# An open dataset on global mining

Simon Jasansky, Mirko Lieber, Stefan Giljum, Victor Maus
Institute for Ecological Economics, Vienna University of Economics and Business

## Abstract

In order to assess the environmental and social impacts of increased mining activities around the world, as well as to trace them along the supply chain from mine to consumer, production data for individual mines is highly valuable. However, such a dataset, with the requirements of covering large parts of the world as well as the majority of important metals, minerals and coal in high detail, is not yet openly available.
Our open dataset on global mining aims to fill that gap. The dataset covers more than **1'100 mines** during the **years 2000-2020**. It includes about **25'000 data points** regarding material extraction, with **83 materials**, extracted in **82 countries**, and more than **7'000 ore grades**. Furthermore, it also covers processing facilities, such as steel mills, and their physical production.
The data was collected from over **2'100 sources**, primarily consisting of company reports of mining firms, such as Annual Reports or Sustainability Reviews. The dataset is not yet published, as we want to ensure sufficient data quality. We expect to publishing it in early 2022.

## Introduction

In environmental assessments, in particular in footprint type analyses, the standard practice is to base analyses on national mineral production data. However, this can lead to distorted outcomes in the analyses, as the impacts of material extraction are not equally distributed across space [1]. Therefore, **there is a need to shift from nationally aggregated mineral production data to sub-national, spatially explicit production data.**
However, it is rare that national statistics offices publish detailed mine-level production data. Furthermore, some academic groups and institutes have compiled mine-level data for specific analyses, although not covering more than a few commodities, and often also with geographic foci [2,3].
Additionally, profit-oriented intelligence providers offer mine-level production data. However, the high cost and strict copyright restrictions create high barriers of entry to quality research in this field.
Therefore, the need for open data on global mining with spatially explicit, detailed information is eminent, and this dataset aims to fill that gap.

**Data Gap identified:**
Openly available, spatially explicit and detailed mineral production accounts

**Solution:**
Compilation of mine-level mineral production from company reports in open database

## Methodology

As explained in the introduction, national statistics offices and international institutions only rarely report mineral production data on a mine-site level. Therefore, the method of data collection chosen for this dataset was to **extract detailed production information of individual mines from company reports of mining corporations**. In addition to production information, general information about the mines, such as their coordinates, were collected. Every data point is linked to a source document, ensuring full transparency and traceability.
After collection of this data in spreadsheets, it was further harmonized and processed using R scripts.
The detail of the reported mine-level production varies between reporting companies. Some firms report highly detailed information, including the material excavated, ore treated, grade of the ore treated, and metal or commodity recovered. Figure 1 provides an example of highly detailed information published by the company Goldcorp Inc. in their 2018 Annual Report, and illustrates how it was accounted for in our dataset.

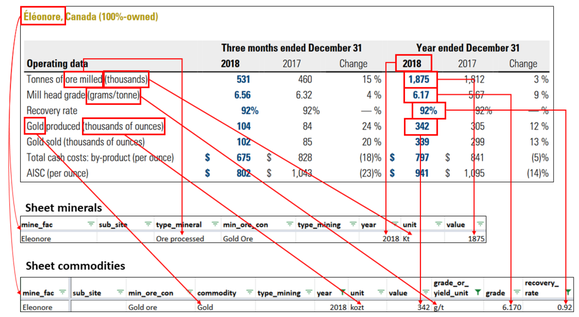


*Figure 1: Mineral accounting scheme from company report to database*

## Results

In order to successfully substitute national or global mineral extraction data with mine-level data, the mine-level dataset must depict the national production volumes as accurately as possible.
**In Figure 2, the global coverages of different materials are displayed**. To assess coverage, the aggregated production of one commodity over all mines in the open dataset on global mining is compared to the global total production, based on data of the IRP [4]. For coal and ferrous ore, the global comparison values are the mass of material mined, while for all other metals the global comparison values are the mass of metal contained in ore mined. As displayed, the coverage of ferrous ores and copper is continuously on a high level of around 60 to 70 percent. For coal and gold, it is relatively steady, although on a lower level of 30 to 40 percent.

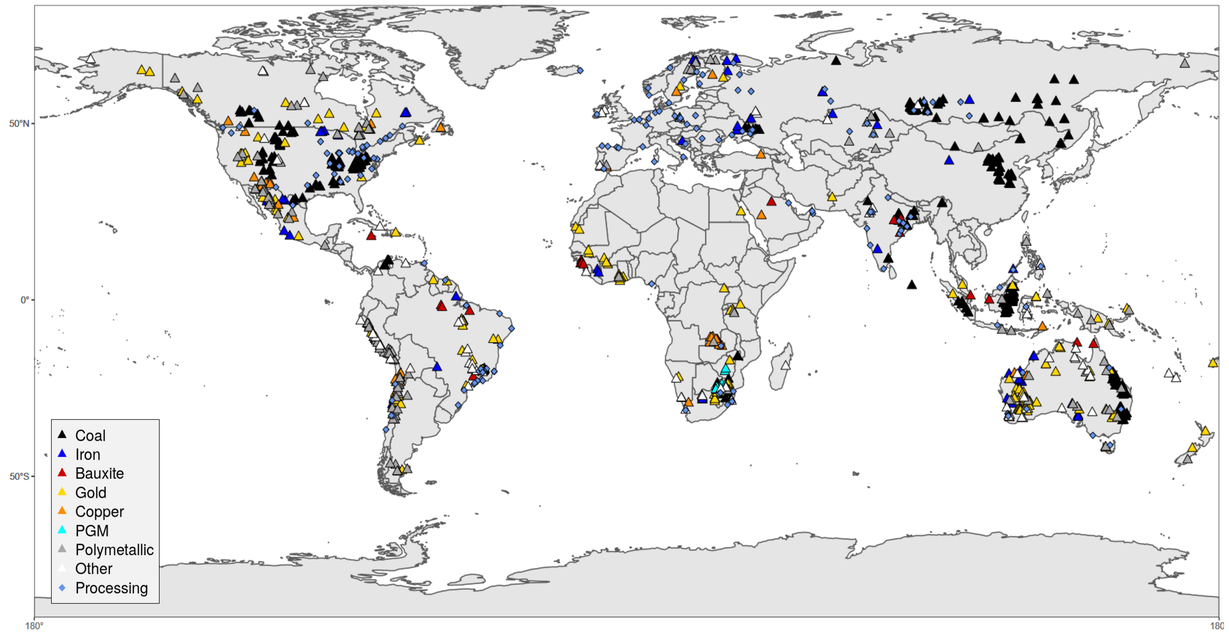


*Figure 3: Map of mines and processing facilities covered in the open dataset on global mining*

**Figure 3 shows the geographic coverage of this open dataset on global mining.** The mines, depicted as triangles, are categorized according to their primary commodity production. Processing facilities, depicted as blue diamond shapes, include steelworks, alumina refineries and aluminium smelters. If production values for processing facilities were available, for example the mass of finished steel products, they were also recorded in the dataset.

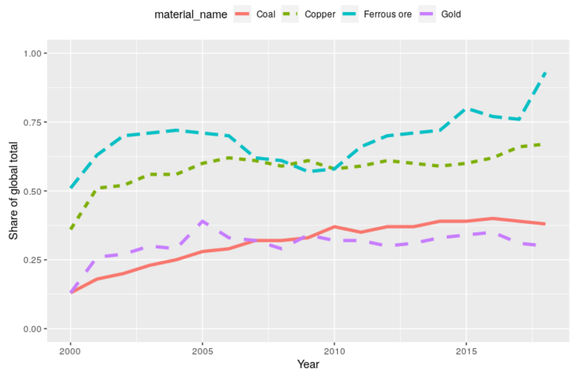


*Figure 2: Share of global production covered in the open dataset on global mining*

As noticeable in Figure 3, coverage of mines is greater in countries with many publicly listed companies, such as the USA and Australia, than it is in China and India. The reason is the greater abundance and better availability of company reports of firms incorporated in such „western" countries. In numbers, the USA is the country with the most mines included in the dataset with 236 mines, followed by Australia (215), China (115) and South Africa (109). Differentiated by primary commodity production, coal mines are the most with a total number of 461, followed by Gold (285) and Copper (184) mines.

## Applications & Outlook

This dataset enables a wide variety of applications, in particular, but not limited to, the fields of industrial ecology, economic geography and development economics. For instance, it can be linked to datasets regarding environmental and social pressures to assess the impacts of mining, or to labor data to find correlations of mining activities and employment.
As mentioned above, **we expect to publish this dataset in early 2022.** However, already thinking beyond that, it is planned to integrate available national and institutional data sources to the open dataset on global mining, to further improve its coverage and enable more accurate analyses.

**Acknowledgement:**
This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme grant number 725525.

**References:**
[1] Koning, A. de, Bruckner, M., Lutter, S., Wood, R., Stadler, K., & Tukker, A. (2015). Effect of aggregation and disaggregation on embodied material use of products in input–output analysis. Ecological Economics, 116, 289–299. https://doi.org/10.1016/j.ecolecon.2015.05.008
[2] Mudd, G.M, Weng, Z., & Jowitt, S.M.(2013). A Detailed Assessment of Global Cu Resource Trends and Endowments. Economic Geology, 108(5), 1163–1183. https://doi.org/10.2113/econgeo.108.5.1163
[3] Oberschelp, C., Pfister, S., Raptis, C.E., & Hellweg, S. (2019). Global emission hotspots of coal power generation. Nature Sustainability, 2(2), 113–121. https://doi.org/10.1038/s41893-019-0221-6
[4] UNEP IRP (2017). Global Material Flows Database. https://www.resourcepanel.org/global-material-flows-database

**Contact:**
Simon Jasansky
E-Mail: simon.jasansky@wu.ac.at

fineprintglobal
www.fineprint.global
www.github.com/fineprint-global