Climate TRACE Ownership Information: Source & Company-Level Ownership Emissions

Peter Thomas[1,5], Ishan Saraswat[1,5], Krsna Raniga[1,5], Christy Lewis[1,5], Jeremy Freeman[1,5], Aaron Davitt[1,5], Lauren Schmeisser[2,5], Joseph Fallurin[2,5], Ashank Sinha[3,5], John Heal[3,5], Verity Crane[3,5], and Matthew Jolley[4,5], and Gavin McCormick[1,5]

1) WattTime, 2) RMI, 3) TransitionZero, 4) Hypervine Limited, 5) Climate TRACE

Overview

This document describes how source-level and company-level ownership data were produced for the following sectors: Cement, Oil and Gas Production, Oil and Gas Transport, Oil and Gas Refining, Petrochemicals steam cracking, Aluminum, Iron, Bauxite, Copper and Coal Mining, International and Domestic Aviation, Electricity Generation, Iron and steel manufacturing, and Cattle Operations.

1. Introduction

Independent reports quantifying the emissions generated by private companies and state-owned enterprises (for-profit enterprises owned by governments) are few and far between. For instance, in 2016, Science.org published a report estimating that just 90 companies generated close to two thirds of industrial greenhouse gas (GHG) emissions for the year 2013 (Heede, 2014; Starr, 2016). The underlying datasets for this report, which estimated the companies' yearly emissions based on historical production numbers, took nearly a decade to collect. With regard to state-owned enterprises (SOEs), in 2022, Columbia's SIPA Center for Global Energy Policy published a report estimating that 300 SOEs emitted 7.49 gigatons of CO_2 equivalent in Scope 1 emissions during 2017 (Clark & Benoit, 2022). This report included the caveat that "the true scale of SOE-related emissions is likely to be substantially higher, particularly when accounting for national oil companies and iron and steel manufacturers that do not currently report their emissions". As these two examples demonstrate, aggregating and analyzing the datasets necessary to produce independent emissions estimates for companies and SOEs is incredibly challenging and time-consuming. As a result, such reports are often out of date by the time they are published, and/or limited in the scope of their coverage. Currently, the most frequent and up-to-date emissions estimates for private companies and SOEs are self-reported inventories. There is a plethora of incentives that can potentially bias self-reporting or discourage doing so at all – such as regulatory consequences, public outcry, and loss of investors. The companies whose GHG emissions are disproportionately large are also disproportionately unlikely to self-report their emissions unless they are required to do so. Therefore, having up-to-date, independent emissions estimates with extensive, worldwide coverage for private companies and SOEs is an essential component of the actionable information that investors, policymakers, and activists need in the fight to reduce carbon emissions.

In 2021, Climate TRACE generated the first ever independent, openly accessible source and company-level emissions database – providing ownership information for 4,342 private companies and SOEs, with comprehensive coverage of emissions-generating activity in 8 sectors. For 2022, we expanded our ownership database to include 13,159 private companies and SOEs, located in 234 countries and administrative regions, and with comprehensive coverage for 17 sectors. In Climate TRACE's 2023 data release, source-level ownership was provided for 38% of global estimated emissions. For our release this year, we are again providing source-level ownership data for 38% of global estimated emissions.

Since 2021, we have made ownership data openly available for download directly from our website. This data included the names of direct owners and ultimate parents, as well as the percent share of ownership attributable to direct owners. This year, direct downloads from our website include the names of all identified owners, the percent shares of all direct owners and ultimate parents, and citations for all available primary sources for each ownership claim (either directly in the download, or available via a link in the download to a page on the GEM wiki).

2. Materials

Identification of company-level ownership occurred in two stages. First, datasets containing source-level owners were aggregated into sector-specific datasets. For this analysis, a 'source-level owner' is defined as the company whose name is attached to a single individual source in a dataset. Once source-level ownership datasets were aggregated, research was conducted to identify the source-level owner's ultimate parent company, subsidiaries, joint ventures, and sibling companies. In particular, identifying *higher-level* rungs in the ownership chain was prioritized, with lower levels being included only when they incidentally turned up during research on higher levels. Often, the lowest-level identified owner is not actually the lowest-level subsidiary in the ownership chain for a source, but merely the lowest level that was reported in a primary dataset. Since the goal of the ownership database is to attribute responsibility for emissions to institutions, it was a priority to follow chains all the way up, but not necessarily all the way down. In the download packages, the column "percent_company_datasource" contains citations that link an emissions source to direct owners, and the column "percent_parent_datasource" contains citations that link the direct owner to its ultimate parent. To generate company-level emissions estimates, sources that share an ultimate parent were aggregated together and their emissions were summed. Emissions (and, where applicable, production) estimates for each ultimate parent include the sum of all estimates for associated sources, in proportion to the ultimate parent's percentage of ownership for each one. Download packages contain raw data that can be re-aggregated in this way to reproduce company-level estimates. Pre-aggregated company level estimates can be generated on request.

2.1 Source-level Ownership Datasets by Sector & Methodology

Assembling the Climate TRACE ownership database was a coalition-wide effort with contributions from many members. Information on direct owners (as in, the lowest-level identified owner in the chain of ownership) was provided by RMI, TransitionZero, Hypervine, OceanMind, WattTime, and Global Energy Monitor (GEM). Meanwhile, information connecting these lowest-level, direct owners up to their highest-level ultimate parents (e.g. corporations, investment firms, and governments) was generated by GEM and WattTime. A diverse set of methodologies and data sources were used both to identify direct owners and map them up to ultimate parents. This document provides a broad overview of the methodologies and data sources used for the entire Climate TRACE ownership dataset, with particular specificity and analysis pertaining to the automated mapping method employed by WattTime specifically. For a more detailed discussion of Global Energy Monitor's methods, see Global Energy Monitor's methodology in the [Asset Ownership methodology folder on The Climate TRACE GitHub](#).

For this section, the following source-level ownership terms were used in each sector:

- 'Real property' is defined as a parcel of land along with any permanent fixtures attached to it.
- 'Property' is defined as a tangible object that is not land, nor affixed to it.
- A 'business concern' is a private legal entity formed for the purpose of engaging in commercial activity. A 'state-owned enterprise' (SOE) is defined as a for-profit business owned by a government.
- A 'government agency' is a non-profit entity owned by a government.
- 'Percent financial interest' is defined as the proportion of the monetary value of an asset or of an organization to which the holder of the interest is legally entitled

2.1.1 Global Energy Monitor (GEM) Ownership-Mapping Sectors

Global Energy Monitor (GEM) specializes in providing the highest standard of verification for each individual ownership claim contributed to the Climate TRACE ownership database. GEM employs teams of sector specialists, language experts, and trained ownership researchers to generate both direct and ultimate parent ownership data through desk research. Because GEM derives ownership data from vetted, official primary sources wherever possible, and manually verifies each claim, users can be confident that GEM ownership data is high quality in terms of accuracy and recency. For all sectors in this section, GEM aggregated both direct ownership and mapped direct owners up to ultimate parents.

Electricity Generation. Sources were defined at the level of individual power plants. Ownership was defined in terms of percent financial interest in the asset as a piece of real property, a business concern, state-owned enterprise, or government agency. Ownership data for 4492 plants was derived from the GEM Wiki's Global Coal Plant Tracker (GCPT) and Global Gas Plant Tracker (GGPT) (GEM, 2022). Ownership for an additional 527 plants was reported using data from the U.S. Energy Information Administration (EIA, 2022) and aggregated by WattTime. In some cases, there were different owners for specific units within each source. In these cases, the unit level ownership data was aggregated to the asset level by summing the ownership of each unit in the asset weighted by its capacity. See the Climate TRACE electricity methodology for more information on plant capacity.

Iron and Steel. Sources were defined at the level of individual iron and steel manufacturing facilities. Ownership was defined in terms of percent financial interest in the asset as a piece of real property, a business concern, state-owned enterprise, or government agency. The source-level ownership data source for steel is the Global Energy Monitor's (GEM) Global Steel Plant Tracker (GSPT). GSPT provides facility-level data for all steel plants that produce at least 0.5 million metric tons of crude steel per annum (GEM, 2022).

Coal Mining. Sources were defined at the level of individual mines as reported by the GEM Wiki Global Coal Mine Tracker (GCMT) and CoalSwarm project (GEM, 2022). Ownership was defined in terms of percent financial interest in the asset's coal reserves (GEM, 2021).

Oil and Gas Refining – Teapot Refineries. 'Teapot refineries' are a subset of small, privately-owned oil and gas refineries in China. Sources were defined at the level of individual facilities engaged in oil and gas refining. Ownership was defined in terms of percent financial interest in the asset as a piece of real property, a business concern, state-owned enterprise, or government agency.

2.1.2 WattTime Ownership-Mapping Sectors

For all sectors in this section except for Cattle Operations, WattTime mapped direct ownership datasets up to ultimate parents, for 77% of sources. Mapping was performed using WattTime's novel, automated ownership mapping algorithm that collects webscraped and API-derived ownership information from large, freely available entity relationship datasets. The algorithm matches records in relationship databases based on international entity-id systems and linguistic pattern matching. Manual review and desk research were performed for low-confidence matches and unmatched entities leftover after automated mapping. Manual, quantitative validation was performed on a subset of the results of the mapping algorithm to ensure accuracy and high-quality. Direct ownership for these sectors was generated from the methods and data aggregators listed below.

Aluminum. Sources were defined at the level of individual facilities engaged in the production of aluminum. Ownership was defined in terms of percent financial interest in the source as a piece of real property, a business concern, state-owned enterprise, or government agency. Direct ownership for aluminum facilities was derived from industry publication, Light Metal Age (Light Metal Age, 2022). Direct ownership was aggregated and sources were defined by TransitionZero.

Cattle Operations. Sources were defined at the level of individual cattle operations (i.e. beef and dairy farms). Ownership was derived from desk research using government sources and Freedom of Information Act (FOIA) requests.

Cement. Sources were defined at the level of individual cement manufacturing facilities. Ownership was defined in terms of percent financial interest in the source as a piece of real property, a business concern, state-owned enterprise, or government agency. The Spatial Finance Initiative (SFI) provides a Global Database of Cement Production Assets (GDCPA) that includes facility-level ownership data. The database is part of the GeoAsset Project developed by SFI, Oxford Sustainable Finance Programme, Satellite Applications Catapult, and The Alan Turing Institute (McCarten *et al*., 2021). Climate TRACE provides emissions estimates for clinker-producing plants identified by the GDCPA (see the Climate TRACE cement methodology for discussion of cement production methods and emissions model). Ownership data was available from the GDCPA for 1,504 of these plants (Armstrong *et al*., 2021; Global Cement Directory, 2021). Direct ownership was aggregated and sources were defined by TransitionZero.

Copper, Iron, & Bauxite Mining. Sources were defined at the level of individual mines as reported by Hypervine. Ownership was defined in terms of percent financial interest in the mine's mineral reserves. Ownership for chemical facilities was derived from desk research using company websites, government sources, and news articles. Direct ownership was aggregated and sources were defined by Hypervine.

International and Domestic Aviation. Sources were defined at the level of airports. Ownership was defined in terms of the percentage of the source's emissions attributable to each airline or 'operator' at the airport. For each operator, Climate TRACE calculated the proportion between the emissions attributable to the operator's flights vs. all flights at the asset overall. Half of the total emissions for each individual flight were attributed to the flight's airports of departure and arrival, respectively. Emissions for international and domestic flights were estimated separately. The Official Airline Guides (OAG) Historical Flight Status Data identifies flights by airline and airport for all domestic and international flights –including passenger, commercial, private, and general aviation. Direct ownership was aggregated and sources were defined by WattTime.

Oil and Gas Production & Oil and Gas Transport. Sources were defined at the level of oil and gas fields, or macro-geological formations, and then aggregated to groups of at least 20 fields or more to protect proprietary data intel. Source boundaries were derived from shapefiles acquired from proprietary data providers, and latitude and longitudes were averaged within an aggregation group. Where available, percent ownership for each source was defined based on the owner's percent financial interest in the total oil and gas production for the source. Owner name and ownership percentage for most Climate TRACE sources were derived from the publicly available Global Energy Monitor (GEM) Global Oil and Gas Extraction Tracker (GOGET: https://globalenergymonitor.org/projects/global-oil-gas-extraction-tracker/). See Climate TRACE oil and gas production and transport methodology for more information about how production and assets were defined. Direct ownership was aggregated and sources were defined by RMI.

Oil and Gas Refining. Sources were defined at the level of individual facilities engaged in oil and gas refining. Ownership was defined in terms of percent financial interest in the source as a piece of real property, a business concern, state-owned enterprise, or government agency. Ownership for refineries was derived from desk research using company websites, government sources, and news articles. Direct ownership was aggregated and sources were defined by RMI.

Petrochemicals steam cracking. Sources were defined at the level of individual facilities engaged in the production of petrochemicals. Ownership was defined in terms of percent financial interest in the source as a piece of real property, a business concern, state-owned enterprise, or government agency. Ownership for petrochemical facilities was derived from desk research using company websites, government sources, and news articles. Direct ownership was aggregated and sources were defined by RMI.

3. WattTime's Automated Entity Relationship Mapping Algorithm

3.1 Automated Mass Mapping Datasets

GLEIF. The Global Legal Entity Identifier Foundation (GLEIF) is an open, global, industry database of legal entities that is curated by McKinsey & Company. Legal Entity Identifiers (LEIs) are international business credentials that companies can apply to receive. The typical reason companies apply for an LEI is to gain access to global financial markets outside their country of origin. As such, the GLEIF database is limited only to such companies that have applied for and received an LEI credential, resulting in a dataset that primarily includes large, international conglomerates and their mid-level international shell companies. GLEIF corporate relationship data was downloaded in bulk from the GLEIF website. An example of GLEIF relationship data for Mitsubishi Group, a parent company of several steel and coal mining assets, can be found here. GLEIF entities from bulk data were pre-processed and cleaned for the purposes of matching (as described in section 4.1 below). LEI relationships were identified at the level of ultimate parents, direct parents, and child entities. GLEIF is especially useful for identifying large corporations headquartered in the global south.

PermID. The PermID database is administered by financial market data provider, LSEG. On their website, PermID is described as "a machine readable identifier that provides a unique reference for data item(s). Unlike most identifiers, PermID provides comprehensive identification across a wide variety of entity types including organizations, instruments, funds, issuers and people. PermID never changes and is unambiguous, making it ideal as a reference identifier… intended to enable interoperability." PermID provides users with "a set of descriptive metadata for each entity to facilitate disambiguation to PermID with or without explicit mapping." PermID data was accessed using the OpenPermID API with the OpenPermID package for Python. PermID is the primary provider of registered address and official website url data for the Climate TRACE ownership database.

Wikidata. Wikidata is a crowdsourced database that contains information from Wikipedia that has been transformed by users into an analyzable format. Using the SPARQL query service, users can generate relevant datasets and download them en-masse. For applicable sectors, SPARQL queries were constructed to return a dataset of all relevant entities that were instances of the following: business, company, conglomerate, enterprise, holding company, corporation, multinational corporation, government agency, public company, state-owned enterprise, and township-level division. In order to ensure the entities returned by the query were relevant to the sector, additional properties were specified. For example, queries were constructed such that the entity needed to either be involved in a specific industry (ex. oil refining) or make a specific product (ex. cement). An example SPARQL query that returns the Wikidata dataset for the steel sector can be found here. A subset of this data was then used to perform corporation mapping.

First, entities whose Wikidata entries included alternate company names and/or relationships such as 'owner of', 'has subsidiary', 'owned by', 'parent organization', and 'partnership with' were added to the mapping dataset for the sector. Additional entities and relationships were added to each dataset through web scraping the actual text of the Wikipedia page links returned by the SPARQL query. Web Scraping was performed using the Python package for Selenium Webdriver with Chrome. The scraper returned in-text lists of subsidiaries. SPARQL queries return Wikidata results from pages in every language available on Wikipedia, with relevant data types transliterated into English. Wikidata entries from non-English language pages were included in the dataset. Meanwhile, scraping of page text was limited to English language pages only. Because of its non-English language coverage and ability to query information about state-owned entities, Wikidata was an especially useful source for international companies, especially in China. Because both LEIs and PermIDs are Wikidata properties, all entities identified through those datasets were queried based on those properties to gather additional information. Although Wikidata is crowdsourced, datasets include citation links so that it is possible to track down the source for each data point. All primary sources linked in Wikidata articles were extracted from SPARQL query results and are listed (along with the Wiki Page they were found on) as sources in the data source table.

OpenCorporates. OpenCorporates is the largest open, crowdsourced database of business entities in the world. Although the data is crowdsourced, it is also curated such that sources only include public official information. OpenCorporates provides many types of current and historical information, such as company officers, registered addresses, agents, statements of control, subsidiaries, branches, and similarly named companies. An example of an OpenCorporates entry for Berkshire Hathaway, the parent company for three of the top 500 emitting power plants in the Climate TRACE dataset, can be found here.

OpenCorporates has a partnership with OpenRefine, a tool created by Code for Science & Society. OpenRefine is designed for working with messy, publicly available data and mapping it onto other open data sources. The OpenCorporates Reconciliation service available on OpenRefine was used to generate matches for asset-level owners based on the owner's name and the country of operation. The reconciliation service provides a scoring system for matches, and initial matches with scores below 50 were manually reviewed for accuracy. Once the initial matches were generated, the list was input to the OpenCorporates API using the Ropencorporate package for R. The API was queried to return a list of parents, subsidiaries, branches, and controlled companies for each match where such information was available. Corporate relationships were identified through the explicitly listed data points returned through the API, and also by identifying entities that shared CEO's, presidents, managers, and registered addresses in common. Because OpenCorporates is derived from local, official business registries, it is an especially useful source for low-to-mid level shell companies. However, its coverage for companies, especially ultimate parent companies, outside the global north is limited.

SEC Filings. Securities and Exchange Commission (SEC) Filings were used as a mass automated mapping dataset for oil and gas sectors only. For the top 50 global oil and gas producers, PDFs were downloaded from the SEC database. Specifically, these PDFs contained an appendix titled 'Subsidiaries of the Registrant' from 10-K and 20-K annual reports for 2021. An example from ExxonMobil can be found here. The text from these PDFs was parsed using the PyPDF2 package for Python and aggregated into a single dataset. This dataset was aggregated exclusively for oil and gas sectors because international oil and gas companies are especially likely to be massive conglomerates or SOE's that do business in the USA. In 2021, the USA was the second largest net oil importer in the world, and the largest oil producer (WorldsTopExports, EIA). SEC filings also provide information about percentage ownership for subsidiaries and joint ventures. SEC filings data was pre-processed and mapped as described in Section 4.1 below.

3.2 Validation Datasets

For three sectors, it was possible to compare company-level estimates for production with company-level production estimates derived from industry or government datasets.

Steel. For last year's ownership database, company-level Climate TRACE estimates of production were compared with self-reported production estimates from the WorldSteel Association for the top 114 steel producers in 2021 (WorldSteel Association, 2021). This year, automated mapping for steel was unnecessary because GEM provided a full, manually mapped dataset for steel. However, this provided a good opportunity to ensure that automated mapping produced results that were equivalent to results produced entirely through human vision. Although GEM's definition for ultimate parents differs slightly from the definition used for the database last year and from how producers self-define themselves to the WorldSteel Association, an analysis comparing estimates for production based on manually mapped, automatically mapped, and self-reported steel production estimates was performed as a validation for this year's ownership database. However, the published data available for download for steel this year, reflects the manual mapping performed by GEM and GEM's definition of ultimate parents (see Global Energy Monitor's ownership methodology for more details).

Oil and Gas Production and Transport. For 2022, percent ownership for oil and gas production from all assets in Texas was compared with equivalent estimates based on data from the Texas Railroad Commission (RRC). Texas assets alone account for 8.7% of global emissions for the sector. Two datasets were aggregated from the RRC. First, annual production data paired with lease ids were scraped from the Oil & Gas Production Data Query system. Then, a dataset matching lease ids to owner names was parsed from PDFs (see Oil & Gas Lease Name Index). These datasets were then joined by lease id. The owner names were pre-processed and mapped as described in Section 4.1.

Cement. Company-level Climate TRACE estimates of production were compared with self-reported production estimates from the Global Cement Directory for the top 50 cement producers in 2022 (Global Cement Directory, 2022).

4. Methods

4.1 Pre-Processing

Prior to mapping corporate networks, owner names from asset-level datasets, Wikidata, PermID, and GLEIF were consolidated, cleaned, and transformed into several standard formats for the purposes of matching. Relationship databases such as OpenCorporates and Mergr.com were queried using LEIs and PermID's to identify high-confidence record matches and based on unique identifiers from these systems. Then, further mapping and consolidation was performed using a combination of algorithmically generated groupings of potentially related entities that were verified using manual desk research where necessary.

Consolidation. Raw ownership data from all source-level datasets frequently contained several name variants for individual entities (ex. 'Mitsui & Co. Ltd', 'Mitsui and Company', 'Mitsui &Co'). Any one of these variants could potentially produce a match from one of the corporate mapping datasets. For this reason, instead of transforming all of these into a single standardized format for matching (which could eliminate useful variations), the first step was to identify which name variants within the dataset likely refer to the same entity and assign them a key. That way, if a match is returned for one variant, it can instantly be applied to the others that share its key.

Variants were identified using OpenRefine's Cluster and Edit functions. The Cluster and Edit functions are approximate string matching algorithms designed to find text strings that contain the same content, even if they are spelled differently. OpenRefine's clustering methods include: Fingerprint, N-gram Fingerprint, Metaphone3, Cologne-Phonetic, Daitch-Mokotoff, Beider-Morse, Levenshtein Distance, and Prediction by Partial Matching (PPM). See OpenRefine's technical reference for an in-depth explanation of each method. Clusters identified through these functions were manually reviewed for accuracy. Each cluster was then assigned a key. The key was based on the most common name variant in the cluster and was constructed using a customized version of OpenRefine's default fingerprint method. The steps in OpenRefine's default fingerprint method are shown in Table 2 (OpenRefine User Guide, 2022).

Table 2 Processing Steps to Produce OpenRefine's Default Fingerprint

Processing Step	**Example String 1**	**Example String 2**
Remove leading and trailing whitespace	Companhia Siderúrgica Nacional (CSN)	Mitsui & Co. Ltd
Change characters to lowercase	companhia siderúrgica nacional (csn)	mitsui & co. ltd.
Remove all punctuation and control characters	companhia siderúrgica nacional csn	mitsui co ltd
Normalize extended western characters to ASCII	companhia siderurgica nacional csn	mitsui co ltd
Split the string into whitespace-separated tokens	['companhia'] ['siderurgica'] ['nacional'] ['csn']	['mitsui'] ['co'] ['ltd']
Sort tokens alphabetically and remove duplicates	['companhia'] ['csn'] ['nacional'] ['siderurgica']	['co'] ['ltd'] ['mitsui']
Join tokens back together into fingerprint string	companhia csn nacional siderurgica	co ltd mitsui

For this analysis, OpenRefine's default fingerprint method had to be modified to increase accuracy during clustering. For each sector, there were specific words that were common across the dataset but were only rarely relevant to identifying unique entities. These irrelevant portions of the string were often longer than the relevant ones. Often, they were shared between unrelated companies, producing false positives. Alternatively, these irrelevant strings varied too much between genuinely related entities, producing false negatives. Table 3 shows the default fingerprints for a cluster of similarly named owners from the steel assets dataset, alongside the portions of these strings that are relevant for clustering vs. the irrelevant portions that tend to create errors.

Table 3 Default vs. Customized OpenRefine Fingerprints

Original Owner Name	Default Fingerprint	Relevant	Irrelevant
Anyang Iron & Steel Co., Ltd.	anyang co iron ltd steel	anyang	co iron ltd steel
ANYANG IRON AND STEEL GROUP CO., LTD.	and anyang co group ltd steel	anyang	and co group ltd steel
Anyang Xinpu Steel Co., Ltd.	anyang co ltd steel xinpu	anyang xinpu	co ltd steel
Anyang Iron and Steel Co Ltd	and anyang co iron ltd steel	anyang	and co iron ltd steel
Angang Steel Co., Ltd.	angang co ltd steel	angang	co ltd steel

Customized fingerprints included only the relevant parts of the default fingerprint string (Table 3). In order to identify irrelevant parts of strings, OpenRefine's word facet was used to review the 100 most common words in the dataset. Then, irrelevant sector-specific words were identified and removed (ex. 'steel', 'cement', 'plant', 'power', 'exploration'). Additionally, a list of common country-specific company abbreviations and terms for business entities were removed for the relevant countries automatically (ex. 'Ltd', 'PT', 'Companhia'), as well as the word 'and'. Exceptions for the word 'and' as well as '&' characters were made in cases where those strings appeared between consecutive single consonants (ex. 'S & N Drilling'). In such cases, 'S & N' was treated as a single string. With these terms removed, the clustering functions were applied, and the resulting clusters were assigned a modified fingerprint as a key.

In the example in Table 3, three unique keys – 'anyang', 'anyang xinpu', and 'angang' –were produced from the original owner's names. Although they were similar, they were not consolidated down further. 'Anyang' could be a company name, or a province name, or both –and 'angang' could be a typo for 'anyang', or it could be a separate entity entirely. As it turned out, matches uncovered during the mapping process confirmed that all three of these keys are distinct shell companies that fall under the same corporate parent: Ansteel Group. However, in other cases (ex. 'Shaanxi' vs. 'Shanxi') slight variants referred to entirely different entities. Hence, the clustering process errs on the side of keeping assumptions conservative.

4.2 Reconciliation among Corporation Mapping Datasets

After consolidation, the original owner names from the datasets were transformed into several formats for matching. Table 4 shows an example of each format for the owner 'Pangang Group Jiangyou Changcheng Special Steel Co., Ltd.'.

Table 4 Examples of Owner Name Formatted for Matching

Variant	Format Type	Formatted Name
1	Original Owner Name	Pangang Group Jiangyou Changcheng Special Steel Co., Ltd.
2	Company terms removed at end of string only	Pangang Group Jiangyou Changcheng Special Steel
3	Key	changcheng jiangyou pangang
4	First two words, lowercase, no punctuation	pangang group
5	Key, sector specific words intact	changcheng jiangyou pangang special steel

For reconciliation with Wikidata and GLEIF, matching was performed using all variants (1-5) – like for like – between the mapping dataset and the asset-level dataset. Information from Wikidata and GLEIF were also cross-referenced with each other, so that ultimate parent groupings would include matches from both. The most similar match for the longest string in a cluster was considered the best match for the group. For reconciliation with OpenCorporates, it is not practical to download bulk data and match transformed names. So, the standard reconciliation process was used to match only variants 1 and 2 in Table 4. Then, these matches were expanded into groupings based on results from the API, and only the novel owner names generated by the API were matched and cross-referenced using the other three variants (3-5). OpenCorporates reconciliation was attempted first while specifying the country of operations. For unmatched entities, it was attempted again without specifying a region.

2.2 Desk Research for Unmatched Owners

In cases where the mapping process uncovered ambiguous (equally similar and conflicting) matches, or if it failed to return a match, desk research was conducted. Desk research included custom, manual searches for government and industry news sources, and within the mapping datasets. It also included searching the following sources: SEC filings, official company websites (ex. Baosteel), and industry news and data sources (ex. mergr). Companies that still remained unmatched or lacked additional nodes in their corporate network at the completion of mapping were considered to be their own ultimate parents.

2.3 Criteria for Assigning Direct Parents and Final Owner Groupings

Once the data were fully mapped, it was necessary to establish criteria for reporting ultimate parents for the purposes of inclusion in the Climate TRACE database. The current analysis was designed to produce reliable data en masse that would enable reasonably accurate and independent emissions estimates at the company-level. It was also designed to produce accessible, summary information for users of the Climate TRACE website. In some cases, some of the nuance and complexity of the corporate relationships data that was returned had to be simplified down in a standardized way. In other cases, the datasets available did not specify components that relate directly to calculating emissions (i.e., percentage ownership). To handle this issue, general and sector-specific criteria were established to simplify complex and incomplete data. These criteria include the following:

General. For assets in which over 50% of the owners were individual persons, the owner is listed as ‘Unknown’ or ‘persons’ –even if their identity was listed in the original source-level dataset. As a policy, Climate TRACE does not publish the identities of private individuals.

Steel. According to their 2020-21 report, the WorldSteel Association calculates the list of top steel producers as follows: “In case of more than 50% ownership, 100% of the subsidiary's tonnage is included, unless specified otherwise. In cases of 30%-50% ownership, pro-rata tonnage is included. Unless otherwise specified in the declaration, less than 30% ownership is considered a minority and therefore, not included.” The same accounting scheme was used for the purposes of validating Climate TRACE production estimates and assessing company-level emissions. For owners with more than one direct parent where the parent’s percent interest was unknown, whichever direct parent was largest (produced the most steel) was chosen.

Oil and Gas Production and Transport. In the United States, there are far more private oil and gas production enterprises than is typical in other countries. The main reasons for this are that most other countries have more nationalized oil and gas assets, and the United States has a uniquely venture-capital friendly system for oil and gas production. Private equity firms often provide the seed money for multiple, small, independent, private oil and gas producers. As a result, SEC filings may not list these companies as subsidiaries, and they are less likely to appear in the mapping datasets. For companies operating in the United States, if an oil and gas company website or industry news source listed a private equity firm as an owner’s sole financier, the firm was listed as the company’s parent. Although parent is not always the correct technical term to describe such a relationship (i.e., a seed investor, etc.), this choice was made to increase the utility of the dataset in facilitating the reduction of emissions in the sector. This approach provides the advantage of quantifying emissions in a way that puts the spotlight on the actions of large institutional actors as opposed to small-to-mid-level businesses.

For SOE’s, data on which entities are SOE’s and the governments they are associated with can be provided as additional data on request. Because the Climate TRACE ownership database provides enhanced insight into the roles of both private equity firms and world governments in the oil and gas sectors, Climate TRACE’s dataset is well-suited to provide more actionable information for investors, policymakers, and activists alike.

3. Discussion, Limitations, and Future Directions

These preliminary analyses suggest Climate TRACE’s mapping algorithm has great potential as a time-efficient, up-to-date, and reliable method for generating company-level emissions estimates. Because Climate TRACE's production estimates correspond well to self-reported production numbers, this creates a common reference point from which companies and SOE's who do not currently track their emissions can receive actionable information. Companies who currently report their emissions could also benefit by comparing their internal estimates to those of Climate TRACE. For investors, policymakers, and activists, the Climate TRACE company-level ownership dataset shows potential to provide insight into economic trends that drive changes in emissions. Results from the oil and gas sector may reflect a phenomenon colloquially referred to as 'hot potato', where high-emitting assets are passed around to avoid accountability for implementing real emission reductions. These results suggest when companies play 'hot potato' with their assets, they may not only end up reporting artificial reductions in their own emissions, but their actions may also make emissions from these assets harder to track. Because private equity firms are not required to abide by the same reporting and regulatory standards as oil and gas companies, it is hard to determine why marginal emissions are rising in the US assets where private equity investors tend to invest. The fact that more oil and gas assets are being transferred into the possession of owners who are harder to regulate and less likely to self-report their emissions underscores the need for better independent emissions reporting at the company-level.

Although these results are promising, there are some important limitations. One is the lack of coverage for all assets and sectors. Additional partial releases of ownership data for additional sectors are planned. The corporate mapping datasets still have limited coverage for companies and SOEs in the global south. Climate TRACE intends to increase coverage, replicate and validate more results, and identify new mapping datasets. Other future directions include: identifying how specific investment models and policies impact the generation of emissions by private companies and SOE's.

6. Supplemental Information

Citation format: Thomas, P., Saraswat, I., Raniga, K., Lewis, C., Freeman, J., Davitt, A., Schmeisser, L., Fallurin, J., Sinha, A., Heal, J., Crane, V., Jolley, M., and McCormick, G. (2024). *Climate TRACE Ownership Information- Source & Company-Level Ownership Methodology.* WattTime, USA, RMI, USA, TransitionZero, UK, and Hypervine, UK. Climate TRACE Emissions Inventory. https://climatetrace.org [Accessed date]

Disclaimer: The emissions provided for this sector are our current best estimates of emissions, and we are committed to continually increasing the accuracy of the models on all levels. Please review our terms of use and the sector-specific methodology documentation before using the data. If you identify an error or would like to participate in our data validation process, please contact us.

7. References

1) Clark, A. and Benoit, P. (2022) 'Greenhouse Gas Emissions from State-Owned Enterprises: A Preliminary Inventory'. Sipa Center on Global Energy Policy. Available at: https://www.energypolicy.columbia.edu/research/report/greenhouse-gas-emissions-state-owned-enterprises-preliminary-inventory
2) Heede, R. (2014) 'Tracing anthropogenic carbon dioxide and methane emissions to fossil fuel and cement producers 1854–2010'. Climatic Change 122, 229–241. https://doi.org/10.1007/s10584-013-0986-y
3) Starr, D. (2016) "Just 90 Companies Are to Blame for Most Climate Change, This 'Carbon Accountant' Says." AAAS Articles DO Group. Available at: https://www.science.org/content/article/just-90-companies-are-blame-most-climate-change-carbon-accountant-says
4) GEM contributors (2022) 'Steel inventory data retrieved from: GEM Global Steel Plant Tracker'. Global Energy Monitor (GEM). Available at: https://globalenergymonitor.org/projects/global-steel-plant-tracker/.
5) McCarten, M., Bayaraa, M., Caldecott, B., Christiaen, C., Foster, P., Hickey, C., Kampmann, D., Layman, C., Rossi, C., Scott, K., Tang, K., Tkachenko, N., and Yoken, D. (2021) 'Global Database of Cement Production Assets'. Spatial Finance Initiative (SFI). Available at:Database of Cement Production Assets'. Spatial Finance Initiative (SFI). Available at: https://www.cgfi.ac.uk/spatial-finance-initiative/geoasset-project/geoasset-databases/ (Accessed: 26 August 2022).
6) Armstrong, T. *et al*. (2021) 'The Global Cement Report™'. 14. Tradeship Publications Ltd, UK. Available at: www.CemNet.com
7) Global Cement Directory (2022). Available at: https://www.globalcement.com/directory The Official Airline Guides Historical Flight Status Data (2022). Available at: https://www.oag.com/
8) GEM Contributors (2021) 'Coal reserves'. Global Energy Monitor. Available at: https://www.gem.wiki/w/index.php?title=Coal_reserves&oldid=239981
9) EIA: United States Energy Information Administration (2022). "Form EIA-860 detailed data with previous form data (EIA-860A/860B)" Available at: https://www.eia.gov/electricity/data/eia860/
10) Light Metal Age (2022). Available at: https://www.lightmetalage.com/