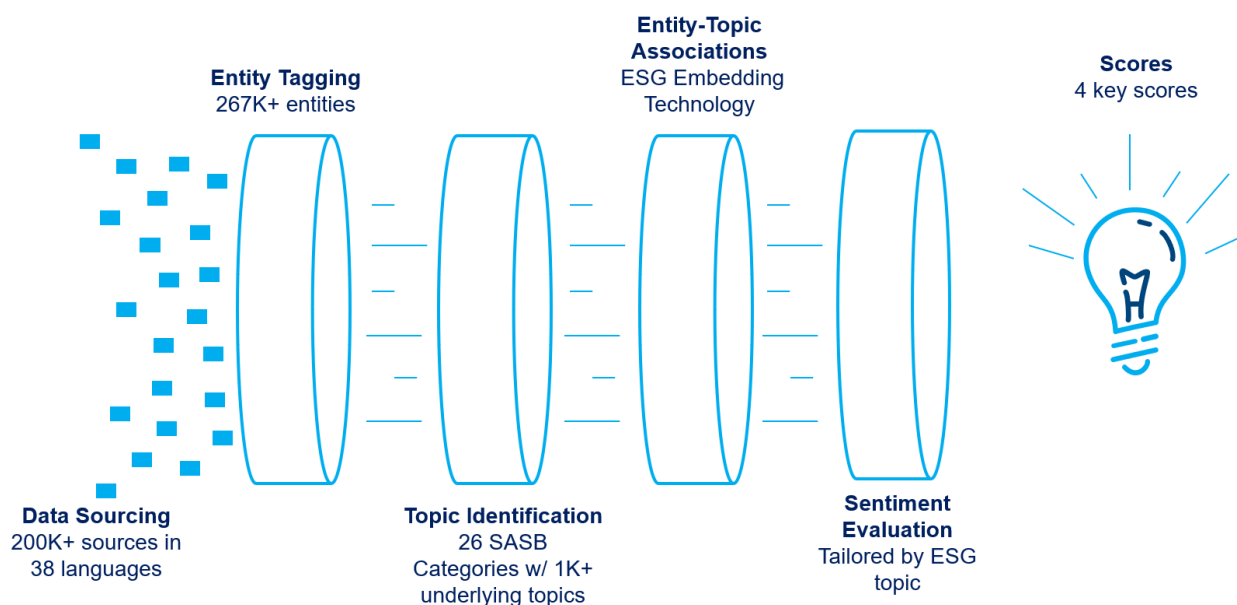


# Truvalue Pipeline 3

➤ A complete refactoring of the Truvalue dataset delivering improved quality, more robust history, and better scoring.

FactSet reprocessed our existing article set and mined over 340 million articles, to add ~900K+ net new articles spanning back to 2007. Reprocessed articles were run through the most up-to-date models governing everything from entity recognition to sentiment evaluation, leading to a significant improvement in overall data quality. Further, our updated scoring algorithms better capture emerging events within the market while increasing the frequency of Spotlight generation to deliver insights to clients faster. Major improvements include:

- Latest Natural Language Processing (NLP) models applied to full history;
- Extended history for ~267K public and private entities dating back to 2007;
- Improved Pulse & Insight scores preserve information over quiet periods and offer speedy reactions to major events;
- Intra-day Spotlights highlighting major news stories & events sooner;
- Daily and Accumulative Article Volume [total count of articles across full history] available painting the full picture; and
- Added controls removing company-reported news.

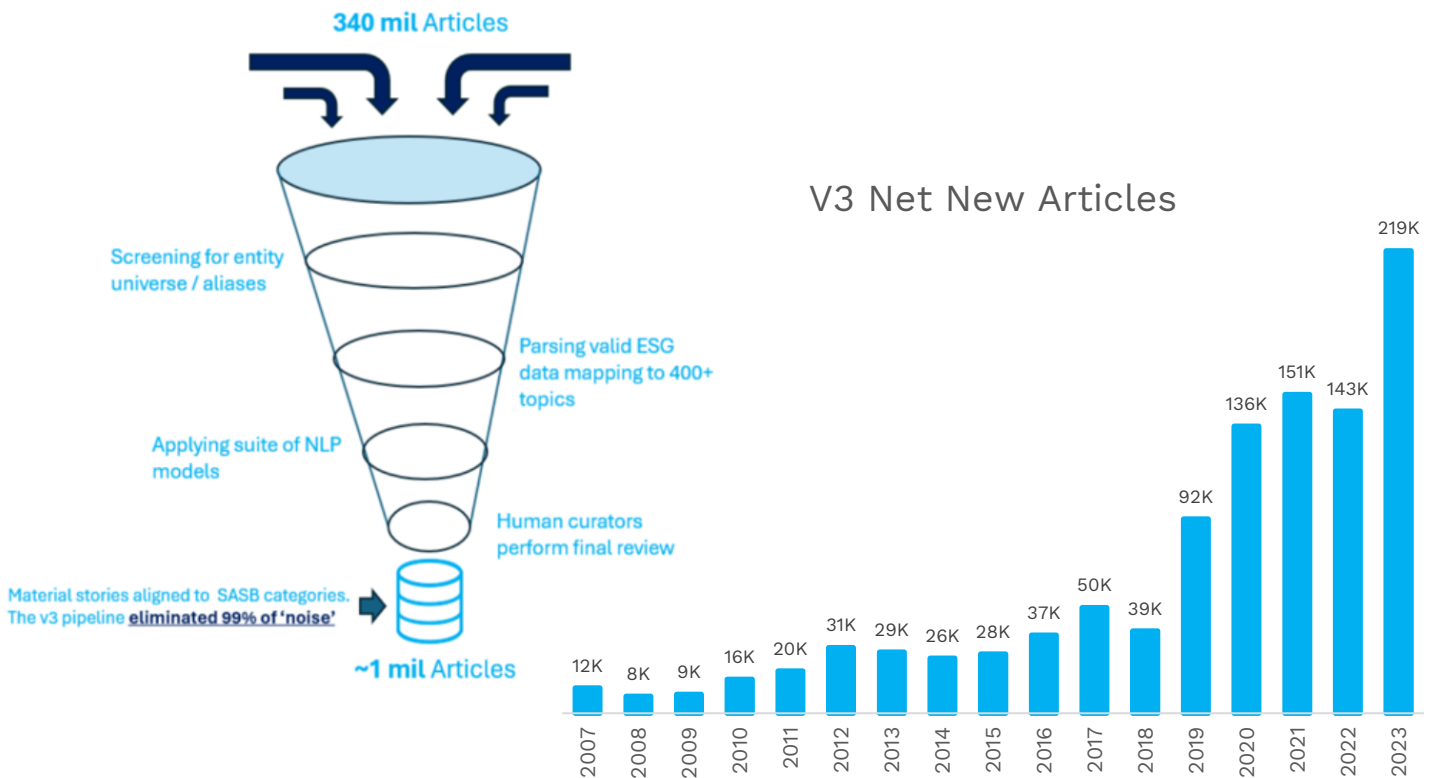


## Enhanced Data Quality & Expanded Historical Coverage for both Public & Private Entities

FactSet reprocessed our current time series, running it through the latest NLP suite. This significantly improved the accuracy of entity and topic tags, the associations between them, and the sentiment assigned. This led to an overall improvement in data quality of over 28% with improvements realized throughout history. Additionally, FactSet re-ran algorithms for newly added firms over its entire article history, mining over 340 million articles, to add ~900K+ net new articles spanning back to 2007. The added history covers both public and private entities, a 50-50 split between them with the largest increase in coverage occurring in EMEA.

In 90% of instances v3 matches or outperforms v2.

Of which, the average quality improvement is 28%



## Improved Scoring Algorithms Incorporate Critical News into the Time Series Faster

*FactSet reevaluated Truvalue legacy scoring algorithms, considering internal research and client feedback, to upgrade Pulse and Insight Score, along with other improvements.*

*Together these improvements make Truvalue's scores responsive to current events, and more consistent in behavior for higher and lower-volume companies and score series, such as individual categories.*

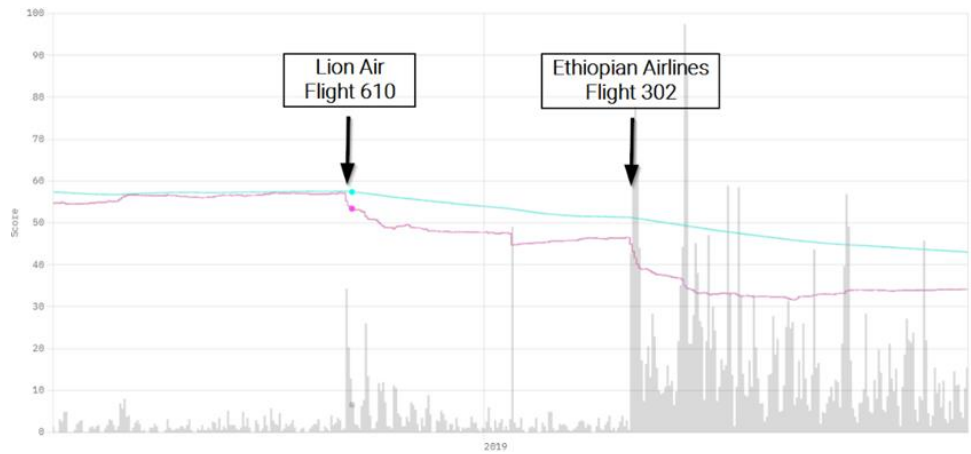
- The Insight Score will be faster to reflect major score moves with high article volume;
- The Pulse Score will be less reactive to single articles following quiet periods due to 'remembered' history of the time series, while still able to swiftly react to events when article volumes spike;
- With Insight and Pulse including a richer record of information, Total Volume over the entire history is now considered a more reliable indication of volume status than Trailing 12-month (TTM) volume. TTM Volume will still be available as a data item, but Truvalue's new volume levels are as follows:
  - High: 100+ articles of Total Volume
  - Medium: 10-99 articles of Total Volume
  - Low: 1-9 articles of Total Volume
- The Spotlight Algorithm will run multiple times a day, intra-day.

# Frequently Asked Questions

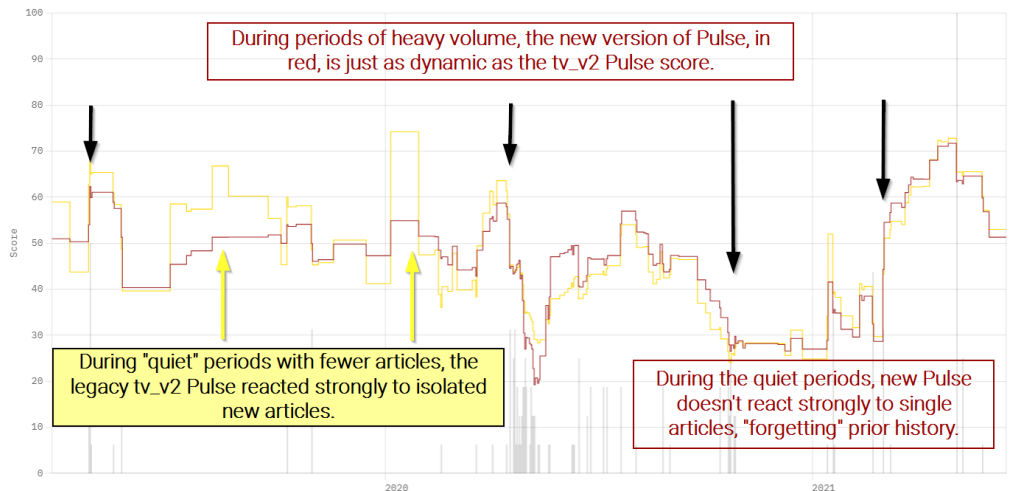
## What are the methodology and scoring changes?

The new scoring methodology changes include key improvements to Insight and Pulse calculations along with Spotlights:

- Insight:** Now an Exponentially Weighted Moving Average (EWMA) of underlying article scores – instead of an average of daily Pulse scores as in tv\_v2. This leads to an active Insight score that responds more quickly during periods of high volume of articles and stays stable during periods of low or no volume of articles.



- Pulse:** A new volume-weighted decay rate is applied based on how many newer articles have been found. This leads to a Pulse score that "remembers" more information in a sparse time series when articles may be months apart. The new Pulse does not continue incrementally reducing the influence of older articles – unless there are new ones. The new Pulse algorithm can still respond quickly to incoming signals.



- Spotlights:** Intra-day Spotlights generating and delivering insights sooner.

## Are there additional data attributes now available?

Yes, FactSet will now be providing **Daily and Accumulative Volume** figures, offering a view into the current day's article volume along with the full history for a given entity.

- **Daily Volume** captures the daily article volume ingested by FactSet for a given company.
- **Accumulative Volume** captures the count of articles for a given company across its full history, extending beyond the Trailing Twelve Month (TTM) Volume, painting a more complete picture.

---

## How were the data quality improvements assessed?

FactSet executed a two-pronged approach to evaluating data quality improvement:

- **Big Data Review:** applied generative artificial intelligence to the content to deeply extract 'benchmark' entities, topics, and sentiment measurements and then applied a [mathematical evaluation technique](#).
- **Sample Review:** pulled a subset of reprocessed articles (v3) and had subject matter experts on the ESG AI Insights team compare results against the current pipeline (v2) in an independent labeling exercise.

---

## How will the data be delivered?

The target delivery date is **September 2024** with all front-end clients and those utilizing formulas switched over to v3 in a single day. Standard Data Feed clients will also have access to the v3 data set, with full history, on the same day with our Client Solutions team ready to support migration.

---

## How has the data expanded in coverage?

FactSet continues to expand coverage across multiple channels. Highlights include:

- **Languages:** broadened foreign language coverage to 38 total languages, covering all official languages spoken in the EU;
- **Entity Universe:** expanded coverage to 268,000+ companies;
- **Data Sources:** partnered with new vendors and expanded our reach to over 200,000 news sources;
- **Source Types:** started ingesting ESG news from podcast transcripts and newsletters in addition to our existing coverage of traditional media, NGOs, Industry & Trade Publications, etc.

What new languages have been added and when?

FactSet strives to present a truly global review within our time series and thus has worked to ingest articles both in native and foreign languages, even when the same article is published in multiple languages. Since 2022 we have added the below languages and as part of the v3 pipeline we mined over 330 million articles across these languages along with our existing set to find new, meaningful ESG data the company level.

**2022:**

- Polish
- Ukrainian
- Hungarian
- Hebrew
- Croatian
- Slovenian
- Lithuanian
- Icelandic
- Thai
- Romanian
- Slovakian
- North Macedonian
- Greek
- Serbian

**2023:**

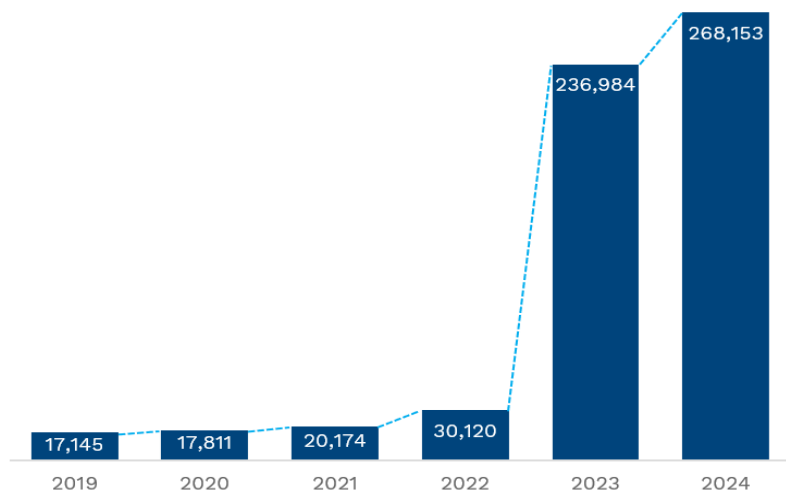
- Russian
- Vietnamese
- Korean
- Turkish
- Arabic
- Czech

**2024:**

- Hindi
- Latvian
- Estonian
- Bosnian

What is the timeline of entity expansion?

FactSet continually looks to expand entity coverage through client requests, but also internal monitoring of corporate actions and research on companies in the broader FactSet ecosystem. FactSet did add entities at scale in 2022 to significantly expand our private entity offering.



What are major private entities getting a boost to coverage?

FactSet added many historical articles for high-volume companies such as **Ikea, Huawei, Cargill**; but also, for medium and low-volume companies such as **Stabia Oil Co. Srl, Clif Bar & Co., and Kleiner Perkins Caufield & Byers, LLC**. The average number of net new articles for these companies is in the range of 200-1000, making a valuable addition to the time series.

How has the makeup or count of sources covered changed?

On an ongoing basis FactSet reviews and refines the aggregation of sources that provide ESG data points. **We've increased source coverage from ~100k+ in 2021 to 200K+ in 2024** due to in-house source additions and strategic partnerships with new data vendors.

Articles are ingested into our system from various source types. The proportion of each type of source in the data varies depending on the industry, ESG category, region, and time frame. We still carry the same source types with the addition of podcast transcripts. An updated high-level curated source type breakout is included below.

**Source Types:** NGO, News, Local News, Industry, Trade Union, Government, Podcast Transcripts, Academic & Networks, Legal News.

Types of Curated Sources 2024

