RPNA 4

RavenPack News Analytics

# User Guide and Service Overview for WRDS Users

Powered by


**CONFIDENTIAL INFORMATION**

Version 4.0
Last Updated: June 10, 2016

# Table of Contents

# Introduction

RavenPack News Analytics is a unique source of explanatory and predictive inputs derived from news. The product includes a data set rich with structured information and potential signals and creates new trading opportunities on both scheduled and unscheduled news events. This data is used to power a number of applications ranging from high frequency trading systems requiring low latency inputs to risk and asset management models requiring factors whose time resolution may be daily, weekly, and monthly. RavenPack News Analytics data is also used across a broad range of academic research into the media and event effects on financial markets.

RavenPack automatically tracks and monitors relevant information on nearly 200,000 companies, government organizations, influential people, key geographical locations, and all major currencies and traded commodities. Among the many benefits, RavenPack delivers sentiment analysis and event data most likely to impact financial markets and trading around the world – all in a matter of milliseconds and in a consistent, structured format. The RavenPack News Analytics real-time data feed is specifically designed for direct integration into business and financial applications.

RavenPack quantifies positive and negative perceptions on facts and opinions reported in the news. It continuously analyzes relevant information from all major real-time newswires, online media and trustworthy sources to produce real-time news analytics. These analytics not only allow market participants to capture alpha opportunities in the market, but they also help to improve risk management and provide for better trading execution. Financial firms can use this data to lead portfolio allocations, which help improve the average life and profit of existing trading models. They can use RavenPack's data to enhance risk-adjusted returns from trading or investing, to manage event risk when investing or market-making, to assist compliance and surveillance analysts, and to encourage trading activity.

## Key Benefits

RavenPack News Analytics can protect portfolio managers or traders from the consequences of missing important news that has an impact on their position or portfolio. Also, news events on natural disasters, economic indicators, earnings, product recalls, layoffs, stock or credit ratings, and many others can be precursors to changes in volatility of securities. RavenPack News Analytics enables traders to get an edge by acting in advance of these changes. [1]

Traders also boost their gains using news-based algorithms to speed their response time to breaking events. They build defensive applications to ensure that key news events are factored in and use RavenPack's automated news analysis as a more effective form of low latency decision support. Firms reduce the time required for low-frequency fundamental traders to assess their options manually and execute their responses more effectively. RavenPack News Analytics can also help in post-trade analysis to explain why a traditional algorithm or strategy did not work.

With RavenPack, it is now possible to measure and instantly incorporate business, economic, and geopolitical events that are difficult to predict and often destabilizing. Wars, elections, terrorist attacks, airspace closures, or volcanic ash clouds – RavenPack monitors every event, all the time, so you're never caught off guard.

[1] As demonstrated in many studies and research reports listed in Appendix E: Links of Interest on News-Based Trading

# Key Information

RavenPack News Analytics delivers sentiment analysis and event data most likely to impact financial markets and trading around the world. The service includes analytics on nearly 200,000 entities in over 200 countries and covers over 98% of the investable global market. All relevant news items about entities are classified and quantified according to their sentiment, relevance, topic, novelty, and market impact; the result is a data product that can be segmented into many distinct benchmarks and used in a variety of applications.

| Entity Type Coverage | |
|---|---|
| Places: | 138,000+ |
| Companies: | 40,000+ |
| Organizations: | 3,000+ |
| Currencies: | 150+ |
| Commodities: | 80+ |

| Equity Entity Coverage by Region | |
|---|---|
| Americas: | 47.2% |
| Asia: | 25.5% |
| Europe: | 22.4% |
| Oceania: | 4.3% |
| Africa: | 0.6% |

For the most updated list of supported entities download the Entity Mapping file which can be accessed via RavenPack's Manuals and Overviews section of the WRDS portal.

# Datasets

RavenPack News Analytics is comprised of several packages or editions as described below:

### Dow Jones Edition

RavenPack News Analytics – Dow Jones Edition analyzes relevant information from Dow Jones Newswires, regional editions of the Wall Street Journal, Barron's and MarketWatch. With more than 5,000 employees around the world, including more than 2,000 journalists in 58 countries, Dow Jones publishes the world's best business and financial news.

| | |
|---|---|
| Source: | Dow Jones Financial Wires, Wall Street Journal, Barron's and MarketWatch |
| Archive Range: | Since Jan 1, 2000 |

### Web Edition

RavenPack News Analytics – Web Edition automatically processes hundreds of thousands of articles a day from leading publishers and web aggregators. Over 19,000 sources are continuously monitored including industry and business publishers, national and local news, blog sites, government and regulatory updates.

| | |
|---|---|
| Source: | Business publishers, national and local news, blog sites, government and regulatory updates. |
| Archive Range: | Since Jan 1, 2007 |

### PR Edition

RavenPack News Analytics – PR Edition analyzes news and information from the leading global media organizations. More than 100,000 press releases and regulatory disclosures are processed on a daily basis from a variety of newswires and press release distribution networks, including exclusive content from PRNewswire, Canadian News Wire, LSE Regulatory News Service, and others.

Source: Press releases, regulatory, corporate and news services.
Archive Range: Since Jan 1, 2004

### Full Edition

RavenPack also provides a dataset that is composed of Dow Jones Edition, Web Edition and PR Edition. This combined data set makes loading and using the data simpler for people that wish to subscribe to all RavenPack products. The content is the union of Dow Jones Edition, Web Edition, and PR Edition and is separated into Equities and Global Macro packages.

Source: Combines the Dow Jones Edition, Web Edition, and PR Edition into one dataset
Archive Range: Since Jan 1, 2000

## Data File Contents

RavenPack News Analytics is updated monthly with new data available to WRDS users on or shortly after the 5th of each month. The RavenPack News Analytics data is split into two different packages. The first contains equity (company) related analytics, and the second one contains global macro analytics. Any given file downloaded from the WRDS platform in .csv format will conform to a typical relational model; each row has an RP_STORY_ID which is a key for all entity records derived from the same news story. These files may be loaded into a relational database as separate tables with a foreign key relationship on the RP_STORY_ID and RP_ENTITY_ID.

Whenever an entity such as a company or currency is mentioned in the news, RavenPack produces an entity-level record. A single news story can yield multiple records if more than one entity is mentioned.

### Global Macro News Analytics

Each record will contain up to 34 fields depending on the variables selected including a timestamp, reference identifiers, scores for relevance, novelty and sentiment, and unique identifiers for each news story analyzed. In the historical data files, each row in the file represents an entity-level record. Here is one example:

Headline: **Malaysian Passenger Plane 'Crashes in Ukraine'**
Published Date: 2014-07-17 15:12:36.957

| | |
|---|---|
| **TIMESTAMP_UTC:**<br>**RPNA_DATE_UTC:**<br>**RPNA_TIME_UTC:** | 2014-07-17 15:12:36.957<br>2014-07-17<br>15:12:36.957 |
| **ENTITY_NAME:** | Ukraine (UA) |
| **ENTITY_TYPE:** | PLCE (Place) |
| **RP_ENTITY_ID:** | 573B99 |
| **POSITION_NAME:** | |
| **RP_POSITION_ID:** | |
| **COUNTRY_CODE:** | UA |
| **RELEVANCE:** | 100 |
| **TOPIC:** | business |

| | |
|---|---|
| **GROUP:** | industrial-accidents |
| **TYPE:** | aircraft-accident |
| **SUB_TYPE:** | |
| **PROPERTY:** | |
| **EVALUATION_METHOD:** | |
| **MATURITY:** | |
| **CATEGORY:** | aircraft-accident |
| **EVENT SENTIMENT (ESS):** | 29 |
| **AGGREGATE EVENT SENTIMENT (AES):** | 25 |
| **AGGREGATE EVENT VOLUME (AEV):** | 12 |
| **SIMILARITY KEY (EVENT_SIMILIARITY_KEY):** | A93EBE321D0E45A39346… |
| **NOVELTY (ENS):** | 100 |
| **NOVELTY SIMILIARITY GAP (ENS_SIMILARITY_GAP):** | 100 |
| **NOVELTY ID (ENS_KEY):** | 27B1DAC300588626280FA… |
| **NOVELTY ELAPSED TIME (ENS_ELAPSED):** | 0 |
| **GLOBAL NOVELTY (G_ENS):** | 100 |
| **GLOBAL NOVELTY SIMILARITY GAP (G_ENS_SIMILARITY_GAP):** | 100 |
| **GLOBAL NOVELTY ID (G_ENS_KEY):** | 27B1DAC300588626280FA… |
| **GLOBAL NOVELTY ELAPSED TIME (G_ENS_ELAPSED):** | 0 |
| **NEWS_TYPE:** | FULL-ARTICLE |
| **SOURCE:** | B5569E (DJ Newswires) |
| **PRODUCT_KEY:** | DJ-GM |
| **RP_STORY_ID:** | 27B1DAC300588626280FA … |
| **RP_STORY_EVENT_INDEX:** | 1 |
| **RP_STORY_EVENT_COUNT:** | 2 |

## Equity News Analytics

Each record contains up to 46 fields depending on the variables selected including a timestamp, company identifiers, scores for relevance, novelty and sentiment, and unique identifiers for each news story analyzed. Here is an example:

Headline: **Malaysia Airlines passenger plane crashes in Ukraine**
Published Date: 2014-07-17 15:28:28.897

| | |
|---|---|
| **TIMESTAMP_UTC:**<br>**RPNA_DATE_UTC:**<br>**RPNA_TIME_UTC:** | 2014-07-17 15:28:28.897<br>2014-07-17<br>15:28:28.897 |
| **ENTITY_NAME:** | Malaysian Airline System Bhd |
| **ENTITY_TYPE:** | COMP (Company) |
| **RP_ENTITY_ID:** | 06FACB |
| **POSITION_NAME:** | |
| **RP_POSITION_ID:** | |
| **COUNTRY_CODE:** | MY |
| **ISIN:** | MYL3786OO000 |
| **COMPANY:** | MY/3786 |

| | |
|---|---|
| **RELEVANCE:** | 100 |
| **TOPIC:** | business |
| **GROUP:** | industrial-accidents |
| **TYPE:** | aircraft-accident |
| **SUB_TYPE:** | |
| **PROPERTY:** | |
| **EVALUATION_METHOD:** | |
| **MATURITY:** | |
| **CATEGORY:** | aircraft-accident |
| **EVENT SENTIMENT (ESS):** | 29 |
| **AGGREGATE EVENT SENTIMENT (AES):** | 14 |
| **AGGREGATE EVENT VOLUME (AEV):** | 81 |
| **SIMILARITY KEY (EVENT_SIMILARITY_KEY):** | 46E95CC959105F15977152… |
| **NOVELTY (ENS):** | 100 |
| **NOVELTY SIMILIARITY GAP (ENS_SIMILARITY_GAP):** | 100 |
| **NOVELTY ID (ENS_KEY):** | 4A460C7A48522B666495790… |
| **NOVELTY ELAPSED TIME (ENS_ELAPSED):** | 0 |
| **GLOBAL NOVELTY (G_ENS):** | 100 |
| **GLOBAL NOVELTY SIMILARITY GAP (G_ENS_SIMILARITY_GAP):** | 100 |
| **GLOBAL NOVELTY ID (G_ENS_KEY):** | 4A460C7A48522B666495790… |
| **GLOBAL NOVELTY ELAPSED TIME (G_ENS_ELAPSED):** | 0 |
| **COMPOSITE SENTIMENT (CSS):** | 50 |
| **NEWS IMPACT PROJECTION (NIP):** | 54 |
| **GLOBAL EQUITIES (PEQ):** | 50 |
| **EARNINGS EVALUATIONS (BEE):** | 50 |
| **EDITORIALS & COMMENTARY (BMQ):** | 50 |
| **VENTURE, COMPANY, M&A (BAM):** | 50 |
| **REPORTS CORP. ACTIONS (BCA):** | 50 |
| **EARNINGS RELEASES (BER):** | 50 |
| **ANALYST CHANGES (ANL_CHG):** | 50 |
| **MULTI CLASSIFIER FOR EQUITIES (MCQ):** | 50 |
| **NEWS_TYPE:** | NEWS-FLASH |
| **SOURCE:** | BE528D (Daily Express) |
| **PRODUCT_KEY:** | WE-EQ |
| **RP_STORY_ID:** | 4A460C7A48522B666495790… |
| **RP_STORY_EVENT_INDEX:** | 1 |
| **RP_STORY_EVENT_COUNT:** | 1 |

# Data Field Descriptions

**TIMESTAMP_UTC**
The Date/Time (YYYY-MM-DD hh:mm:ss.sss) at which the news item was received by RavenPack servers in Coordinated Universal Time (UTC).

**RPNA_DATE_UTC**
The Date (YYYY-MM-DD) at which the news item was received by RavenPack servers in Coordinated Universal Time (UTC).

**RPNA_TIME_UTC**
The Time (hh:mm:ss.sss) at which the news item was received by RavenPack servers in Coordinated Universal Time (UTC).

**RP_ENTITY_ID**
A unique and permanent entity identifier assigned by RavenPack. Every entity tracked is assigned a unique identifier comprised of 6 alphanumeric characters.

The RP_ENTITY_ID field consistently identifies entities throughout the historical archive. RavenPack's entity detection algorithms find only references to entities by information that is accurate at the time of story publication (point-in-time sensitive). A full list of RP_ENTITY_IDs is available in the Entity Mapping file.

**ENTITY_TYPE**
The type of entity associated with a particular RP_ENTITY_ID. Currently RavenPack supports the following 5 entity types:

1. **COMP** (Company): Business organization that may be traded directly on an exchange.
2. **ORGA** (Organization): Non-business organization such as a government, central bank, not-for-profit, terrorist organization, etc.
3. **CURR** (Currency): Currencies of all financial and industrial countries.
4. **CMDT** (Commodity): Exchange traded commodities such as crude oil and soy.
5. **PLCE** (Place): Towns, cities and countries.

**ENTITY_NAME**
The official canonical name of the entity identified by the RP_ENTITY_ID.

**POSITION_NAME**
The position held by an individual within the entity involved in a specific news event. A full list of POSITION_NAMEs is available in the Entity Mapping file.

**RP_POSITION_ID**
A unique and permanent identifier for positions assigned by RavenPack. Every position tracked is assigned a unique entity identifier comprised of 6 alphanumeric characters. A full list of RP_POSITION_IDs is available in the Entity Mapping file.

**COUNTRY_CODE**
The two character ISO-3166 country code associated with an entity. Companies and organizations are associated with the country of incorporation, currencies are associated with the country where the central bank resides, and commodities are global and not associated with specific countries, so their COUNTRY_CODE label is 'XX'.

**RELEVANCE**
A score between 0-100 that indicates how strongly related the entity is to the underlying news story, with higher values indicating greater relevance. For any news story that mentions an entity, RavenPack provides a relevance score. A score of 0 means the entity was passively mentioned while a score of 100 means the entity was prominent in the news story. Values above 75 are considered significantly relevant.

Specifically, a value of 100 indicates that the entity identified plays a key role in the news story and is considered highly relevant. RavenPack's analysis is not limited to keywords or mentions when calculating relevance. Automated classifiers look for meaning by detecting the roles entities play in specific events like acquisitions or legal disputes or when announcing corporate actions, executive changes, product launches or recalls, among many other categories. An entity will be assigned a high mark of 100 if it plays a main role in these types of stories (context-aware).

If an entity is referenced in the headline or story body, it will receive a value between 0 and 99 (context-unaware). The score is assigned by a proprietary text positioning algorithm based on where the entity is first mentioned (i.e. headline, first paragraph, second paragraph, etc.), the number of references in the text, and the overall number of entities mentioned in the story. Usually, a relevance value of at least 90 indicates that the entity is referenced in the main title or headline of the news item, while lower values indicate references further in the story body.

For example, a news story about IBM where the company is referenced in the headline of the story receives a minimum value of 90. If the headline read "*IBM In Software Pact With Raytheon Unit For Navy Program*" then IBM and Raytheon would receive a relevance score of 100 since they both play a key role in the story.

If a headline reads "Bank of Spain: Data Points To 2Q GDP Contraction", the system automatically infers this story is about the country "Spain". Since this story would match the event category "gdp-guidance-down" designed to match a country, the entity "Spain" would receive a relevance score of 100 and the entity "Bank of Spain" a score of 90 or above.

If an entity is detected in a so-called "low-relevance" role, then it automatically gets a score of 20. For example, a brokerage or analyst firm making a recommendation on a company's stock (i.e. upgrade or downgrade) plays a low-relevance role and therefore receives a default relevance score of 20.

If an entity is identified in a "source" role, then it's given a lower score of 10. A source may be a publisher, data provider, or firm that authored, originated, or is referenced in the story. However, if an entity is identified in a source role but also detected as a non-source role within the story, then the source role is disregarded (for the purpose of computing relevance), and it's treated the same as any other entity described above.

Entities not detected as explicitly mentioned in a story are not given a relevance score. While a story about Yahoo! might be considered in some other context to be relevant to Google, the company Google (US/GOOG) will not be given a relevance score unless that story explicitly mentions Google.

The classifier detecting entities has access to information about each entity including short-names, long names, abbreviations, securities identifiers, subsidiaries information, and up-to-date corporate actions data. This allows for "point-in-time" detection of entities in the text.

A news story relevant to multiple entities generates scores for each entity in separate "entity-level" records, each with their own relevance score.

## RavenPack Taxonomy Fields

Relevant stories about entities are classified into a set of predefined event categories following the RavenPack taxonomy.

RavenPack automatically detects key news events and identifies the role played by the entity. Both the topic and the entity's role in the news story are tagged and categorized. For example, in a news story with the headline "IBM Completes Acquisition of Telelogic AB" the category field includes the tag "*acquisition-acquirer*" (since IBM is involved in an acquisition and is the acquirer company). Telelogic would receive the category tag "*acquisition-acquiree*" in its corresponding record since the company is also involved in the acquisition but as the acquired company.

Similarly, a story published as "Xerox Sues Google over Search-Query Patents" is categorized as a *patent-infringement*. Xerox receives the tag *patent-infringement-plaintiff* while Google gets *patent-infringement-defendant*.

Typically, an entity linked to an event category given its role receives a RELEVANCE score of 100. When playing a low relevance role in an event (ex. a rater), an entity receives a relevance score of 20.

There are over 2,000 types of event categories automatically detected by RavenPack. A full list is available in the "RavenPack Event Taxonomy" file.

**TOPIC**
A subject or theme of events detected by RavenPack. The highest level of the RavenPack Event Taxonomy.

**GROUP**
A collection of related events. The second highest level of the RavenPack Event Taxonomy.

**TYPE**
A class of events, the constituents of which share similar characteristics.

**SUB_TYPE**
A subdivision of a particular class of events.

**PROPERTY**
A named attribute of an event such as an entity, role, or string extracted from a matched event type. When applicable, the role played by the entity in the story is detected and tagged.

**EVALUATION_METHOD**
A period of time used to measure changes from previous levels in an event. Currently RavenPack supports the following 3 evaluation methods as named attributes:

1. **YOY**: Year-over-Year change
2. **QOQ**: Quarter-over-Quarter change
3. **MOM**: Month-over-Month change

**MATURITY**
For events related to debt, this named attribute indicates the period of time for which a financial instrument remains outstanding. The time period is represented by the following formats:

1. **{1-7}-DAY**: Maturity in days. Prefix is a number from 1 to 7, e.g. 2-DAY
2. **{1-4}-WK**: Maturity in weeks. Prefix is a number from 1 to 4, e.g. 4-WK
3. **{1-12}-MTH**: Maturity in months. Prefix is a number from 1 to 12, e.g. 9-MTH
4. **{1-4}-Q**: Maturity in quarters. Prefix is a number from 1 to 4, e.g. 1-Q

5. **{1-50}-YR**: Maturity in years. Prefix is a number from 1 to 50, e.g. 40-YR

**CATEGORY**
A unique tag to label, identify, and recognize a particular type and property of an entity-specific news event.

**ESS – EVENT SENTIMENT SCORE**
A granular score between 0 and 100 that represents the news sentiment for a given entity by measuring various proxies sampled from the news. The score is determined by systematically matching stories typically categorized by financial experts as having short-term positive or negative financial or economic impact. The strength of the score is derived from a collection of surveys where financial experts rated entity-specific events as conveying positive or negative sentiment and to what degree. Their ratings are encapsulated in an algorithm that generates a score ranging from 0-100 where 50 indicates neutral sentiment, values above 50 indicate positive sentiment and values below 50 show negative sentiment.

ESS probes many different sentiment proxies typically reported in financial news and categorized by RavenPack. The algorithm produces a score for more than 2,000 types of business, economic, and geopolitical events ranging from earnings announcements to terrorist attacks. The score is determined by systematically detecting entities and the roles played by that those entities in a story using RavenPack's proprietary technology and extensive database of time sensitive information about entities. The algorithms then can dynamically assign an ESS score based on score ranges assigned by the experts and by performing analysis and computation when factors such as magnitudes, comparative values or ratings are disclosed in the story.

For example, the algorithm is capable of interpreting actual figures, estimates, ratings, revisions, magnitudes, and recommendations disclosed in news stories. It can compare actual vs. estimated figures about earnings, revenues or dividends and produce an ESS score based on the comparisons. It calculates percentage differences between financial figures and identifies and interprets stock and credit ratings disclosed by analysts. The ESS algorithms can factor information such as the Richter scale in the case of an earthquake or the number of casualties in a suicide bombing event. The use of emotionally charged language by authors is also factored when shaping the strength component of the ESS.

The ESS algorithm has embedded information on rating scales from all major brokerage firms, investment banks, and credit rating agencies. It uses this information to differentiate and assess the various actions taken by analysts. For example, the algorithm generates a lower (more negative) ESS score for stories about an analyst downgrade from a "Strong Buy to a Strong Sell" than from a "Buy to a Neutral". In the case of stories about financial results or economic indicators, it computes the percentage change between the disclosed actual figures vs. the street consensus or any other benchmarks disclosed in the story. For example, a company beating earnings by 70% will receive a higher (more positive) ESS score than a company exceeding a benchmark by 1%.

ESS leverages RavenPack's event detection technology and produces an entity specific sentiment score every time an event category is matched. See appendix D for more information on Factors in the Event Sentiment Score.

**AES – AGGREGATE EVENT SENTIMENT**
A granular score between 0 and 100 that represents the ratio of positive events reported on an entity compared to the total count of events (excluding neutral ones) measured over a rolling 91-day window in a particular package (Dow Jones, Web or PR Editions). Only news items that match a RavenPack event category receiving an ESS score are included in the computation of AES. An event with ESS>50 is counted as a positive entry whereas one with ESS<50 is counted as a negative entry. Events with ESS=50 are considered neutral and excluded from the computation.

Events matching “Order Imbalance” and “Insider Trading” categories are filtered out as these tend to add noise given their lack of sentiment, high volume, and frequency. An AES score is published every time an entity is mentioned in the news. Changes in the AES score, however, are observed only when an event category is matched or when one drops out of the 91-day window calculation. AES leverages the RavenPack Taxonomy and is based on RavenPack's Expert Consensus methodology.

**AEV – AGGREGATE EVENT VOLUME**

A value that represents the count of events for an entity (excluding neutral ones) measured over a rolling 91-day window in a particular package (Dow Jones, Web, or PR Editions). Only news items that match a RavenPack event category receiving an ESS score are included in the computation of AEV. Both events with an ESS score above and below 50 are counted by AEV, effectively signaling the volume of highly relevant news on the entity over the past 91 days. Events with ESS=50 are considered neutral and excluded from the computation.

Events matching “Order Imbalance” and “Insider Trading” categories are filtered out as these tend to add noise given their lack of sentiment, high volume, and frequency. An AEV value is published every time an entity is mentioned in the news. Changes in the AEV value, however, are observed only when a new event category is matched or when it drops out of the 91-day window calculation. AEV leverages the RavenPack Taxonomy and is based on RavenPack's Expert Consensus methodology.

**ENS – EVENT NOVELTY SCORE**

A score between 0 and 100 that represents how "new" or novel a news story is within a 24-hour time window across all news stories in a particular package (Dow Jones, Web or PR Editions). Any two stories that match the same event for the same entities will be considered similar according to ENS. The first story reporting a categorized event about one or more entities is considered to be the most novel and receives a score of 100. Subsequent stories from the same package about the same event for the same entities receive scores following a decay function whose values are (100 75 56 42 32 24 18 13 10 8 6 4 3 2 2 1 1 1 1 0 ...) based on the number of stories in the past 24-hour window. If a news story is published more than 24 hours after any other similar story, it will again be considered novel and start a separate chain with a score of 100.

Note that for any particular story, the ENS score is based on the number of similar stories in the most recent 24-hour window preceding that story. However, a chain of similar stories can span more than 24 hours, provided no two similar stories are more than 24 hours apart. Occasionally, the ENS score of a story which arrives more than 24 hours after the first story in the chain can be equal to or greater than the ENS score of some story earlier in the chain.

Consider the following simplified example:

| Story | Timestamp | ENS | ENS_KEY |
|---|---|---|---|
| **AAA** | 2011-02-14 08:00:00 | 100 | AAA |
| **BBB** | 2011-02-14 09:00:00 | 75 | AAA |
| **CCC** | 2011-02-14 20:00:00 | 56 | AAA |
| **DDD** | 2011-02-15 10:00:00 | 75 | AAA |

Notice that no two consecutive stories are more than 24 hours apart, so all the stories are part of the same chain. Story DDD gets a score of 75 because in the 24-hour window preceding DDD, there are only two stories: CCC and DDD itself. Therefore, DDD is the second story whose count maps to a score of 75 according to the decay function (Count 1=100, 2=75, 3=56, 4= 42 …).

**ENS_SIMILARITY_GAP**
The number of days since a similar story was detected in this RPNA edition (Dow Jones Edition, Web Edition, or PR Edition). Values range between 0.00000 and 100.00000 inclusive. The value 100.00000 means that the most recent similar story occurred 100 or more days in the past. The value 0.00000 means a similar story exists with the exact same timestamp.

**ENS_KEY – EVENT NOVELTY KEY**
An alphanumeric identifier that provides a way to chain or relate stories about the same categorized event for the same entities. The ENS_KEY corresponds to the RP_STORY_ID of the first news story in the sequence of similar events. The identifier allows a user to track similar stories reporting on the same event about the same entities.

As with ENS, for two stories to receive the same ENS_KEY they must be published within 24 hours of one another. However, the overall time range of a chain may extend beyond 24 hours.

**ENS_ELAPSED – EVENT NOVELTY ELAPSED TIME**
The number of milliseconds between the first story and the current story in an event novelty chain. The first story in a chain will always be given a value of zero milliseconds (ENS_ELAPSED = 0). Subsequent stories in the same event novelty chain will receive higher values indicating the number of milliseconds elapsed since the first news story reporting the event. ENS_ELAPSED is only based on news stories from a particular package (Dow Jones Edition, Web Edition, or PR Edition).

**G_ENS – GLOBAL EVENT NOVELTY SCORE**
A score between 0 and 100 that represents how "new" or novel a news story is within a 24-hour time window across all news providers covered by RavenPack. Any two stories that match the same event for the same entities will be considered similar according to G_ENS. The first story reporting a categorized event about one or more entities is considered to be the most novel and receives a score of 100. Subsequent stories from any news provider covered by RavenPack about the same event for the same entities receive scores following a decay function whose values are (100 75 56 42 32 24 18 13 10 8 6 4 3 2 2 1 1 1 1 0 ...) based on the number of stories in the past 24-hour window. If a news story is published more than 24 hours after any other similar story, it will again be considered novel and start a separate chain with a score of 100.

Note that for any particular story, the G_ENS score is based on the number of similar stories in the most recent 24-hour window preceding that story across all news providers covered by RavenPack. A chain of similar stories can span more than 24 hours, provided no two similar stories are more than 24 hours apart. Occasionally, the G_ENS score of a story which arrives more than 24 hours after the first story in the chain can be equal to or greater than the ENS score of some story earlier in the chain. Consider the following simplified example:

| Story | Timestamp | G_ENS | G_ENS_KEY |
|---|---|---|---|
| **AAA** | 2011-02-14 08:00:00 | 100 | AAA |
| **BBB** | 2011-02-14 09:00:00 | 75 | AAA |
| **CCC** | 2011-02-14 20:00:00 | 56 | AAA |
| **DDD** | 2011-02-15 10:00:00 | 75 | AAA |

While an event in a particular package (Dow Jones, Web, or PR Editions) can have an ENS score of 100, it is possible that the G_ENS score is lower indicating that some source from another package (Dow Jones, Web, or PR Editions) reported this news event earlier. G_ENS can be helpful to determine whether a news event was first published within the package or by some other source covered in a different RavenPack subscription package. For more information about other available subscription packages, contact your RavenPack account representative.

In this first example, the initial event in the chain is covered first by a source in the DJ Edition:

| Story | Timestamp | ENS | ENS_KEY | G_ENS | G_ENS_KEY |
|---|---|---|---|---|---|
| **AAA** | 2011-01-01 10:00:00 | 100 | AAA | 100 | ZZZ |
| **BBB** | 2011-01-01 11:00:00 | 75 | AAA | 75 | ZZZ |
| **CCC** | 2011-01-01 12:00:00 | 56 | AAA | 56 | ZZZ |
| **DDD** | 2011-01-01 14:00:00 | 42 | AAA | 42 | ZZZ |

In this second example, the initial event in the chain is covered first by a source in a different RavenPack subscription package and then followed by a source in the DJ Edition:

| Story | Timestamp | ENS | ENS_KEY | G_ENS | G_ENS_KEY |
|---|---|---|---|---|---|
| **AAA** | 2011-01-01 10:00:00 | 100 | AAA | 75 | ZZZ |
| **BBB** | 2011-01-01 11:00:00 | 75 | AAA | 56 | ZZZ |
| **CCC** | 2011-01-01 12:00:00 | 56 | AAA | 42 | ZZZ |
| **DDD** | 2011-01-01 14:00:00 | 42 | AAA | 32 | ZZZ |

**G_ENS_SIMILARITY_GAP**
The number of days since a similar story was detected across all product editions (Dow Jones Edition, Web Edition, and PR Edition). Values range between 0.00000 and 100.00000 inclusive. The value 100.00000 means that the most recent similar story occurred 100 or more days in the past. The value 0.00000 means a similar story exists with the exact same timestamp.

**G_ENS_KEY – GLOBAL EVENT NOVELTY KEY**
An alphanumeric identifier that provides a way to chain or relate stories about the same categorized event for the same entities across all news providers covered by RavenPack. The G_ENS_KEY corresponds to the RP_STORY_ID of the first news story in the sequence of similar events, across all news providers covered by RavenPack. The identifier allows a user to track similar stories reporting on the same event about the same entities across all news providers covered by RavenPack.

As with G_ENS, for two stories to receive the same G_ENS_KEY they must be published within 24 hours of one another. However, the overall time range of a chain may extend beyond 24 hours.

**G_ENS_ELAPSED – GLOBAL EVENT NOVELTY ELAPSED TIME**
The number of milliseconds between the first story and the current story in an event novelty chain across all news providers covered by RavenPack. The first story in a chain will always be given a value of zero milliseconds (G_ENS_ELAPSED = 0). Subsequent stories in the same event novelty chain will receive higher values indicating the number of milliseconds elapsed since the first event.

G_ENS_ELAPSED is based on news stories from all providers covered by RavenPack.

**EVENT_SIMILARITY_KEY**
A unique 32 character key that identifies similar stories in the RPNA data. All similar stories across the entire archive and those arriving on the real-time feed share the same EVENT_SIMILARITY_KEY. Stories are similar when they are categorized with the same event and entities.

**NEWS_TYPE – TYPE OF NEWS STORY**
Classifies the type of news story into one of five categories:

1. HOT-NEWS-FLASH**:** A news article composed of a headline and no body text marked as breaking news during the editorial process.

2. NEWS-FLASH: A news article composed of a headline and no body text.
3. FULL-ARTICLE: A news article composed of both a headline and one or more paragraphs of mostly textual material.
4. PRESS-RELEASE: A corporate announcement originated by an entity and distributed via a news provider.
5. TABULAR-MATERIAL: A news article composed of both a headline and one or more segments of mostly tabular data.

**SOURCE**
A unique and permanent news source identifier assigned by RavenPack. Every news provider tracked is assigned a unique identifier comprised of 6 alphanumeric characters.

This field can be linked to the "Source Mapping" file for additional details about the publication name and type, coverage dates, and trustworthiness of each provider. More details about these characteristics are available in the "Source Mapping File" section of this User Guide.

**RP_STORY_ID – RAVENPACK UNIQUE STORY IDENTIFIER**
An alphanumeric character identifier to uniquely identify each news story analyzed. This value is unique across all records. *Example: 1FB2B3F5E99C4D3BCF59FDB3E8C8C9BD*

**RP_STORY_EVENT_INDEX**
Represents the order in which entity records are published by RavenPack per news story. This integer can be equal to or less than the RP_STORY_EVENT_COUNT.

**RP_STORY_EVENT_COUNT**
Represents the total entity records published by RavenPack per news story.

**PRODUCT_KEY**
Identifies which subscription package contains the record. Its value can be one of the following:

- DJ-EQ – Dow Jones Edition – Equities
- DJ-GM – Dow Jones Edition – Global Macro
- WE-EQ – Web Edition – Equities
- WE-GM – Web Edition – Global Macro
- PR-EQ – Press Release Edition – Equities
- PR-GM – Press Release Edition – Global Macro

**COMPANY** (Equities Package Only)
This field includes a company Identifier in the format ISO_CODE / TICKER. The ISO_CODE is based on the company's original country of incorporation and TICKER on a local exchange ticker or symbol. If the company detected is a privately held company, there will be no ISO_CODE/TICKER information, only an RP_ENTITY_ID.

A full list of tracked entities including tickers and common identifiers is available for download on the Product Area.

**ISIN** (Equities Package Only)
An International Securities Identification Number (ISIN) to identify the company referenced in a story. The ISINs used are accurate at the time of story publication. Only one ISIN is used to identify a company, regardless of the number of securities traded for any particular company. The ISIN used will be the Primary ISIN for the company at the time of the story.

**CSS – COMPOSITE SENTIMENT SCORE** (Equities Package Only)
A sentiment score between 0 and 100 that represents the news sentiment of a given story by combining various sentiment analysis techniques. The direction of the score is determined by looking at emotionally charged words and phrases and by matching stories typically rated by experts as having short-term positive or negative share price impact. The strength of the score (values above or below 50, where 50 represents neutral strength) is determined from intraday stock price reactions modeled empirically using tick data from approximately 100 large cap stocks.

CSS combines 5 sentiment analytics (PEQ, BEE, BMQ, BCA, and BAM) using an intuitive set of rules while ensuring no sentiment disagreement exists amongst the analytics. One way of using CSS scores may involve rules like:

```
If CSS > 50 Then ' (Positive Signal)
' Go long
ElseIf CSS < 50 Then ' (Negative Signal)
' Go short
ElseIf CSS = 50 Then ' (Neutral Signal)
' Hold position
End If
```

CSS was trained on market data using a portfolio of large cap stocks and evaluating intraday fluctuations to determine "strength" or how positive or negative a story is. Using the "strength" aspect of this score may depend on your investment strategy and trading horizon. Typically, CSS scores hover between 40-60 so higher or lower values are assigned only in cases where confidence is high on short term signals. This score combines RavenPack's Traditional, Expert Consensus, and Market Response methodologies.

**NIP – NEWS IMPACT PROJECTIONS** (Equities Package Only)
A score taking values between 0 and 100 that represents the degree of impact a news flash has on the market over the following two-hour period. The training set for this classifier used tick data for a test set of large cap companies and looked at the relative volatility of each stock price measured in the two hours following a news flash. The relative volatility is the volatility divided by the mean of volatilities of all companies in the test set during the same period. The classifier is trained to predict whether relative volatility is high or low given the language used by journalists in news flashes, typically about corporate actions and analyst revisions.

Whether something is considered high or low depends on the time of day when the story arrived. The score is centered at 50, which represents zero impact; values above 50 indicate higher impact in terms of volatility. The more extreme the impact value, the higher the confidence of the score. Scores below 50 indicate low or unknown impact and lower confidence in the score. The best performance of the score is obtained when filtering for RELEVANCE above 90. This NIP score is based on RavenPack's Market Response Methodology.

**PEQ – GLOBAL EQUITIES** (Equities Package Only)
A score that represents the news sentiment of the given news item according to the PEQ classifier, which specializes in identifying positive and negative words and phrases in articles about global equities. Scores can take values of 0, 50, or 100 indicating negative, neutral, or positive sentiment, respectively. This sentiment score is based on RavenPack's Traditional Methodology.

**BEE** – **EARNINGS EVALUATIONS** (Equities Package Only)
A score that represents the news sentiment of the given story according to the BEE classifier, which specializes in news stories about earnings evaluations. Scores can take values of 0, 50, or 100 indicating negative, neutral, or positive sentiment, respectively. This sentiment score is based on RavenPack's Traditional Methodology.

**BMQ – EDITORIALS & COMMENTARY** (Equities Package Only)
A score that represents the news sentiment of the given story according to the BMQ classifier, which specializes in short commentary and editorials on global equity markets. Scores can take values of 0, 50, or 100 indicating negative, neutral, or positive sentiment, respectively. This sentiment score is based on RavenPack's Expert Consensus Methodology.

**BAM – VENTURE, COMPANY, MERGERS & ACQUISITIONS** (Equities Package Only)
A score that represents the news sentiment of the given story according to the BAM classifier, which specializes in news stories about mergers, acquisitions, and takeovers. Scores can take values of 0, 50, or 100 indicating negative, neutral, or positive sentiment, respectively. This sentiment score is based on RavenPack's Expert Consensus Methodology, and has been trained on stories that lead up to a pre-identified mergers, acquisitions, and takeover events.

**BCA – REPORTS ON CORPORATE ACTIONS** (Equities Package Only)
A score that represents the news sentiment of the given news story according to the BCA classifier, which specializes in reports on corporate action announcements. Scores can take values of 0, 50, or 100 indicating negative, neutral, or positive sentiment, respectively. This sentiment score is based on RavenPack's Expert Consensus Methodology and has been trained on stories that lead up to a pre-identified corporate action announcement.

**BER – EARNINGS RELEASES** *(Equities Package Only)*
A score that represents the news sentiment of the given story according to the BER classifier, which specializes in news stories about earnings releases. Scores can take values of 0, 50, or 100 indicating negative, neutral, or positive sentiment, respectively. This sentiment score is based on RavenPack's Expert Consensus Methodology.

**ANL_CHG – ANALYST RECOMMENDATIONS & CHANGES** (Equities Package Only)
A score that represents a change in recommendation by an analyst firm in the form of a numerical score. When the mention of a company in a story matches the criteria for ANL-CHG, scores can take values of 0, 50 or 100, indicating a downgrade, neutral, or upgrade rating, depending on the recommendation change by the analyst. This analytic is based on the event category results of *a) analyst-ratings-change-positive and b) analyst-ratings-change-negative*. This analytic contains values only for news stories that disclose changes in analyst recommendations.

**MCQ – MULTI CLASSIFIER FOR EQUITIES** (Equities Package Only)
A score that represents the news sentiment based on the tone; applicable only towards the most relevant companies mentioned in a story. The score is derived from a combination of analytics values produced by the BMQ, BEE, BCA, and ANL-CHG classifiers. An MCQ score is present when the relevance score for a company is 90 or higher and either there is an ANL-CHG score or all of BMQ, BEE, and BCA scores are positive (100) and neutral (50) or negative (0) and neutral (50). The logic behind this analytic is to detect consistent sentiment classifications, discarding combinations where these classifiers may have contradictory scores. When the mention of a company in a story matches the criteria for MCQ, scores can take values of 0, 50, or 100 indicating negative, neutral, or positive sentiment, respectively.

# RavenPack Reference Service

RavenPack supplies a proprietary reference data service for all entities it tracks and monitors in the media. The service indicates the links and relationships between entities in business, economic, and political structures within a geographic space. Entities include companies (stocks), organizations, geographical locations, currencies, and commodities. The service is available in RavenPack's Manuals and Overview section on the WRDS portal for RavenPack subscribers.

Entities tracked by RavenPack are keyed using the RP_ENTITY_ID and for each entity the mapping file lists identifying information such as names and securities along with the validity dates for each value. The entity mapping file allows for cross-referencing a client's proprietary universe of entities (e.g. securities) to the hundreds of thousands tracked by RavenPack over time.

The RavenPack Reference Service also provides unique relationships between entities. For example, it maps countries to membership organizations such as the European Union, the Euro-zone, G20, OPEC, OECD, and NATO or to trading blocs such as NAFTA, MERCOSUR, or The League of Arab States, among others. Relationships are point in time sensitive in that it keeps track of when entities join (or leave) any given membership. By delineating a country to a supra-organization or trading bloc for example, at any point in time, this mapping data helps expedite and simplify macroeconomic and geopolitical analysis by clients.

In addition, RavenPack maps all currencies to their corresponding countries and governments. In the case of multi-government denominated currencies such as the Euro, we map it only to the member states of the European Union that form part of the Eurozone.

The entity mapping dataset is updated daily and has the following format:

**RP_ENTITY_ID**
A permanent primary key identifier assigned by RavenPack for uniquely identifying entities over time. This should be used for linking any data in the analytics data files.

**ENTITY_TYPE**
The type of entity associated with a particular RP_ENTITY_ID. Currently RavenPack supports the following 7 entity types:

1. **COMP** (Company): Business organization that may be traded directly on an exchange.
2. **ORGA** (Organization): Non-business organization such as a government, central bank, non-for-profit, terrorist organization, etc.
3. **CURR** (Currency): Currencies of all financial and industrial countries.
4. **CMDT** (Commodity): Exchange traded commodities such as crude oil or soy.
5. **PLCE** (Place): Towns, cities, and countries
6. **ORGT** (Organization Type): A categorization of related organizations.
7. **POSI** (Position): A position held by an individual within an entity

***NOTE:*** *To provide as much details as possible about entities, new data types may appear as and when they are added to our database. We advise that any code used to parse this document should allow for these unspecified values.*

**DATA_TYPE**

Categorizes 26 types of data relevant to this version of the product:

1. **COMPANY**: A company identifier comprised of a code based on the company's country of incorporation and a local exchange ticker or symbol. This is used primarily for visual representation in the data files and will change over time.
2. **COUNTRY**: A country associated with this entity.
3. **COUNTRY_ID**: An RP_ENTITY_ID of the entity of the country associated with this entity.
4. **CUSIP**: A known CUSIP identifier for a stock of the detected company.
5. **DESCRIPTION**: A description of the entity where applicable.
6. **ENTITY_NAME**: The official canonical name for the entity as listed in the data files.
7. **GEONAME_ID**: For a place, this is the geoname from the geonames data source.
8. **GOVERNMENT**: For a place, this is the associated central government organization.
9. **HAS_MEMBER_ID**: Indicates that the RP_ENTITY_ID listed in the DATA_VALUE field is a constituent or a member of this entity.
10. **ISIN**: A known ISIN (International Securities Identifier Number) for the company.
11. **IS_MEMBER_ID**: Indicates that this entity is a constituent or member of the RP_ENTITY_ID listed in the DATA_VALUE field.
12. **ISO_CODE**: Defines the two-letter ISO country code for the country.
13. **LATITUDE:** A geographic coordinate for latitude of the place represented in decimal degrees (DD).
14. **LISTING**: A combination of the MIC and a local exchange ticker/symbol for the company. For example, the NASDAQ ticker for Apple Inc. would be represented as XNAS: APPL.
15. **LONGITUDE:** A geographic coordinate for longitude of the place represented in decimal degrees (DD).
16. **MIC**: The ISO 10383 Market Identifier Code (MIC) is a four-character code used to identify stock markets and other trading exchanges within global trading and referencing systems.

17. **NAME**: The official registered name for the entity.
18. **ORGANIZATION_TYPE**: The type of ORGANIZATION entity.
19. **ORGANIZATION_TYPE_ID**: The unique ID of the ORGANIZATION_TYPE.
20. **PARENT_ORG_ID**: Indicates that this entity is a child of the RP_ENTITY_ID listed in the DATA_VALUE field.
21. **PLACE_TYPE**: The type of PLACE. This can be one of COUNTRY, CITY, REGION_1 or REGION_2. The REGION_2 and REGION_1 fields can vary according to the regional differences within the country. For example, in the US, REGION_2 will be a county and REGION_1 a state. However, in Spain, REGION_2 will be a province and REGION_1 an autonomous region, etc.
22. **PROVINCE**: For a place, this is the region where the place is found.
23. **REGION_ID**: An RP_ENTITY_ID of the entity of the region associated with this entity.
24. **SEDOL**: A known SEDOL identifier for a stock of the detected company.
25. **SYMBOL**: For a currency, this is the associated symbol.
26. **TICKER**: The local exchange ticker or symbol for the company.

***NOTE:*** *To provide as much details as possible about entities, new data types may appear as and when they are added to our database. We advise that any code used to parse this document should allow for these unspecified values.*

**DATA_VALUE**
The value of a particular piece of information during the time indicated by the RANGE_START and RANGE_END.

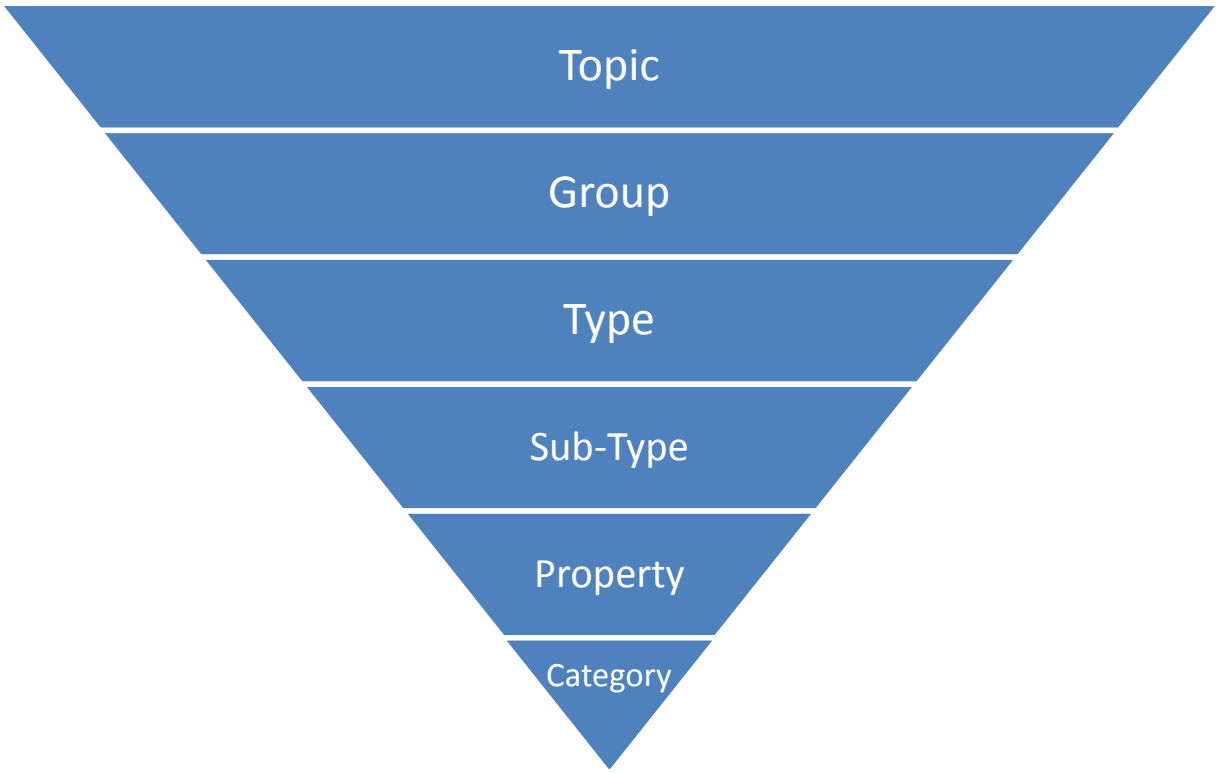
**RANGE_START**
A DATA_VALUE is valid from this date.

**RANGE_END**
A DATA_VALUE is valid up to this date. If this field is empty, it means that the value is valid at the time of generation of the entity-mapping file.

The most recent mapping file is available on the Product Area.

# RavenPack Event Taxonomy

RavenPack tracks more than 2,000 entity-specific news announcements or formal events by role reported in the news. Relevant stories about entities are classified into a set of predefined event categories following the proprietary RavenPack Taxonomy described below. When applicable, the role played by the entity in the story is also detected and tagged.




The RavenPack Event Taxonomy dataset includes the following elements:

**TOPIC**
A subject or theme of events. The highest level of the RavenPack Event Taxonomy.

**GROUP**
A collection of related events.

**TYPE**
A class of events, the constituents of which share similar characteristics.

**SUB_TYPE**
A subdivision of a particular class of events.

**PROPERTY**
A named attribute of an event such as an entity, role, number, or string extracted from a matched event type.

**CATEGORY**
A tag to label, identify, or recognize a particular type and property of an event.

**DESCRIPTION**
An explanation describing the meaning of the category tag and the context where it can match.

**SCHEDULED**
If the announcement of the event was anticipated, arranged or planned according to some schedule or timetable, then the value will be TRUE otherwise it will be FALSE.

**VALID_ENTITY_TYPES**
The ENTITY_TYPE elements, which are valid matches for that particular category.

The RavenPack Event Taxonomy can be access in RavenPack's Manual and Overviews section on the WRDS portal.

# Source Mapping File

RavenPack provides a supporting "Source" dataset of all news and social media providers with additional details such as their type of publication, coverage dates, and their level of trustworthiness and influence. The details in the source file can be cross-referenced with the “SOURCE” field in the main RavenPack News Analytics dataset.

The source mapping file is updated daily and is available on RavenPack’s Manuals and Overviews section of the WRDS portal:

**SOURCE**: A unique identifier assigned for each news source. All Dow Jones, Web, and PR Edition sources are represented as 6 alphanumeric characters.

**DATA_TYPE:** Underlying characteristics for a particular news source. Currently, RavenPack supports the following 5 data types:

1. ENTITY_NAME: The official canonical name for the news source.

2. PUBLICATION_TYPE: Identifies the general type of the news source. Its value can be one of the following:
    a. BLOG: A discussion or informational website consisting of discrete entries or postings.
    b. NEWS: A source providing up-to-the-minute news stories, financial market updates, and other newly received or noteworthy information.
    c. JOURNAL: A scholarly publication containing articles written by researchers, professors and other experts.
    OTHER: A variety of other information sources beyond blog or news providers. Sources can include company websites, government and regulatory agencies, among others.

3. RANK: A categorization of influence and trustworthiness of a news provider. The ranking is based on a range from 1 to 10 where rank 1 is the highest (i.e. most trusted source). Below are the descriptions of the 10 possible ranks:

| Rank | News Source Classification |
|---|---|
| 1 | Fully accountable, reputable, and impartial |
| 2 | Official, reliable, and honest |
| 3 | Acknowledged, formal, and credible |
| 4 | Known and reasonable credibility |
| 5 | Satisfactory credibility |
| 6 | Limited influence and sincerity |
| 7 | Biased and seemingly unreliable |
| 8 | Unknown identity, strongly biased, and unconvincing |
| 9 | Illusory and misleading |
| 10 | Unverifiable, fake, and fabricating |

4. VALIDITY: Indicates the period of time where a news source is actively providing coverage. The earliest possible coverage date for any Web Edition source is 2007-01-01, for any Dow Jones source is 2000-01-01 and any PR Edition source is 2004-01-01

5. PARENT_SOURCE: Indicates that this source is a child of the SOURCE listed in the DATA_VALUE field.

6. PRODUCT: Identifies which subscription package contains the record. Its value can be one of the following:
    - DJ - Dow Jones Edition
    - WE - Web Edition
    - PR - PR Edition

***Note:*** *To provide as much details as possible about sources, new data types may appear as and when they are added to our database. We advise that any code used to parse this document should allow for these unspecified values.*

**DATA_VALUE:** The value of a particular piece of information during the time indicated by the RANGE_START and RANGE_END.

**RANGE_START:** A DATA_VALUE is valid from this date. If this field is empty, it means that the value was valid from at least the start of the archive forward.

**RANGE_END:** A DATA_VALUE is valid up to this date. If this field is empty, it means that the value is valid at the time of generation of the source file.

# Getting Started with RavenPack News Analytics

This section outlines a quick five-step process to help you streamline the evaluation and testing of RavenPack's News Analytics (RPNA) data:

1. **Find relevant entities**
    - RavenPack provides a unique and permanent RP_Entity_ID for each tracked entity.
    - The 'Code Lookup' tool on the query builder pages for all Equities and Global Macro editions provides an easy way to find the RP_ENTITY_IDs for the required entities. Users can search company name as well as other identifiers, allowing for quick lookup of RP_ENTITY_IDs.
    - For more complex or programmatic mappings, we recommend that users take advantage of the entity mapping file builder. Map your securities to relevant RP_Entity_IDs using the following procedures:
        - Equities: Filter for the ISIN, CUSIP, LISTING values in the DATA_TYPE field and use the corresponding identifiers located in the DATA_VALUE field.
        - Global Macro: Find entities that may affect the price of your assets. Examples include companies and countries that produce or export raw materials, government agencies that make regulatory changes, central banks, currencies etc. In addition, use the "HAS_MEMBER_ID", "IS_MEMBER_ID" and "PARENT_ORG_ID" fields to find further related entities.
    - Once a given entity has been successfully mapped, you should then use your own security data to find which securities are (or were) available for trading at any point in time.

2. **Download the data from the WRDS portal or access via your preferred WRDS API**
   We strongly recommend downloading data only over the exact time period required in order to make the file sizes as manageable as possible. The shorter the time frame requested, the faster the file will be prepared and the easier it will be to work with in your analysis platform of choice.

3. **Browse our Event Taxonomy**
   The taxonomy may be used to filter for the events that play an important role in your investment decisions. Refer to the *RavenPack Taxonomy Fields* section in this User Guide for data field descriptions.

4. **Read our "Quick Guide to Trading and Investment Applications Using News Analytics"** *(you will need to request access to this from your account manager)***:**
   https://ravenpack.com/secure/newsanalytics/doc/api/papers/introduction/20111215/RavenPack%20-%20Quick%20Guide%20to%20Trading%20and%20Investment%20Applications.pdf

5. **Find inspiration in our extensive library of research and case-studies** *(you will need to request access to this from your account manager)***:**
   https://ravenpack.com/secure/newsanalytics/doc/api/resources.html

# How and Where Are News Analytics Produced?

People have been "manually" extracting patterns from news for many years, but the increasing volume of news in modern times has called for more automated approaches. In RavenPack News Analytics, news is received and processed through an automated real-time data stream made up of several software components, some running in parallel. Once content analysis and sentiment calculations have been made, information about the news is immediately sent on to subscribers as real-time events, which they use to drive trading models and applications.

RavenPack maintains two data centers to host and maintain the service. A primary center is located in the US (Secaucus, NJ) and a secondary back-up center in RavenPack's redundant facility. A data center consists of a set of racks equipped with dozens of multiprocessor servers, storage area networks, power management, and automated backup systems all located in a secure facility with 24/7 monitoring and maintenance. Communication with news publishers and clients is made over dedicated fiber lines or a Level-3 Internet backbone connection.

The following sections describe the process involved in producing news analytics including details on RavenPack's proprietary sentiment analysis techniques.

## News Sentiment Analysis Techniques

The news analytics or metrics represented in the data are based on three proprietary methodologies for identifying sentiment and market impact: Traditional, Expert Consensus, and Market Response.

The Traditional Tagging methodology underpins the PEQ and BEE scores and is based on a Rule Base that maps key words, phrases, combinations, and other word-level definitions to pre-defined sentiment values (positive, negative, or neutral). Each story may match several of these rules. The Rule Base contains more than 12,000 of these combinations, specifically designed to match in financial news articles. Two Rule Bases are defined: one that identifies positive sentiment and another that identifies negative sentiment. A news item that neither matches the positive nor negative sentiment rule bases is classified as neutral.

The Expert Consensus methodology underpins the ESS, AES, AEV, BMQ, BCA, BER and BAM scores and entails training classification algorithms on the results of financial experts manually tagging stories. Financial experts are given large sets of news articles to tag as likely to have a positive, negative, or neutral effect in the stock price of a given company in the hours ahead. The manually constructed training sets are used as the basis for automated computer classification including using a proprietary classifier such as a Bayes network.

The Market Response methodology underpins the NIP score and measures the degree of impact a news item has on the market over the following two hour period. The classifier was trained on several years of news archives related to a set of global companies. For each news item about a company, the relative volatility of its stock price was measured in the hours following story arrival. The relative volatility is the volatility divided by the mean of volatilities of all companies during the same period. The classifier is trained to predict whether relative volatility is high or low given key information in the story. Whether something is considered high or low depends on the company in question and also on the time of day when the story arrived. The classifier is trained based on how markets tend to respond to news without human judgment or input.

The CSS score represents the news sentiment of a given story by combining the three sentiment analysis techniques described above. The direction of the score is determined by combining the Traditional Tagging and Expert Consensus. The strength of the score (values above or below 50, which represents neutral strength) is determined by Market Response through empirical investigation of abnormal returns over a finite set of stocks.

See **Appendices A-D** for more detailed information on the construction of RavenPack's news sentiment classifiers.

## Entity Event and Topic Detection

When dealing with the tens of thousands of stories published about entities every day, it makes sense to try to classify them into a set of predefined categories. RavenPack has approached this problem by applying technological advancements acquired through many years of experience and came up with a solution to categorize stories into a simple set of themes which are fundamental to today's investment environment.

The technique of producing this kind of analysis came about by performing a careful study of the types of stories available on entities and by extracting the primary categories that would allow meaningful interpretation of a story. Once the categories had been determined, the goal was to implement technology that could perform the classification automatically. Some categories are more straightforward than others, so different techniques were applied.

Events as part of the RavenPack's Taxonomy are defined using thousands of proprietary template programs and Part-of-Speech tagging. Templates are compositions of language tokens or values taken in a specific context. Tokens may be a type of language marker, such as a number or date. They may be words or phrases, perhaps broken down to their root form or taken only for a given tense.

Part-of-speech tagging involves marking up the words in a text corpus as corresponding to a particular part of speech, based on both its definition, as well as its context - i.e. relationship with adjacent and related words in a phrase, sentence, or paragraph. This makes templates more scalable, modular, and effective.

For example, one template for the event category *bankruptcy* is defined as:

> **Citadel Broadcasting Files for Bankruptcy in New York**
> ($COMPANY %FILE FOR %BANKRUPTCY IN $PLACE)

Likewise, a template for the event category *sec-investigation* might be defined as:

> **Xerox Sues Google, Yahoo over Search-Query Patents –Report**
> ($ENTITY %SUE $ENTITY $ENTITY %PREPOSITION %PATENT-TYPE)

The classification is broken up into different steps:

1. **Identification of entities mentioned in the story:** This involves applying RavenPack's technology to the full story text, taking into account any available metadata provided with the story.

2. **Extraction of the story theme:** In order to be more precise, this is limited to the headline although using hints from other parts of the story including metadata. The headline is broken up into its component tokens, each token being analyzed to find out what the exact contextual meaning is. Once the components have been broken apart, labeling all appropriate information, the headline is compared to sets of templates, each template placing it into a theme. A story can have more than one theme. This approach is extensible to the sentence level.

3. **Role Detection:** Next, the context in which the entities are mentioned is analyzed to determine what role they played in the story. Events can have more than one participant with only one being principal.

4. **Determination of Categories:** If the same story carries the first mention of a particular theme, then the story is deemed to be an event for that category, otherwise it is deemed to be chatter and excluded. This minimizes the noise caused by duplicate stories.

A full list of categories and definitions that form part of the RavenPack's Taxonomy is available on the Product Area.

## Automated News Story Analysis

Many of the RavenPack's classifiers are effectively source-independent and will work without problems on a textual story retrieved from any publisher service or website. Others are trained or configured for one particular source and need to be customized in order to get the best results. Some classifiers are designed on particular topics or themes and these will target specific sources only. Classifiers are RavenPack's proprietary natural language processing software deployed across multiple high-speed servers in virtual machines.

The result of automated news story analysis is an aggregated data feed comprised entirely of structured news information, enriched with RavenPack's metadata including tags on entities, events, entity roles, novelty, sentiment, and other analytics - ready for consumption by all types of applications.

## Process Latency

The Latency of the RavenPack News Analytics Process is defined as the average time it takes a story to get through the system from the time RavenPack receives it to the time at which it leaves the Event Distribution Server at the RavenPack's Data Center.

RavenPack has calculated the average overall latency of a processed news story at any time as 250ms. Approximate, cumulative latency figures below:

**T 0ms** : Story arrives in collector and the text story is parsed
**T 50ms** : Story arrives in classifiers and processed in parallel
**T 200ms** : Story arrives in Event Distribution Server with analytics
**T 250ms** : Analytics have been published to all real-time subscribers

Individual latency of stories may vary depending on a variety of different factors:

**Story Size:** The most frequent cause of variations in the latency is larger stories which contain more text and take a few milliseconds longer to process. Roughly 70% of news stories are less than 5 Kb in size and take less than 200 milliseconds on average. A 15 Kb story can take up to 600ms to process.

**Publication Frequency:** RavenPack can handle many stories in the same second with little or no noticeable lag. If the number of stories received per second becomes very large over a period of time, then latency could be affected (in the milliseconds range).

# Frequently Asked Questions (FAQ)

**What is the most important section in this user guide?**

The critical section to review is the "Data Field Descriptions." It explains the exact functionality of each analytic. The "Getting Started with RavenPack News Analytics" section is a close second as it provides a roadmap to streamline the evaluation and testing process.

**How many entities does RavenPack track?**

RavenPack tracks more than 175,000 entities worldwide since Jan 1, 2000.

**How does your entity detection system work?**

The RavenPack Entity Detection System is powered by a proprietary entity database containing information about each entity including:

- **Registered Name**
- **Common Abbreviations/Aliases** (i.e. *Bank of America = BofA, Bk of America, BAC, etc.)*
- **Popular Affiliates** (subsidiaries, acquisitions, popular brands, official people)
- **Publisher Codes** (tags applied by publishers to identify entities)
- **Positions** (Executive roles, cabinet positions, memberships, etc.)
- **Securities Identifiers for Equities** (ISINs, Cusips, Sedols, Tickers, MIC Codes)

The RavenPack entity database is updated daily with corporate actions information from various sources. Most updates are applied automatically as new information about an entity becomes available. Other updates are applied manually by the RavenPack team who periodically add new entities requested by clients, or new aliases for entities, and other identifier information, which makes our entity detection more effective.

**What are some common pitfalls when initially working with the data?**

Early missteps can occur in the following five fields:

1. TIMESTAMP_UTC: Not converting the Coordinated Universal Time (UTC) time zone if market data is represented in a local time zone.
2. COMPANY: Using this field to historically backtest data instead of RP_ENTITY_ID. The COMPANY field changes over time based on country of incorporation or stock ticker adjustments. In contrast, RP_ENTITY_ID field consistently identifies entities throughout the historical archive.
3. RELEVANCE: Not filtering on relevance can cause poorer results. In the paper "Construction of Market Sentiment Indices Using News Sentiment" (Hafez, Nov. 2009), filtering for relevance improved results up to three times.
4. CATEGORY: Not filtering certain categories. Several categories focus on events which have already taken place (ex. "stock-loss"). Other categories have high frequency in short periods of time (ex. "mkt-close-buy-imbalance") and could also be filtered.
5. NOVELTY: Not filtering for novelty. Consider filtering for only first instances of a CATEGORY event by using the ENS score of 100, since the price impact of another story announcing the same event may be less important.

**Why are some RP_ENTITY_IDs less than six characters in Excel?**

Excel has a tendency to display certain data in a particular way. Values in CSV files can be misinterpreted by Excel and consequently some fields (ex. RP_ENTITY_ID or TIMESTAMP_UTC) can be displayed incorrectly or converted into numerical values. For example, the company “A. Friedrich Flender AG” has an RP_ENTITY_ID of 9918E9 but Excel displays this value as 9.92E+12. Another example is the TIMESTAMP_UTC which can be displayed as mm:ss.0 (00:01.3) rather than YYYY-MM-DD hh:mm:ss.sss (2015-05-20 00:00:01.01). It is therefore recommended that any CSV file be imported as an external data range in Excel and set the column data format to text.

**Is event novelty derived from publisher tags or meta-data?**

No. The Event Novelty Score (ENS) is not dependent on publisher tags. The main reasons for this include: a) publisher novelty tags are applied manually and at times inconsistently b) 85% of all news articles lack publisher novelty tags while the other 15% may chain events together that have different characteristics but may appear related to the publisher.

**Can scores have conflicting sentiment results?**

Yes. Although most sentiment analytic results correlate highly, there are many cases where they differ. For example, the sentiment direction of ESS & CSS scores agrees in about 95% of the cases. However, most analytics are independent from one another in order to provide diversification and a different perspective of sentiment for the same news story. Some analytics have dependencies in the values or algorithms of another analytics while most don't.

For example, ESS is produced independently from CSS or any other score for that matter. The algorithm producing ESS values depends on an event CATEGORY detected and other information it teases out of the news story like financial figures, analyst ratings, and directional language. CSS is dependent on the classification results of 5 other story level analytics and combines the data in a unique way to yield its sentiment score.

**What news sources are covered?**

*Dow Jones Edition* analyzes relevant information from Dow Jones Newswires, regional editions of the Wall Street Journal, Barron's and MarketWatch. It includes more than 5,000 employees around the world, including more than 2,000 journalists in 58 countries.

*Web Edition* automatically processes hundreds of thousands of articles a day from leading publishers and web aggregators. Over 19,000 sources are continuously monitored including industry and business publishers, national and local news, blog sites, government and regulatory updates.

*Press Release Edition* analyzes news and information from the leading global media organizations. More than 100,000 press releases and regulatory disclosures are processed on a daily basis from a variety of newswires and press release distribution networks, including exclusive content from PRNewswire, Canadian News Wire, Regulatory News Service, and others.

To obtain the full list of sources covered in each edition, download the “Source List” file and filter the DATA_TYPE field for PRODUCT and DATA_VALUE for one of the following options:

- Dow Jones Edition: DJ
- Web Edition: WE

- Press Release Edition: PR

**Do more “insider-buy” events occur after the market hours than “insider-sell” events?**

Yes, a large portion of the insider-buy events do occur between 10pm and 12am EDT while relatively few insider-sell events take place during the same period. 37% of all the insider-buy events occur between 10pm and 12am EDT across the archive. In contrast, only 1% of all "insider-sell" events happen in the same timeframe.

Insider-sell events display a more disbursed pattern: 36% of all insider-sell events occur between 9am and 4pm EDT. Another 53% take place within the next 4 hours of the market close (4pm EDT and 8pm EDT). The balance of these events occurs overnight.

**What countries produce insider-buy and insider-sell events?**

Over 95% of all insider-buy and insider-sell events originate from US companies. These events are also detected in 52 other countries including Canada, Bermuda, and the United Kingdom.

**When can ENS and ENS_KEY contain null values?**

It is possible for the ENS and ENS_KEY analytics to contain null values when a news or system outage has occurred. This indicates that the novelty of the event could not be determined.

# Appendix A: Traditional Tagging Methodology

RavenPack's Traditional Tagging methodology underpins the PEQ and BEE scores and is based on a Rule Base that maps specified words, phrases, combinations, and other word-level definitions to pre-defined sentiment values. Each story may match several of these rules. Two Rule Bases are defined as: one that identifies positive sentiment and another that identifies negative sentiment.

Step One: A Classification Base is defined
When developing a new sentiment series that uses the Traditional Tagging methodology, the first step is to develop a Classification Base, or define the types of stories that contain the content relevant for tagging. The ideal Classification Base contains only stories that contain news that affect the target market or asset class.

Step Two: A large sample is analyzed to create a Rule Base
A sample of up to 75,000 stories in the Classification Base developed in step one is drawn from RavenPack's news database for a fixed date range. Stories are randomly selected for review. Experts read and extract key language from this sample. This can include individual words, phrases, template phrases, and complex linguistic rules that account for diverse types of language. Entries in the Rule Base contain names of key actors, corporations, rating agencies, commentary vocabulary, considerations of tense and positioning, and countless other types of language that is specifically relevant to the target market or asset class. The two Rule Bases, positive and negative are equally balanced and carefully reviewed to not favor specific types of stories, be too general, or be prone to unexpected variations.

Step Three: The Rule Base is tested on a large sample
Once a Rule Base has been established, it undergoes extensive review and expansion through testing on large sample sets. During this process, methodological consistency is maintained by not applying any weights or sentiment values to individual stories. Rather, testing always involves identifying how many rules each story matches. This allows experts to accurately identify inconsistencies and make additions.

When developing a Sentiment Classifier, this step is repeated up to six or seven times in order to reach a high level of confidence in the Rule Base. Key statistics such as the number of stories which match at least one rule, the number of stories which match rules from both lists, and the number of stories which match one list but not the other are carefully evaluated and used to inform further development.

Step Four: Generate historical analysis and enable real-time tagging
Using the Rule Base on the target story type, historical analysis is generated and real-time tagging is enabled. This process involves several consistency checks of historical data and generation of volume statistics. When this process is complete, the series is published.

Step Five: Quarterly re-evaluation
Because story sampling is based on a limited data range, there always exists the possibility that new economic terminology, trends, types of reporting, market forces, etc. may emerge after the sample period used in step three. In order to account for these trends, classifiers are re-evaluated on a quarterly basis.

This process involves completing step three for stories sampled outside of the date range of the original sample or most recent quarterly re-evaluation. Although it is not possible to quantify trends in the accuracy of these types of classifiers, experts who analyze the original Rule Base carefully search for new language patterns that are not detected. If the original Rule Base expands by more than 10% as a result of this process, a new series is developed.

# Appendix B: Expert Consensus Tagging Methodology

RavenPack's Expert Consensus Methodology underpins the BMQ, BCA, BER and BAM scores and entails a group of financial experts manually tagging a set of stories that is later used as a basis for automated computer classification using a Bayes Classifier.

Step One: A Classification Base is defined
When developing a new sentiment series that uses the Expert Consensus methodology, the first step is to develop a Classification Base, or define the types of stories that contain the content relevant for tagging. The ideal Classification Base contains only stories that contain news that affect the target market or asset class.

Step Two: Experts build an internal Tagging Guide
When developing a new sentiment series that requires manual tagging, a team of in-house experts with extensive backgrounds in linguistics, finance, and economics first develop and agree upon a set of parameters and basic assumptions that will guide sentiment tagging.
This Tagging Guide ensures that the assumptions used in identifying story sentiment are consistent and agreed upon. It provides rules for when to identify stories as positive, negative, or neutral.

Step Three: A large sample is tagged
A sample of up to 28,000 stories in the Classification Base developed in step one is drawn from RavenPack's news database for a fixed date range. Stories are randomly selected for tagging.
Up to ten experts read and classify all story headlines in the sample using the Tagging Guide developed in step two.

The Tagging Guide is built to avoid disagreements in story sentiment among experts. Even so, stories that were not classified by 80% of expert taggers as having the same sentiment are automatically given the NEU code. Stories that contain both positive and negative sentiment are judged based on the story's overall effect on the market. Stories with equal amounts of positive and negative sentiment are tagged as neutral.

Step Four: Software is trained from sample to automate tagging
Once an appropriate sample of stories has been tagged, a Bayes Classifier uses supervised learning to discern patterns in expert tagging and establish rules for future automation. This automated tagging process must meet exceptional levels of accuracy in order to be made available to clients. In cases when accuracy is not sufficiently high, step three is repeated with a larger sample set. Accuracy levels vary by classifier but range from 80% to 96%.

Step Five: Generate historical analysis and enable real-time tagging
After the classifier has been trained to reach acceptable levels of accuracy, historical analysis is generated and real-time tagging is enabled. This process involves several consistency checks of historical data and generation of volume statistics. When this process is complete, the series is published.

Step Six: Quarterly re-evaluation
Because training is based on a limited data range, there always exists the possibility that new economic terminology, trends, types of reporting, market forces, etc. may emerge after the sample period used in step three. In order to account for these trends, classifiers are re-evaluated on a quarterly basis. This process involves completing step three for stories sampled outside of the date range of the original sample or most recent quarterly re-evaluation. The results of this expert classification are compared to the results of automated classification. If the accuracy level is 10% lower than the level when the series was originally released, a new series is developed.

# Appendix C: Market Response Methodology

RavenPack's Market Response methodology underpins the CSS and NIP scores and is based on a Rule Base that identifies and maps individual words or word combinations in the story headline to the price impact on stocks of companies mentioned in the headline. The price impact is measured in the hours ahead of the arrival of the news item and is transformed into an impact score using advanced machine learning techniques.

Step One: A Classification Base is defined
When developing a new impact series that uses the Market Response methodology, the first step is to develop a Classification Base, or define the types of stories that contain the content relevant for measuring the market impact. The ideal Classification Base contains only stories that contain news that affect the target market or asset class.

Step Two: A large sample is analyzed to create a Rule Base
A sample of up to 30,000 headline stories in the Classification Base developed in step one is drawn from RavenPack's news database for a fixed date range. The headlines of these stories are extracted and parsed into words to form a list of candidates of individual words and word combinations that are typical for such headline stories. Based on statistical measures, about 4,000 of these words and word combinations have been identified as promising in relations to predicting market impact.

Step Three: Create an Impact Score using the Rule Base
Once a Rule Base has been established, different advanced machine learning techniques are applied with the objective of creating an Impact Score that identifies the probability of the volatility of a particular stock to be either higher or lower than the volatility of the market.

Step Four: Generate historical analysis and enable real-time tagging
Applying step two and three on the target story type, historical analysis is generated and the real-time creation of Impact Scores is enabled. This process involves several consistency checks of historical data. When this process is complete, the series is published.

Step Five: Quarterly re-evaluation
Because story sampling is based on a limited data range, there always exists the possibility that new economic terminology, trends, types of reporting, market forces, etc. may emerge after the sample period used in step two and three. In order to account for these trends, scores are re-evaluated on a quarterly basis.

This process involves completing step two for stories sampled outside of the date range of the original sample or most recent quarterly re-evaluation. Statistical measures are used with the purpose of identifying additional words or word combinations that, from a statistical perspective, seem promising in terms of market impact. Such words and word combinations are added to the Rule Base and are then used under step three to continuously maintain and improve the impact series to reflect current market conditions.

# Appendix D: Factors in the Event Sentiment Score

In addition to the expert consensus survey data, the Event Sentiment Score (ESS) has a strength component that is influenced by a variety of factors, depending on the type of event. RavenPack systematically extracts information from every news story to model these factors and determine how positive or negative each event should be. Here is a list of some of these factors:

**Emotional Factor:**

There are 5 scales containing groups of words and phrases in the RavenPack's emotional magnitude component of ESS. Each component contains words that signify the magnitude of an event as described by the author of the story.

1. *Low Magnitude:* Contains adjective words such as "low, minor, small, or inconsequential" and phrases such as "below the mark or not meaningful"

2. *Moderate Magnitude:* Contains words such as "moderate, mellow, or dainty" and phrases such as "nothing much or fairly flat"

3. *Substantial Magnitude:* Contains words such as "substantial, durable, considerable or extensive" and phrases such as "fairly considerable or significantly large"

4. *Severe Magnitude:* Contains words such as "severe, commanding, destructive, or excruciating" and phrases such as "extremely high or highly elevated"

5. *Critical Magnitude:* Contains words such as "critical, devastation, massacre, or catastrophic" and phrases such as "super colossal or most damaging"

**Weather and Climate Factor**

Tracks official scales to measure extreme weather such as the Richter scale or the Volcanic Eruption Index.

**Analyst Rating Factor**

Covers over 150 different broker and analyst rating scales for stocks (e.g. strong buy, buy, hold, sell, strong sell).

**Credit Rating Factor**

Consolidates the three main credit ratings scales by Moody's, Fitch, and S&P (e.g. AAA, AA, BB, C, etc.) into one normalized scale.

**Fundamental Comparison Factor**

Extracts and calculates numerical differences between actual or estimated values in earnings, revenues, dividends, macroeconomic indicators, and any other financial or economic announcement. Performs arithmetic and translates fundamental percentage changes into a normalized score within the ESS ranges.

**Casualties Factor**

Identifies how many people are dead or injured as a result of an event and uses this as sentiment strength factor, particularly for natural disasters and industrial accidents.

# Appendix E: Links of Interest on News-Base Trading

***Sentiment News Blog:*** http://www.sentimentnews.com/

*White Papers available from RavenPack:*

**Using Sentiment to Create Theme-based Alphas**
Author: RavenPack (2015)
http://www.ravenpack.com/company/news-and-events/details/research-sentiment-theme-based-alphas/

**A Machine Learning-Based Trading Strategy Using Sentiment Analysis Data**
Author: Lucena (2015)
http://www.ravenpack.com/research/white-papers/document_detail/enhancing-equity-trading-models-corporate-macro-sentiment-abnormality/

**Enhancing Equity Trading Models with Corporate & Macro Sentiment Abnormality**
Author: RavenPack (2015)
http://www.ravenpack.com/research/white-papers/document_detail/enhancing-equity-trading-models-corporate-macro-sentiment-abnormality/

**Bond over Big Data - Trading bond futures (& FX) with RavenPack News Data**
Author: The Thalesians (2015)
http://www.ravenpack.com/research/white-papers/document_detail/bond-over-big-data-trading-bond-futures-fx-ravenpack-news-data/

**Predictive Media Content Analytics: 24/7 Information Has Forever Changed Financial Market Strategies**
Author: TrendPointers (2014)
http://www.ravenpack.com/research/white-papers/document_detail/predictive-media-content-analytics-247-information-has-forever-changed-financial-market-strategies/

**Improved Stock Market Returns from Systematically Trading Infrequent News**
Author: RavenPack (2014)
http://www.ravenpack.com/research/white-papers/document_detail/improved-stock-market-returns-systematically-trading-infrequent-news/

**Commonality in News Around the World**
Author: UNSW (2014)
http://www.ravenpack.com/research/white-papers/document_detail/commonality-news-around-world/

**Filtering FX Carry Using RavenPack News Analytics**
Author: The Thalesians (2014)
http://www.ravenpack.com/research/white-papers/document_detail/filtering-fx-carry-using-ravenpack-news-analytics/

**Exploring Global Variations in News Impact on Equities**
Author: Ravenpack (2014)
http://www.ravenpack.com/research/white-papers/document_detail/exploring-global-variations-news-impact-equities/

**Incorporating Bidirectional Momentum Effects & Media Attention to Profitably Trade Linked Companies**
Author: FactSet (2014)
http://www.ravenpack.com/research/white-papers/document_detail/incorporating-bidirectional-momentum-effects-media-attention-profitably-trade-linked-companies/

**Tactical Equity Portfolio Formation Using News Analytics**
Author: Arialytics (2014)
http://www.ravenpack.com/research/white-papers/document_detail/tactical-equity-portfolio-formation-using-news-analytics/

**Web News Analytics Enhance Stock Portfolio Returns**
Author: RavenPack (2013)
http://www.ravenpack.com/research/white-papers/document_detail/understanding-illiquidity-corporate-bonds-arrival-public-news/

**Understanding the Illiquidity of Corporate Bonds: The Arrival of Public News**
Author: CA / TX Univ (2013)
http://www.ravenpack.com/research/white-papers/document_detail/understanding-illiquidity-corporate-bonds-arrival-public-news/

**Quantamentals - The Price is Right**
Author: Macquarie (2013)
http://www.ravenpack.com/research/white-papers/document_detail/macquarie-quantamentals-price-right/

**Media-Driven High Frequency Trading: Evidence from News Analytics**
Author: Wharton/INSEAD (2013)
http://www.ravenpack.com/research/white-papers/document_detail/media-driven-high-frequency-trading-evidence-news-analytics/

**Seeking a Behavioral Response to Information: News Analytics and VIX Futures Price**
Author: Diff.Research (2013)
http://www.ravenpack.com/research/white-papers/document_detail/seeking-behavioral-response-information-news-analytics-and-vix-futures-prices/

**Enhancing Short Term Reversal Strategies with News Analytics**
Author: RavenPack (2013)
http://ravenpack.com/research/resources.htm?id=Paper37

**Trading Relative Value Based on News Indicators**
Author: RavenPack (2013)
http://ravenpack.com/research/resources.htm?id=Paper36

**Sentiment Derived from News Predicts EURUSD Movements**
Author: RavenPack (2013)
http://ravenpack.com/research/resources.htm?id=Paper35

**Country News Sentiment Factors Predict Forex Prices**
Author: RavenPack (2013)
http://ravenpack.com/research/resources.htm?id=Paper33

**Attention Conditions Stock Market Reaction to News Sentiment**
Author: RavenPack (2013)
http://ravenpack.com/research/resources.htm?id=Paper32

**How does News Sentiment Impact Asset Volatility**
Author: ANU (2013)
http://ravenpack.com/research/resources.htm?id=Paper34

**Time-Varying Relationship of News Sentiment, Implied Volatility and Stock Returns**
Author: UNSW (2012)
http://ravenpack.com/research/resources.htm?id=Paper28

**Intraday Forex Trading Based on Sentiment Inflection Points**
Author: RavenPack (2012)
http://ravenpack.com/research/resources.htm?id=Paper27

**Capital Market Consequences of Business Press Coverage of Management Earnings Guidance**
Author: Texas A&M (2012)
http://ravenpack.com/research/resources.htm?id=Paper29

**Sell on the News - The Impact of News Sentiment on Stocks**
Author: Nomura (2012)
http://ravenpack.com/research/resources.htm?id=Paper30

**Sentiment Derived from Economic and Geopolitical News Predicts Real GDP Growth**
Author: RavenPack (2012)
http://ravenpack.com/research/resources.htm?id=Paper25

**CEO Resignation - Impact Study**
Author: Macquarie (2012)
http://ravenpack.com/research/resources.htm?id=Paper26

**Size Matters in Sentiment Trading**
Author: RavenPack (2012)
http://ravenpack.com/research/resources.htm?id=Paper24

**Quantamentals - Quantifying Events**
Author: Macquarie (2012)
http://ravenpack.com/research/resources.htm?id=Paper23

**Quantamentals - Macquarie Events Compendium**
Author: Macquarie (2012)
http://ravenpack.com/research/resources.htm?id=Paper22

**News Sensitivity in Sector Rotation Models**
Author: RavenPack (2012)
http://ravenpack.com/research/resources.htm?id=Paper15

**Asia Pac Dynamics, Quant Strategy - Macro risk and the news**
Author: Macquarie (2012)
http://ravenpack.com/research/resources.htm?id=Paper21

**Short-Term Stock Selection Using News Based Indicators**
Author: RavenPack (2012)
http://ravenpack.com/research/resources.htm?id=Paper14

**News Arrival and Cross-Asset Correlation Breakdown**
Author: UNSW (2012)
http://ravenpack.com/research/resources.htm?id=Paper31

**Behavioral Trends and Market Neutrality**
Author: Kittrell (2012)
http://ravenpack.com/research/resources.htm?id=Paper13

**Factoring Sentiment Risk into Quant Models**
Author: RavenPack (2012)
http://ravenpack.com/research/resources.htm?id=Paper12

**RavenPack Sentiment and Macro-Economic Indicators**
Author: RavenPack (2012)
http://ravenpack.com/research/resources.htm?id=Paper11

**Introducing the RavenPack Sentiment Index**
Author: RavenPack (2011)
http://ravenpack.com/research/resources.htm?id=Paper10

**Event Trading Using Market Response**
Author: Hafez (2011)
http://ravenpack.com/research/resources.htm?id=Paper09

**Aussie News Flow - Extra! Extra! Read all about it**
Author: Macquarie (2011)
http://ravenpack.com/research/resources.htm?id=Paper20

**News Movers and Shakers in Finance**
Author: Hafez (2011)
http://ravenpack.com/research/resources.htm?id=Paper08

**Japan Dynamics, Quant Analysis - Momentum in the news**
Author: Macquarie (2010)
http://ravenpack.com/research/resources.htm?id=Paper19

**News Beta - A New Measure for Risk & Stock Analysis**
Author: Hafez (2010)
http://ravenpack.com/research/resources.htm?id=Paper07

**Making the trend your friend - Momentum and News Flow**
Author: Macquarie (2010)
http://ravenpack.com/research/resources.htm?id=Paper18

**How News Events Impact Market Sentiment**
Author: Hafez (2010)
http://ravenpack.com/research/resources.htm?id=Paper06

**Sentiment Reversals as Buy Signals**
Author: Kittrell (2010)
http://ravenpack.com/research/resources.htm?id=Paper05

**Construction of Market Sentiment Indices Using News Sentiment**
Author: Hafez (2009)
http://ravenpack.com/research/resources.htm?id=Paper03

**Eventful investing - Harnessing the power of event-driven strategies**
Author: Macquarie (2009)
http://ravenpack.com/research/resources.htm?id=Paper17

**Impact of News Sentiment on Abnormal Stock Returns**
Author: Hafez (2009)
http://ravenpack.com/research/resources.htm?id=Paper04

**Breaking news - How to use news sentiment to pick stocks**
Author: Macquarie (2009)
http://ravenpack.com/research/resources.htm?id=Paper16

**Equity Portfolio Risk (volatility) Estimation Using Market Information and Sentiment**
Author: Northfield (2008)
http://ravenpack.com/research/resources.htm?id=Paper02

**Sector Rotation Strategies Driven By News Sentiment Indices**
Author: Hafez (2009)
http://ravenpack.com/research/resources.htm?id=Paper01

***Other Sources:***

**Does Public Financial News Resolve Asymmetric Information?**
*Author: Paul Tetlock - Columbia Business School - 1 Nov 2008*
http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1303612

**Equity portfolio risk (volatility) estimation using market information and sentiment**
*Author: Leela Mitra, Gautam Mitra, Dan diBartolomeo - 1 Dec 2008*
http://www.northinfo.com/documents/313.pdf

**More Than Words: Quantifying Language to Measure Firms' Fundamentals**
*Author: Paul C Tetlock*
http://papers.ssrn.com/sol3/papers.cfm?abstract_id=923911

**Giving Content to Investor Sentiment: The Role of Media in the Stock Market**
*Author: Paul C Tetlock*
http://www0.gsb.columbia.edu/faculty/ptetlock/papers/Tetlock_JF_07_Giving_Content_to_Investor_Sentiment.pdf

**Trading Strategies To Exploit News Sentiment**
*Authors: Wenbin Zhang and Steven Skiena - 2009*
http://www.cs.sunysb.edu/~skiena/lydia/blogtrading.pdf

**Do Stock Market Investors Understand the Risk Sentiment of Corporate Annual Reports?**
*Author: Feng Li*
http://papers.ssrn.com/sol3/papers.cfm?abstract_id=898181

**Do US Stock Markets Typically Overreact to Corporate News Stories?**
*Author: Werner Antweiler, Murray Z. Frank*
http://papers.ssrn.com/sol3/papers.cfm?abstract_id=878091

**Intraday Public Information - The French Evidence**
*Author: Alen Vukic*
http://ethesis.unifr.ch/theses/downloads.php?file=VukicA.pdf

**Intraday Market Dynamics Around Public Information Arrivals**
*Author: Angelo Ranaldo*
http://www.ranaldo.net/pdf/Work_in_progress/working-paper_06_11.pdf

**Investor Sentiment and Stock Market Sensitivity to Corporate News**
*Author: G. Mujtaba Mian, Srinivasan Sankaraguruswamy*
http://www.isb.edu/faculty/upload/Doc27920101052.pdf

**Language Models for Financial News Recommendation**
*Author: e-Analyst: (Victor Lavrenko, Matteo Schmill, Dawn Lawrie, Paul Ogilvie)*
http://ciir.cs.umass.edu/pubfiles/ir-207.pdf

**The Implications of Annual Report's Risk Sentiment for Future Earnings and Stock Returns**
*Author: Feng Li*
http://papers.ssrn.com/sol3/papers.cfm?abstract_id=890586

**Stock Price Reaction to News and No-News: Drift and Reversal After Headlines**
*Author: Wesley S Chan - 2001*
http://jfe.rochester.edu/02207.pdf

**Investor Sentiment and the Cross-Section of Stock Returns**
*Author: Malcolm Baker, Jeffery Wurgler*
http://pages.stern.nyu.edu/~jwurgler/papers/sentiment.pdf

**Share Prices and Trading Activity Over the Corporate Action Processing Cycle**
*Author: Oxera Consulting - 2006*
http://www.dtcc.com/downloads/leadership/whitepapers/2006_oxera.pdf

**Impact of Public Information on the Stock Market**
*Author: Mark L. Mitchell, Harold Mulherin - 1994*
http://ideas.repec.org/a/bla/jfinan/v49y1994i3p923-50.html

**Beyond the Numbers: An Analysis of Optimistic and Pessimistic Language in Earnings Press Releases**
*Author: Angela K. Davis, Jeremy M. Piger, Lisa M. Sedor*
http://papers.ssrn.com/sol3/papers.cfm?abstract_id=875399

**Market Reactions to Tangible and Intangible Information**
*Author: Kent Daniel, Titman Seridan - 2005*
http://papers.ssrn.com/sol3/papers.cfm?abstract_id=274204

**News and Trading Rules**
*Author: James D. Thomas - 2003*
http://www.e-m-h.org/Thom03.pdf

**Daily Stock Market Forecast From Textual Web Data, The Hong Kong University of Science and Technology**
*Author: B. Wuthrich, V. Cho, S. Leung, D. Permunetilleke, K. Sankaran, J. Zhang, W. La*
http://www.cwu.edu/~borisk/finance/smc98.pdf

**Consumer Sentiment, the Economy, and the News Media**
Author: Mark Doms and Norman Morin
http://www.federalreserve.gov/pubs/feds/2004/200451/200451pap.pdf

# Appendix F: Dataset Schemas

This section explains the data types and maximum lengths for the fields included in the Equity and Global Macro datasets as well as the company and entity mapping files and the event taxonomy file.

## Equity Dataset

**TIMESTAMP_UTC**
The Date/Time (YYYY-MM-DD hh:mm:ss.sss) at which the news item was received by RavenPack servers in Coordinated Universal Time (UTC).This field cannot contain a null value.

**RPNA_DATE_UTC**
The Date (YYYY-MM-DD) at which the news item was received by RavenPack servers in Coordinated Universal Time (UTC).This field cannot contain a null value.

**RPNA_TIME_UTC**
The Time (hh:mm:ss.sss) at which the news item was received by RavenPack servers in Coordinated Universal Time (UTC).This field cannot contain a null value.

**RP_ENTITY_ID**
A fixed-width character field of 6 characters. Oracle example: VARCHAR2(6). This field cannot contain a null value.

**ENTITY_TYPE**
A fixed-width character field of 4 characters. Oracle example: VARCHAR2(4). This field cannot contain a null value.

**ENTITY_NAME**
A variable-width character field with a maximum of 400 characters. Oracle example: VARCHAR2(400)

**POSITION_NAME**
A variable-width character field with a maximum of 400 characters. Oracle example: VARCHAR2(400)

**RP_POSITION_ID**
A fixed-width character field of 6 characters. Oracle example: VARCHAR2(6)

**COUNTRY_CODE**
A fixed-width character field of 2 characters. Oracle example: VARCHAR2(2)

**RELEVANCE**
A numerical field that stores an integer value. Oracle example: NUMBER. This field cannot contain a null value.

**TOPIC**
A variable-width character field with a maximum of 50 characters. Oracle example: VARCHAR2(50)

**GROUP**
A variable-width character field with a maximum of 50 characters. Oracle example: VARCHAR2(50)

**TYPE**
A variable-width character field with a maximum of 50 characters. Oracle example: VARCHAR2(50)

**SUB_TYPE**
A variable-width character field with a maximum of 50 characters. Oracle example: VARCHAR2(50)

**PROPERTY**
A variable-width character field with a maximum of 50 characters. Oracle example: VARCHAR2(50)

**EVALUATION_METHOD**
A variable-width character field with a maximum of 50 characters. Oracle example: VARCHAR2(50)

**MATURITY**
A variable-width character field with a maximum of 50 characters. Oracle example: VARCHAR2(50)

**CATEGORY**
A variable-width character field with a maximum of 100 characters. Oracle example: VARCHAR2(100)

**ESS**
A numerical field that stores an integer value. Oracle example: NUMBER

**AES**
A numerical field that stores an integer value. Oracle example: NUMBER

**AEV**
A numerical field that stores an integer value. Oracle example: NUMBER

**ENS**
A numerical field that stores an integer value. Oracle example: NUMBER

**ENS_SIMILARITY_GAP**
A numerical field that stores fixed and floating-point numbers. Oracle Example: NUMBER(8,5) 8= total number of digits (precision). 5= Number of digits to the right of the decimal point (scale). 5 is not mandatory.

**ENS_KEY**
A variable-width character field with a maximum of 32 characters. Oracle example: VARCHAR2(32)

**ENS_ELAPSED**
A numerical field that stores an integer value. Oracle example: NUMBER

**G_ENS**
A numerical field that stores an integer value. Oracle example: NUMBER

**G_ENS_SIMILARITY_GAP**
A numerical field stores fixed and floating-point numbers. Oracle Example: NUMBER(8,5) 8= total number of digits (precision). 5= Number of digits to the right of the decimal point (scale). 5 is not mandatory.

**G_ENS_KEY**
A variable-width character field with a maximum of 32 characters. Oracle example: VARCHAR2(32)

**G_ENS_ELAPSED**
A numerical field that stores an integer value. Oracle Example: NUMBER

**EVENT_SIMILARITY_KEY**
A variable-width character field with a maximum of 32 characters. Oracle example: VARCHAR2(32)

**NEWS_TYPE**
A variable-width character field with a maximum of 50 characters. Oracle example: VARCHAR2(50). This field cannot contain a null value.

**SOURCE**
A fixed-width character field of 6 characters. Oracle example: VARCHAR2(6). This field cannot contain a null value.

**RP_STORY_ID**
A variable-width character field with a maximum of 32 characters. Oracle example: VARCHAR2(32). This field cannot contain a null value.

**RP_STORY_EVENT_INDEX**
A numerical field that stores an integer value. Oracle example: NUMBER. This field cannot contain a null value.

**RP_STORY_EVENT_COUNT**
A numerical field that stores an integer value. Oracle example: NUMBER. This field cannot contain a null value.

**PRODUCT_KEY**
A fixed-width character field with a maximum of 5 characters. Oracle example: VARCHAR2(5). This field cannot contain a null value.

**COMPANY**
A variable-width character field with a maximum of 50 characters. Oracle example: VARCHAR2(50)

**ISIN**
A variable-width character field with a maximum of 12 characters. Oracle example: VARCHAR2(12)

**CSS**
A numerical field that stores an integer value. Oracle Example: NUMBER. This field cannot contain a null value.

**NIP**
A numerical field that stores an integer value. Oracle Example: NUMBER. This field cannot contain a null value.

**PEQ**
A numerical field that stores an integer value. Oracle Example: NUMBER. This field cannot contain a null value.

**BEE**
A numerical field that stores an integer value. Oracle Example: NUMBER. This field cannot contain a null value.

**BMQ**
A numerical field that stores an integer value. Oracle Example: NUMBER. This field cannot contain a null value.

**BAM**
A numerical field that stores an integer value. Oracle Example: NUMBER. This field cannot contain a null value.

**BCA**
A numerical field that stores an integer value. Oracle Example: NUMBER. This field cannot contain a null value.

**BER**
A numerical field that stores an integer value. Oracle Example: NUMBER. This field cannot contain a null value.

**ANL_CHG**
A numerical field that stores an integer value. Oracle Example: NUMBER. This field cannot contain a null value.

**MCQ**
A numerical field that stores an integer value. Oracle Example: NUMBER. This field cannot contain a null value.

The primary key for this dataset would be the following fields: RP_STORY_ID + RP_STORY_EVENT_INDEX

## Global Macro Dataset

**TIMESTAMP_UTC**
The Date/Time (YYYY-MM-DD hh:mm:ss.sss) at which the news item was received by RavenPack servers in Coordinated Universal Time (UTC). This field cannot contain a null value.

**RPNA_DATE_UTC**
The Date (YYYY-MM-DD) at which the news item was received by RavenPack servers in Coordinated Universal Time (UTC).This field cannot contain a null value.

**RPNA_TIME_UTC**
The Time (hh:mm:ss.sss) at which the news item was received by RavenPack servers in Coordinated Universal Time (UTC).This field cannot contain a null value.

**RP_ENTITY_ID**
A fixed-width character field of 6 characters. Oracle example: VARCHAR2(6). This field cannot contain a null value.

**ENTITY_TYPE**
A fixed-width character field of 4 characters. Oracle example: VARCHAR2(4). This field cannot contain a null value.

**ENTITY_NAME**
A variable-width character field with a maximum of 400 characters. Oracle example: VARCHAR2(400)

**POSITION_NAME**
A variable-width character field with a maximum of 400 characters. Oracle example: VARCHAR2(400)

**RP_POSITION_ID**
A fixed-width character field of 6 characters. Oracle example: VARCHAR2(6)

**COUNTRY_CODE**
A fixed-width character field of 2 characters. Oracle example: VARCHAR2(2)

**RELEVANCE**
A numerical field that stores an integer value. Oracle example: NUMBER. This field cannot contain a null value.

**TOPIC**
A variable-width character field with a maximum of 50 characters. Oracle example: VARCHAR2(50)

**GROUP**
A variable-width character field with a maximum of 50 characters. Oracle example: VARCHAR2(50)

**TYPE**
A variable-width character field with a maximum of 50 characters. Oracle example: VARCHAR2(50)

**SUB_TYPE**
A variable-width character field with a maximum of 50 characters. Oracle example: VARCHAR2(50)

**PROPERTY**
A variable-width character field with a maximum of 50 characters. Oracle example: VARCHAR2(50)

**EVALUATION_METHOD**
A variable-width character field with a maximum of 50 characters. Oracle example: VARCHAR2(50)

**MATURITY**
A variable-width character field with a maximum of 50 characters. Oracle example: VARCHAR2(50)

**CATEGORY**
A variable-width character field with a maximum of 100 characters. Oracle example: VARCHAR2(100)

**ESS**
A numerical field that stores an integer value. Oracle example: NUMBER

**AES**
A numerical field that stores an integer value. Oracle example: NUMBER

**AEV**
A numerical field that stores an integer value. Oracle example: NUMBER

**ENS**
A numerical field that stores an integer value. Oracle example: NUMBER

**ENS_SIMILARITY_GAP**
A numerical field that stores fixed and floating-point numbers. Oracle Example: NUMBER(8,5) 8= total number of digits (precision). 5= Number of digits to the right of the decimal point (scale). 5 is not mandatory.

**ENS_KEY**
A variable-width character field with a maximum of 32 characters. Oracle example: VARCHAR2(32)

**ENS_ELAPSED**
A numerical field that stores an integer value. Oracle example: NUMBER

**G_ENS**
A numerical field that stores an integer value. Oracle example: NUMBER

**G_ENS_SIMILARITY_GAP**
A numerical field stores fixed and floating-point numbers. Oracle Example: NUMBER(8,5) 8= total number of digits (precision). 5= Number of digits to the right of the decimal point (scale). 5 is not mandatory.

**G_ENS_KEY**
A variable-width character field with a maximum of 32 characters. Oracle example: VARCHAR2(32)

**G_ENS_ELAPSED**
A numerical field that stores an integer value. Oracle Example: NUMBER

**EVENT_SIMILARITY_KEY**
A variable-width character field with a maximum of 32 characters. Oracle example: VARCHAR2(32)

**NEWS_TYPE**
A variable-width character field with a maximum of 50 characters. Oracle example: VARCHAR2(50). This field cannot contain a null value.

**SOURCE**
A fixed-width character field of 6 characters. Oracle example: VARCHAR2(6). This field cannot contain a null value.

**RP_STORY_ID**
A variable-width character field with a maximum of 32 characters. Oracle example: VARCHAR2(32) This field cannot contain a null value.

**RP_STORY_EVENT_INDEX**
A numerical field that stores an integer value. Oracle example: NUMBER. This field cannot contain a null value.

**RP_STORY_EVENT_COUNT**
A numerical field that stores an integer value. Oracle example: NUMBER. This field cannot contain a null value.

**PRODUCT_KEY**
A fixed-width character field with a maximum of 5 characters. Oracle example: VARCHAR2(5). This field cannot contain a null value.

The primary key for this dataset would be the following fields: RP_STORY_ID + RP_STORY_EVENT_INDEX

## Entity Mapping Dataset

**RP_ENTITY_ID**
A fixed-width character field of 6 characters. Oracle example: VARCHAR2(6). This field cannot contain a null value.

**ENTITY_TYPE**
A fixed-width character field of 4 characters. Oracle example: VARCHAR2(4). This field cannot contain a null value.

**DATA_TYPE**
A variable-width character field with a maximum of 200 characters. Oracle example: VARCHAR2(200).This field cannot contain a null value.

**DATA_VALUE**
A variable-width character field with a maximum of 400 characters. Oracle example: VARCHAR2(400).This field cannot contain a null value.

**RANGE_START**
A date field with the following format (YYYY-MM-DD). Oracle example: DATE

**RANGE_END**
A date field with the following format (YYYY-MM-DD). Oracle example: DATE

## Event Taxonomy Dataset

**TOPIC**
A variable-width character field with a maximum of 50 characters. Oracle example: VACHAR2(50)

**GROUP**
A variable-width character field with a maximum of 50 characters. Oracle example: VACHAR2(50)

**TYPE**
A variable-width character field with a maximum of 50 characters. Oracle example: VACHAR2(50)
**SUB_TYPE**
A variable-width character field with a maximum of 50 characters. Oracle example: VACHAR2(50)

**PROPERTY**
A variable-width character field with a maximum of 50 characters. Oracle example: VACHAR2(50)

**CATEGORY**
A variable-width character field with a maximum of 100 characters. Oracle example: VARCHAR2(100)

**DESCRIPTION**
A variable-width character field with a maximum of 4000 characters. Oracle example: VARCHAR2(4000)

**SCHEDULED**
A variable-width character field with a maximum of 5 characters. Oracle example: VARCHAR2(5). This field cannot contain a null value.

**VALID_ENTITY_TYPES**
A variable-width character field with a maximum of 4000 characters. Oracle example: VARCHAR2(4000)

## Source List Dataset

**SOURCE**
A fixed-width character field with a maximum of 6 characters. Oracle example: VARCHAR2(6). This field cannot contain a null value.

**DATA_TYPE**
A variable-width character field with a maximum of 200 characters. Oracle example: VARCHAR2(200).This field cannot contain a null value.

**DATA_VALUE**
A variable-width character field with a maximum of 4000 characters. Oracle example: VARCHAR2(4000).This field cannot contain a null value.

**RANGE_START**
A date field with the following format (YYYY-MM-DD). Oracle example: DATE

**RANGE_END**
A date field with the following format (YYYY-MM-DD). Oracle example: DATE

## ABOUT RAVENPACK

RavenPack develops and distributes structured data products from unstructured content. The firm is the leader in news analytics which involves turning news into numbers so they can be easily manipulated and consumed by quantitative models and trading programs.

RavenPack classifies news items using multiple sophisticated sentiment detection algorithms. In addition, RavenPack generates a number of non-sentiment analytics including information about companies, events, relevance, and market impact. Outputs are often in the form of numerical news scores that can be used as inputs in the calculation of company, sector, and industry indicators.

News Analytics are used to power a number of applications ranging from high frequency trading applications requiring low latency inputs, to risk and asset management applications requiring factors whose time resolution may be daily, weekly, and monthly.

There is strong empirical evidence that RavenPack News Analytics have explanatory and predicative power in three dimensions: market direction, volume, and volatility. Beyond some of the world's top industry clients, RavenPack is working with a number of partners and academic researchers who are using news analytics in their applications and research endeavors.