company asset scored by News Analytics. Prefixed by "R:".

2.2.5.1. Company PermIDs

Companies in TRNA are identified with an organizational-level identifier called the "PermID". Since RICs are quote identifiers, they are less stable over time than a company identifier. Furthermore, the many-to-one relationship between quotes and companies makes it possible also for multiple RICs to be tagged to news simultaneously. Thomson Reuters is thus exposing its privately mastered information model and making the PermID publicly available. See https://permid.org for more information on the PermID.

All live and archive News Analytics and News Sentiment Indices data use the PermID as the primary company identifier. In addition, most company RICs in the live Real-time News data will be supplemented with a PermID.

Example: 4295904307 for International Business Machines Corporation

Identification:

- assetId field, as the company asset scored by News Analytics. Not prefixed.
- assetCodes field, among identifiers for the company asset scored by News Analytics. Prefixed by "P:".
- subjects field, as one of the entities tagged to the news item. Prefixed by "P:".

CHAPTER 3 MRN MESSAGE ENVELOPE: PRESENTATION AND CONSUMPTION

3.1 OVERVIEW

News Analytics (TRNA) data falls under the Machine Readable News (MRN) model for data presentation and feed consumption.

MRN data is published over Elektron using an Open Message Model (OMM) envelope in News Text Analytics domain RSSL messages.

All MRN messages carry the same FIDs, regardless of the content set. The content-specific information is contained in a FRAGMENT (BUFFER type) FID that has been compressed, and potentially fragmented across multiple messages, in order to reduce bandwidth and RSSL message size. The TRNA data model is described in Chapter 4.

3.2 NEWS TEXT ANALYTICS DOMAIN PUBLISHING

The News Text Analytics domain is designed for publishing large complex nested data structures over Elektron and TREP using a FieldList-based envelope. Each item of data can be fragmented over multiple envelope messages.

3.2.1. Message Fragmentation and Fragment Compression

3.2.1.1. Overview

In order to fit the data into the restricted size of the RSSL update messages allowed over Elektron and through TREP, the data goes through a series of transformations.

1. The core MRN data item is converted into a JSON UTF-8 string
2. This JSON string is then compressed using gzip
3. The compressed JSON is split into a number of fragments which each fit into a single RSSL update
4. The data fragments are added to an update message as the FRAGMENT FID value in a FieldList envelope

3.2.1.2. Fragmentation

Most MRN data items will fit inside a single message, but the message format does allow for a large data item to be split across multiple messages.

Five FIDs, as well as the RIC itself, are necessary to determine whether an entire data item has been received in its various fragments and how to concatenate the fragments to construct a data item:

- MRN_SRC: identifier of the scoring/processing system that published the FRAGMENT
- GUID: globally unique identifier for the data item. All messages for this data item will have equal GUID values.
- FRAGMENT: compressed data item fragment, itself
- TOT_SIZE: total size in bytes of the fragmented data
- FRAG_NUM: sequence number of fragments within a data item. This is set to 1 for the first fragment of each item published and is incremented for each subsequent fragment for the same item.

A single MRN data item publication is uniquely identified by the combination of RIC, MRN_SRC and GUID.

For a given RIC-MRN_SRC-GUID combination, when a data item requires only a single message, then TOT_SIZE will equal the number of bytes in the FRAGMENT and FRAG_NUM will be 1.

When multiple messages are required, then the data item can be deemed as fully received once the sum of the number of bytes of each FRAGMENT equals TOT_SIZE. The consumer will also observe that all FRAG_NUM range from 1 to the number of fragment, with no intermediate integers skipped. In other words, a data item transmitted over three messages will contain FRAG_NUM values of 1, 2 and 3.

3.2.1.3. Compression of FRAGMENT FID

The FRAGMENT FID is compressed with gzip compression technology, thus requiring the consumer to decompress to reveal the JSON plain-text data in that FID.

When an MRN data item is sent in multiple messages, all the messages must be received and their FRAGMENTs concatenated before being decompressed. In other words, the FRAGMENTs should not be decompressed independently of each other.

The decompressed output is encoded in UTF-8 and formatted as JSON. The data model varies by content type.

3.2.2. Permissioning and Subscription

Data published using the News Text Analytics domain is controlled on a per-feed and per-item basis. Initial refresh responses contain a single PE in their header and PROD_PERM FID which controls access to the feed as a whole. Updates have DACS locks attached that control access on a per-item basis.

3.2.2.1. Service

Elektron Connect users connect with the ELEKTRON_DD or ELEKTRON_EDGE service. Clients using a local TREP (Thomson Reuters Enterprise Platform) might see the service under a different service, according to how their local TREP is configured.

3.2.2.2. Data Domain

MRN data is published on the News Text Analytics domain, number 33.

3.2.2.3. Subscription RICs

Each of the MRN content sets is made available over a different RIC, according to the table below.

CONTENT SET	RIC
Real-time News	MRN_STORY
News Analytics: Company and C&E assets	MRN_TRNA
News Analytics: Macroeconomic News & Events	MRN_TRNA_DOC
News Sentiment Indices	MRN_TRSI

Note: The “Macroeconomic News & Events” package of News Analytics is composed of “document-level” scores, and is referred to by this name in the rest of this document. See Section 4.1 for information on the distinction between “asset-level” and “document-level” scores.

3.2.3. Outage Detection

If all the messages on a fragmented MRN data item are not received, then this signifies an outage. A safe rule of thumb is to expect all such fragments to appear within 60 seconds of each other, although in practice they should all appear within one second.

See Section 3.2.1.2 for more information about how to determine whether an entire MRN data item has been received.

3.2.4. Envelope

Each envelope has a field that holds a fragment of the item, which often is the entire item, as well as additional fields that hold metadata about the MRN data item as a whole or that particular fragment within the envelope.

The fields that appear in the envelope vary by the kind of message. Sixteen fields appear upon the initial refresh. For an actual MRN message, ten fields appear in the first fragment of an MRN data item while only four are necessary for a "subsequent" update. A subsequent update carries additional fragments of an MRN data item, beyond the first such fragment.

The table below demonstrates which fields are present in the three types of messages. A blank value denotes that the field does not appear in that kind of message. In most cases, the table describes what values to expect, when it is simple to express what kinds of values can appear. When describing that the value(s) is complicated, "Present" is written instead.

NAME	FID	RWF TYPE (SIZE)	DESCRIPTION	INITIAL REFRESH: VALUES	FIRST UPDATE: VALUES	SUBSEQUENT UPDATE: VALUES
PROD_PERM	1	UINT64	Product permissions information for the RIC. Note that full permissioning is stored in Message.permData outside the FieldList.	Range from 10000 to 10003		
ACTIV_DATE	17	DATE	The UTC date when the time in TIMACT_MS was updated. Denotes when the update was published by the MRN_SRC. Stamped upon publication to the Elektron network.	Present, not useful	UTC date, in *yyyy-mm-dd* format	
RECORDTYPE	259	UINT32	Field which indicates the type of record and also the type of data in that record.	30		
RDN_EXCHD2	1709	ENUM	Identifier for the source from where the data originates	1370, denoting MRN		
TIMACT_MS	4148	UINT64	UTC time of last activity in milliseconds. Relative to ACTIV_DATE, and thus it resets to 0 at the beginning of each UTC day. Denotes when the update was published by the MRN_SRC. Stamped upon publication to the Elektron network.	Present, not useful	Range from 0 *(00:00:00.000 UTC)* to 86399999 *(23:59:59.999 UTC)*	
GUID	4271	RMTES_STRING (255)	Globally unique ID for the MRN data item	<empty>	Present	Present
CONTEXT_ID	5357	REAL64	The numeric identifier for a group of instruments with a common field list and market processing rules.	3752 or 3929		

NAME	FID	RWF TYPE (SIZE)	DESCRIPTION	INITIAL REFRESH: VALUES	FIRST UPDATE: VALUES	SUBSEQUENT UPDATE: VALUES
DDS_DSO_ID	6401	UINT32	Elektron (DDS) equivalent of the IDN FID DSO_ID. Has its own set of values.	Present		
SPS_SP_RIC	6480	ASCII_STRING (32)	Populated in each underlying instrument by the provider as a reference to the appropriate SPS sub-provider level RIC	Present		
MRN_V_MAJ	8506	RMTES_STRING (4)	Major version of data model used in FRAGMENT. See also MRN_V_MIN.	<empty>	“2”, for data models described in this document	
MRN_TYPE	8593	RMTES_STRING (16)	Type of data contained in FRAGMENT	<empty>	STORY, TRNA, or TRSI	
MRN_V_MIN	11787	RMTES_STRING (10)	Minor version of data model used in FRAGMENT. See also MRN_V_MAJ.	<empty>	“10” , for data models described in this document	
MRN_SRC	12215	RMTES_STRING (40)	Identifies the scoring/processing system that published the FRAGMENT	<empty>	Present	Present
FRAG_NUM	32479	UINT64	Number of the current FRAGMENT, within the overall MRN data item	Present, not useful	1	2 or more
TOT_SIZE	32480	UINT64	Total size of the compressed MRN data item in bytes after all fragments are concatenated	0	Present	
FRAGMENT	32641	BUFFER (11000)	Fragment of data	<empty>	Present	Present

3.2.5. Initial Responses and Subsequent Updates

3.2.5.1. Initial Refresh Responses

The initial refresh responses contain all the fields from the envelope. All the fields related to the item and to the fragment are empty or 0. Alternatively, they contain other stale information that should be ignored. The constant and per-feed fields contain meaningful values. The initial refresh responses are cached.

Example refresh for MRN_STORY RIC

```
Message.permData: <empty>
Message.data: <FieldList>
PROD_PERM (1): 10001
ACTIV_DATE (17): 21 OCT 2015
```

```
RECORDTYPE (259): 30
RDN_EXCHD2 (1709): 1370
TIMACT_MS (4148): 60413133
GUID (4271): <empty>
CONTEXT_ID (5357): 3752
DDS_DSO_ID (6401): 12424
SPS_SP_RIC (6480): ".[SPSML1L1"
MRN_V_MAJ (8506): "2"
MRN_TYPE (8593): "STORY"
MRN_V_MIN (11787): "10"
MRN_SRC (12215): "dtc_prd_A"
FRAG_NUM (32479): 0
TOT_SIZE (32480): 0
FRAGMENT (32641): <empty>
```

3.2.5.2. Updates

As mentioned above, the updates contain only fields related to the item and the fragment. They do not contain any of the static or per-feed fields. The updates are not cached or conflated.

3.2.5.2.1. First Update

The first update contains all the fields related to the item and the first fragment, subsequent updates only contain the fields relating to the fragment they contain.

Example first update for MRN_STORY RIC

```
Message.permData: <DACS lock for the current piece of data>
Message.data: <FieldList>
ACTIV_DATE (17): 20 AUG 2015
TIMACT_MS (4148): 10157563
GUID (4271): "BSE4njFBT_1509242kv2m5neJzQ52U7adOPFd2fc4P6PMZ/X8yPsDxw"
MRN_V_MAJ (8506): "2"
MRN_TYPE (8593): "STORY"
MRN_V_MIN (11787): "10"
MRN_SRC (12215): "dtc_prd_A"
FRAG_NUM (32479): 1
TOT_SIZE (32480): 4436
FRAGMENT (32641): <fragment of compressed JSON>
```

3.2.5.2.2. Subsequent Update

The subsequent update contains the fields necessary to identify the MRN data item, the order of this fragment among all the fragments for this item, and the fragment itself.

Example subsequent update

```
Message.permData: <DACS lock for the current piece of data>
Message.data: <FieldList>
GUID (4271): "BSE4njFBT_1509242kv2m5neJzQ52U7adOPFd2fc4P6PMZ/X8yPsDxw"
MRN_SRC (12215): "dtc_prd_A"
FRAG_NUM (32479): 2
FRAGMENT (32641): <fragment of compressed JSON>
```

3.3 DEVELOPER RESOURCES

The Thomson Reuters Developer Community portal offers open access to APIs used to consume MRN Elektron feeds. Users may register freely.

The following sections list relevant Developer Portal code, descriptions, and resources.

3.3.1. MRN-specific Sample Code and Tutorials

- Elektron Message API (EMA)
  - C++: here
  - Java: here
- Elektron Transport API (ETA). Formerly known as Ultra Performance API (UPA).
  - C: here
  - Java: here

For the optimal balance of speed and simplicity, Thomson Reuters recommends the EMA APIs.

If you require an API in another programming language, please contact your Technical Account Manager.

3.3.2. Other Developer Resources

- MRN Developer Webinar, recorded January 2017: here
- MRN EMA Overview article: here
- The following tools are available on Github and Developer Portal. Note that they are not officially supported by Thomson Reuters and may be removed in the future:
  - MRN Console: here
  - MRN GUI Viewer, for viewing various feeds side-by-side: here
  - Sample implementation on storing News Analytics in MySQL: here

CHAPTER 4 NEWS ANALYTICS DATA MODEL

4.1 ASSET-LEVEL VS. DOCUMENT-LEVEL SCORING

Until V4, all TRNA scores were assigned only to content that contained a scorable asset, i.e., a company or C&E topic on the TRNA coverage list. Content lacking such an asset was not scored. Further, the Analytics Score Group (Section 4.3.1) contains metadata that is specific to individual assets in the data.

The V4 release adds TRNA scoring on documents as a whole. This means that the asset filter described above is not applied, thus allowing for TRNA scores on a superset of the content scored at the asset level. In addition, the data model itself is smaller in two key ways.

A. Most of the fields in the Analytics Score Group (Section 4.3.1) are irrelevant and thus not present. See Notes in individual fields for more information.

B. Related, there is no Linked Id Group (Section 4.3.2)

Sections 4.5 and 4.6 show examples of asset-level and document-level scores, respectively.

4.2 OVERVIEW

The structure of the data within each data feed is defined in the following sections. After assembly and decompression, the data appears as JSON in UTF-8.

The TRNA feed has three top-level items.

- sub-group containing the analytics scores
- id value that identifies the news item
- sub-group containing information about the news item that was used to generate the analytics scores

Envelope Data Type (MRN_TYPE): TRNA

FIELD	TYPE	SAMPLE	NOTES
analytics	Analytics Group: Section 4.3		
id	String	"tr:BSE4njFBT_1509242kv2m5neJzQ52U7adOPFd2fc4P6PMZ/X8yPsDxw"	[*feedFamilyCode*]:[*sourceId*] Note that in the initial release of TRNA V4, id and sourceId of feed and archive do not match.
newsItem	News Item Group: Section 4.4		This group contains metadata sourced directly from the STORY item, in contrast to the *newsItem* group also inside the *analytics* group that contains data derived from the TRNA scoring.

4.3 ANALYTICS GROUP

The analytics sub-group has three top-level items.

- array of scores for the assets the news item relates to
- sub-group containing analytics relating directly to the news item and not any specific asset

- version for the system which generated the analytics scores

FIELD	TYPE	SAMPLE	NOTES
analyticsScores	Array of Analytics Score Groups: Section 4.3.1		Contains an item per asset related to the news item
newsItem	Group		Information derived from the news item not present in the STORY feed itself
systemVersion	String	"TS:40060083"	Version of the system that generated the data. The prefix denotes the type of system. “TS” refers to the Thomson Reuters Text Analytics System (TRTS) that generates TRNA and TRSI scores. A TRTS system version is an 8-digit number of the format: ABCCDDDD ABCC is the version of the newest component in the system with the final digit zero-padded. DDDD is the zero-padded company coverage list version.

4.3.1. Analytics Score Group

Each analytics score group contains all the analytics information derived from the news item for a specific asset as a simple group of named values.

FIELD	TYPE	SAMPLE	NOTES
assetClass	String	"CMPNY"	The broad class that the asset belongs to. Also describes the type of TRTS sentiment engine used in the scoring. Either "CMPNY" for a company or "COM" for a commodity. Set to “CMPNY” for document-level scores because of use of the same scoring engine as used for company-level scores.

FIELD	TYPE	SAMPLE	NOTES
assetCodes	String Array	[“R:MSFT.O”, "P:4295907168"]	List of prefixed codes, in conjunction with *assetId* field below, which identify the asset within various symbologies. By *assetClass* value: “CMPNY”: “P:” prefix for PermID and “R:” for RIC. Can contain multiple RICs for a single company, including the primary one and those tagged to the news item. Note that it will be empty for all document-level scores. “COM”: “N2” for Thomson Reuters topic code
assetId	String	"4295907168" or “CRU”	Primary identifier for the asset. PermID for company and topic code for commodity. Set to “R:DOC.CMPNY” for document-level scores.
assetName	String	"Microsoft Corp"	A human readable name for the asset, used as an identifier for unknown entity scoring. Blank for all document-level scores.
brokerAction	String	"UPGRADE"	Denotes whether the news item is reporting the action of a broker recommendation for a security issued by the company. One of "UPGRADE", "DOWNGRADE", "MAINTAIN", "BROKER", "INITIATE", "UNDEFINED" Blank for all document-level scores.

FIELD	TYPE	SAMPLE	NOTES
firstMentionSentence	Int	15	The first sentence, starting with the headline, in which the scored asset is mentioned. Thus, a value of 1 denotes the headline, 2 the first sentence of the story body, 3 the second sentence, etc. This is a relevance proxy because typically more relevant assets are mentioned towards the beginning of a news item. Can be used in conjunction with *sentenceCount* in the *newsItem* group to determine the relative position of the first mention in the item. Note: When this value equals 0, this means that the asset being scored was not found in the news item's headline or body text. As a result, the entire news item's text (headline + body) will used to determine the sentiment score. In effect, this means that the asset being scored receives a document-level sentiment score. In this scenario *sentimentWordCount* (below) will equal *wordCount* (Section 4.3.3). Set to 0 for document-level scores.
linkedIds	Array of Linked Id Groups: Section 4.3.2		The five most recent and five oldest linked articles for the longest of the historical periods, i.e., 7 days. If there are 10 or fewer linked articles, then the array contains the full list. See also *noveltyCounts*, above. Not populated for document-level scores.

FIELD	TYPE	SAMPLE	NOTES
noveltyCounts	Array of Windowed Count Groups: Section 4.3.3		The novelty of the content within a news item on a particular asset is calculated by comparing it with the asset-specific text over a cache of previous news items that contain the asset. The comparison between items is done using a linguistic fingerprint. If the news items are similar, they are termed as being “linked”. As a result, a content item can “link” only to an item of the same language. There are five historical periods that are used in the comparison. The default periods are 12 hours, 24 hours, 3 days, 5 days and 7 days prior to the news item’s timestamp. See also *linkedIds*, below. Automatically set to 0 when *firstMentionSentence* = 0. (See below.) Thus, it is not calculated for document-level scores.
priceTargetIndicator	String	"UNDEFINED"	When the news item is a price target indicator for the asset. One of "INCREASE", "DECREASE", "MAINTAIN", "BROKER", "INITIATE", "UNDEFINED" Set to “UNDEFINED” for all Japanese-language and document-level scores.
relevance	Double	0.0294118	A decimal number indicating the relevance of the news item to the asset. It ranges from 0 to 1. It is calculated by comparing the relative number of occurrences of the asset with the number of occurrences of other organizations and commodities within the text of the item. In addition, if the asset is mentioned in the headline, the relevance is set to 1.0. For stories with multiple assets, the asset with the most mentions will have the highest relevance. An asset with a lower amount of mentions will have a lower relevance score. When the item is an alert (urgency = 1, Section 4.4), relevance should be gauged by *firstMentionSentence* (above) instead. Set to 0.000000 for document-level scores.

FIELD	TYPE	SAMPLE	NOTES
sentimentClass	Int	-1	This field indicates the predominant sentiment class for this news item with respect to this asset. The indicated class is the one with the highest probability. 1: Positive, 0: Neutral, -1: Negative
sentimentNegative	Double	0.548096	The probability that the sentiment of the news item was negative for the asset
sentimentNeutral	Double	0.12516	The probability that the sentiment of the news item was neutral for the asset
sentimentPositive	Double	0.32674	The probability that the sentiment of the news item was positive for the asset
sentimentWordCount	Int	1369	The number of lexical tokens (words and punctuation) in the sections of the item text that are deemed relevant to the asset. This is the number of words used in the sentiment calculation for this asset. It can be used in conjunction with *wordCount* (Section 4.3.3) to determine the proportion of the news item discussing the asset. See also *firstMentionSentence*, above. Set equal to *wordCount* for document-level scores. (See below.)
volumeCounts	Array of Windowed Count Groups: Section 4.3.3		The volume of news for each asset is calculated. A cache of previous news items is maintained and the number of news items that mention the asset within each of five historical periods is calculated. The cache is language-specific, e.g., a volumeCount on an English-language item measures the number of other English-language items in that historical period. By default the historical periods are 12 hours, 24 hours, 3 days, 5 days and 7 days prior to the news item's timestamp and are the same used in the novelty calculations. Thus direct comparisons between similar and total items within the historical periods can be achieved. Not calculated for document-level scores.

Note: The three sentiment values; *sentimentNegative*, *sentimentNeutral* & *sentimentPositive* sum to 1.0.

4.3.2. Linked Id Group

The linked id group is used to associate an id with its position in a longer list of ids. It is used for the *linkedIds*.

This group is not populated for document-level scores, since novelty is not calculated.

FIELD	TYPE	SAMPLE	NOTES
idPosition	Int	2	Position of the linkedId in the complete list of linked Ids. 0 is the first/oldest, and the largest/most recent is the 7-day itemCount minus 1.
linkedId	String	"tr:BSE4njFBT_1509242dfjBeDgnVLEHNzNZMeMb51mzPaSedU2p+JCz+A"	id of the item at this position

4.3.3. Windowed Count Group

The windowed count group is used to associate a count with the window of time it relates to. It is used for the *noveltyCounts* and *volumeCounts*.

This group is not populated for document-level scores, since novelty and volume are not calculated.

FIELD	TYPE	SAMPLE	NOTES
itemCount	Int	25	Number of items
Window	String	"1D"	Length of time the count covers *n*H (for hours) or *n*D (for days). Default values are “12H”, “24H”, “3D”, “5D”, and “7D”.

4.3.4. News Item Group (Analytics Sub-group)

The TRNA feed contains two news item groups. This group, within the *analytics* group, contains values derived from the news item by the analytics system. The other group (Section 4.4), at the top level, contains values which are contained within the news item being processed.

FIELD	TYPE	SAMPLE	NOTES
bodySize	Int	1328	The size of the current version of the story body in characters
companyCount	Int	6	The number of companies explicitly listed in the news item in the subjects field (Section 4.4)
exchangeAction	String	"IMBALANCE"	One of "IMBALANCE", "HALT", "RESUME", "BLOCK TRADE", "INDICATION", "UNDEFINED". Set to “UNDEFINED” for all Japanese-language scores.

FIELD	TYPE	SAMPLE	NOTES
headlineTag	String		The Thomson Reuters headline tag for the news item. See Appendix 3 for the list of supported headline tags for English-language TRNA, and Appendix 4 for the list for Japanese-language TRNA.
marketCommentary	Boolean	TRUE	Indicator that the item is discussing general market conditions, such as “After the Bell” summaries.
sentenceCount	Int	34	The total number of sentences in the news item. Can be used in conjunction with *firstMentionSentence* (Section 4.3.1) to determine the relative position of the first mention in the item.
wordCount	Int	308	The total number of lexical tokens (words and punctuation) in the news item. Can be used in conjunction with *sentimentWordCount* (Section 4.3.1) to determine the proportion of the news item discussing an asset.

4.4 NEWS ITEM GROUP (TOP-LEVEL GROUP)

The TRNA feed contains two news item groups. This top-level group contains values which are contained within the news item being processed; the other group (Section 4.3.3) within the *analytics* group contains values derived from the news item by the analytics system.

Because the fields below are sourced from the incoming news item data and mapped to the below fields, those mappings can vary by the *feedFamilyCode* value. Those mappings are distinguished in the Notes section in the below table.

FIELD	TYPE	SAMPLE	NOTES
dataType	String	"News"	The broad type of data the news item belongs to. One of "News", "Social"
feedFamilyCode	String	"tr"	A code that identifies the family of feeds the news item came from. Thomson Reuters feeds = "tr" Moreover feeds = "mrvr". Not currently available.
headline	String	"PRESS DIGEST - Financial Times - April 17"	The headline text of the news item.
language	String	"en"`	ISO-639 language code. “en” for English, “ja” for Japanese

FIELD	TYPE	SAMPLE	NOTES
metadata	Metadata Group: Section 4.4.1		
provider	String	"NS:RTRS" "BBC"	Identifier for the organisation which provided the news item. The source of this data varies by the *feedFamilyCode* value. "tr": from *provider* field "mrvr": from *sourceName or publisher* field
sourceId	String	"BSE4njFBT_1509242dfjBeDgnVLEHNzNZMeMb51mzPaSedU2p+JCz+A"	The id of the news item in its feed. Note that in the initial release of TRNA V4, id and sourceId of feed and archive do not match.
sourceTimestamp	DateTime	2013-04-16T23:00:14.000Z	UTC timestamp of this news item. Millisecond precision. The source of this data varies by the *feedFamilyCode* value. "tr": set to *versionCreated* in underlying data "mrvr": set to *harvest date* in underlying data
subjects	String Array	["N2:COM", "A:4", "N2:PREC", "B:21", "R:VOD.L"]	Topic codes and company identifiers that relate to this news item. The source of this data varies by the *feedFamilyCode* value. "tr": from, *subjects* field "mrvr": set to “OneCalais” Company RIC codes are prefixed "R:", company PermIDs by “P:”, and topic codes by"N2:". Other code prefixes describe alternate representations of topic codes.
urgency	Int		Differentiates story types. 1: alert, 3: article

4.4.1. Metadata

The metadata sub-group contains additional fields some of which are only present in news items from specific feeds. The presence of certain fields depends upon the *feedFamilyCode* value.

FIELD	TYPE	SAMPLE	NOTES
altId	String	"nL3N0D3D3H"	Depends on *feedFamilyCode* value "tr": from *altId* field. For Thomson Reuters sourced news this contains the PNAC value. See Section 2.1 for more information. "mrvr": from *id* field

FIELD	TYPE	SAMPLE	NOTES
feedTimestamp	DateTime	2013-04-16T23:00:14.000Z	UTC timestamp showing when the data was available on the feed. Millisecond precision.
isArchive	Boolean	TRUE	Flag to show whether this is archive data.
Only from "tr" feeds			
audiences	String Array	["NP:M", "NP:T"]	News products for this news item. See Section 2.2.1 for more information.
firstCreated	DateTime	2013-04-16T23:00:14.000Z	UTC timestamp for the first version of the story. Millisecond precision. See also section 2.1 for more information on stories.
instancesOf	String Array	["NI:CRU/MED"]	Named Items for this news item. See Section 2.2.3 for more information.
takeSequence	Int	2	The take sequence number of the news item, starting at 1. For a given story, alerts and articles have separate sequences.
Only from "mrvr" feeds			
author	String		Only appears if it is present in source data
autoRank	String		Only appears if it is present in source data
autoRankOrder	String		Only appears if it is present in source data
countryCode	String	"GB"	ISO-3166 two-letter country code
editorialRank	String		Only appears if it is present in source data
feedGenre	String		Only appears if it is present in source data
loginStatus	String		Only appears if it is present in source data
publisher	String	"Bayt.com"	The parent company of an online publication
sourceCategory	String		Only appears if it is present in source data
sourceName	String	"Jobs in Saudi Arabia | Bayt.com"	The common household name of the publication
topics	String Array		Only appears if it is present in source data

4.5 JSON EXAMPLE OF TRNA FEED DATA: ASSET-LEVEL SCORING


```
{
   "analytics":{
      "analyticsScores":[
         {…}
         {
            "assetClass":"COM",
            "assetCodes":["N2:MTAL"],
            "assetId":"MTAL",
            "assetName":"Non-Ferrous Metals",
            "brokerAction":"UNDEFINED",
            "firstMentionSentence":79,
            "linkedIds":[
               {"idPosition":0,"linkedId":"tr:MKW0LsGla_1705302560+anis+dl88E6R+uZxPaEOG/Lx75R2+TrWa"},
               {"idPosition":1,"linkedId":"tr:MKWvyBJ1a_1705302DPUAg8FOGt3HpqlOtP10cyzLYmerenqahRK8I"},
               {"idPosition":2,"linkedId":"tr:CCNVh1JSa_1705302JcCwYBye/yKzaejZmXyLsvfpSXvsLXV5Om54D"}
            ],
            "noveltyCounts":[
               {"itemCount":3,"window":"12H"},
               {"itemCount":3,"window":"24H"},
               {"itemCount":3,"window":"3D"},
               {"itemCount":3,"window":"5D"},
               {"itemCount":3,"window":"7D"}
            ],
            "priceTargetIndicator":"UNDEFINED",
            "relevance":0.0860663,
            "sentimentClass":1,
            "sentimentNegative":0.142598,
            "sentimentNeutral":0.0791933,
            "sentimentPositive":0.778209,
            "sentimentWordCount":146,
            "volumeCounts":[
               {"itemCount":320,"window":"12H"},
               {"itemCount":530,"window":"24H"},
               {"itemCount":855,"window":"3D"},
               {"itemCount":1838,"window":"5D"},
               {"itemCount":3261,"window":"7D"}
            ]
         }
      ],
      "newsItem":{
         "bodySize":16953,
         "companyCount":1,
         "exchangeAction":"UNDEFINED",
         "headlineTag":"",
         "marketCommentary":false,
         "sentenceCount":113,
         "wordCount":2676
      },
      "systemVersion":"TS:40060082"
   },
"id":"tr:CCNNK6tCa_1705302WS5wDmNOk02EmeV5s6hz4aYbZnRfIXXlqognP",
"newsItem":{
   "dataType":"News",
   "feedFamilyCode":"tr",
```

```
   "headline":"Osisko Intersects 97.4 g/t Au Over 8.4 Metres at Windfall",
   "language":"en",
   "metadata":{
      "altId":"nCCNNK6tCa",
      "audiences":["NP:CCN"],
      "feedTimestamp":"2017-05-30T12:00:00.431Z",
      "firstCreated":"2017-05-30T12:00:00.000Z",
      "isArchive":false,
      "takeSequence":1
   },
   "provider":"NS:MKW",
   "sourceId":"CCNNK6tCa_1705302WS5wDmNOk02EmeV5s6hz4aYbZnRfIXXlqognP",
   "sourceTimestamp":"2017-05-30T12:00:00.407Z",
   "subjects":["B:12","B:19","B:20","B:21","G:4","G:8W","G:9","M:1NP","M:1QD","M:32","M:B6","M:Z","R:OSK.TO",
   "N2:AMERS","N2:BMAT","N2:CA","N2:CLIST1","N2:CMPNY","N2:LEN","N2:MIN","N2:MINE","N2:MTAL","N2:NAMER","N2:N
   EWR","N2:PREC","N2:PRECAL","P:5037933384"],
   "urgency":3
   }
}
```

4.6 JSON EXAMPLE OF TRNA FEED DATA: DOCUMENT-LEVEL SCORING


```
{
   "analytics":{
      "analyticsScores":[
         {
            "assetClass":"CMPNY",
            "assetCodes":[],
            "assetId":"R:DOC.CMPNY",
            "assetName":"",
            "brokerAction":"",
            "firstMentionSentence":0,
            "linkedIds":[],
            "noveltyCounts":[
               {"itemCount":0,"window":"12H"},
               {"itemCount":0,"window":"24H"},
               {"itemCount":0,"window":"3D"},
               {"itemCount":0,"window":"5D"},
               {"itemCount":0,"window":"7D"}
            ],
            "priceTargetIndicator":"UNDEFINED",
            "relevance":0.000000,
            "sentimentClass":1,
            "sentimentNegative":0.258225,
            "sentimentNeutral":0.173744,
            "sentimentPositive":0.568031,
            "sentimentWordCount":683,
            "volumeCounts":[
               {"itemCount":0,"window":"12H"},
               {"itemCount":0,"window":"24H"},
               {"itemCount":0,"window":"3D"},
               {"itemCount":0,"window":"5D"},
               {"itemCount":0,"window":"7D"}
            ]
         }
      ],
      "newsItem":{
         "bodySize":4430,
         "companyCount":1,
         "exchangeAction":"UNDEFINED",
         "headlineTag":"",
         "marketCommentary":false,
         "sentenceCount":28,
         "wordCount":683
      },
      "systemVersion":"TS:40060082"
   },
"id":"tr:MKWM400Pa_1705302zVk9faoxe0WtsZ21QOS2lbnxCzfGCMX4i9FYc",
"newsItem":{
   "dataType":"News",
   "feedFamilyCode":"tr",
   "headline":"Shenzhen Fruit Industry Association to Support Marketing of Kiwa Products",
   "language":"en",
   "metadata":{
      "altId":"nMKWM400Pa",
      "audiences":["NP:CNR","NP:MKW"],
```

```
      "feedTimestamp":"2017-05-30T12:00:00.326Z",
      "firstCreated":"2017-05-30T12:00:00.000Z",
      "isArchive":false,
      "takeSequence":1
   },
   "provider":"NS:MKW",
   "sourceId":"MKWM400Pa_1705302zVk9faoxe0WtsZ21QOS2lbnxCzfGCMX4i9FYc",
   "sourceTimestamp":"2017-05-30T12:00:00.313Z",
   "subjects":["A:4","B:106","B:107","B:112","B:113","B:12","B:14","B:16","G:4","G:6J","G:9","M:1QD","M:32","
   M:Z","U:8","R:KWBT.PK","N2:AGRC","N2:AGRI","N2:AMERS","N2:BMAT","N2:CHEM","N2:CLIST1","N2:CMPNY","N2:COM",
   "N2:FARM","N2:FOBE","N2:FOTB","N2:LEN","N2:NAMER","N2:NCYC","N2:NEWR","N2:US","P:4295953887"],
   "urgency":3
   }
}
```

CHAPTER 5 MRN FTP SITE: ARCHIVE FILES, ASSET LISTS, MAPPINGS

Thomson Reuters provides archive TRNA data and reference information over FTP.

Data is formatted as tab-delimited text files, except for the JSON-formatted archives and the Changes file, which is listed in Section 0. Character encoding is UTF-8. Archive files are compressed, in .gz format.

Production clients can access the full history, while trial clients may access a more limited period. Files for the two client types are stored in different directories.

The FTP site also provides useful reference data. It lists live and active assets, along with useful reference information. The Companies reference section also contains the following:

- Mappings to common identifiers, including CUSIP, ISIN, and SEDOL
- Mapping between PermID and the deprecated orgID
- Changes in coverage

5.1 ACCESS

Production and most trial users are granted access to the FTP site. Trial users who convert to paid customers should have their credentials upgraded to full-history access. Users should contact their sales specialist or account manager to obtain login credentials.

The FTP site is available at ftp://mrn-ftp.thomsonreuters.com, or 54.243.148.106. The site is accessible via FTP client only. Please set your FTP client to passive mode, although in some cases active mode will work instead.

Clients may use their credentials to connect via plain FTP and also secure FTP (FTPS) via explicit FTP over TLS, using TLS version 1.2.

5.2 ARCHIVES

5.2.1. Directories

/TRNA/Archives/TR_News/[Scored Asset: CMPNY_ALL/CMPNY_AMER/CMPNY_APAC/CMPNY_EMEA/ COM/DOC]/[Language: EN/JA]/[File Format: Flat/JSON]/[Historical/Recent]

Notes:

- The “TR_News” level denotes content based on news aggregated by Thomson Reuters. This level is added to distinguish from other content sets that may be added in the future.
- Scored Asset
  - CMPNY_ALL replaces the former “EQU” designation. COM represents commodities, as before.
  - DOC represents document-level scoring
- File Format
  - “Flat” file format presents the data in text tab-delimited format for easy import into standard database packages. The data is separated into two files, one containing the score-specific data and a second containing the news-specific data. See Appendix 1 for illustration of how the two files are designed.
    - Note that Flat archives were marked with TRTS V4.0.2. Field values are equivalent to those of TRTS V4.0.1.
  - JSON format corresponds to that of the live feed, as described in Chapter 4
- Historical/Recent
  - Historical contains all files from the previous calendar year, generally. Files are annual in duration.
  - Recent contains files from the current year. Files may be monthly or daily in duration.
  - Exception to the previous/current year rule: Recent will continue to hold files from the previous year that just passed, until that year’s archives have been regenerated based on the most recent TRNA build.

5.2.2. Files

TRNA.TR.News.[Scored Asset: CMPNY_ALL|CMPNY_AMER|CMPNY_EMEA|CMPNY_APAC|COM|DOC].[EN|JA].[Time Period: yyyy|yyyymm|yyyymmdd].[System Version excluding the "TS:" prefix].[Format: JSON/Flat-News/Flat-Scores].txt.gz

- *TRNA*: content set
- *TR* : denotes Thomson Reuters as the aggregator of the underlying textual content
- *News*: type of underlying textual content
- *Scored Asset*: assets being scored for sentiment
  - *CMPNY_ALL*: all companies
  - *CMPNY_AMER*: companies domiciled in the Americas
  - *CMPNY_APAC*: companies domiciled in Asia-Pacific
  - *CMPNY_EMEA*: companies domiciled in Europe, Middle East, or Africa
  - *COM*: commodities and energies
  - *DOC*: document-level scoring
- *Language*: language of underlying content
  - *EN*: English
  - *JA*: Japanese
- *Time period*
  - *yyyy*: annual
  - *yyyymm*: monthly
  - *yyyymmdd*: daily
- *Format*
  - *JSON*
  - *Flat-News*: Note that Flat archives were marked with TRTS V4.0.2. Field values are equivalent to those of TRTS V4.0.1. See Appendix 1 for illustration of how the two Flat files are designed.
  - *Flat-Scores*
- *System Version:* describes the build number of the system used to create the TRNA scores. It represents a combination of scoring system and asset list versions. The system version is constructed as *AAAABBBB*, where…
  - "AAAA" denotes the version of the installed TRTS system without the delimiting periods. If the third number in the TRTS version lacks a leading zero, then here it is added. Thus an TRTS version value of "4.0.1" is expressed here as "4001"
  - "BBBB" denotes the version of the TRNA coverage list, also known as the "asset store".

5.3 COMPANY REFERENCE FILES

Mappings to common identifiers will be provided for companies. These include ticker, MIC, and a Thomson Reuters Identifier Code (RIC).

In addition, third-party identifiers – CUSIP, ISIN, and SEDOL – are available upon request in conjunction with a license from the issuer. CUSIP and ISIN require a license with Standard & Poor's for CUSIPs. ISINs are included because some ISINs are CUSIP-based. SEDOL access requires a license with the London Stock Exchange. Please contact your Account Manager or Sales Specialist if you are interested in viewing these third-party identifiers and have a requisite license. Thomson Reuters will contact the identifier issuer(s) to verify the license(s).

5.3.1. Directories

/TRNA/Companies/[EN/JA]/[BASIC/CUSIPISIN/SEDOL/CUSIPISINSEDOL]/

Notes:

- There are separate files for English (EN) and Japanese (JA) because the coverage differs between these language-specific variants of TRNA
- The fourth-level directory is permissioned according to user's combination of licenses for third-party identifiers. Access will be given to exactly one such directory. By default, users are granted access to the BASIC directory.

5.3.2. Files

5.3.2.1. Asset Lists

TRNA.Companies.[EN/JA].[BASIC/CUSIPISIN/SEDOL/CUSIPISINSEDOL].[Asset Store Version, e.g., 038].txt

Notes:

- There are separate files for English (EN) and Japanese (JA) because the coverage differs between these language-specific variants of TRNA
- The Asset Store Version currently corresponds to the last three digits of the systemVersion. It denotes a version of the TRNA coverage list, which typically changes monthly.
- In the future, the TRNA coverage list may be updated more regularly, weekly or daily. In such a case, the string may be of the format "*yyyymmdd*", corresponding to a date.

Fields:

- PermID: used to identify the company in the *assetId* field. See Section 2.2.5.1 for more information.
- companyName
- countryOfDomicile: two-character ISO country code
- TRBCEconomicSector: plain-text description of Thomson Reuters Business Classification (TRBC) economic sector
- status: "active" if the company may be scored in a live feed. Otherwise, "inactive".
- RIC: corresponds to the Editorial RIC, which typically is the one most commonly tagged to news. Note that for Nasdaq RICs ending in ".OQ", the RIC typically tagged to news ends in ".O". However, the ".OQ" RIC will appear in the *assetCodes* field.
- ticker
- marketMIC: ISO-10383 code for market or exchange identification. Value may differ from similar value maintained by London Stock Exchange.
- CUSIP: only available in files with "CUSIPISIN" in the file name
- ISIN: only available in files with "CUSIPISIN" in the file name
- SEDOL: only available in files with "SEDOL" in the file name

Sort order: by PermID, in ascending order

5.3.2.2. Changes

The second is a file of changes since the previous asset store version.

Changes.txt

Contents:

- First line: "Asset Store [*Asset Store Version*] was created on [*yyyy-mm-dd*]."
- Further lines:
  - "PermID [*xxx.x*], [*companyName*], became active."
  - "PermID [*xxx.x*], [*companyName*], became inactive."

5.3.2.3. PermID-OrgID Mapping

Archives for TRNA V3 used a company-level orgID field as an organizational identifier. In TRNA V4, this is deprecated in favor of the PermID. There is a one-to-one relationship between these two identifiers. This file helps clients to convert to the new identifier.

This file pertains only to the English-language coverage list.

TRNA.Mapping.orgID.PermID.[Asset Store Version, e.g., 038].txt

Notes:

- The Asset Store Version currently corresponds to the last three digits of the systemVersion. It denotes a version of the TRNA coverage list, which typically changes monthly.
- In the future, the TRNA coverage list may be updated more regularly, weekly or daily. In such a case, the string may be of the format "*yyyymmdd*", corresponding to a date.

Fields:

- PermID
- orgID

Sort order: by PermID, in ascending order

5.4 COMMODITY & ENERGY REFERENCE FILES

The commodity & energy coverage list changes seldom, less than once per year in recent years. The most recent list as of this writing is in Appendix 2.

5.4.1. Directories

/TRNA/Commodities

5.4.2. File

There will be one file, listing the codes and describing them. Due to the infrequency of changes, there is no Changes file.

TRNA.Commodities.[Asset Store Version, e.g., 038].txt

Notes:

- The Asset Store Version currently corresponds to the last three digits of the systemVersion. It denotes a version of the TRNA coverage list, which typically changes monthly.
- In the future, the TRNA coverage list may be updated more regularly, weekly or daily. In such a case, the string may be of the format "*yyyymmdd*", corresponding to a date.

Fields:

- topicCode: corresponds to assetId
- description: short description
- status: "active" if the code may be scored in a live feed. Otherwise, "inactive".
- topicDefinition: long definition
- TRCSConceptType: category of topic code

Sort order: by topicCode, alphabetically

5.5 TRIAL ACCESS COMPARED TO PRODUCTION ACCESS

As mentioned in the overview to this chapter, trial clients can access less data than can production clients, and from a different directory. Thus the files are the same, but the time duration is less.

Following are the key differences in directory structure for trial clients compared to production clients:

- Root directory is /TRNA_TRIAL, instead of /TRNA
- For archives, the lowest-level directory denotes the amount of data. It replaces the directory level of Recent and Historical.

APPENDIX 1 FLAT FILE ARCHIVE FIELDS

The flat file archives come in two files, linkable via the “id” field. In general, “Flat-News” corresponds to news-specific data, while “Flat-Scores” corresponds to asset-specific data.

FLAT NEWS FIELDS

Field	Sample Value
id	tr:20160301-000137000-nASC08E94-4-1
dataType	News
feedFamilyCode	tr
headline	SUMMIT FINANCIAL GROUP INC <SMMF.O> SAYS ESTIMATED AGGREGATE MERGER CONSIDERATION IN TRANSACTION IS ABOUT $21.8 MLN
language	en
altId	nASC08E94
feedTimestamp	2016-03-01T00:01:37.430Z
isArchive	TRUE
audiences	NP:E NP:U
firstCreated	2016-03-01T00:00:14.000Z
instancesOf	
takeSequence	4
provider	NS:RTRS
sourceId	20160301-000137000-nASC08E94-4-1
sourceTimestamp	2016-03-01T00:01:37.000Z
subjects	N2:BANK N2:BISV N2:BSVC N2:FINS N2:MRG N2:BACT N2:DEAL1 N2:RESF N2:RES N2:US N2:AMERS N2:BLR N2:CMPNY N2:LEN N2:RTRS R:SMMF.O
urgency	1
bodySize	0
companyCount	1
exchangeAction	UNDEFINED
headlineTag	
marketCommentary	FALSE
sentenceCount	1
wordCount	17
systemVersion	TS:40060082

FLAT SCORES FIELDS

Field	Sample Value
id	tr:20160301-000137000-nASC08E94-4-1
assetClass	CMPNY
assetCodes	P:4295915061 R:SMMF.OQ
assetId	4295915061
assetName	Summit Financial Group Inc
brokerAction	UNDEFINED
firstMentionSentence	1
priceTargetIndicator	UNDEFINED
relevance	1
sentimentClass	1
sentimentNegative	0.00932229
sentimentNeutral	0.462078
sentimentPositive	0.528599
sentimentWordCount	17
windows	12H 24H 3D 5D 7D
noveltyCountValues	3 3 3 3 3
volumeCountValues	4 8 8 8 8
linkedIdPositions	0 1 2
linkedIdValues	tr:20160301-000130000-nASC08E94-3-1 tr:20160301-000125000-nASC08E94-2-1 tr:20160301-000002000-nGNXMNQYJa-1-2

APPENDIX 2 COMMODITY & ENERGY CODE COVERAGE

Since Commodity & Energy coverage rarely changes, this document will list it as of May 2016, Asset Store Version 70. (See Section 5.4 for more information and for how to find the most up-to-date coverage.) There were 43 active codes, and 6 retired codes that are comparable to certain active codes. These codes appear in the *assetId* field in the data.

Active Code	Description	Topic Definition	TRCS Concept Type	Former Code
BIOF	Biofuels	Production, transport, processing, trading / broking, demand for and use of biofuels, including ethanol (derived from sugar cane), biodiesel (from animal or vegetable oils) and other fuels derived from living organisms.	Commodity	
BUN	Bunker Fuel	Production, transport, processing, trading / broking, demand for and use of bunker fuel, the heavy fuel oil used by ships, and marine diesel.	Commodity	
CHEM	Chemicals (TRBC)	Producers and refiners of agricultural, commodity and specialty chemicals.	Business Sector	CHE
CO2	Carbon / Emissions Markets	Trading in permits to emit carbon dioxide and other greenhouse gases. Policy, allowances, regulations, monitoring and enforcement governing greenhouse gas emissions as well as related technology in energy intensive industries, including carbon capture and storage.	Commodity	
COA	Coal	Extraction, transport, processing, trading / broking, demand for and use of coal. Organisations associated with coal.	Commodity	
COC	Cocoa	Growth, transport, processing, trading / broking, demand for and use of cocoa / cacao and chocolate. ICCO and organisations associated with cocoa.	Commodity	
COF	Coffee	Growth, transport, processing, trading / broking, demand for and use of coffee. Including arabica, robusta, roastings, milds, conillon. ICO, IACO and organisations associated with coffee.	Commodity	
COR	Corn	Growth, transport, processing, trading / broking, demand for and use of Corn-Maize. Organisations associated with Corn-Maize.	Commodity	
COT	Cotton	Production, transport, processing, trading / broking, demand for and use of cotton. Organisations associated with cotton.	Commodity	
CPPR	Copper	Extraction, transport, processing, trading / broking, demand for and use of copper. Organisations associated with copper.	Commodity	
CRU	Crude Oil	Exploration, extraction, transport, processing, trading / broking, demand for and use of crude oil and gas condensate. Organisations associated with crude oil.	Commodity	
DIAM	Diamond Markets	Extraction, transport, processing, trading / broking, demand for and use of diamonds. Organisations associated with diamonds.	Commodity	
GMO	Genetically Modified Organisms	Genetically modified organisms their development, deployment and any products made from them. Also including government policy and opposition to.	Environment / Accident / Disaster	
GOL	Gold	Extraction, transport, processing, trading / broking, demand for and use of gold. Organisations associated with gold.	Commodity	
GRA	Grains	Grains - wheat; barley; oats; corn; maize; sorghum; cereal; rice; IWC; HGCA; FAO; EEP; COGECA; COCERAL; all grains commodity market reports; organisations associated with grains.	Commodity	
HOIL	Heating Oil / Gasoil	Production, transport, processing, trading / broking, demand for heating oil, the gas oil or middle distillate chiefly used as a heating fuel or in industry.	Commodity	
JET	Jet Fuel / Kerosene	Production, transport, processing, trading / broking, demand for and use of jet kerosene as a jet fuel and for heating.	Commodity	
LIV	Livestock	Livestock - meat; beef; pork bellies; hogs; cattle; poultry; fish and fish oil; all livestock commodity market reports; organizations associated with livestock.	Commodity	
LNG	Liquefied Natural Gas	Production, transport, processing, trading / broking, demand for and use of liquefied natural gas.	Commodity	
LPG1	Liquefied Petroleum Gas	Production, transport, processing, trading / broking, demand for and use of liquefied petroleum gas, including propane and butane.	Commodity	LPG
MEAL	Meals / Feeds / Pulses	Feed grains; expeller; soya meal; soy meal; rape-meal; fishmeal; sun meal; sunflower meal; tapioca; manioc; citrus pulp; pulses; lentils; corn gluten; all meals and feeds commodity market reports; organizations associated with meals and feeds.	Commodity	
MOG	Gasoline	Production, transport, processing, trading / broking, demand for and use of gasoline / petrol.	Commodity	
MTAL	Metals & Mining (TRBC)	Miners and processors of precious metals and minerals, steel, aluminum and specialty metals and minerals.	Business Sector	MET
NAP	Naphtha	Production, transport, processing, trading / broking, demand for and use of naphtha.	Commodity	

NGL	Natural Gas Liquids	Production, transport, processing, trading / broking, demand for and use of ethane, propane, normal butane, isobutane, pentane and combinations of these products including condensate, lease condensate, y-grade and natural gasoline.	Commodity	
NGS	Natural Gas	Production, transport, processing, trading / broking, demand for and use of natural gas. Organisations associated with natural gas.	Commodity	
NSCRU	North Sea Crudes	North Sea crude oil streams	Commodity	
NSEA	North Sea Oil	North Sea crude oil streams	Commodity	
OILS	Oilseeds	Oils and oilseeds - oilseed; soya bean; soybean; soy oil; rape seed; rape oil; colza; canola; palm oil; palm kernel; coconut; coconut oil; sunflower; sunoil; groundnut; peanut; linseed; linoil; olein; stearin; castor; tung; castor oil; tungoil.	Commodity	
OPEC	Organization of the Petroleum Exporting Countries	OPEC policy, statements, meetings, output and exports; compliance by member countries with the crude oil producer group's decisions.	Organization	
ORJ	Orange Juice	Growth, transport, processing, trading / broking, demand for and use of orange juice as juice or frozen concentrate. Organisations associated with orange juice.	Commodity	
PLAS	Polymers / Plastics	Production, transport, trading / broking, demand for and use of polymer materials, typically thermoplastic or thermosetting resins. Processing new or recycled plastics resins into intermediate or final products including packaging.	Commodity	
PROD	Refined Products	Production, transport, processing, trading / broking, demand for and use of refined oil products including fuel oil, naphtha, gasoline, petroleum, gas oil, jet kerosene and petrochemicals. News on refineries and on policy affecting petroleum products, such as new specifications for sulphur content.	Commodity	
RFO	Fuel Oil	Production, transport, processing, trading / broking, demand for and use of heavy or residual fuel oil, including high-sulphur, low sulphur and bunker fuel oil as well as vacuum gasoil.	Commodity	
RUB	Rubber	Growth, transport, processing, trading / broking, demand for and use of natural and synthetic rubber. INRO, INRA, organisations associated with rubber.	Commodity	
SOY1	Soybean	Growth, transport, processing, trading / broking, demand for and use of Soybean. Organisations associated with Soybean.	Commodity	SOY
STEE	Iron & Steel (TRBC)	Miners and processers of iron ore and coke coal for the processing of iron and steel. Includes providers of metal ore exploration and mining services, operators of iron and steel foundries for beneficiating, smelting, rolling, forging, spinning, recycling, stamping, polishing and plating of iron and steel products such as pipes, tubes, wire, spring, rolls and bars, and providers of metal processing services and wholesalers of iron and steel products.	Business Sector	STL
SUG	Sugar	Growth, transport, processing, trading / broking, demand for and use of sugar from cane or beet. Also high fructose corn syrup, HFCS, GEPLACEA, COGECA. Organisations associated with sugar.	Commodity	
TEA	Tea	Growth, transport, processing, trading / broking, demand for and use of all teas. Including CTC, pekoe, fannings. Organisations associated with tea.	Commodity	
TREE	Paper & Forest Products (TRBC)	Operators of forests, pulp and paper mills, producers and manufacturers of wood and paper products, and providers of related services.	Business Sector	TIM
URAN	Uranium	The mining, production, processing, export and supply of uranium. The uranium market and uranium trade. Included on corporate news about uranium miners and regulatory news about government policy relating to the supply of uranium. Not included on stories about uranium enrichment for nuclear weapons.	Commodity	
WHT	Wheat	Growth, transport, processing, trading / broking, demand for and use of Wheat. Organisations associated with Wheat.	Commodity	
WOO	Wool	Production, transport, processing, trading / broking, demand for and use of wool, merino or mohair. Organisations associated with wool.	Commodity	

APPENDIX 3 HEADLINE TAGS: ENGLISH

The headlineTag field in TRNA can contain the news item's headline tag. The list of possible English-language (language = "en", as per Section 4.4) headline tag values appear on the following page.

ADR Report	GLOBAL MARKETS	NORWAY GAS	Taiwan Stocks
ADVISORY	GRAINS	NSE	TAKE A LOOK
AFRICA INVESTMENT	GRAPHIC	NY Euro Deposits	TAKE-A-LOOK
After the Bell	GUIDE	NY Euro Deposits Close	TECHNICALS
AIRSHOW	GULF STOCKS	NY Exchanges Forwards	TEXT
AMEX CONTINUING HALT	HEADLINE STOCKS	NYMEX	Thai Hot Stocks
AMEX OPENING DELAY	HIGHLIGHTS	NYSE CONTINUING HALT	Thai Stocks
ANALYSIS	Hot Stock Highlights	NYSE HALT	TIMELINE
ANALYST VIEW	India Money Market Rates	NYSE NEW INDICATION	Tokyo Stocks
ANALYSTS' VIEW	Indian Debt	NYSE OPENING DELAY	TOP NEWS
Asia Forex	Indian Government Securities	NYSE RESUMPTION	TOPWRAP
ASIA GRAINS	Indian IPOs	OBITUARY	TRADE IDEA
Asia Markets	INDICATOR	OPINION	TREASURIES
Assets (US)	INDICATORS	OTC HALT	U.S. Cash Product
BEFORE THE BELL	Indonesian Stocks	OTC OPENING DELAY	U.S. Corporate Bonds
Block Trade	INSIGHT	OTC RESUMPTION	U.S. Headline Stocks
Bombay Shares	INSTANT VIEW	OUTLOOK	U.S. Municipals
Bond Markets	INSTANT VIEW 1	Philippine Stocks	U.S. OPTIONS
Brazil Forex Week	INSTANT VIEW 2	POLL	U.S. REFINERY FILING
BREAKINGVIEWS	INSTANT VIEW 3	PRECIOUS	U.S. RESEARCH ROUNDUP
BRIEF	INSTANT VIEW 4	Presale Cattle	U.S. Yield Curve
BUY OR SELL	INSTANT VIEW 5	PRESS DIGEST	UK GAS
BUZZ	INSTANT VIEW 6	PREVIEW	UK Gilts Weekahead
CANADA FX	INSTANT VIEW 7	Q&A	UK Stocks
CANADA RESEARCH ROUNDUP	INTERVIEW	REFILE	UPDATE
CANADA STOCKS	IPO Pricing	RESEARCH ALERT	UPDATE 1
CHRONOLOGY	IPO VIEW	REUTERS EXCLUSIVE	UPDATE 2
COLUMN	IPOVIEW	Reuters Money Graph	UPDATE 3
Commodities	Jakarta Stocks	REUTERS POLL	UPDATE 4
COMMODITIES	Japan Hot Stocks	REUTERS SUMMIT	UPDATE 5
CORRECTED	Lifting the Lid	REUTERS SURVEY	UPDATE 6
CORRECTED-(OFFICIAL)	LIVESTOCK	RLPC	UPDATE 7
DIARY	LOOKAHEAD	RUBBER	UPDATE 9
Dutch Options	LPC	SCENARIO	URGENT
EARNINGS POLL	Malaysia Stocks	SEALED BIDS	US CREDIT
Emerging Debt Emerging Markets FX	MARKET EYE	Singapore Stocks	US CREDIT OUTLOOK
EMERGING EUROPE STOCKS	MARKET EYE WEEKAHEAD	SNAP ANALYSIS	US RESEARCH NEWS
EMERGING MARKETS	MARKET PULSE	SNAPSHOT	US RESEARCH SUMMARY
EURO CORP	MERGER TALK	SOFTS	US STOCKS
Euro Debt	METALS	South Korea Stocks	US STOCKS SNAPSHOT
EURO GOVT	METALS INSIDER	Special Report	VEGOILS
EUROPE MINOR METALS	MIDDLE EAST NEWS HIGHLIGHTS	SPECIAL REPORT	Wall Street Week Ahead
EUROPE RESEARCH ROUNDUP	MIDEAST DEBT	STOCKS NEWS AFRICA	WEEK AHEAD
EXCLUSIVE	MIDEAST MONEY	STOCKS NEWS EUROPE	WITNESS
FACTBOX	MIDEAST STOCKS	STOCKS NEWS INDONESIA	WORLD BONDS
Factors to watch	MMNEWS	STOCKS NEWS MALAYSIA	WORLD NEWS HIGHLIGHTS
FEATURE	MONEY MARKETS	STOCKS NEWS MIDEAST	WRAPUP
FINEWS	MOVES	STOCKS NEWS SINGAPORE	WRAPUP 1
FOREX	NASDAQ HALT	STOCKS NEWS UK SMALL	WRAPUP 2
FUND SCORE	NASDAQ RESUMPTION	STOCKS NEWS US	WRAPUP 3
FUND VIEW	NATGAS PIPELINE CRITICAL NOTICE	STOCKS NEWS VIETNAM	WRAPUP 4
FX Cross	New Issue	STXNEWS LATAM	WRAPUP 5
FX in Europe	NEWSMAKER	SUMMIT	WRAPUP 6
FXNEWS	NORDIC POWER	Swiss stocks	YOUR MONEY
German Data Due	NORDIC STOCKS	TABLE	

APPENDIX 4 HEADLINE TAGS: JAPANESE

The headlineTag field in TRNA can contain the news item's headline tag. The list of possible Japanese-language (language = "ja", as per Section 4.4) headline tag values appear below.

〔焦点〕	UPDATE 1	ロイター短観こうみる	米ＦＲＢ議長証言こうみる
〔表〕ロイター短観	UPDATE 2	中国インフレ指標こうみる	米ＦＲＢ議長講演こうみる
10 年債入札こうみる	UPDATE 3	中国利上げこうみる	米ＧＤＰこうみる
10 月ロイター短観	UPDATE 4	中国指標こうみる	米ＩＳＭ製造業指数こうみる
11 月ロイター短観	UPDATE 5	中国貿易収支こうみる	米債務上限問題こうみる
12 月ロイター短観	UPDATE 6	中国貿易統計こうみる	米新規失業保険申請件数こうみる
1 月ロイター短観	UPDATE1	中国預金準備率引き上げこうみる	米経済指標こうみる
20 年債入札こうみる	UPDATE2	人民元柔軟化こうみる	米雇用統計こうみる
2 年債入札こうみる	UPDATE3	債券格付け情報	経常収支こうみる
2 月ロイター短観	UPDATE4	債券格付情報	緊急インタビュー
30 年債入札こうみる	UPDATE5	円債こうみる	訂正
3 月ロイター短観	UPDATE6	再送	訂正（会社側の申し出）
40 年債入札こうみる	WRAPUP 1	展望リポートこうみる	訂正(発表者側の申し出)
4 月ロイター短観	WRAPUP 2	新規上場企業の横顔	訂正（発表者側の申し出）
5 年債入札こうみる	WRAPUP 3	新規公開企業の横顔	豪ＣＰＩこうみる
5 月ロイター短観	WRAPUP 4	新設ファンド概要	豪ＧＤＰこうみる
6 月ロイター短観	WRAPUP 5	日本国債見通し引き下げこうみる	豪中銀据え置きこうみる
7 月ロイター短観	WRAPUP1	日銀人事こうみる	豪中銀金利据え置きこうみる
8 月ロイター短観	WRAPUP2	日銀会合こうみる	豪利上げこうみる
9 月ロイター短観	WRAPUP3	日銀決定会合こうみる	豪利下げこうみる
BUZZ	WRAPUP4	日銀短観こうみる	豪金利据え置きこうみる
COLUMN	WRAPUP5	日銀総裁会見こうみる	豪雇用統計こうみる
ＣＰＩこうみる	アップル決算こうみる	日銀追加緩和こうみる	貿易収支こうみる
ＥＣＢこうみる	アナリスト情報	株式こうみる	貿易統計こうみる
ＥＣＢ総裁会見こうみる	インタビュー	機械受注こうみる	起債評価
ＦＯＭＣこうみる	お知らせ	為替こうみる	追加
Ｇ２０こうみる	クレジットこうみる	独ＧＤＰこうみる	追加再送
Ｇ７こうみる	サムスン業績予想こうみる	短観こうみる	追加緩和こうみる
ＧＤＰこうみる	ホットストック	米ＦＯＭＣこうみる	鉱工業生産こうみる
ＪＦＥ	ロイター・テクニカル分析	米ＦＯＭＣ議事録こうみる	銘柄速報
Top News			

Document Version: 1.0
Date of Issue: 08 June 2017

THOMSON REUTERS