# Explainable and Transparent ESG Investment Methodologies

With Databricks and GCP

By Antoine Amend

While 83% of C-suite leaders and investment professionals expect their environmental, social and governance (ESG) programs will contribute more shareholder value 5 years from now,[1] there's little agreement on what data companies should collect and how they should disclose and analyze it. Proliferation of company disclosures, rankings and metrics and a lack of data standards make it hard for asset managers to measure ESG actions and goals at scale. For far too many actors, ESG verification is essentially an honor code by which market participants pledge to follow the rules. However, this lack of accountability has recently led regulators such as the Securities and Exchange Commission to step up scrutiny[2] with the climate disclosure proposal, creating urgency for capital markets to take a more comprehensive and quantitative approach to analyzing ESG data.
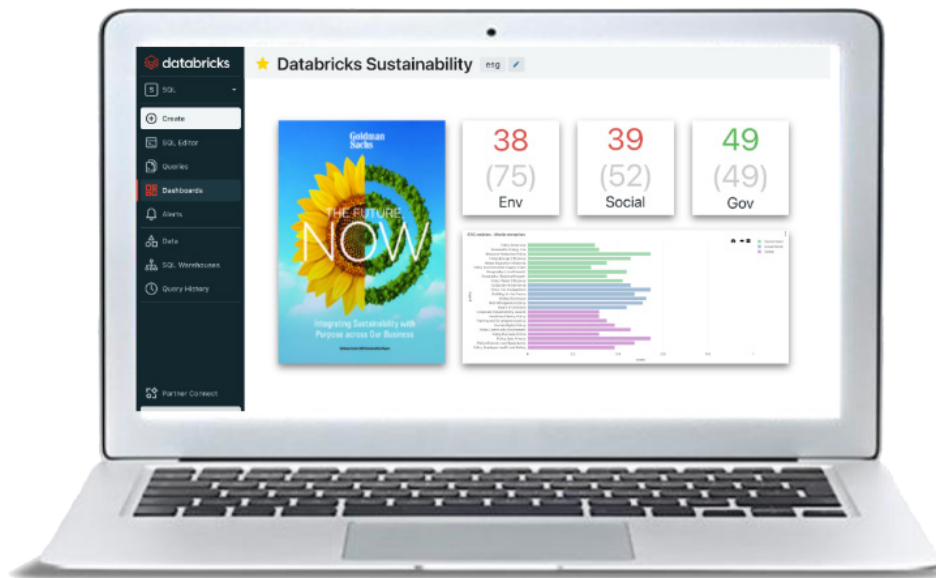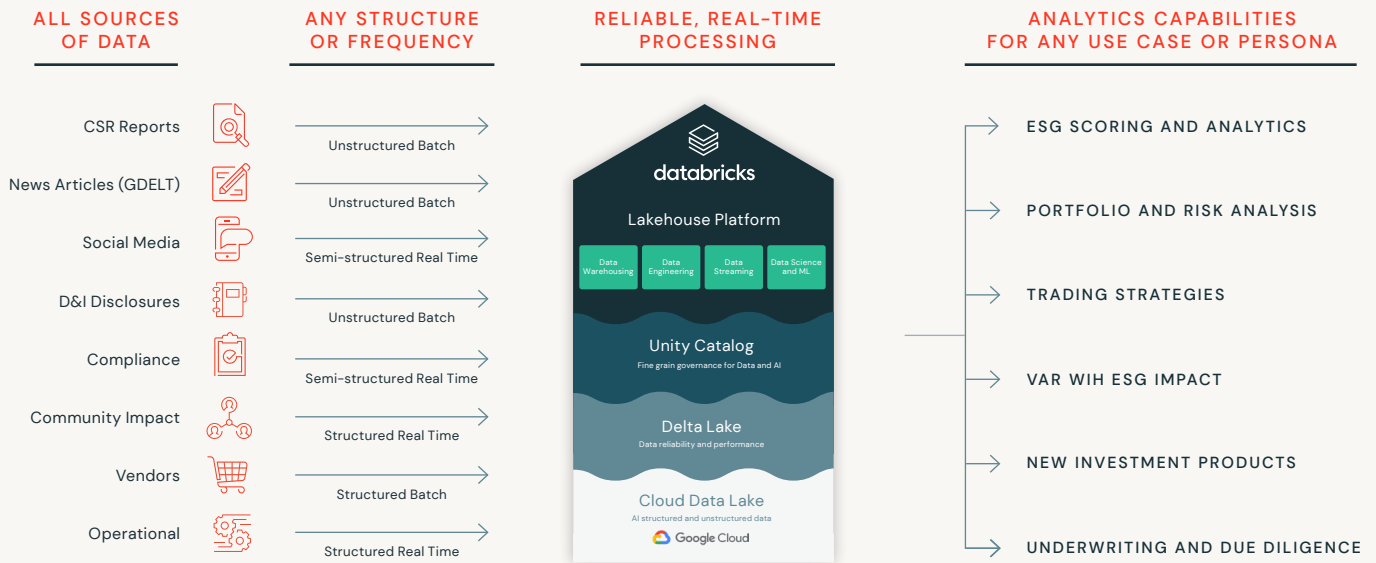

Analyze ESG performance

With over 40,000 downloads, our ESG Solution Accelerator was designed with the asset management community in mind to help them move beyond subjective ratings and build an objective, 360-degree view of their ESG exposure and that of their multi-asset portfolios. Our solution focuses on learning strategic initiatives from corporate disclosures using natural language processing (NLP) techniques that can be quantified and tracked using news analytics and public data feeds such as the Global Database of Event, Location and Tones (GDELT). This approach makes ESG analytics possible for both private and public companies.

---

[1] The ESG premium: New perspectives on value and performance
[2] Goldman Sachs Is Being Investigated Over E.S.G. Funds

ALL SOURCES OF DATA

- CSR Reports
- News Articles (GDELT)
- Social Media
- D&I Disclosures
- Compliance
- Community Impact
- Vendors
- Operational

ANY STRUCTURE OR FREQUENCY

- Unstructured Batch
- Unstructured Batch
- Semi-structured Real Time
- Unstructured Batch
- Semi-structured Real Time
- Structured Real Time
- Structured Batch
- Structured Real Time

RELIABLE, REAL-TIME PROCESSING

databricks

Lakehouse Platform

| Data Warehousing | Data Engineering | Data Streaming | Data Science and ML |

Unity Catalog
Fine grain governance for Data and AI

Delta Lake
Data reliability and performance

Cloud Data Lake
AI structured and unstructured data

Google Cloud

ANALYTICS CAPABILITIES FOR ANY USE CASE OR PERSONA

- ESG SCORING AND ANALYTICS
- PORTFOLIO AND RISK ANALYSIS
- TRADING STRATEGIES
- VAR WIH ESG IMPACT
- NEW INVESTMENT PRODUCTS
- UNDERWRITING AND DUE DILIGENCE

Lakehouse for ESG investing

Embracing our [Lakehouse for Financial Services](#) and the following Solution Accelerators, many of our customers are able to combine market data with alternative data on one platform to successfully quantify, monitor and communicate their ESG strategies in a more transparent and objective manner.

ESG Scoring

Take a quantitative view into sustainability and ensure companies are accountable for their actions

Available on GitHub

Modern Risk Management

Adopt a more agile approach to risk management by unifying data and AI in the lakehouse

Available on GitHub

Regulatory Reporting

Ensure reliability and timeliness in the collection, processing and exchange of regulatory data

Available on GitHub

With Databricks, customers like Credit Suisse and S&P are able to evaluate the relevance and efficacy of their ESG data engineering and analytics in a matter of days instead of months, strategies that they later turned into financial products and monetization vehicles.

CREDIT SUISSE
Customer story →

S&P Global
Customer story →

ESG on the Google Cloud Platform

Building upon the success of hundreds of customers adopting the Lakehouse for Financial Services, we extended our ESG solution to our GCP customers by taking full advantage of the Google Cloud ecosystem with additional capabilities such as Google BigQuery and Looker. Public data sets such as GDELT, already available on BigQuery, can now be accessed out of the box alongside internal data to supercharge risk models with news analytics covering the "three quarters of a trillion emotional scores"[3] required to fully tackle the subjective challenge of corporate disclosures.
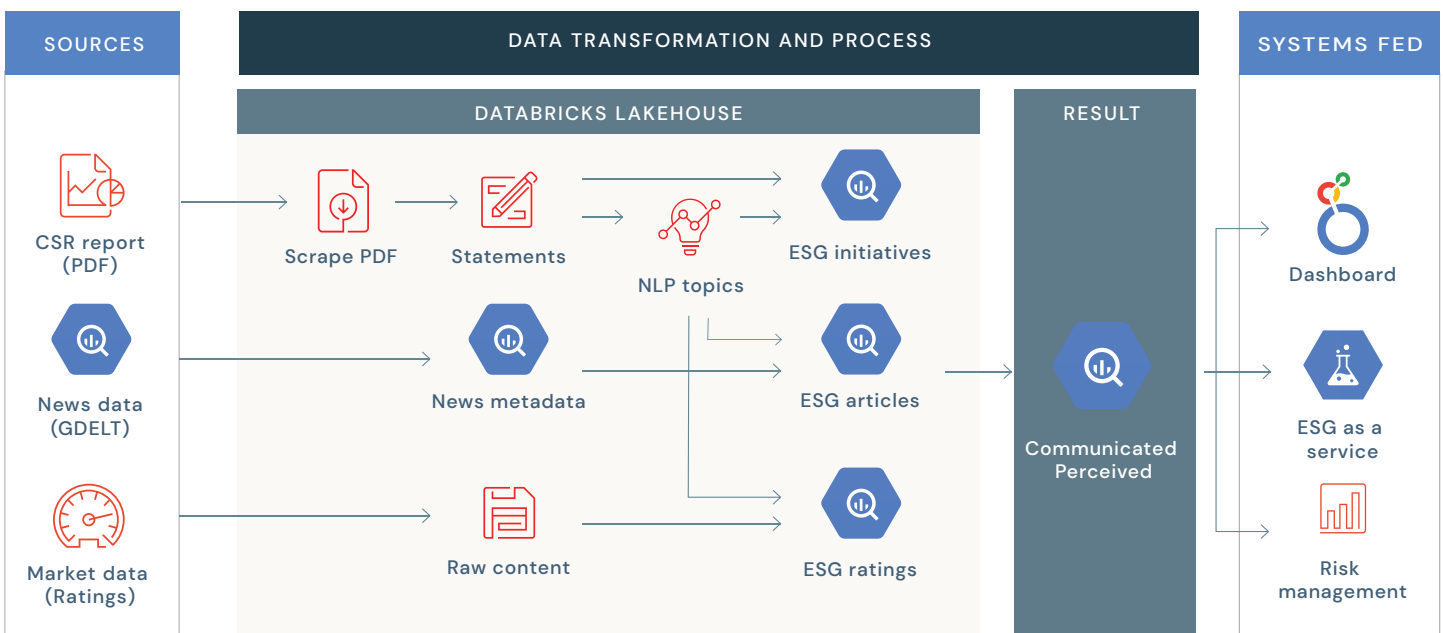

The availability of those news feeds (updated every 15 minutes) coupled with high-performance compute offered by the Databricks Lakehouse on the Google Cloud Platform allows asset managers to not only access ESG insights in near real-time through production dashboards, but also turn those insights into actionable signals that can be fed back to risk models and backtesting of trading strategies.

Adopting industry standards

Together with 50+ financial institutions such as Goldman Sachs, Deutsche Bank and RBC Capital Markets as well as technology providers such as Google Cloud, Databricks joined the Fintech Open Source Foundation (FINOS) to drive collaboration and standardization through open source initiatives and open data standards. Goldman's flagship data management and data governance platform, Legend, open sourced through FINOS[4] and now adopted as an industry standard for the extension of the CDM, shows great potential and interest for the modeling of ESG reference data.

With the Legend framework integration to the Lakehouse for Financial Services[5] and its native support for the recently announced open-sourced Delta Lake,[6] industry data models used for the collection, processing and exchange of ESG data can be converted into efficient operation pipelines.
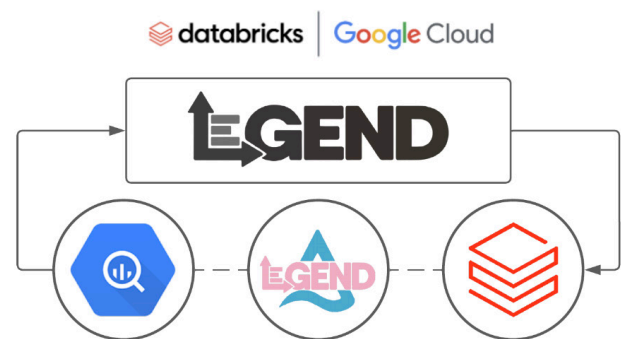
> "The Legend Delta Lake integration is the latest example of how open source collaboration between financial institutions and technology firms, without the need for bilateral proprietary agreements, can deliver immense value-add to the financial services industry in the AI/ML field."
>
> — Gabriele Columbro, Executive Director, FINOS

Coupled with both BigQuery and Delta Sharing capabilities, this framework closes the loop of ESG investing by allowing data interoperability between systems and simplifying regulatory reporting, creating an environment where every market constituent becomes accountable for their ESG initiatives.

Learn more about the Databricks Lakehouse for Financial Services and ESG investing: dbricks.co/fiserv

[4] Goldman Sachs Open Sources its Data Modeling Platform through FINOS
[5] github.com/finos/legend-delta
[6] Delta Lake project announces the availability of 2.0 Release Candidate