



eurofidai
— UAR CNRS / ESSEC 3390

EUROPEAN FINANCIAL DATA INSTITUTE

Powered by



CLARITY AI

DATA DESCRIPTIONS GUIDE

ESG RAW DATA ON CORPORATIONS

User guide



The European Financial Data Institute (EUROFIDAI – UAR3390) is a public academic institute attached to the CNRS (French National Center for Scientific Research) and the ESSEC Business School

Overview

1. Presentation¹

The Clarity AI ESG Risk – Raw Data on Corporations database, distributed by EUROFIDAI, is designed for academic research purposes and can only be used by researchers of academic institutions. A great deal of effort and extensive resources go into making the database accurate. Our added value is to provide academics with verified, controlled and homogeneous data.

The Clarity AI ESG Risk – Raw Data on Corporations database provides:

- 105 Risk Raw Data metrics on corporations at the issuer level

The ESG data is group in three pillars:

- Environmental (E)
- Social (S)
- Governance (G)

2. Coverage

The coverage varies per metric, as it is composed on company's reports and estimations. The database covers over 70,000 companies for some metrics, but generally Clarity AI bases its coverage in the collection universe, which are the companies that Clarity actively look for. The coverage begins in Q1 2016.

3. Data sourcing

Clarity AI builds its database from a wide and varied collection of sources:

- Data sourced from specialist **data providers** and generalist financial data providers.
- **Global news** monitoring service, which provides Clarity AI with access to a universe of 8,500+ media publishers from 200 countries, with 1.4M+ articles added per day, from 33,000+ sources.
- Data parsed directly from **public sources** (e.g. public reports, company websites).
- Network of **corporate** partners providing own first-hand data.
- Data **estimated** through machine learning to complement information not disclosed by organizations.

4. Estimated data

Clarity AI provides high-accuracy estimates thanks to proprietary machine learning models. Their main characteristics are the following:

- Metric-specific models analyze more than 300 company features (e.g., sub-industry, location, supply chain, revenues, business lines...).

¹ References: Clarity AI ESG Risk Methodology, Nov. 2021, and Clarity AI Raw Data Dictionary, July 2023, © Clarity AI.

- Machine learning algorithms account for non-linear effects and feature interactions (i.e. relations between features).
- Unlike other providers, Clarity AI's models are optimized for intensity estimation (e.g., tCO₂/M\$) rather than output metrics (e.g., tCO₂), and calibrated to get optimal rankings.
- Model output assessed through cross-validations over different out-of-sample data sets.

5. Data types

Raw data integrated in Clarity AI's database can be of two types.

Data type	Description	Clarity AI metric type	Example
Structured quantitative data	ESG data that is numeric (absolute numbers, ratio and indexes)	Quantitative	Tons of Scope 1 CO ₂ emissions
Structured qualitative data	ESG data that indicates whether or not a company has a specific policy related to an ESG subject	Policy	Board Gender Diversity Policy

6. Granularity and Year Methodology

Clarity AI Raw dataset has organization and Year level granularity. An organization is a legal entity, also called "Issuer" in the industry. Clarity AI does not apply inheritance in Raw Data, meaning that subsidiaries will not have data from their parents (companies with controlling interest) imputed. If an organization has data, it's because it directly reports it, or it was estimated for the company itself. Year level granularity: Clarity AI considers the fiscal year a "year" (or metric year). Clarity AI applies a logic for its consideration of fiscal year to help normalizing data and comparability of companies.

Fiscal Year Methodology:

If a Non-US company's fiscal year ends on or before January 15th, the associated figures will be attributed to the previous year i.e. if a Non-US company's period ends on January 15th 2022, then the metrics of that company will be assigned to the 2021 metric year.

If a US company's fiscal year ends on or before February 10th, then the associated figures will be attributed to the previous year i.e. if a US company's period ends on January 10th 2022, then the metrics of that company will be assigned to the 2021 metric year.

7. Data reliability

Datasets are combined and cleaned using reliability algorithms. Sourced data goes through advanced quality checks and standardization techniques that allow Clarity AI to select the most reliable source for each metric.

The reliability process is specific to each data type used by Clarity AI, and its core principles are specified below:

- Use of multiple sources to contrast underlying data and enable **consistency checks**;
- Machine learning and domain expertise for **automatic best source selection**, correcting inconsistencies in quantitative data sources for data reported by companies themselves;

- Data-driven approach to build a trustworthy database free of **data errors**: cleaning, standardization and normalization
- For Clarity AI proprietary data resulting from machine learning models, such as estimated quantitative metrics and controversies, quality is ensured by validating the predictive accuracy of the models against datasets already curated and checked by experts:
 - Quantitative metrics are validated based on whether the models are able to predict available, out-of-sample reported data with accuracy;
 - Controversies are validated based on whether the models are able to detect, classify and score controversies at scale at least as well as a human ESG controversies expert.

8. Frequency of data

The data is reported annually with monthly updates.

9. Metric selection

Clarity AI has selected its metric catalogue taking into account two criteria: consensus across the industry (i.e., significant use and acceptance by industry standards and players), and coverage (i.e., high data availability to enable meaningful comparisons across companies). In order to calculate ESG Risk scores for companies, Clarity AI makes available 105 different metrics. To view the complete list, please consult Table 2.

Data

Data extraction

The extraction of the Clarity AI ESG Risk – Raw Data on Corporations database can be done by specifying a number of criteria:

1. Period: The period covered by your analysis. This is optional. By default, the period covered is set from Q1, 2016 to today.
2. Codes: Search can be done using three different criteria:
 - a. company's name; or
 - b. company's Perm_ID; or
 - c. ISIN code²

Extraction results

Table 1 provides lists the variables in the extraction result file:

Table 1: Extraction variables	
Variable	Description
company_name	Name of the company
data_type	Type of data: <ul style="list-style-type: none"> - Float (numeric) - Int (boolean)
disclosure	Methodology for calculating the metric <ul style="list-style-type: none"> - Calculated: Datapoint computed by Clarity AI on the basis of other datapoints; - Estimated: Datapoint with no info but calculated using estimation models provided by Data Science or using NLP; - Reported: Datapoint as reported by the company; - Adjusted: The value is established based on a publicly communicated datapoint with some corrections from other similar datapoints; - Imputed: the aforementioned (ADJUSTED) plus internal Clarity AI research.
headquarter_country	Country of the firm's headquarter.
metric_description	Description of the metric.
metric_name	Name of the metric.
metric_period	Frequency of the metric (i.e. PY1 refers to one year of coverage).
metric_unit	Measurement unit of the metric.
metric_value	Value of the metric.
metric_year	Fiscal year of the company for which the metric is calculated / gathered.
nb_points_of_observations	Number of metric_value for a given perm_id.
perm_id	Permanent identifier number of the company.

² A company may have numerous ISIN codes, one for each financial instruments issued by the company.

pillar	ESG flag that identifies the scope concerned by the metric. Environmental (E), Social (S), and Corporate Governance (G).
provider_name	Name of the data provider.
reported_date	Datetime that identifies the end of the reporting period for the metric. Clarity AI uses ISO 8601 and all the date-time are UTC(i.e: 2020-01-30T00:00:00Z).

Metric name

Table 2 reports the name and the description of the raw 105 metrics for corporation.

Table 2: ESG Metrics			
Metrics	Pillar	Description	Units
AIRPOLLUTANTS_DIRECT	E	External cost of pollutants released to air by the consumption of fossil fuels and production processes which are owned or controlled by the company. Direct external environmental impacts are those impacts that a company has on the environment through its own activities. Trucost applies monetary values to air pollutant quantities, which represents the global average damage of each environmental impact. All values employed are secondary - the synthesis of existing published and unpublished literature.	USD (000)
AIRPOLLUTANTS_INDIRECT	E	External cost of indirect emissions of pollutants to the air by the consumption of fossil fuels and production processes.	USD (000)
ANALYTICAUDITCOMMIND	G	Percentage of independent board members on the audit committee as stipulated by the company.	%
ANALYTICBOARDFEMALE	G	Percentage of females on the board.	%
ANALYTICCEO_CHAIRMAN_SEPARATION	G	Does the CEO simultaneously chair the board or has the chairman of the board been the CEO of the company?	Yes / No
ANALYTICCOMPCOMMIND	G	Percentage of independent board members on the compensation committee as stipulated by the company.	%
ANALYTICCSR_COMP_INCENTIVES	G	Is the senior executive's compensation linked to CSR/H&S/Sustainability targets?	Yes / No
ANALYTICEMPLOYMENTCREATION	G	Employment growth over the last year.	% growth in number of employees over last year
ANALYTICESTIMATEDCO2TOTAL	E	The estimated total CO2 and CO2 equivalents emission in tons. Total	Tons CO2e

		CO2e Emissions (Scope 1 + Scope 2 + Scope 3)	
ANALYTICINDEPBOARD	G	Percentage of independent board members as reported by the company.	%
ANALYTICNOMINATIONCOMMIND	G	Percentage of non-executive board members on the nomination committee.	%
ANALYTICNONAUDITAUDITFEESRATIO	G	All non-audit fees divided by the audit and audit-related fees paid to the group auditor.	% (Non-audit USD / audit USD)
ANALYTICNONEXECBOARD	G	Percentage of non-executive board members.	%
ANALYTICQMS	G	Does the company claim to apply quality management systems, such as ISO 9000, Six Sigma, Lean Manufacturing, Lean Sigma, TQM or any other similar quality principles?	Yes / No
ANALYTICTOTALDONATIONS	S	Total amount of all donations divided by net sales or revenue.	% (USD donated per million of revenue USD)
ANALYTICWASTERECYCLINGRATIO	E	Total recycled and reused waste produced in tons divided by total waste produced in tons.	%
ANALYTIC_ANTI_TAKEOVER_DEVICES	G	The number of anti-takeover devices in place in excess of two.	Yes / No
ANALYTIC_AUDIT_COMM_EXPERTISE	G	Does the company have an audit committee with at least three members and at least one "financial expert" within the meaning of Sarbanes-Oxley?	Yes / No
ANALYTIC_VOTING_RIGHTS	G	Are all shares of the company providing equal voting rights?	Yes / No
ANIMAL_TESTING_REDUCTION	S	Has the company established a program or an initiative to reduce, phase out or substitute for animal testing?	Yes / No
ANNUAL_MEDIAN_COMPENSATION	G	Median annual compensation for the entire company, excluding the highest paid employee.	USD
AUDITCOMMNONEXECCMEMBERS	G	Percentage of non-executive board members on the audit committee as stipulated by the company.	%
AVGTRAININGHOURS	G	Average hours of training per year per employee.	Hours/ employee
BIODIVERSITY_IMPACT_REDUCTION	E	Does the company report on its impact on biodiversity or on activities to reduce its impact on the native ecosystems and species, as well as the biodiversity of protected and sensitive areas?	Yes / No

BOARDMEETINGATTENDANCEAVG	G	The average overall attendance percentage of board meetings as reported by the company.	%
BRIBERY_AND_CORRUPTION_PA_INSUFFICIENT_ACTIONS	G	Cases of insufficient action taken to address breaches of standards of anti-corruption and anti-bribery.	Yes / No
CALL_MEETINGS_LIMITED_RIGHTS	G	Has the company limited the rights of shareholders to call special meetings?	Yes / No
CEO_ANNUAL_COMPENSATION	G	Total annual compensation for the highest paid individual (normally the CEO).	USD
CEO_PAY_RATIO_MEDIAN	G	Already reported ratio of CEO compensation over median annual employee compensation.	Ratio
CLIMATE_CHANGE_RISKS_OPP	E	Is the company aware that climate change can represent commercial risks and/or opportunities?	Yes / No
CO2DIRECTSCOPE1	E	Scope 1 emissions that occur within a company's organizational boundary from sources that the company owns or controls in tons of CO2e.	Tons CO2e
CO2INDIRECTSCOPE2	E	Scope 2 emissions that result from the generation of purchased electricity, heating, cooling, and steam in tons of CO2e.	Tons CO2e
CO2INDIRECTSCOPE3	E	Scope 3 emissions that occur in the value chain of the reporting company in tons of CO2e.	Tons CO2e
CO2_NO_EQUIVALENTS	E	The estimated total CO2 emission in tons (without counting CO2 equivalents).	Tons CO2
COMMMEETINGSATTENDANCEAVG	G	The average overall attendance percentage of board committee meetings as reported by the company.	%
COMPCOMMONONEXECCMEMBERS	G	Percentage of non-executive board members on the compensation committee as stipulated by the company.	%
CONFORMANCE_OECD_MNE	S	Does the company claim to follow the OECD Guidelines for Multinational Enterprises?	Yes / No
CONFORMANCE_UN_GUID	S	Does the company have clear reference to conformance with the UN Guiding Principles on Business and Human Rights?	Yes / No
CSR_REPORTINGGRI	G	Is the company's CSR report published in accordance with the GRI guidelines?	Yes / No
CSR_REPORTING_EXTERNAL_AUDIT	G	Does the company have an external auditor of its CSR/H&S/Sustainability report?	Yes / No

DAY_CARE_SERVICES	S	Does the company claim to provide day care services for its employees?	Yes / No
ECO_DESIGN_PRODUCTS	E	Does the company report on specific products which are designed for reuse, recycling or the reduction of environmental impacts?	Yes / No
ELECTRICITYPURCHASED	E	Electricity purchased in gigajoules.	GJ
ELIMINATION_CUM_VOTING_RIGHTS	G	Has the company reduced or eliminated cumulative voting in regard to the election of board members?	Yes / No
EMPLOYEEFATALITIES	S	Number of employee fatalities resulting from operational accidents.	Number of fatalities
EMPLOYEE_HEALTH_SAFETY_POLICY	S	Does the company have a policy to improve employee health & safety?	Yes / No
EMS_CERTIFIED_PCT	E	The percentage of company sites or subsidiaries that are certified with any environmental management system.	%
ENERGYPURCHASEDDIRECT	E	Direct energy purchased in gigajoules.	GJ
ENERGYUSETOTAL	E	Total energy consumed by a company within its operational control, including energy produced and purchased (e.g., electricity, steam, natural gas, coal, oil).	GJ
ENV_INVESTMENTS	E	Does the company report on making proactive environmental investments or expenditures to reduce future risks or increase future opportunities?	Yes / No
ENV_SUPPLY_CHAIN_MGT	E	Does the company use environmental criteria (ISO 14000, energy consumption, etc.) in the selection process of its suppliers or sourcing partners?	Yes / No
E_WASTE_REDUCTION	E	Does the company report on initiatives to recycle, reduce, reuse, substitute, treat or phase out e-waste?	Yes / No
GENDER_PAY_GAP_PERCENTAGE	S	Percentage of remuneration of women to men.	%
GLOBAL_COMPACT	G	Has the company signed the UN Global Compact?	Yes / No
GRIEVANCE_REPORTING_PROCESS	S	Does the company have a formal grievance reporting process for concerns reporting misconduct or ethical concerns?	Yes / No
HAZARDOUSWASTE	E	Total amount of hazardous waste produced in tons.	Tons
HUMAN_RIGHTS_CONTRACTOR	S	Does the company report or show to use human rights criteria in the selection or monitoring process of its suppliers or sourcing partners?	Yes / No
HUMAN_RIGHTS_POLICY_DUEDILIGENCE	S	Does the company have a due diligence process in place to prevent/ act upon	Yes / No

		human right abuses? Both in its operations and suppliers' operations?	
HUMAN_RIGHTS_VIOLATION_PAI	S	Number of identified cases of severe human rights issues and incidents.	Number of breaches
IMPROVEMENT_TOOLS_BUSINESS_ETHICS	S	Does the company have appropriate communication tools (whistle blower, ombudsman, suggestion box, hotline, newsletter, website, etc.) to improve general business ethics?	Yes / No
ISO14000	S	Does the company claim to have an ISO 14000 or EMS certification?	Yes / No
LABELED_WOOD	E	Does the company claim to produce, source or distribute wood or forest products that are labeled (e.g., Forest Stewardship Council (FSC))?	Yes / no
LOSTWORKINGDAYS	S	Total working days lost by the workforce (employees and contractors), due to some form of incapacity, including occupational diseases and injuries.	Number of days
NATURAL_RESOURCE_USE_DIRECT	E	External cost of the direct extraction of minerals, metals, natural gas, oil, coal, forestry, agriculture and aggregates by the company. Direct external environmental impacts are those impacts that a company has on the environment through its own activities.	USD (000)
NOXEMISSIONS	E	Total amount of NOx emissions emitted in tons.	Tons of NOx
N_OXS_OX_EMISSIONS_REDUCTION	E	Does the company report on initiatives to reduce, reuse, recycle, substitute, or phase out SOx (sulfur oxides) or NOx (nitrogen oxides) emissions?	Yes / No
ORGANIC_PRODUCTS_INITIATIVES	E	Does the company report or show initiatives to produce or promote organic food or other products?	Yes / No
PARTICULATE_MATTER_EMISSIONS	E	Amount of particulate matter (a complex mixture of small liquid droplets and solid particulates suspended in the air) emitted by the company.	Tons
POLICY_BOARD_DIVERSITY	S	Does the company have a policy regarding the gender diversity of its board?	Yes / No
POLICY_BRIBERYAND_CORRUPTION	S	Does the company have a policy and/or processes in place to avoid bribery and corruption in all its operations?	Yes / No
POLICY_BUSINESS_ETHICS	S	Does the company describe in the code of conduct that it strives to maintain the highest level of general business ethics?	Yes / No

POLICY_CHILD_LABOR	S	Does the company have a policy to avoid the use of child labor?	Yes / No
POLICY_DATA_PRIVACY	S	Does the company have a policy to protect customer and general public privacy and integrity?	Yes / No
POLICY_EMISSIONS	E	Does the company have a policy to improve emission reduction?	Yes / No
POLICY_FORCED_LABOR	S	Does the company have a policy to avoid the use of forced labor?	Yes / No
POLICY_FREEDOMOF_ASSOCIATION	S	Does the company describe, claim to have or mention the processes in place to ensure the freedom of association of its employees?	Yes / No
POLICY_HUMAN_RIGHTS	S	Does the company have a policy to ensure the respect of human rights in general?	Yes / No
POLICY_SUSTAINABLE_PACKAGING	E	Does the company have a policy to improve its use of sustainable packaging?	Yes / No
POLICY_WATER_EFFICIENCY	E	Does the company have a policy to improve its water efficiency?	Yes / No
RENEWENERGYCONSUMED	E	Total renewable energy consumed in Gigajoules.	GJ
RENEWENERGYPRODUCED	E	Total renewable energy produced for self-consumption in Gigajoules.	GJ
RENEWENERGYPURCHASED	E	Total renewable energy purchased for self-consumption in Gigajoules.	GJ
SOXEMISSIONS	E	Total amount of SOx emissions emitted in tons.	Tons of SOx
SUPPLY_CHAINHS_POLICY	S	Does the company have a policy to improve employee health & safety in its supply chain?	Yes / No
SUSTAINABLE_BUILDING_PRODUCTS	E	Does the company develop products and services that improve the energy efficiency of buildings?	Yes / No
TAKEBACK_RECYCLING_INITIATIVES	E	Does the company reports about take-back procedures and recycling programs to reduce the potential risks of products entering the environment?	Yes / No
TARGETS_DIVERSITY_OPPORTUNITY	S	Has the company set targets or objectives to be achieved on diversity and equal opportunity?	Yes / No
TARGETS_EMISSIONS	E	Has the company set targets or objectives to be achieved on emission reduction?	Yes / No
TARGETS_WATER_EFFICIENCY	E	Has the company set targets or objectives to be achieved on water efficiency?	Yes / No
TIRTOTAL	S	Rate of total number of injuries (Total Injury Rate TIR) and fatalities in the	Rate

		workforce (employees and contractors) relative to 1 million hours worked.	(Injuries per million of hours worked)
TOXIC_CHEMICALS_REDUCTION	E	Does the company report on initiatives to reduce, reuse, substitute or phase out toxic chemicals or substances?	Yes / No
TRADEUNIONREP	S	Percentage of employees represented by independent trade union organizations or covered by collective bargaining agreements.	%
TRANALYTICRENEWENERGYUSE	E	Total energy consumed from primary renewable energy sources divided by total energy consumed.	%
TURNOVEREMPLOYEES	S	Percentage of employee turnover.	%
VOCEMISSIONS	E	Amount of Volatile Organic Compounds (VOC) emitted by the company.	Tons of VOC
VOC_EMISSIONS_REDUCTION	E	Does the company report on initiatives to reduce, substitute, or phase out volatile organic compounds (VOC)?	Yes / No
WASTETOTAL	E	Total amount of waste produced in tons.	Tons
WASTE_RECYCLED	E	Total waste that is generated by the company and is recycled, reused or composted.	Tons
WASTE_REDUCTION_TOTAL	E	Does the company report on initiatives to recycle, reduce, reuse, substitute, treat or phase out total waste?	Yes / No
WATERWITHDRAWALTOTAL	E	Total water withdrawal in cubic meters.	Cubic meters
WATER_TECHNOLOGIES	E	Does the company develop products or technologies that are used for water treatment, purification or that improve water use efficiency?	Yes / No
WATER_USE_PAI_M10	E	Average amount of water consumed and reclaimed by the investee companies.	Cubic meters / million EUR of revenue of investee companies
WHISTLEBLOWER_PROTECTION	S	Does the company have a provision or comply with regulations protecting whistleblowers?	Yes / No
WOMENEMPLOYEES	S	Percentage of women employees.	%
WOMENMANAGERS	S	Percentage of women managers.	%