Straive
Connecting the dots

Extracting ESG Data

Extracting ESG data from company disclosures like annual reports, CSRs, 10K filings and transforming them to fit into the desired ESG framework like SASB and TCFD - needs advanced extraction capabilities.

Objective

Extracting ESG data from disclosures is the primary methodology for ESG data strategy today. While quantitative metrics are easy to track and are similar across strategies, qualitative data points vary depending on multiple factors. The data extracted also needs to fit into an ESG framework, including SASB, TCFD, and custom frameworks, depending on the overall ESG and investment strategy. Sometimes this may also take the form of scores, in which case, the metrics must be transformed accordingly.

Business Challenges

Qualitative data points pose a unique challenge. They can produce different results. Variations can occur due to the data owner's definition, the engine's training context, or the engine or analyst extracting the data.
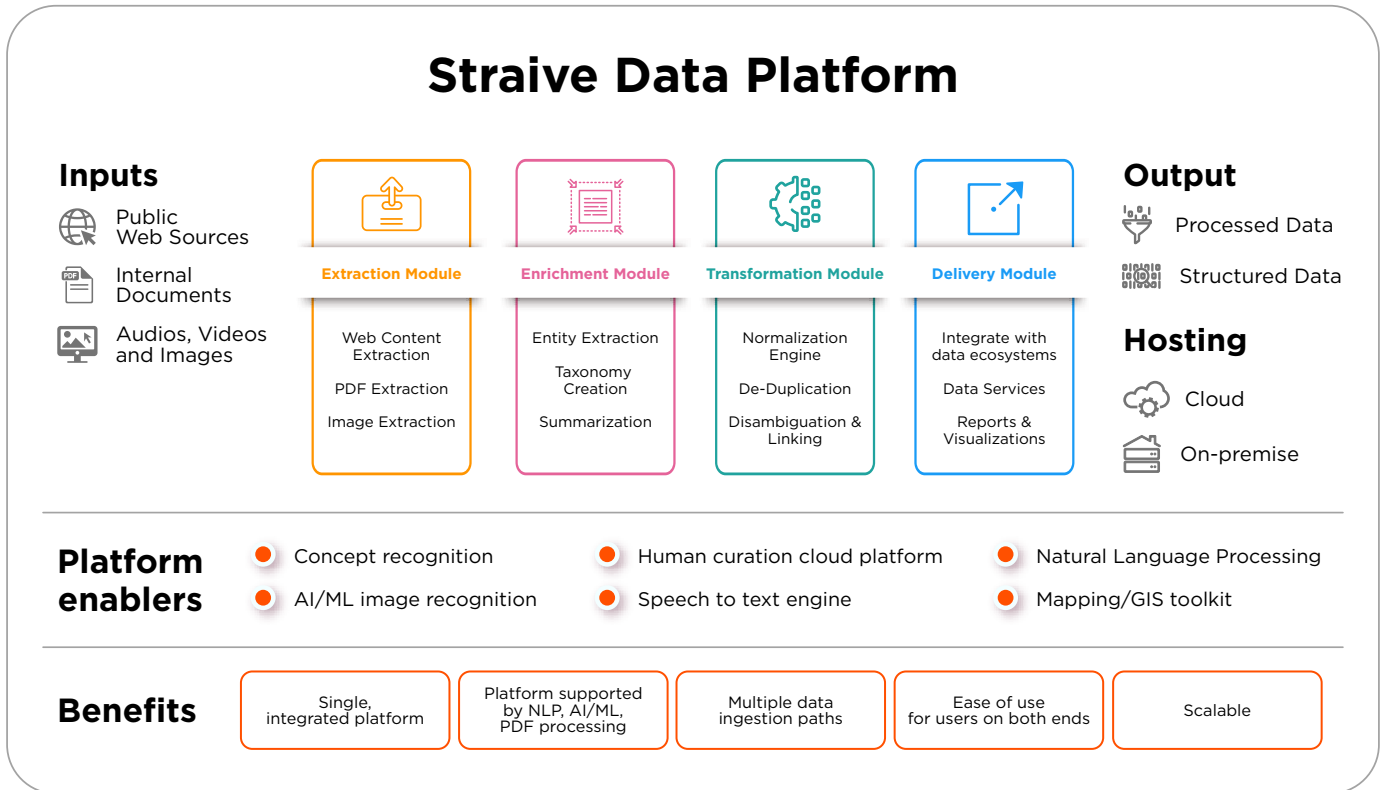
Other challenges companies face while extracting the core ESG data from disclosures include:

- Removing analyst's bias
- Timely tracking of multiple sources that are published throughout the year
- Transforming data to meet the enterprise ESG frameworks, which requires a level of nimble customization
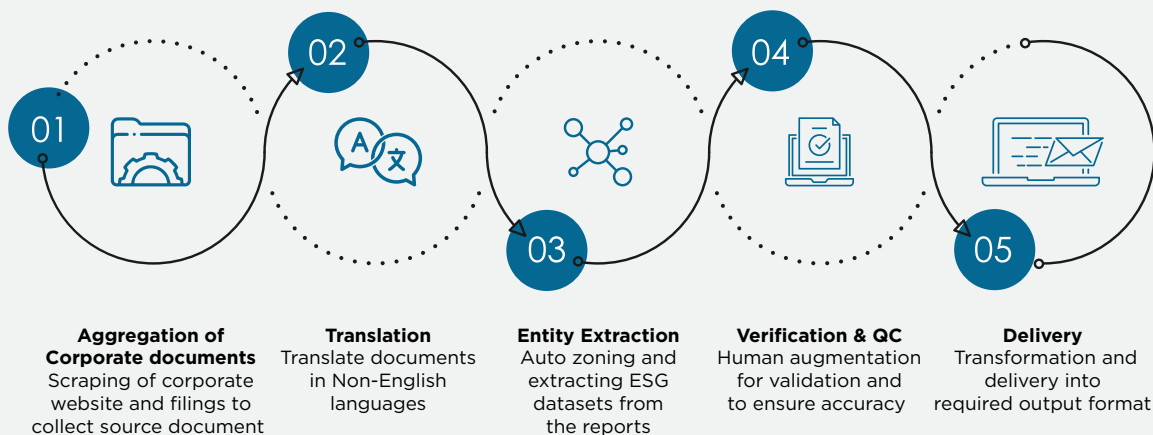
Straive Data Platform (SDP) uses an advanced auto extraction data engine to pull data from Annual Reports, CSR Reports, etc.

Straive Data Platform

Inputs
- Public Web Sources
- Internal Documents
- Audios, Videos and Images

Extraction Module
- Web Content Extraction
- PDF Extraction
- Image Extraction

Enrichment Module
- Entity Extraction
- Taxonomy Creation
- Summarization

Transformation Module
- Normalization Engine
- De-Duplication
- Disambiguation & Linking

Delivery Module
- Integrate with data ecosystems
- Data Services
- Reports & Visualizations

Output
- Processed Data
- Structured Data

Hosting
- Cloud
- On-premise

Platform enablers
- Concept recognition
- AI/ML image recognition
- Human curation cloud platform
- Speech to text engine
- Natural Language Processing
- Mapping/GIS toolkit

Benefits
- Single, integrated platform
- Platform supported by NLP, AI/ML, PDF processing
- Multiple data ingestion paths
- Ease of use for users on both ends
- Scalable

ESG Core Data & Financials

01 Aggregation of Corporate documents
Scraping of corporate website and filings to collect source document

02 Translation
Translate documents in Non-English languages

03 Entity Extraction
Auto zoning and extracting ESG datasets from the reports

04 Verification & QC
Human augmentation for validation and to ensure accuracy

05 Delivery
Transformation and delivery into required output format

Business Benefits

- Customized data exploration UI for individual analyst needs
- Can be set up for any publicly available source or internal sources
- ESG Taxonomy defined for every use case
- Automatic update-tracking of sources