THOMSON REUTERS MARKETPSYCH INDICES

USER GUIDE

MARKETPSYCH INDICES VERSION 2.2

Document Version: 1.0
Date of Issue: 23 March 2016

THOMSON REUTERS

CONTENTS

ABOUT THIS DOCUMENT

INTENDED READERSHIP

This guide is primarily intended for individuals with responsibility to develop applications to consume Thomson Reuters MarketPsych Indices (TRMI) messages. The information contained within this document is also of use to anyone working with the associated archive files.

IN THIS GUIDE

This document describes the format of TRMI output images. It is applicable to both the live engine's published output images and TRMI archive files.

CHAPTER 1 INTRODUCTION

OVERVIEW

Thomson Reuters MarketPsych Indices (TRMI) analyze news and social media in real-time to convert the volume and variety of professional news and the internet into manageable information flows that drive sharper decisions. The indices are delivered as real-time data series that can easily be incorporated into your investment and trading decision processes – quantitative or qualitative.

Three key types of indicator are provided:

- Emotional indicators such as **Anger**, **Fear** and **Joy**
- Macroeconomic metrics including **Earnings Forecast**, **Interest Rate Forecast**, **Long vs. Short**
- Buzz metrics on the asset level, i.e., **Buzz**, and on market-moving topics for that asset, such as **Litigation**, **Mergers**, and **Volatility**

The indicators are updated every minute for companies, sectors, regions, countries, commodities and energy topics, indices and currencies. They can be translated directly into spreadsheets or charts that can be monitored by traders, risk managers or analysts – or they can be plugged straight into your algorithms for low frequency or longer term asset allocation or sector rotation decisions.

CONTACT INFORMATION

For Production Customers

Production customers should direct their TRMI questions or support issues to the News Feed Direct support team, via one of the following:

- Phone: Call +1 877-814-3571 and choose option 2.
- Electronic: Visit the “CONTACT US” link in the top left of the page at http://customers.reuters.com. In filling out the form, type in “Thomson Reuters Machine Readable News-News Feed Direct” as the product name.

For Trial Customers

Trial customers are assigned a dedicated Presales Technical Engineer to assist with technical questions during the trial. During the trial, all questions should be sent directly to the Presales Technical Engineer.

DOCUMENTATION AND NOTIFICATIONS

Client-facing documentation may be found on the TRMI Customer Zone page, located here. Other information may be found on the MRN FTP site, as described in Chapter 14.

For updates on releases of new versions of TRMI or of the companies coverage list, 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 MarketPsych Indices, go to the PCN subscription page, here. Check the box at All Products -> Reuters Enterprise Information -> Thomson Reuters MarketPsych Indices. Toward the bottom, make sure that the box for “All Client impacts”, or at least “For Action” under that, is checked. Then click the submit → link in the lower or upper right to confirm your selection.

CHAPTER 2 WINDOW LENGTHS AND UPDATE FREQUENCIES

OVERVIEW

TRMIs are asset-level scores on a collection of content. There are two time-related metrics that determine TRMI scores. The **window length** determines what range of content is scored in generating a set of TRMIs. For live content, all content collected by MarketPsych within that time period will be incorporated into the corresponding window. The **update frequency** determines the time between consecutive TRMI scores. (See Chapter 5 for more information on content types.)

FILE NAMING CONVENTION

For files on the FTP site, the convention for denoting the combination of window length and update frequency shall be

W[Period Length]_U[Period Length], where *W* denotes window length and *U* denotes update frequency.

Period lengths can be any of the following:

- 01M: 1 minute
- 05M: 5 minutes
- HOU: hourly (60 minutes)
- DAI: daily (1440 minutes)
    - Note that daily update frequency refers to TRMI calculation at 3:30pm Eastern. When a daylight savings hour is added or subtracted in the previous 24 hours, the time since the last update can be 25 or 23 hours, respectively.

ACROSS TRMI ASSET CLASSES

The below table summarizes the options across the various TRMI asset classes.

Asset Class(es)	Window Length	Update Frequency	FTP File Abbreviation	Live or Archive	Delivery
All	1 minute	1 minute	W01M_U01M	Both	FTP
All	1440 minutes / 24 hours	1 hour	WDAI_UHOU	Both	FTP
All except for Companies	1440 minutes / 24 hours	1 minute	N/A	Live	News Feed Direct
All except for Companies	1440 minutes / 24 hours	5 minutes	WDAI_U05M	Archive	FTP
All	1440 minutes / 24 hours	Daily, at 3:30 Eastern time	WDAI_UDAI	Archive	FTP

EXAMPLE

To illustrate with the WDAI_U05M data, one series of TRMI scores occurs with data collected between December 22, 2014 15:00:00 UTC and December 23, 2014 15:00:00 UTC, a 1440-minute / 24-hour **window length**. The next set of scores would occur with data collected between December 22, 2014 15:05:00 UTC and December 23, 2014 15:05:00 UTC, because of the 5-minute **update frequency**.

CHAPTER 3 FEED DELIVERY OPTIONS

NEWS FEED DIRECT

All live TRMIs with 24-hour window lengths and 1-minute update frequencies (See Chapter 2) are distributed over the News Feed Direct delivery platform. The TRMI messages arrive in xml format inside a “content envelope” structure. The TRMI-specific data is inside the <Data> tag. See Chapter 6 for sample xml.

For more information on subscribing to News Feed Direct and the content envelope structure, see the News Feed Direct Developer’s Guide.

FTP

The 1-minute window data and the hourly updating data are delivered over FTP. See Chapter 14 for more information on this delivery method.

CHAPTER 4 GENERAL COMMENTS ABOUT INDEX VALUES

”NA” INDEX VALUES

As mentioned in Chapter 2, all TRMIs are based on relevant text collected over a window of content.

If over that window there was no relevant text identified, then the correct value is “NA”, not zero, and the xml tag will not be published at all. For example, this could happen for the carryTrade index on AUD (Australian Dollar) in news if there were no discussions on carry trade with the Australian dollar found in news-based text over the content collected in that window.

An NA differs in meaning from true zero in that true zero represents the presence of text corresponding to positive and negative values that add up to zero. In other words, a zero value reflects that relevant text was found and its sentiment implications net to zero. In contrast, NA represents the absence of any relevant text and of any resultant measurement.

Note that when the Buzz is zero, this means that no values were detected for any of the indices and thus all index values necessarily will be NA. See Chapter 6 for more information on Buzz.

NEGATIVE INDEX VALUES

The indices are marked as ranging from either -1 to 1 or 0 to 1, corresponding to bipolar and unipolar indices, respectively. In practice, those denoted as “unipolar” can in fact range below 0, although not below -1. This occurs because unipolar indices reflect the orthogonal nature of many single emotions and topics. A negative comment such as, “I don’t enjoy owning this stock” is not emotively equivalent to, “I am pessimistic about the stock’s prospects” or “I am angry with the company’s management.” The initial statement is specifically one of negative Joy, which decreases the overall Joy index for assets related to that company. When there are many such negative Joy comments for an asset, the Joy index itself may show negative values.

Nonetheless, in practice unipolar indices are positive over 90% of the time, because language typically reflects positive assertions.

Thus, in the sections below we will mark this range as “0 to 1*”.

CHAPTER 5 UNDERLYING CONTENT SOURCES

OVERVIEW

TRMIs are evaluated on three different content sets: news, social media, and the combined content. History on all content dates back to the beginning of 1998. Only English-language text is used. This section will outline those sets and their composition.

NEWS

Reuters news is present in the entire historical news dataset, as are a host of mainstream news sources collected by MarketPsych Data. During 2005, the archive began including Internet news content collected by Moreover Technologies. The Moreover content is restricted to those from top international and business news sources, top regional news sources, and leading industry sources.

SOCIAL MEDIA

The social media collection process is less diverse. It starts in 1998 with content collected by MarketPsych Data. Internet forums and finance-specific tweets compose this space. Starting in late 2008, Moreover Technologies social media content is included. Using popularity ranks measured by incoming links, this includes generally the top 30% of blogs, microblogs, and other social media content.

Note that selected Moreover social media is included in the company groups social media dataset. The company groups data is composed of a subset of finance-specific Moreover content and the MarketPsych-based social media collection.

CHAPTER 6 NEWS FEED DIRECT CONTENT

XML EXAMPLE

For illustration, below is an abbreviated sample of a TRMI message on currencies data that was published over News Feed Direct. The instances of "…" were inserted for brevity.

```
ContentEnvelope xmlns="http://data.schemas.tfn.thomson.com/Envelope/2008-05-01/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:e="http://company.schemas.tfn.thomson.com/2008-
11-13/" majVers="1" minVers="0.7" pubStyle="Message">
  <Header>
    <Info>
      <Id>urn:uuid:f803cac7-ebef-43a9-bdfa-d74cb5099b50</Id>
      <Timestamp>2015-12-16T18:13:06.810Z</Timestamp>
    </Info>
  </Header>
  <Body contentSet="FILE" majVers="1" minVers="2.0">
    <ContentItem action="Insert">
      <Entitlements>
        <Product>
          <Code>MUA3</Code>
        </Product>
        <Source>
          <Group>NONE</Group>
        </Source>
      </Entitlements>
      <Data>
         <id>mp:2015-12-16_18.13.News_Social.CUR</id>
         <windowTimestamp>2015-12-16T18:13:00.000Z</windowTimestamp>
         <feedFamilyCode>mp</feedFamilyCode>
         <dataTypes>
            <dataType>News</dataType>
            <dataType>Social</dataType>
         </dataTypes>
         <systemVersion>MP:2.2.0</systemVersion>
         <mPsychIndices>
            <mPsychIndex>
               <assetId>
                  <assetClass>CUR</assetClass>
                  <assetCodes>
                     <assetCode>ISO-4217-Alpha:ARS</assetCode>
                  </assetCodes>
               </assetId>
               <buzz>585.800000</buzz>
               <relativeBuzz>0.016536</relativeBuzz>
               <sentiment>-0.158757</sentiment>
               <optimism>-0.020955</optimism>
               <gloom>0.061454</gloom>
               <stress>0.124616</stress>
               <timeUrgency>-0.011949</timeUrgency>
               ...
               <volatility>0.018436</volatility>
            </mPsychIndex>
            <mPsychIndex>
               <assetId>
                  <assetClass>CUR</assetClass>
                  <assetCodes>ISO-4217-Alpha:AUD</assetCodes>
               ...
               <priceMomentum>0.001822</priceMomentum>
```

```
            </mPsychIndex>
            ...
         </mPsychIndices>
      </Data>
    </ContentItem>
  </Body>
</ContentEnvelope>
```

STRUCTURE OF THE TRMI-SPECIFIC DATA

As discussed in Chapter 2, all the TRMI-specific data is in the <Data> tag.

FIELDS GENERIC TO ALL TRMI INDICES

The opening section contains five tags and one other group tag.

```
         <id>mp:2015-12-16_18.132013-02-21_20.31.00.News_Social.CUR</id>
         <windowTimestamp>2015-12-16T18:132013-02-21T20:31:00.000Z</windowTimestamp>
         <feedFamilyCode>mp</feedFamilyCode>
         <dataTypes>
            <dataType>News</dataType>
            <dataType>Social</dataType>
         </dataTypes>
         <systemVersion>MP:2.2.0</systemVersion>
```

Id

The <id> tag identifies the transmission. It is composed according to the following:

> *{feedFamilyCode}:{windowTimestamp in yyyy-mm-dd_hh.mm.ss}.{"News"|"Social"|"News_Social" according to dataTypes}.{assetClass}*

Window Timestamp

As mentioned in Chapter 2, TRMIs are based on content aggregated over a trailing window. The <windowTimestamp> tag represents the endpoint of that period.

It is expressed in ISO-8601 format: "yyyymmddThh:mm:ss.000Z", where "000" represents the milliseconds and "Z" represents the time zone. The time zone will always be GMT/UTC.

Feed Family Code

The <feedFamilyCode> tag describes the feed that delivers the raw content being scored by the TRMI system.

For TRMI it will always be "mp", because MarketPsych centrally collects and filters the content scored by TRMI.

Incoming Sources

The <dataTypes> group tag represents the type of content source, or Data Type Set, on which these TRMIs are based.

There are three possible values:

- "News" for news
- "Social" for social media
- "News Social" for the combined content. In the xml they will be separate tags, as in the example above. In the archives, they will be space-separated in one column.

See Chapter 5 for more information on these sources.

System Version

The <systemVersion> tag represents an overall build for the MarketPsych Indices. Different versions may lead to different scores as enhancements are made and bugs are fixed. It is thus used for version control.

It consists of a three-level version string preceded by an “MP” prefix denoting the MarketPsych scoring engine, e.g., “MP:2.2.0”.

Monthly updates to the companies coverage list will result in the third digit of the system version incrementing by one, for all asset classes.

FIELDS DESCRIBING THE TRMI INDICES

The <mPsychIndices> group tag contains TRMI data on all the assets in this ContentEnvelope. For each asset and corresponding set of TRMI, there is a <mPsychIndex> tag. Following those are identifiers for the asset and the Buzz.

```
<mPsychIndices>
   <mPsychIndex>
      <assetId>
         <assetClass>CUR</assetClass>
         <assetCodes>
            <assetCode>ISO-4217-Alpha:ARS</assetCode>
         </assetCodes>
      </assetId>
      <buzz>585.800000</buzz>
      <relativeBuzz>0.016536</relativeBuzz>
```

Asset Class

The <assetClass> tag is common to all assets in the ContentEnvelope. It can take one of six values.

- CMPNY: individual companies
- CMPNY_GRP: company groups
- CUR: currencies
- COM:AGR: agricultural commodities
- COM:ENM: energy & material commodities
- COU: countries

Asset Code

The <assetCode> tag represents the code used to denote the asset scored. They are of the form

{asset code type}:{code}

where the asset code type varies by Asset Class:

- CMPNY: “P” – Thomson Reuters organizational PermID
- CMPNY_GRP: “MP” – a MarketPsych-specific code
- CUR: “ISO-4217-Alpha” – three-character ISO-4217 code
- COM:AGR, COM:ENM, COU: “N2” – Thomson Reuters topic code. For COU, these also correspond to the two-character ISO 3166-1 codes. In the future, other geographies not in ISO 3166-1 may be added.

Buzz

Buzz, tagged as <buzz>, represents a sum of entity-specific words and phrases used in TRMI computations. It can be non-integer when any of the words/phrases are described with a “minimizer”, which reduces the intensity of the primary word or phrase. For example, in the phrase “less concerned” the score of the word “concerned” is minimized by “less”. Additionally, common words such as “new” may have a minor but significant contribution to the Innovation TRMI. As a result, the scores of common words/phrases with minor TRMI contributions can be minimized.

Relative Buzz

Relative Buzz, tagged as <relativeBuzz>, represents the fraction of all Buzz, for a particular Window Timestamp and Data Type(s), which belongs to a particular asset. Thus Relative Buzz is between 0 and 1, and the sum of Relative Buzz across all assets for a particular Window Timestamp and Data Type Set always sum to 1.00, with one exception.

This exception is that for the company groups asset class, Relative Buzz sums to 2.00. It sums to 1.00 for the ten assets based on TRBC global economic sectors and 1.00 for the remaining company groups assets separately, Also for the company groups asset class, one should note that Relative Buzz is computed on the asset/group level like for the other asset classes. That is, it is not a sum of buzz scores for individual companies in a group, divided by the sum of buzz scores for companies in any group. Because many companies are in multiple company groups, if Relative Buzz were computed in this way, then the sum of the Relative Buzz would exceed 1 instead of equal 1.

Relative Buzz is computed differently for individual companies (Chapter 7). Because TRMI Companies covers thousands of companies, Relative Buzz is made more meaningful by restricting the denominator to the domicile level, so that the sum of Relative Buzz across all companies from a specific domicile is 1. See the TRMI Companies reference file, described in Chapter 14, for information on companies reference data.

For companies, Relative Buzz is generated for DAI and HOU data but not for 01M. The volatility of Buzz scores in the 01M data reduces the intuitive value of this measure.

When the Buzz is zero, Relative Buzz can be either zero or indeterminate, depending upon whether any of the other assets for that Window Timestamp and Data Type Set have non-zero buzz. If any have non-zero Buzz, then the Relative Buzz for this asset will be zero, and otherwise indeterminate. Indeterminate Relative Buzz is expressed as a blank value in the live feed and the archive.

CHAPTER 7 COMPANIES

COMPANIES ASSETS

TRMI Companies data covers more than 8,000 active companies from over 30 countries, across the feeds and archives. In general, these companies were selected for inclusion based on domicile, listed market capitalization, and/or economic/industry significance. Coverage includes all companies that are or were constituents of a company groups asset. Coverage is updated monthly.

A list of the covered companies and monthly changes in coverage is available on the MRN FTP site. See Chapter 14 for information on accessing the file and its contents.

COMPANIES TRMI INDICES

The 31 TRMI indices for the companies carry six significant digits past the decimal point. Negative numbers have a leading minus (-) sign. The table below summarizes these fields. For more information on those with range "0 to 1*", see Chapter 4 above.

These 31 are the same as those for company groups.

Index	**Description:** *24 hour rolling average score of references in news and social media to...*	**Range**
sentiment	overall positive references, net of negative references	-1 to 1
optimism	optimism, net of references to pessimism	-1 to 1
fear	fear and anxiety	0 to 1*
joy	happiness and affection	0 to 1*
trust	trustworthiness, net of references connoting corruption	-1 to 1
violence	violence and war	0 to 1*
conflict	disagreement and swearing net of agreement and conciliation	-1 to 1
gloom	gloom and negative future outlook	0 to 1*
stress	distress and danger	0 to 1*
timeUrgency	urgency and timeliness, net of references to tardiness and delays	-1 to 1
uncertainty	uncertainty and confusion	0 to 1*
emotionVsFact	all emotional sentiments, net of all factual and topical references	-1 to 1
longShort	buying, net of references to shorting or selling	-1 to 1
longShortForecast	forecasts of buying, net of references to forecasts of shorting or selling	-1 to 1
priceDirection	price increases, net of references to price decreases	-1 to 1
priceForecast	forecasts of asset price rises, net of references to forecasts of asset price drops	-1 to 1
volatility	volatility in market prices or business conditions	0 to 1*

loveHate	love, net of references to hate	-1 to 1
anger	anger and disgust	0 to 1*
debtDefault	debt defaults and bankruptcies	0 to 1*
innovation	innovativeness	0 to 1*
marketRisk	positive emotionality and positive expectations net of negative emotionality and negative expectations. Includes factors from social media found characteristic of speculative bubbles – higher values indicate greater bubble risk. Also known as the “Bubbleometer.”	-1 to 1
analystRating	upgrade activity, net of references to downgrade activity	-1 to 1
dividends	dividends rising, net of references to dividends falling	-1 to 1
earningsForecast	expectations about improving earnings, less those of worsening earnings	-1 to 1
fundamentalStrength	positivity about accounting fundamentals, net of references to negativity about accounting fundamentals	-1 to 1
layoffs	staff reductions and layoffs	0 to 1*
litigation	litigation and legal activity	0 to 1*
managementChange	changes in a company's management team, net of references to stability in the management team	-1 to 1
managementTrust	trust expressed in a company's management team, net of references to reports of unethical behavior amongst the management team	-1 to 1
mergers	merger or acquisition activity	0 to 1*

CHAPTER 8 COMPANY GROUPS

COMPANY GROUPS ASSETS

Data is reported on 61 assets corresponding to groups of companies assembled according to domicile, market capitalization, and/or business classification.

Many of these correspond to grouping in the hierarchical Thomson Reuters Business Classification (TRBC) system. In descending order of hierarchy, the four levels are economic sector, business sector, industry group and industry. For more information on TRBC, please see http://thomsonreuters.com/products_services/financial/thomson_reuters_indices/trbc/. All TRBC codes below are extant in the TRBC 2012 system.

The 61 assets can be characterized in a few ways. 26 are composed of US-based companies, while 17 are for non-US companies and another 18 are global. 21 resemble equity indices and are filtered chiefly by market cap ranks, while the other 40 are composed according to a combination of TRBC code, domicile, and market cap above $100 million USD.

Note that because these groups are calculated with approximate point-in-time composition, all the asset codes used here were invented for this application. They are not RICs.

US-based Company Groups (26)

Index-oriented Assets (4)

All internal rankings are by market capitalization.

Asset Code	Description	Resembling Index
MPTRXUS30	Top 30 US-based companies	Dow Jones Industrial Average
MPTRXUS500	Top 500 US-based companies	S&P 500
MPTRXUSMID2000	Ranks 2001-3000 of US-based companies	Russell 2000
MPTRXUSNAS100	Top 100 Nasdaq-based companies	Nasdaq 100

TRBC Economic Sector-based Assets (10)

Asset Code	TRBC Description	TRBC Code
MPTRXUSENE	Energy	50
MPTRXUSMAT	Basic Materials	51
MPTRXUSIND	Industrials	52
MPTRXUSYCY	Cyclical Consumer Goods & Services	53
MPTRXUSNCY	Non-Cyclical Consumer Goods & Services	54
MPTRXUSFIN	Financials	55
MPTRXUSHLC	Healthcare	56
MPTRXUSTEC	Technology	57
MPTRXUSCOM	Telecommunications Services	58
MPTRXUSUTL	Utilities	59

TRBC Business Sector-based Assets (5)

Asset Code	TRBC Description	TRBC Code
MPTRXUSI4	Transportation	5240
MPTRXUSY3	Cyclical Consumer Services	5330
MPTRXUSY4	Retailers	5340
MPTRXUSN1	Food & Beverages	5410
MPTRXUSF4	Real Estate	5540

TRBC Industry Group-based Assets (2)

Asset Code	TRBC Description	TRBC Code
MPTRXUSOILS	Oil & Gas Related Equipment and Services	501030
MPTRXUSAERO	Aerospace & Defense	521010

TRBC Industry-based Assets (5)

Asset Code	TRBC Description	TRBC Code
MPTRXUSOILE	Oil & Gas Exploration and Production	50102020
MPTRXUSWAST	Environmental Services	52203010
MPTRXUSAIRL	Airlines	52406010
MPTRXUSHBLD	Homebuilding & Construction Supplies	53203010
MPTRXUSBANK	Banks	55101010

Non-US- based Company Groups (17)

Index-oriented Assets (17)

All internal rankings are by market capitalization.

Asset Code	Description	Resembling Index
MPTRXAU500	Top 500 Australia-based companies	ASX All Ordinaries
MPTRXBR50	Top 50 Brazil-based companies	IBRX 50
MPTRXCA250	Top 250 Canada-based & Toronto-listed companies	S&P/TSX Composite
MPTRXCH20	Top 20 Switzerland-based companies	Swiss Market
MPTRXCN300	Top 300 China-based companies	CSI 300
MPTRXDE30	Top 30 Germany-based companies	Deutsche Börse DAX 30
MPTRXEM50	Top 50 emerging markets companies	MSCI 50
MPTRXES35	Top 35 Spain-based companies	IBEX 35
MPTRXEU50	Top 50 pan-European companies	EURO STOXX 50
MPTRXFR40	Top 40 France-based companies	CAC 40
MPTRXGB100	Top 100 UK-based & LSE-listed companies	FTSE 100
MPTRXGBMID250	Ranks 101-350 of UK-based & LSE-listed companies	FTSE Mid 250
MPTRXHK50	Top 50 Hong Kong-listed companies based in Hong Kong and China	Hang Seng
MPTRXIN50	Top 50 India-based companies	Nifty 50
MPTRXJP225	Top 225 Japan-based companies	Nikkei 225
MPTRXRU50	Top 50 Russia-based companies	RTS
MPTRXSG30	Top 30 Singapore-based companies	FTSE Straits Times

Global Company Groups (18)

TRBC Economic Sector-based Assets (10)

Asset Code	TRBC Description	TRBC Code
MPTRXENE	Energy	50
MPTRXMAT	Basic Materials	51
MPTRXIND	Industrials	52
MPTRXYCY	Cyclical Consumer Goods & Services	53
MPTRXNCY	Non-Cyclical Consumer Goods & Services	54
MPTRXFIN	Financials	55
MPTRXHLC	Healthcare	56
MPTRXTEC	Technology	57
MPTRXCOM	Telecommunications Services	58
MPTRXUTL	Utilities	59

TRBC Business Sector-based Assets (1)

Asset Code	TRBC Description	TRBC Code
MPTRXE14	Renewable Energy	5020

TRBC Industry Group-based Assets (3)

Asset Code	TRBC Description	TRBC Code
MPTRXCOAL	Coal	501010
MPTRXBIOT	Biotechnology & Medical Research	562020
MPTRXT11	Semiconductors & Semiconductor Equipment	571010

TRBC Industry-based Assets (2)

Asset Code	TRBC Description	TRBC Code
MPTRXGOLD	Precious Metals & Minerals	51201010
MPTRXGAMI	Casinos & Gaming	53301030

TRBC Hybrid-driven Assets (2)

Asset Code	TRBC Description	TRBC Code
MPTRXPKB	Construction Materials	512020
	Construction & Engineering	522010
	Homebuilding & Construction Supplies	532030
MPTRXPEJ	Leisure Products	532050
	Hotels & Entertainment Services	533010

Constitution of Company Groups Assets in the TRMI Archives and Feed

The constituent companies of these assets change over time. In the archives, the TRBC-based constituent lists are refreshed quarterly starting on January 1, 2006. The remaining groups are refreshed quarterly since the beginning of the archives, January 1, 1998. All groups have their constituents update monthly starting from September 2014.

In the live data, the systemVersion value will increase with each monthly update.

COMPANY GROUPS TRMI INDICES

The 31 TRMI indices for the company groups asset class carry six significant digits past the decimal point. Negative numbers have a leading minus (-) sign. The table below summarizes these fields. For more information on those with range “0 to 1*”, see Chapter 4 above.

These 31 are the same as those for companies.

Index	**Description:** *24 hour rolling average score of references in news and social media to...*	**Range**
sentiment	overall positive references, net of negative references	-1 to 1
optimism	optimism, net of references to pessimism	-1 to 1
fear	fear and anxiety	0 to 1*
joy	happiness and affection	0 to 1*
trust	trustworthiness, net of references connoting corruption	-1 to 1
violence	violence and war	0 to 1*
conflict	disagreement and swearing net of agreement and conciliation	-1 to 1
gloom	gloom and negative future outlook	0 to 1*
stress	distress and danger	0 to 1*
timeUrgency	urgency and timeliness, net of references to tardiness and delays	-1 to 1
uncertainty	uncertainty and confusion	0 to 1*
emotionVsFact	all emotional sentiments, net of all factual and topical references	-1 to 1
longShort	buying, net of references to shorting or selling	-1 to 1
longShortForecast	forecasts of buying, net of references to forecasts of shorting or selling	-1 to 1
priceDirection	price increases, net of references to price decreases	-1 to 1
priceForecast	forecasts of asset price rises, net of references to forecasts of asset price drops	-1 to 1
volatility	volatility in market prices or business conditions	0 to 1*
loveHate	love, net of references to hate	-1 to 1
anger	anger and disgust	0 to 1*
debtDefault	debt defaults and bankruptcies	0 to 1*
innovation	innovativeness	0 to 1*
marketRisk	positive emotionality and positive expectations net of negative emotionality and negative expectations. Includes factors from social media found characteristic of speculative bubbles – higher values indicate greater bubble risk. Also known as the “Bubbleometer.”	-1 to 1
analystRating	upgrade activity, net of references to downgrade activity	-1 to 1

dividends	dividends rising, net of references to dividends falling	-1 to 1
earningsForecast	expectations about improving earnings, less those of worsening earnings	-1 to 1
fundamentalStrength	positivity about accounting fundamentals, net of references to negativity about accounting fundamentals	-1 to 1
layoffs	staff reductions and layoffs	0 to 1*
litigation	litigation and legal activity	0 to 1*
managementChange	changes in a company's management team, net of references to stability in the management team	-1 to 1
managementTrust	trust expressed in a company's management team, net of references to reports of unethical behavior amongst the management team	-1 to 1
mergers	merger or acquisition activity	0 to 1*

CHAPTER 9 CURRENCIES

CURRENCIES ASSETS

There are 30 currencies, listed in the table below.

Asset Code	Currency
ARS	Argentine Peso
AUD	Australian Dollar
BRL	Brazilian Real
CAD	Canadian Dollar
CNY	Chinese Yuan Renminbi
EGP	Egyptian Pound
EUR	Euro
HKD	Hong Kong Dollar
INR	Indian Rupee
IDR	Indonesian Rupiah
IRR	Iranian Rial
ILS	Israeli Shekel
JPY	Japanese Yen
MYR	Malaysian Ringgit
MXN	Mexican Peso
NZD	New Zealand Dollar
NOK	Norwegian Krone
PLN	Polish Zloty
RUB	Russian Ruble
SGD	Singapore Dollar
ZAR	South African Rand
KRW	South Korean Won
SEK	Swedish Krona
CHF	Swiss Franc
TWD	Taiwanese Dollar
THB	Thai Baht
TRY	Turkish Lira
USD	U.S. Dollar
AED	United Arab Emirates Dirham
GBP	United Kingdom Pound Sterling

CURRENCIES TRMI INDICES

The 21 TRMI indices for the currencies asset class carry six significant digits past the decimal point. Negative numbers have a leading minus (-) sign. The table below summarizes these fields. For more information on those with range “0 to 1*”, see Chapter 4 above.

Index	**Description:** *24 hour rolling average score of references in news and social media to...*	Range
sentiment	overall positive references, net of negative references	-1 to 1
optimism	optimism, net of references to pessimism	-1 to 1

fear	fear and anxiety	0 to 1*
joy	happiness and affection	0 to 1*
trust	trustworthiness, net of references connoting corruption	-1 to 1
violence	violence and war	0 to 1*
conflict	disagreement and swearing net of agreement and conciliation	-1 to 1
gloom	gloom and negative future outlook	0 to 1*
stress	distress and danger	0 to 1*
timeUrgency	urgency and timeliness, net of references to tardiness and delays	-1 to 1
uncertainty	uncertainty and confusion	0 to 1*
emotionVsFact	all emotional sentiments, net of all factual and topical references	-1 to 1
longShort	buying, net of references to shorting or selling	-1 to 1
longShortForecast	forecasts of buying, net of references to forecasts of shorting or selling	-1 to 1
priceDirection	price increases, net of references to price decreases	-1 to 1
priceForecast	forecasts of asset price rises, net of references to forecasts of asset price drops	-1 to 1
volatility	volatility in market prices or business conditions	0 to 1*
loveHate	love, net of references to hate	-1 to 1
carryTrade	carry trade	0 to 1*
currencyPegInstability	the instability of a currency peg, net of references to the stability of a currency peg	-1 to 1
priceMomentum	currency price trend strength, net of references to trend weakness	-1 to 1

CHAPTER 10 AGRICULTURAL COMMODITIES

AGRICULTURAL COMMODITIES ASSETS

There are 12 agricultural commodities being scored. The asset codes correspond to Thomson Reuters topic codes found in news stories.

Asset Code	Commodity
CTTL	Cattle
COC	Cocoa
COF	Coffee
COR	Corn
COT	Cotton
HOGS	Hogs
ORJ	Orange Juice
POIL	Palm Oil
RICE1	Rice
SOY1	Soybeans
SUG	Sugar
WHT	Wheat

AGRICULTURAL COMMODITIES TRMI INDICES

The 27 TRMI indices for the agricultural commodities asset class carry six significant digits past the decimal point. Negative numbers have a leading minus (-) sign. The table below summarizes these fields. For more information on those with range “0 to 1*”, see Chapter 4 above.

Index	**Description:** *24 hour rolling average score of references in news and social media to...*	Range
sentiment	overall positive references, net of negative references	-1 to 1
optimism	optimism, net of references to pessimism	-1 to 1
fear	fear and anxiety	0 to 1*
joy	happiness and affection	0 to 1*
trust	trustworthiness, net of references connoting corruption	-1 to 1
violence	violence and war	0 to 1*
conflict	disagreement and swearing net of agreement and conciliation	-1 to 1
gloom	gloom and negative future outlook	0 to 1*
stress	distress and danger	0 to 1*
timeUrgency	urgency and timeliness, net of references to tardiness and delays	-1 to 1
uncertainty	uncertainty and confusion	0 to 1*
emotionVsFact	all emotional sentiments, net of all factual and topical references	-1 to 1

longShort	buying, net of references to shorting or selling	-1 to 1
longShortForecast	forecasts of buying, net of references to forecasts of shorting or selling	-1 to 1
priceDirection	price increases, net of references to price decreases	-1 to 1
priceForecast	forecasts of asset price rises, net of references to forecasts of asset price drops	-1 to 1
volatility	volatility in market prices or business conditions	0 to 1*
consumptionVolume	factors leading to increased consumption, net of references to factors leading to decreased consumption	-1 to 1
productionVolume	increased production, net of references to factors leading to decreased production	-1 to 1
regulatoryIssues	regulatory issues	0 to 1*
supplyVsDemand	surplus supply and lack of demand, net of references to supply shortage and high demand	-1 to 1
supplyVsDemandForecast	expectations of supply outstripping demand, net of references to expectations of demand outstripping supply	-1 to 1
acreageCultivated	increases in acreage and crop cultivation, net or references to decreases in acreage and crop cultivation	-1 to 1
agDisease	commodity disease	0 to 1*
subsidies	subsidies affecting commodity prices	0 to 1*
subsidiesSentiment	increases in subsidies, net of references to decreases in subsidies	-1 to 1
weatherDamage	commodity weather damage	0 to 1*

CHAPTER 11 ENERGY & MATERIAL COMMODITIES

ENERGY & MATERIAL COMMODITIES ASSETS

There are 22 energy & material commodities being scored. The asset codes correspond to Thomson Reuters topic codes found in news stories.

Asset Code	Commodity
ALU	Aluminum
BIOF	Biofuels
COA	Coal
CPPR	Copper
CRU	Crude Oil
BIOETH	Ethanol
MOG	Gasoline
GOL	Gold
HOIL	Heating Oil
IRN	Iron
JET	Jet Fuel
LNG	Liquefied Natural Gas
NAP	Naphtha
NGS	Natural Gas
NKL	Nickel
NSEA	North Sea Oil
PALL	Palladium
PLAT	Platinum
RAREE	Rare Earths
SLVR	Silver
STEE	Steel
URAN	Uranium

ENERGY & MATERIAL COMMODITIES TRMI INDICES

The 24 TRMI indices for the energy & material commodities asset class carry six significant digits past the decimal point. Negative numbers have a leading minus (-) sign. The table below summarizes these fields. For more information on those with range “0 to 1*”, see Chapter 4 above.

Index	**Description:** *24 hour rolling average score of references in news and social media to...*	Range
sentiment	overall positive references, net of negative references	-1 to 1
optimism	optimism, net of references to pessimism	-1 to 1
fear	fear and anxiety	0 to 1*
joy	happiness and affection	0 to 1*
trust	trustworthiness, net of references connoting corruption	-1 to 1
violence	violence and war	0 to 1*
conflict	disagreement and swearing net of agreement and	-1 to 1

	conciliation	
gloom	gloom and negative future outlook	0 to 1*
stress	distress and danger	0 to 1*
timeUrgency	urgency and timeliness, net of references to tardiness and delays	-1 to 1
uncertainty	uncertainty and confusion	0 to 1*
emotionVsFact	all emotional sentiments, net of all factual and topical references	-1 to 1
longShort	buying, net of references to shorting or selling	-1 to 1
longShortForecast	forecasts of buying, net of references to forecasts of shorting or selling	-1 to 1
priceDirection	price increases, net of references to price decreases	-1 to 1
priceForecast	forecasts of asset price rises, net of references to forecasts of asset price drops	-1 to 1
volatility	volatility in market prices or business conditions	0 to 1*
consumptionVolume	factors leading to increased consumption, net of references to factors leading to decreased consumption	-1 to 1
productionVolume	increased production, net of references to factors leading to decreased production	-1 to 1
regulatoryIssues	regulatory issues	0 to 1*
supplyVsDemand	surplus supply and lack of demand, net of references to supply shortage and high demand	-1 to 1
supplyVsDemandForecast	expectations of supply outstripping demand, net of references to expectations of demand outstripping supply	-1 to 1
newExploration	new ventures/exploration	0 to 1*
safetyAccident	safety accidents	0 to 1*

CHAPTER 12 COUNTRIES

COUNTRIES ASSETS

There are 133 countries or regions being scored. The asset codes correspond to Thomson Reuters topic codes for geopolitical units.

The table below is sorted by Country/Region.

Asset Code	Country/Region
AF	Afghanistan
DZ	Algeria
AO	Angola
AR	Argentina
AM	Armenia
AU	Australia
AT	Austria
BH	Bahrain
BD	Bangladesh
BY	Belarus
BE	Belgium
BO	Bolivia
BA	Bosnia and Herzegovina
BR	Brazil
BG	Bulgaria
KH	Cambodia
CA	Canada
KY	Cayman Islands
CL	Chile
CN	China
CO	Colombia
CD	Congo, Democratic Republic of the
CR	Costa Rica
CI	Côte d'Ivoire
HR	Croatia
CU	Cuba
CY	Cyprus
CZ	Czech Republic
DK	Denmark
DO	Dominican Republic
EC	Ecuador
EG	Egypt
SV	El Salvador
ER	Eritrea
EE	Estonia
ET	Ethiopia
EZ	Euro-zone
FI	Finland
FR	France
GA	Gabon

GE	Georgia
DE	Germany
GH	Ghana
GR	Greece
GT	Guatemala
GY	Guyana
HT	Haiti
HN	Honduras
HK	Hong Kong
HU	Hungary
IS	Iceland
IN	India
ID	Indonesia
IR	Iran
IQ	Iraq
IE	Ireland
IL	Israel
IT	Italy
JP	Japan
JO	Jordan
KZ	Kazakhstan
KE	Kenya
KW	Kuwait
KG	Kyrgyzstan
LV	Latvia
LA	Laos
LB	Lebanon
LR	Liberia
LY	Libya
LT	Lithuania
MO	Macau
MK	Macedonia
MG	Madagascar
MY	Malaysia
ML	Mali
MT	Malta
MX	Mexico
MN	Mongolia
MA	Morocco
MM	Myanmar
NP	Nepal
NL	Netherlands
NZ	New Zealand
NI	Nicaragua
NG	Nigeria
KP	North Korea
NO	Norway
OM	Oman
PK	Pakistan
PS	Palestinian Territories

PA	Panama
PG	Papua New Guinea
PY	Paraguay
PE	Peru
PH	Philippines
PL	Poland
PT	Portugal
QA	Qatar
RO	Romania
RU	Russia
SA	Saudi Arabia
RS	Serbia
SC	Seychelles
SL	Sierra Leone
SG	Singapore
SO	Somalia
ZA	South Africa
KR	South Korea
SS	South Sudan
ES	Spain
LK	Sri Lanka
SD	Sudan
SE	Sweden
CH	Switzerland
SY	Syria
TW	Taiwan
TJ	Tajikistan
TZ	Tanzania
TH	Thailand
TN	Tunisia
TR	Turkey
TM	Turkmenistan
UG	Uganda
UA	Ukraine
AE	United Arab Emirates
GB	United Kingdom
US	United States
UY	Uruguay
UZ	Uzbekistan
VE	Venezuela
VN	Vietnam
YE	Yemen
ZW	Zimbabwe

COUNTRIES TRMI INDICES

The 48 TRMI indices for the countries asset class carry six significant digits past the decimal point. Negative numbers have a leading minus (-) sign. The table below summarizes these fields. For more information on those with range “0 to 1*”, see Chapter 4 above.

Index	**Description:** *24 hour rolling average score of references in news and social media to...*	Range
sentiment	overall positive references, net of negative references	-1 to 1
optimism	optimism, net of references to pessimism	-1 to 1
fear	fear and anxiety	0 to 1*
joy	happiness and affection	0 to 1*
trust	trustworthiness, net of references connoting corruption	-1 to 1
violence	violence and war	0 to 1*
conflict	disagreement and swearing net of agreement and conciliation	-1 to 1
gloom	gloom and negative future outlook	0 to 1*
stress	distress and danger	0 to 1*
timeUrgency	urgency and timeliness, net of references to tardiness and delays	-1 to 1
uncertainty	uncertainty and confusion	0 to 1*
emotionVsFact	all emotional sentiments, net of all factual and topical references	-1 to 1
loveHate	love, net of references to hate	-1 to 1
anger	anger and disgust	0 to 1*
debtDefault	debt defaults and bankruptcies	0 to 1*
innovation	innovativeness	0 to 1*
marketRisk	positive emotionality and positive expectations net of negative emotionality and negative expectations. Includes factors from social media found characteristic of speculative bubbles – higher values indicate greater bubble risk. Also known as the “Bubbleometer.”	-1 to 1
budgetDeficit	a budget deficit, net of references to a surplus	-1 to 1
businessExpansion	businesses expanding, net of references to contraction	-1 to 1
centralBank	the central bank of a country	0 to 1*
commercialRealEstateSentiment	positive references to commercial real estate, net of negative references	-1 to 1
consumerSentiment	positive consumer sentiment, net of references to negative consumer sentiment	-1 to 1
creditEasyVsTight	credit conditions being easy, net of references to credit conditions being tight	-1 to 1
economicGrowth	increased business activity, net of references to decreased business activity	-1 to 1
economicUncertainty	uncertainty about business climate net of confidence and certainty	-1 to 1
economicVolatility	increasing economic volatility, net of economic stability	-1 to 1

financialSystemInstability	financial system instability, net of references to financial system stability	-1 to 1
fiscalPolicyLooseVsTight	fiscal policy being loose, net of references to fiscal policy being tight	-1 to 1
governmentAnger	anger and disgust about government officials and departments	0 to 1*
governmentCorruption	fraud and corruption in government, net of references to trust in government	-1 to 1
governmentInstability	governmental instability, net of references to governmental stability	-1 to 1
inflation	consumer price increases, net of references to consumer price decreases	-1 to 1
inflationForecast	forecasts of consumer price increases, net of forecasts of consumer price decreases (deflation)	-1 to 1
interestRates	interest rates rising, net of references to rates falling	-1 to 1
interestRatesForecast	forecasts of interest rates rising, net of forecasts of rates falling	-1 to 1
investmentFlows	investment inflows, net of references to investment outflows	-1 to 1
monetaryPolicyLooseVsTight	monetary policy being loose, net of references to monetary policy being tight	-1 to 1
naturalDisasters	natural disasters	0 to 1*
regimeChange	regime change	0 to 1*
residentialRealEstateGrowth	residential real estate expansion, net of references to contraction	-1 to 1
residentialRealEstateSales	residential real estate sales rising, net of references to sales decreasing	-1 to 1
residentialRealEstateSentiment	positive references to residential real estate, net of negative references	-1 to 1
residentialRealEstateValues	residential real estate values rising, net of references to declining values	-1 to 1
sanctions	sanctions or embargoes emanating from or against a country	0 to 1*
socialInequality	social inequality	0 to 1*
socialUnrest	social unrest and calls for political change	0 to 1*
tradeBalance	exports, net of references to imports	-1 to 1
unemployment	unemployment rising, net of references to unemployment falling	-1 to 1

CHAPTER 13 CONSTRUCTING TRMI WITH GREATER WINDOW LENGTHS

BACKGROUND

TRMIs are constructed internally as Buzz-weighted averages across various News and Social Media content sources.

CREATING CUSTOM TRMI WINDOW LENGTHS

Users may construct custom TRMIs of varying window lengths from Buzz-weighted averages of minutely TRMI data.

For example, for a given company (*assetCode*), content source (*dataType*) and datetime (*windowTimestamp*), let $Buzz_0$, $Buzz_{-1}$, ..., $Buzz_{-(N-1)}$ and $TRMI_0$, $TRMI_{-1}$, …, $TRMI_{-(N-1)}$ represent the corresponding Buzz and TRMI minutely data over the trailing *N* minutes. Then the Buzz-weighted average TRMI over the trailing *N*-minute window length may be explicitly calculated as:

$$(Buzz_0 * TRMI_0 + Buzz_{-1} * TRMI_{-1} + \ldots + Buzz_{-(N-1)} * TRMI_{-(N-1)}) \,/\, (Buzz_0 + Buzz_{-1} + \ldots + Buzz_{-(N-1)})$$

COMPARING CONSTRUCTED TRMI DATA TO ACTUAL TRMI DATA

Note that the daily (WDAI_UDAI) and hourly (WDAI_UHOU) TRMI data, which use a 24-hour/1440-minute trailing window, cannot be reproduced perfectly using minutely TRMI data (W01M_U01M). This is because textual content is aggregated in the daily and hourly feeds based in part on their online availability (*publication timestamp*). Articles are excluded from the daily and hourly feeds 24 hours after publication, whereas content in the minutely feed is aggregated solely by its time of acquisition by MarketPsych crawlers (*crawl timestamp*). See “Timing of Social Media Articles in Archive” for more information about timestamps. The greater the delay between the publication and crawl timestamps, the fewer 1440-minute TRMI windows will include that content. As a result, the Buzz of the artificially reconstructed 24-hour TRMI will always match or exceed that of the actual 24-hour TRMI.

CHAPTER 14 MRN FTP SITE: DATA AND REFERENCE FILES

OVERVIEW

Thomson Reuters provides live and archive TRMI data over FTP.

Data is formatted as tab-delimited text files. Archive files are compressed, in .zip 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 contains useful reference files for the Companies offering:

- Listings of live and obsolete assets
- Mappings to common identifiers, including CUSIP, ISIN, and SEDOL
- Changes in coverage

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.

WINDOW LENGTHS AND UPDATE FREQUENCIES

Chapter 2 explained that TRMI are produced over varying window lengths and update frequencies. The excerpted table below shows which asset classes are available over which combinations of the above two factors.

Asset Class(es)	Window Length	Update Frequency	FTP File Abbreviation	Live or Archive	Delivery
All	1 minute	1 minute	W01M_U01M	Both	FTP
All	1440 minutes / 24 hours	1 hour	WDAI_UHOU	Both	FTP
All except for Companies	1440 minutes / 24 hours	1 minute	N/A	Live	News Feed Direct
All except for Companies	1440 minutes / 24 hours	5 minutes	WDAI_U05M	Archive	FTP
All	1440 minutes / 24 hours	Daily, at 3:30 Eastern time	WDAI_UDAI	Archive	FTP

Of note, there is one case where there is not archive data to match the live data: the non-Companies 24-hour window TRMI that is updated minutely (third row above). Minutely archives for this option are not provided due to how slowly the data changes. Instead of minutely updates, the archive data is conflated to 5-minute updates (fourth row above). These two combinations will likely be deprecated in a future release.

LIVE FILES

Each live file contains all the scores for an asset class, for a certain window length and update frequency. This section does not discuss live files delivered over News Feed Direct. See Chapter 6 for more information.

Window Length and Update Frequency

- 1-minute window length, updated minutely: W01M_U01M
- 1440-minute (24-hour) window length, updated hourly on the hour: WDAI_UHOU.

Non-zero Buzz Records and Files

Only rows with non-zero Buzz appear in these files. Moreover, if there are no non-zero Buzz rows to include in a file, then that file will not be published.

Directory Structure

/TRMI_LIVE/*{Asset Class}*/*{Time Abbreviation}*/*[LastTwoHours/LastTwoDays/Older]*

- *{Asset Class}* can take one of the following values:
  - CMPNY: individual companies
  - CMPNY_GRP: company groups
  - COM_AGR: agricultural commodities
  - COM_ENM: energy & material commodities
  - COU: countries
  - CUR: currencies
- *{Time Abbreviation}* describes the combination of window length and update frequency, as per the table above
  - W01M_U01M
  - WDAI_UHOU
- *[LastTwoHours/LastTwoDays/Older]*
  - *LastTwoHours* contains files that were published since the beginning of the *previous* calendar hour. For example, if the current time is 18:32 UTC, then it will contain files published since 17:00 UTC. Unless older files are needed, FTP clients may concentrate only on this subdirectory.
  - *LastTwoDays* contains files, not present in *LastTwoHours*, that were published since the beginning of the *previous* calendar day in UTC. For example, if the current date and time is 2016-03-04 18:32 UTC, then it will contain files published since 2016-03-03 00:00 UTC.
  - *Older* contains less recent files, not present in *LastTwoHours* or *LastTwoDays*, that were published since the last monthly archive update. At each monthly archive update, all corresponding files in /TRMI_LIVE are deleted.

File Naming

TRMI.Live.{Asset Class}.{Time Abbreviation}.{Time Period}.{System Version, converted}.txt, such that

- *{Asset Class}* as per above
- *{Time Abbreviation}* as per above
- *{Time Period}* will be in “yyyymmdd-hhmm” format. Note that boundaries are according to UTC and mark the end of the content window.
- *{System Version, converted}* according to the <systemVersion> tag. The systemVersion value, “MP:2.2.0” will be converted to “0220”. Note that this string’s length reaches five characters when the third number in the systemVersion exceeds 9, e.g., “2.2.10” will be converted to “02210”.

Fields

- id
- assetCode, excluding its prefix

- ticker: only for Companies. Note that ticker is generally not point-in-time, although it can vary across System Version values.
- windowTimestamp
- dataType: “News”, “Social”, or “News_Social”
- systemVersion, excluding its prefix
- buzz
- relativeBuzz: not present in Companies minutely-window data, as stated in Chapter 6
- other indices: sentiment, optimism, et al.

Sort Order

- assetCode, excluding its prefix
- dataType

Accessing Old Live Files and Directory Cleanup

Live files remain available until a corresponding archive file has been uploaded. Thus there is typically between 1-2 months of data in the /TRMI_LIVE subdirectories.

ARCHIVE FILES

Each live file contains all the scores for an asset class, for a certain window length and update frequency.

Window Length and Update Frequency

- 1-minute window length, updated minutely: W01M_U01M
- 1440-minute (24-hour) window length, updated hourly on the hour: WDAI_UHOU.
- 1440-minute (24-hour) window length, updated every 5 minutes: WDAI_U05M. This corresponds to the live data published over News Feed Direct, which is updated every minute. This format likely will be deprecated in a future release.
- 1440-minute (24-hour) window length, updated daily at 3:30 Eastern time: WDAI_UDAI.

Timing of New Archive Files

New archive files for the previous month are typically uploaded to FTP between 10 and 15 days after the close of that month.

Timing of Social Media Articles in Archive

As background, there are three timestamps used in assigning timestamps to content.

- Publication time: a timestamp from the data itself that may be assumed to represent when the content was published
- Acquisition time: MarketPsych-assigned timestamp, for when MarketPsych obtained the data. For some content sources, the difference between the publication time and acquisition time is consistently ~1 second or less, but for certain significant social media sources this can average in the tens of minutes. This large difference is especially problematic for certain sources that MarketPsych did not collect live.
- Aggregation time: MarketPsych-assigned timestamp describing when the component TRMI scores from content are finalized. For aligning live feed data and pro forma archives, aggregation time showed the best performance.

Over different periods, the timestamp used for archive calculation changed.

- 1998 - February 2013: publication time

- March 2013 - August 2015: acquisition time. For TRMI data that uses a 1440-minute/24-hour window length, the live feed and all archives are aggregated from article acquisition time to publication time plus 24 hours.
- September 2015 - present: aggregation time.

Now using aggregation time, although single-minute differences between the live feed and archive buzz remain, they affect well under 1% of buzz scores in the minute-window data.

Archive Backfills

As might be expected for data derived from numerous textual sources, the processing of MarketPsych Indices undergoes intermittent changes to its content inputs and filtering algorithms. These changes generally are introduced in the live data and in monthly archive updates, on a pro forma basis. Full-history TRMI archives are backfilled to reflect these content changes on an occasional basis, often in tandem with a major version upgrade.

Non-zero Buzz Records

Only rows with non-zero Buzz appear in these files, except for daily updating files (WDAI_UDAI) for non-Companies data.

Standard and Bulk Archive Files

Archive files come in two types. Standard archive files were designed so that they that can be viewed in Microsoft Excel 2007, which supports 1,048,576 rows. These files are stored in the “Historical” and “Recent” subdirectories described in the next section.

For users who prefer to work with fewer files, archive files are also being provided in a small number of large files, according to the capacity of what can be zipped. These files are stored in the “Bulk” subdirectory.

Bulk files contain scores on an entire asset class over a given time period. Standard files may cover only a single asset, depending on the window length and update frequency.

Directory Structure

Because of the large number of Companies assets, the directory structure for Companies adds two or three extra levels for organization.

/TRMI/*{Asset Class}*/*{Time Abbreviation}*/*[Historical/Recent/Bulk] [[/{Country}/{Economic Sector code}] /{Asset Code ex-prefix}]*

- *{Asset Class}* can take one of the following values:
  - CMPNY: individual companies
  - CMPNY_GRP: company groups
  - COM_AGR: agricultural commodities
  - COM_ENM: energy & material commodities
  - COU: countries
  - CUR: currencies
- *{Time Abbreviation}* describes the combination of window length and update frequency, as per the table above
  - W01M_U01M
  - WDAI_UHOU
  - WDAI_U05M
  - WDAI_UDAI
- *[Historical/Recent/Bulk]*

- Historical contains all files from the previous calendar year, generally. Files are annual in duration.
- Recent contains files from the current year. Files are monthly 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 TRMI build.
- Bulk is an option for most asset class / time abbreviation combinations, as per the table below. It is not available for any of the WDAI_U05M data. This directory contains larger files than the Historical and Recent directories. As a result, it has none of the subdirectories listed below, and the files are different from the standard archive files.

Bulk File Details

Asset Class	Minutely: W01M_U01M	Hourly: WDAI_UHOU	Daily: WDAI_UDAI
CUR	1998-2010, 2011-2013, 2014, 2015	1998-2010, 2011-2013, 2014, 2015	1998-2010, 2011-2013, 2014, 2015
COU	None	None	1998-2010, 2011-2013, 2014, 2015
COM_ENM	1998-2010, 2011-2013, 2014, 2015	1998-2010, 2011-2013, 2014, 2015	1998-2010, 2011-2013, 2014, 2015
COM_AGR	1998-2010, 2011-2013, 2014, 2015	1998-2010, 2011-2013, 2014, 2015	1998-2010, 2011-2013, 2014, 2015
CMPNY_GRP	None	None	1998-2010, 2011-2013, 2014, 2015
CMPNY	Annual	Quarterly	1998-2003, 2004-2007, 2008-2010, 2011-2013, 2014, 2015

- Two intermediate subdirectory levels apply only to Companies non-bulk data
  - *{Country}* is the two-letter ISO 3166-1 country code
  - *{Economic Sector}* is a three-letter code corresponding to the TRBC Economic Sector.
    - ENE: Energy
    - MAT: Basic Materials
    - IND: Industrials
    - YCY: Cyclical Consumer Goods & Services
    - NCY: Non-Cyclical Consumer Goods & Services
    - FIN: Financials
    - HLC: Healthcare
    - TEC: Technology
    - COM: Telecommunications Services
    - UTL: Utilities
- *{Asset Code ex-prefix}* is the Asset Code without the prefix, e.g., "CRU" for "N2:CRU" (Crude Oil). This level only applies to WDAI_U05M data and to Companies non-bulk data.
  - This level is used for the minutely data only, due to the amount of data

File Naming

TRMI.Archive.{Asset Class}.[[.{Country}.{Economic Sector code}] /.{Asset Code ex-prefix}]{Time Abbreviation}.{Time Period}.{System Version, converted}.zip/txt, such that

- *{Asset Class}* as per above
- *{Country}.{Economic Sector code}* as per above. This applies only to CMPNY WDAI_UHOU and WDAI_UDAI Historical/Recent files.
- *{Asset Code ex-prefix}* is the asset's code without the prefix, e.g., "US" for "N2:US" (United States). This level is only present for WDAI_U05M files and CMPNY W01M_U01M non-Bulk.

- *{Time Abbreviation}* as per above
- *{Time Period}* can take one of the following values, depending on its directory:
    - Bulk: “yyyy-yyyy” for multi-year data, as per “Bulk File Details” table above
    - Historical:
        - “yyyy” format for annual data.
        - “yyyyQq” format for quarterly data. This applies only to hourly (WDAI_UHOU) data.
    - Recent: “yyyymm” format for monthly data
- *{System Version, converted}* according to the <systemVersion> tag. The systemVersion value, “MP:2.2.0” will be converted to “0220”. Note that this string’s length reaches five characters when the third number in the systemVersion exceeds 9, e.g., “2.2.10” will be converted to “02210”.

Archive File Structure

Each .zip file holds one .txt file that contains relevant archive data for that asset and time period, across all three Data Types (News, Social, and combined).

Fields

- id
- assetCode, excluding its prefix
- ticker: only for companies. Note that ticker is generally not point-in-time, although it can vary across System Version values.
- windowTimestamp
- dataType: “News”, “Social”, or “News_Social”
- systemVersion, excluding its prefix
- buzz
- relativeBuzz: not present in Companies minutely-window data, as stated in Chapter 6
- other indices: sentiment, optimism, et al.

Sort Order, Single-asset Files

- windowTimestamp
- dataType

Sort Order, Multi-asset Files

- windowTimestamp
- assetCode
- dataType

COMPANIES REFERENCE FILES

Companies reference files serve several purposes. The files list companies that can be found in the live or archive scores. They add supporting information on the company and on its primary quote, In addition, third-party identifiers – CUSIP, ISIN, and SEDOL – are also available. 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).

Directories

/TRMI/CMPNY/[Basic/CUSIPISIN/SEDOL/CUSIPISINSEDOL]/

Notes:

- The third 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.

Asset Lists and Mappings Files

TRMI.Companies.[Basic/CUSIPISIN/SEDOL/CUSIPISINSEDOL].{System Version, converted}.txt

Notes:

- *{System Version, converted}* according to the <systemVersion> tag. The systemVersion value, "MP:2.1.0" will be converted to "0210". Note that this string's length reaches five characters when the third number in the systemVersion exceeds 9, e.g., "2.1.10" will be converted to "02110".

Fields:

- PermID: Thomson Reuters organizational identifier.
- companyName
- countryOfDomicile: two-character ISO 3166-1 country code
- TRBCEconomicSector: plain-text description of Thomson Reuters Business Classification (TRBC) economic sector
- status: "active" if the PermID may be scored in a live feed. Otherwise, "inactive".
- RIC: main RIC for this company.
- ticker: Because ticker is in the score files, it shall be populated for all companies. Private companies shall have a MarketPsych-designated ticker beginning with "PVT-".
- 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

Values on inactive companies shall attempt to represent their most recent values, including after delisting. Some values may be blank.

Changes File

The second is a file of changes since the previous System Version.

Changes.txt

Contents:

- First line: "TRMI Companies list for systemVersion [*systemVersion*] was created on [*yyyy-mm-dd*]."
- Further lines:
  - "PermID [*xxxxxxxxxx*], [*companyName*], became active."
  - "PermID [*xxxxxxxxxx*], [*companyName*], became inactive."

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 /TRMI_TRIAL, instead of /TRMI
- The third-level directory denotes the amount of data
- The directory level of Recent and Historical is absent

For product information or to make a sales enquiry, visit the Machine Readable News website:

http://thomsonreuters.com/machine-readable-news/

Read more about our products at

http://thomsonreuters.com/products_services

Find out how to contact your local office

http://thomsonreuters.com/about/locations

Document Version 1.0
Date of Issue: 23 March 2016

THOMSON REUTERS