THOMSON REUTERS MACHINE READABLE NEWS

MRN DATA MODELS AND ELEKTRON IMPLEMENTATION GUIDE

DATA MODEL VERSION 2.12

NEWS ANALYTICS VERSION 4.0.1

Document Version 1.0
Date of Issue: 23 August 2016

THOMSON REUTERS

CONTENTS

ABOUT THIS DOCUMENT

INTENDED READERSHIP

This document is intended for two groups of readers.

- Users of Machine Readable News (MRN) content interested in the data model that is present in both Elektron feed and archive data[1]
- Developers writing applications to consume this content

This guide also serves as an addendum to Section 9 (Implementation Guidelines) of the Thomson Reuters Elektron Edge Programmer's Guide, downloadable from Customer Zone, here.

News Analytics and News Sentiment Indices clients should refer to the respective User Guides on Customer Zone, listed here and here, respectively. Those guides contain other information that is highly relevant to those products. Readers of those documents will not require this one.

IN THIS GUIDE

This document describes the structure of MRN data and how to consume it from Elektron. It covers the MRN data feeds: Real-time News, News Analytics (TRNA) and News Sentiment Indices (TRSI).

FEEDBACK

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

[1] This document does not describe MRN content sets not available over Elektron, such as MarketPsych Indices and Economic Events.

CHAPTER 1 MACHINE READABLE NEWS OVERVIEW

Thomson Reuters Machine Readable News (MRN) is the industry's most advanced service for automating the consumption and systematic analysis of news. It delivers deep historical news archives, ultra-low latency structured news and leading edge news analytics directly to applications. This enables algorithms to exploit the power of news to seize opportunities, capitalize on market inefficiencies and manage event risk.

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

1.1 MRN DATA FEEDS

Currently there are three MRN data feeds available over Elektron. They are all published as structured text. Common fields across the feeds, e.g., "id" and "subjects", are centrally defined and consistently named.

1.1.1. Real-time News

This feed is sourced from news alerts and stories from Reuters and dozens of third-party news sources. It contains the headline, story body text, and associated metadata available at news publication time.

1.1.2. News Analytics (TRNA)

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.

1.1.3. Sentiment Indices (TRSI)

The TRSI feed provides moving averages of TRNA sentiment scores for easier consumption than the raw TRNA data. Moving average sentiment scores are provided for each of the individual scored entities over different trailing windows. Averages are also calculated for sets of entities, whose scores are rolled up to provide sentiment indices for regions and sectors.

The TRSI feed consists of the sentiment index values and news item volume counts.

CHAPTER 2 STORY CONSTRUCTION AND NEWS CODING

2.1 EVOLUTION OF A STORY

A story[2] 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

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 Real-time News and News Analytics. 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.

Example: “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.

---

[2] 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: values are prefixed by "R:"

- subjects field
- assetCodes field

2.2.6. Company PermIDs

Companies in TRNA, TRSI, and in an increasing number of news attributions in Real-time News 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 News Analytics and News Sentiment Indices data use the PermID as the primary company identifier. PermIDs supplement the RICs tagged to Real-time News for some attributions.

Example: 4295904307 for International Business Machines Corporation

Identification: values are prefixed by "P:"

- subjects field
- assetId field

CHAPTER 3 MRN MESSAGE ENVELOPE: PRESENTATION AND CONSUMPTION

3.1 OVERVIEW

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 data model for each of the content sets varies and is described in subsequent chapters.

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_SUM. 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 RICs

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.

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

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

3.2.4.1. Fields

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		
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	

NAME	FID	RWF TYPE (SIZE)	DESCRIPTION	INITIAL REFRESH: VALUES	FIRST UPDATE: VALUES	SUBSEQUENT UPDATE: VALUES
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

3.2.4.2. 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>
```

3.2.4.3. 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.

3.2.4.3.1. 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>
```

3.2.4.3.2. 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>
```

CHAPTER 4 REAL-TIME NEWS DATA MODEL

As mentioned above, the data appears as JSON in UTF-8 after decompression and assembly of the individual messages.

The Real-time News feed contains the headline, story body text, and associated metadata about the story as a simple group of named values.

Envelope Data Type (MRN_TYPE): STORY

FIELD	TYPE	SAMPLE	NOTES
altId	String	"nL3N0D3D3H"	For Thomson Reuters sourced news, this contains the PNAC value. See Section 2.1 for more information.
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.
headline	String	"PRESS DIGEST - Financial Times - April 17"	The headline text of the news item
id	String	"BSE4njFBT_1509242kv2m5neJzQ52U7adOPFd2fc4P6PMZ/X8yPsDxw"	Uniquely identifies the news item. This is the same value as the GUID in the OMM envelope.
instancesOf	String Array	["NI:PRESS/FT"]	Named Items for this news item. See Section 2.2.3 for more information.
language	String	"en"	ISO-639 language code. When *pubStatus* = “stat:canceled”, this may be set to “und” denoting “undefined”.
provider	String	"NS:RTRS"	Identifier for the source that provided the news item. See Section 2.2.3 for more information.
pubStatus	String	"stat:usable"	Publication status. “stat:usable” for most content. “stat:canceled” is a request to remove the story from active story caches for viewing. Note that standard MRN usage terms do not allow for viewing.
subjects	String Array	["N2:COM", "A:4", "N2:PREC", "B:21", "R:VOD.L", “P:4295896661”]	Topic codes and company identifiers that relate to this news item. See Sections 2.2.2 and 2.2.5 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.

FIELD	TYPE	SAMPLE	NOTES
urgency	Int	3	Differentiates story types. 1: alert, 3: article
versionCreated	DateTime	2013-04-16T23:00:14.000Z	UTC timestamp for this version of the story. Millisecond precision. See also section 2.1 for more information on stories.
body	String		The full body text of the news item.
mineType	String	"text/plain"	Mime type of the body. Currently always “text/plain” or “text/x-bdc-tms”. Other mime types may be supported in the future.

4.1 JSON EXAMPLE OF STORY FEED DATA


```
{
   "altId": "nBSE4njFBT",
   "audiences": [ "NP:BSE" ],
   "body": "Firstsource Solutions Ltd has informed BSE that the members of the Company have passed the
      resolution by way of Postal Ballot, under Clause 35A.\n\n
      \n\nhttp://pdf.reuters.com/pdfnews/pdfnews.asp?i=43059c3bf0e37541&u=urn:newsml:reuters.com:20150924:nBS
      E6yFYfg\n\n \n\n \n\nDouble click on the URL above to view the article.Please note that internet access
      is required. If you experience problem accessing the internet, please consult your network
      administrator or technical support\n\nLatest version of Adobe Acrobat reader is recommended to view PDF
      files.  The latest version of the reader can be obtained from
      http://www.adobe.com/products/acrobat/readstep2.html\n\n",
   "firstCreated": "2015-09-24T15:41:50.000Z",
   "headline": "FIRSTSOURCE SOLUTIONS LTD. - Results of Postal Ballot (Clause 35A) <FISO.NS>",
   "id": "BSE4njFBT_1509242kv2m5neJzQ52U7adOPFd2fc4P6PMZ/X8yPsDxw",
   "instancesOf": [],
   "language": "en",
   "mimeType": "text/plain",
   "provider": "NS:BSE",
   "pubStatus": "stat:usable",
   "subjects": [ "R:FISO.NS", "P:4295873587", "B:195", "B:34", "B:43", "B:49", "BL:52", "G:1", "G:5B", "G:K",
      "M:Z", "N2:BSUP", "N2:INDS", "N2:ISER", "N2:CMSS", "N2:BUS", "N2:EMRG", "N2:IN", "N2:ASIA", "N2:CMPNY"
   ],
   "takeSequence": 1,
   "urgency": 3,
   "versionCreated": "2015-09-24T15:41:50.000Z"
}
```

CHAPTER 5 NEWS ANALYTICS DATA MODEL

As mentioned earlier, News Analytics clients should refer to the News Analytics User Guide on Customer Zone, listed here. That guide contains other information that is highly relevant to this product. Readers of that document will not require this one.

5.1 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 5.3.1) 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 5.3.1) are irrelevant and thus not present. See Notes in individual fields for more information.

B. Related, there is no Linked Id Group (Section 5.3.2)

Sections 5.5 and 5.6 show examples of asset-level and document-level scores, respectively.

5.2 OVERVIEW

As mentioned above, the data appears as JSON in UTF-8 after decompression and assembly of the individual messages.

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 5.3		
id	String	"tr:BSE4njFBT_1509242kv2m5neJzQ52U7adOPFd2fc4P6PMZ/X8yPsDxw"	[*feedFamilyCode*]:[*sourceId*]
newsItem	News Item Group: Section 5.4		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.

5.3 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 5.3.1		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:40010061"	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.

5.3.1. 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, "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.
assetCodes	String Array	[“R:MSFT.O”, "P:4295907168"]	Additional list of prefixed codes which identify the asset within various symbologies, e.g., RIC, OrgId. Empty for all document-level scores.

FIELD	TYPE	SAMPLE	NOTES
assetId	String	"P:4295907168" or “N2:CRU”	Prefixed code with the primary identifier for the asset. PermID (“P:”) for company and topic code (“N2:”) 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.
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 5.3.3). Set to 0 for document-level scores.
linkedIds	Array of Linked Id Groups: Section 5.3.2		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 5.3.3		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 5.4), 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 5.3.3) 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 5.3.3		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.

5.3.2. 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

5.3.3. 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”.

5.3.4. 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 5.4), 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 5.4)
exchangeAction	String	"IMBALANCE"	One of "IMBALANCE", "HALT", "RESUME", "BLOCK TRADE", "INDICATION", "UNDEFINED". Set to “UNDEFINED” for all Japanese-language scores.
headlineTag	String		The Thomson Reuters headline tag for the news item

FIELD	TYPE	SAMPLE	NOTES
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 5.3.1) 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 5.3.1) to determine the proportion of the news item discussing an asset.

5.4 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 5.3.3) 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
metadata	Metadata Group: Section 5.4.1		

FIELD	TYPE	SAMPLE	NOTES
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

5.4.1. 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
feedTimestamp	DateTime	2013-04-16T23:00:14.000Z	When the data was available on the feed, in milliseconds.
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

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


```
{
   "analytics": {
      "analyticsScores": [
         {
            "assetClass": "CMPNY",
            "assetCodes": ["P:4295873587", "R:FISO.NS"],
            "assetId": "4295873587",
            "assetName": "Firstsource Solutions Ltd",
            "brokerAction": "UNDEFINED",
            "firstMentionSentence": 1,
            "linkedIds": [
               {
                  "idPosition": 0,
                  "linkedId": "tr:BSE4njFBT_1509242dfjBeDgnVLEHNzNZMeMb51mzPaSedU2p+JCz+A"
               }
            ],
            "noveltyCounts": [
               { "itemCount": 1, "window": "12H" },
               { "itemCount": 1, "window": "24H" },
               { "itemCount": 1, "window": "3D" },
               { "itemCount": 1, "window": "5D" },
               { "itemCount": 1, "window": "7D" }
            ],
            "priceTargetIndicator": "UNDEFINED",
            "relevance": 1,
            "sentimentClass": -1,
            "sentimentNegative": 0.444419,
            "sentimentNeutral": 0.334288,
            "sentimentPositive": 0.221294,
            "sentimentWordCount": 40,
            "volumeCounts": [
               { "itemCount": 1, "window": "12H" },
               { "itemCount": 1, "window": "24H" },
               { "itemCount": 1, "window": "3D" },
               { "itemCount": 1, "window": "5D" },
               { "itemCount": 1, "window": "7D" }
            ]
         }
      ],
      "newsItem": {
         "bodySize": 655,
         "companyCount": 1,
         "exchangeAction": "",
         "headlineTag": "",
         "marketCommentary": false,
         "sentenceCount": 8,
         "wordCount": 99
      },
      "systemVersion": "TS:40010061"
   },
   "id": "tr:BSE4njFBT_1509242kv2m5neJzQ52U7adOPFd2fc4P6PMZ/X8yPsDxw",
```

```
   "newsItem": {
      "dataType": "News",
      "feedFamilyCode": "tr",
      "headline": "FIRSTSOURCE SOLUTIONS LTD. - Results of Postal Ballot (Clause 35A) <FISO.NS>",
      "language": "en",
      "metadata": {
         "altId": "nBSE4njFBT",
         "audiences": [ "NP:BSE" ],
         "feedTimestamp": "2015-09-24T15:41:51.037Z",
         "firstCreated": "2015-09-24T15:41:50.000Z",
         "isArchive": false,
         "takeSequence": 1
      },
      "provider": "NS:BSE",
      "sourceId": "BSE4njFBT_1509242kv2m5neJzQ52U7adOPFd2fc4P6PMZ/X8yPsDxw",
      "sourceTimestamp": "2015-09-24T15:41:50.000Z",
      "subjects": [ "R:FISO.NS", "P:4295873587", "B:195", "B:34", "B:43", "B:49", "BL:52", "G:1", "G:5B",
         "G:K", "M:Z", "N2:BSUP", "N2:INDS", "N2:ISER", "N2:CMSS", "N2:BUS", "N2:EMRG", "N2:IN", "N2:ASIA",
         "N2:CMPNY" ],
      "urgency": 3
   }
}
```

5.6 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":0.00995381,
            "sentimentNeutral":0.468720,
            "sentimentPositive":0.521327,
            "sentimentWordCount":18,
            "volumeCounts":[
               {"itemCount":0,"window":"12H"},
               {"itemCount":0,"window":"24H"},
               {"itemCount":0,"window":"3D"},
               {"itemCount":0,"window":"5D"},
               {"itemCount":0,"window":"7D"}
            ]
         }
      ],
      "newsItem":{
         "bodySize":0,
         "companyCount":0,
         "exchangeAction":"",
         "headlineTag":"",
         "marketCommentary":false,
         "sentenceCount":1,
         "wordCount":18
      },
      "systemVersion":"TS:40010000"
   },
   "id":"tr:20160101-000000000-nS6N11N01W-1-1-R:DOC.CMPNY",
```

```
   "newsItem":{
      "dataType":"News",
      "feedFamilyCode":"tr",
      "headline":"S.KOREA SAYS DEC EXPORTS -13.8 PCT VS YR EARLIER (REUTERS POLL -10.9 PCT)",
      "language":"en",
      "metadata":{
         "altId":"nS6N11N01W",
         "audiences":["NP:M","NP:E","NP:D","NP:T","NP:C","NP:MTL","NP:GRO","NP:SOF","NP:O"],
         "feedTimestamp":"2016-01-01T00:00:00.078Z",
         "firstCreated":"2016-01-01T00:00:00.000Z",
         "isArchive":true,
         "takeSequence":1
      },
      "provider":"NS:RTRS",
      "sourceId":"20160101-000000000-nS6N11N01W-1-1",
      "sourceTimestamp":"2016-01-01T00:00:00.000Z",
      "subjects":["N2:KR","N2:EMRG","N2:ASIA","N2:MCE","N2:ECI","N2:NEWS1","N2:TRD","N2:LEN","N2:RTRS",
         "R:KRW=","R:0#KRCOMP1=KQ","R:KREXGR=ECI","R:KRIMGR=ECI","R:KRTBAL=ECI"],
      "urgency":1
   }
}
```

CHAPTER 6 NEWS SENTIMENT INDICES DATA MODEL

As mentioned earlier, News Sentiment Indices clients should refer to the News Sentiment Indices User Guide on Customer Zone, listed here. That guide contains other information that is highly relevant to this product. Readers of that document will not require this one.

The TRSI feed has a group of named values including an array of scores.

Envelope Data Type (MRN_TYPE): TRSI

FIELD	TYPE	SAMPLE	NOTES
dataType	String	"News"	The broad type of data the news item belongs to. Set to "News".
eventTimestamp	DateTime	2007-01-01T03:22:29.000Z	UTC timestamp for the event which triggered this update.
eventType	String	"EXIT_1H"	The type of event which has triggered this event to be published. "SCORE": denotes a fresh news item. Adds 1 to each *itemCount* value and a new sentiment score to be incorporated into each *sentimentAverage*. (See Section 6.1.1.1.) "EXIT_1H", "EXIT_4H", "EXIT_24H", "EXIT_168H": denote a news item leaving a trailing window. The corresponding *itemCount* decreases by 1 and *sentimentAverage* is recalculated without the exiting sentiment score. Thus, each news item generates five events.
feedFamilyCode	String	"tr"	A code that identifies the family of feeds the news item came from. Set to “tr”, denoting Thomson Reuters feeds.
id	String	"tr:BSE4njFBT_1509242kv2m5neJzQ52U7adOPFd2fc4P6PMZ/X8yPsDxw.EXIT_1H"	[*feedFamilyCode*]:[*sourceId*].[*eventType*]
language	String	"en"	ISO-639 language code.
siScores	Array of SI Score Groups: Section 6.1.1		

FIELD	TYPE	SAMPLE	NOTES
systemVersion	String	"TS:40010070"	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.

6.1.1. SI Score Group

Each SI score group contains all the rolled up analytics information 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. One of "CMPNY" for a company, “CMPNY_GRP” for a rollup, and "COM" for a commodity
assetCodes	String Array	["P:4295902883", "R:BABY.OQ"]	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. “CMPNY_GRP”: none “COM”: “N2” for Thomson Reuters topic code
assetId	String	"4295902883", “.TRXFLDGLPU”, or “CRU”	Primary TRSI identifier for the asset
assetName	String	"Canadian Imperial Bank of Commerce"	A human readable name for the asset, used as an identifier for unknown entity scoring.
sentimentAverages	Array of SI Average Groups: Section 6.1.1.1		See Section 6.1.1.1.

6.1.1.1. SI Average Group

The SI Average group is used to associate a count and an average with the window of time it relates to. It is used inside the *sentimentAverages* group.

FIELD	TYPE	SAMPLE	NOTES
itemCount	Int	25	Number of items.
sentimentAverage	Double	0.1234	Average sentiment score, based on TRNA positive sentiment minus negative sentiment. If *itemCount* is 0, then this will be set to 0.0, thus representing an indeterminate number due to the absence of any sentiment values to average.
window	String	"1D"	Length of time the count covers *n*H (for hours) or *n*D (for days). Default values are “1H”, “4H”, “1D”, and “7D”.

6.2 JSON EXAMPLE OF TRSI FEED DATA

The below example of TRSI feed data is intended for demonstration. It is not a production-quality TRSI data item.


```
{
   "dataType": "News",
   "eventTimestamp": "2016-07-06T13:33:56.000Z",
   "eventType": "SCORE",
   "feedFamilyCode": "tr",
   "id": "tr:FWN19S0FY_1607061tOFWGjmfZOhzjyjRAVzEcKcfhK3R32oWqntg3.SCORE",
   "language": "en",
   "siScores": [
      {
         "assetClass": "CMPNY",
         "assetCodes": ["P:4295902883","R:BABY.OQ"],
         "assetId": "4295902883",
         "assetName": "Natus Medical Inc",
         "sentimentAverages": [
            { "itemCount": 1, "sentimentAverage": 0.554743, "window": "1H" },
            { "itemCount": 1, "sentimentAverage": 0.554743, "window": "4H" },
            { "itemCount": 1, "sentimentAverage": 0.554743, "window": "1D" },
            { "itemCount": 1, "sentimentAverage": 0.554743, "window": "7D" } ]
      },
      {
         "assetClass": "CMPNY_GRP",
         "assetCodes": [],
         "assetId": ".TRXFLDUSP",
         "assetName": "TR UNITED STATES IDX",
         "sentimentAverages": [
            { "itemCount": 23, "sentimentAverage": 0.229196, "window": "1H" },
            ...
            { "itemCount": 4104, "sentimentAverage": 0.190497, "window": "7D" } ]
      },
      {
         "assetClass": "CMPNY_GRP",
         "assetCodes": [],
         "assetId": ".TRXFLDUSPHLC",
         "assetName": "TR US HLTHCARE IDX",
         "sentimentAverages": [
            { "itemCount": 9, "sentimentAverage": 0.250254, "window": "1H" },
            ...
            { "itemCount": 458, "sentimentAverage": 0.0767925, "window": "7D" } ]
      },
      {
         "assetClass": "CMPNY_GRP",
         "assetCodes": [],
         "assetId": ".TRXFLDGLPU",
         "assetName": "TR GLOBAL IDX",
         "sentimentAverages": [
            { "itemCount": 33, "sentimentAverage": 0.246230, "window": "1H" },
            ...
            { "itemCount": 11561, "sentimentAverage": 0.114180, "window": "7D" } ]
      }
   ],
   "systemVersion": "TS:40010070"
}
```

APPENDIX 1 ALL MRN FIELDS

The following is a list of all the fields used across the feeds.

FIELD	TYPE	STORY	TRNA	TRSI
altId	String	x	x	
analytics	Analytics Group		x	
assetClass	String		x	x
assetCodes	String Array		x	x
assetId	String		x	x
assetName	String		x	x
audiences	String Array	x	x	
author	String		x	
autoRank	String		x	
autoRankOrder	String		x	
body	String	x		
bodySize	Int		x	
brokerAction	String		x	
companyCount	Int		x	
countryCode	String		x	
dataType	String		x	x
editorialRank	String		x	
entitlements	String Array		x	x
eventTimestamp	DateTime			x
eventType	String			x
exchangeAction	String		x	
feedFamilyCode	String		x	x
feedGenre	String		x	
feedTimestamp	DateTime		x	
firstCreated	DateTime		x	

FIELD	TYPE	STORY	TRNA	TRSI
firstCreated	DateTime	x		
firstMentionSentence	Int		x	
headline	String	x	x	
headlineTag	String		x	
id	String	x	x	x
idPosition	Int		x	
instancesOf	String Array	x	x	
isArchive	Boolean		x	
itemCount	Int			x
language	String	x	x	
linkedId	String		x	
linkedIds	Array of LinkedId Groups		x	
loginStatus	String		x	
marketCommentary	Boolean		x	
metadata	Group		x	
mimeType	String	x		
newsItem	Group		x	
noveltyCounts	Array of Windowed Count Groups		x	
priceTargetIndicator	String		x	
provider	String	x	x	
publisher	String		x	
pubStatus	String	x		
relevance	Double		x	
sentenceCount	String		x	
sentimentAverage	Double			x
sentimentAverages	Array of SI Average Groups			x
sentimentClass	Int		x	

FIELD	TYPE	STORY	TRNA	TRSI
sentimentNegative	Double		x	
sentimentNeutral	Double		x	
sentimentPositive	Double		x	
sentimentWordCount	Int		x	
siScores	Array of SI Score Groups			x
sourceCategory	String		x	
sourceId	String		x	
sourceName	String		x	
sourceTimestamp	DateTime		x	
subjects	String Array	x	x	
systemVersion	String		x	x
takeSequence	Int	x		
topics	String Array		x	
transmissionTimestamp	DateTime		x	x
urgency	Int	x	x	
versionCreated	DateTime	x		
volumeCounts	Array of Windowed Count Groups		x	
window	String		x	x
wordCount	Int		x	

For more information

Send us a sales enquiry at reuters.com/salesenquiry

Read more about our products at reuters.com/productinfo

Find out how to contact your local office

reuters.com/contacts

Document Version: 1.0
Date of Issue: 23 August 2016

THOMSON REUTERS