# Inclusive Development International (IDI)

Classifying Environmental and Human Rights Violations in the Palm Oil Industry

## Background

Palm oil is a popular and versatile vegetable oil found in animal feed, biofuels, and nearly 50 percent of packaged supermarket goods and 70 percent of cosmetics. However, its production has caused significant environmental and social harm.

Companies eager to secure fertile, tropical soil on which to build palm fruit plantations have violated local and Indigenous communities’ right to land and self-determination through land seizures, trespassing, coercive tactics, bribes, and failures to fully consult with communities’ chosen representatives during land sales. Following their displacement, these communities are often unjustly compensated and spend years in land disputes. The Consortium for Agrarian Reform (KPA) documented 2,047 such conflicts in Indonesia from 2015 to 2019.

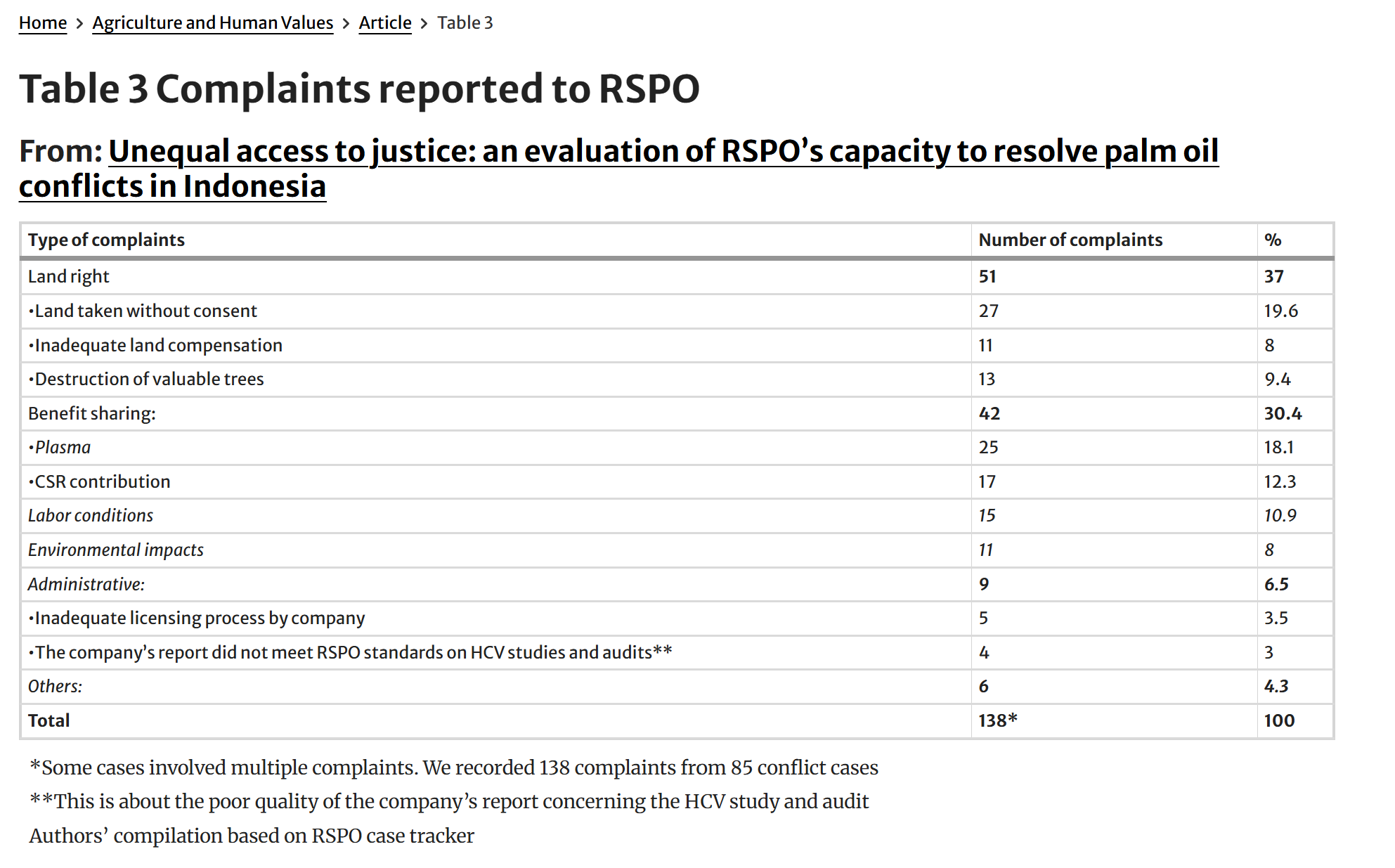
Industrial farming and palm oil refinement have also reduced biodiversity, increased carbon emissions, and polluted nearby water sources due to run-off from fertilizers, pesticides, and palm oil effluents (POME). Communities have borne the brunt of these effects in the form of increased health risks and a loss of food sources, economic livelihoods, and cultural heritage.

It is challenging to assess the true nature and extent of the harms caused by the palm oil industry due to the fact that ***grievance logs***—complaints formally submitted by communities and NGOs to suppliers and watchdog organizations that mediate disputes–are scattered across many different data sources and consist of unstructured text. However, with advances in natural language processing, it is becoming easier to identify trends in grievance and therefore advocate more effectively for reform.

## Goal

The goal of this project is to build a natural language model that classifies grievances by topic (e.g., “labor violation”, “land dispute”, “deforestation”, etc.). (Note: A grievance might fall under more than one topic.) Afterwards, summary statistics can be generated to reveal patterns in the types of complaints that have occurred over time, by country, etc.

As an example, the table below illustrates how a team of researchers coded (categorized) complaints submitted to the Roundtable on Sustainable Palm Oil (RSPO). You should defer to your DSSI instructor when choosing which categories to use.



## Data

The dataset is a CSV file of 430 grievances scraped from three different websites:

* [Golden Agri-Resources (GAR)](https://www.goldenagri.com.sg/sustainability/responsible-sourcing/grievance-list-and-reports/), a vertically-integrated, “seed-to-shelf” agribusiness based in Singapore. Manages oil palm plantations with farmers, sells crude and refined palm oils and kernels, and manufactures consumer goods made from palm derivatives, among other operations.
* [Roundtable on Sustainable Palm Oil (RSPO)](https://rspo.my.site.com/Complaint/s/casetracker), an international non-profit that unites stakeholders across the seven sectors of the palm oil industry – oil palm producers, palm oil processors and traders, consumer goods manufacturers, retailers, banks and investors, environmental/nature conservation non-governmental organisations (NGOs), and social/development NGOs – to develop and implement global standards for sustainable palm oil production.
* [Wilmar International](https://www.wilmar-international.com/sustainability/grievance-procedure), a global processor of agricultural commodities, based in Singapore. Focuses on oil palm cultivation, edible oil refining, oilseed crushing, sugar milling and refining, oleochemicals, production of biodiesel and specialty fats, manufacturing fertilizers, and flour and rice milling.

Each of these organizations offers a formal procedure for settling grievances among stakeholders in their member base or supply chain. It is recommended to peruse these procedures to gain context for the text you will be analyzing (e.g., see [this diagram](https://www.rspo.org/wp-content/uploads/brochure-rspo-complaints-appeals-procedures-english.pdf) published by RSPO).

Within the dataset, each record contains ***metadata*** about the grievance submission and resolution as well as a ***summary*** of the allegation and ***progress updates*** reported by the mediator. Information needed to classify the grievance may be found in one or both of the summary and progress update fields. (Note: While an attempt was made to clean and standardize the data, inconsistencies may still exist.)

### Data Dictionary

#### Metadata

| pk | A manually-created unique identifier for each grievance in the dataset (e.g., 1, 2, 3…). |
| --- | --- |
| internal\_id | The unique identifier of the grievance within its original log (e.g., “500Mg00000L0di6IAB” in the RSPO log). |
| status | The current outcome of the submitted grievance. One of:   * “In Progress” * “In Progress - Investigation” * “Closed” * “Closed - With Monitoring” * “Closed - No Longer in Supply Chain” * “Closed - Not in Supply Chain” * “Unknown”   “Closed - No Longer in Supply Chain” refers to a case in which a stakeholder (a mill, plantation, or middleman) was found to be non-compliant and the supplier removed the company from its supply chain (not always permanently).  “Closed - Not in Supply Chain” refers to cases in which a grievance is closed because the accused company was never in the supplier’s supply chain to begin with. |
| submitted\_on | The date the grievance was submitted to the supplier or watchdog organization (e.g., “2025-02-26”). |
| accepted\_on | The date the grievance was accepted by the supplier or watchdog organization for further investigation (e.g., “2025-04-15”). |
| last\_updated\_on | The date of the most recent status update for the grievance case (e.g., “2025-05-20”). |
| country | The country in which the grievance occurred (e.g., “Indonesia”). |
| region | The region of the country in which the grievance occurred, if available (e.g., “Sumatera Selatan”). |
| complainants | The individual(s) and/or organization(s) submitting the complaint. Sometimes confidential or undisclosed. Multiple entries are separated by semicolon and space (e.g., “Awasmifee; Greenpeace; Forest Peoples Programme (FPP)”). However, inconsistencies may exist. |
| respondents | The companies accused of malpractice. Multiple entries are separated by semicolon and space (e.g., “Palmas y Trabajo S.A.S; Progreso Palmero S.A.S; Oleaginosas de Yuma S.A.S”). However, inconsistencies may exist. |
| respondent\_sector | The sector of the companies. One of “Oil Palm Growers”, “Processors and/or Traders”, “Retailers”, or “Environmental NGO”. |
| respondent\_parent\_company | The respondents’ parent companies (e.g., “Genting Plantations Berhad”). |
| respondent\_supplier | The respondents’ supplier (e.g., “TSH Resources Berhad”). |
| external\_link | The URL to the complaint description, if an individual webpage exists. |
| extras | A JSON object containing additional metadata scraped from grievance’s webpage, or null if no extras exist. |

#### Content

| summary | A summary of the complaint provided by the supplier or nonprofit watchdog mediating the grievance. |
| --- | --- |
| history | The history of the grievance’s submission, mediation, and resolution. |