

An open source database on global coal and metal mine production

Simon Jasansky, Mirko Lieber, Stefan Giljum & Victor Maus

Vienna University of Economics and Business, Austria

FINEPRINT Brief No. 18, April 2023

While the extraction of natural resources has been well documented and analysed at the national level, production trends at the individual mine level are more difficult to uncover, mainly due to poor availability of mining data containing subnational detail. This Brief summarises our work to fill this gap by presenting an open database of global coal and metal mine production at the individual mine level. It is based on information gathered manually from more than 1900 freely available mining companies reports, in which all provided information is cross referenced, ensuring full transparency. The database covers 1171 individual mines and reports minelevel production for 80 different materials from 2000-2021. It also contains mining coordinates, ownership, mineral reserves, mining waste, transportation of mining products, and mineral processing capacities and production.

This FINEPRINT Brief builds on the paper 'An open database on global coal and metal mine production' published by Jasansky and colleagues in Nature Scientific Data. The freely available database is available on Zenodo and can be accessed and visualised on the FINEPRINT website.

The global mining sector grew rapidly in the past two decades, with global production of mineral fuels, metal ores, and industrial minerals amounting to 17.3 billion tonnes in 2020, up 52% from 11.3 billion tonnes in 2000. Some commodities, such as iron ore (151%) and aluminium (166%), substantially outpaced the industry's average growth rate [1]. This trend is expected to continue, with metal ores projected to be among the fastest growing material categories [2], partly due to increased demand induced by the green energy transition and electrification efforts. To analyse the past and current economic and environmental performance of the global mining sector, and to predict fu-

ture trends and impacts, a solid data foundation is required. While robust mineral extraction data is available at the national level [3], availability of open data for individual mine-sites is still poor. Current databases are (a) only available with a paid licence (e.g. the S&P Capital IQ Pro Metals and Mining database [4]) or not comprehensive, comprising records covering (b) only one country (e.g. mineral production data for Peru) or (c) only single domestic materials (e.g. coal data for the USA), or (d) limited in time range (e.g. the Global Coal Mine Tracker database [5]). We contribute to filling this knowledge gap by presenting an open source database on global coal and metal mine production.

Accounting framework and data processing steps

To record all information in a detailed and consistent format, we developed an accounting framework for mine-specific data, structured as a relational database. Relevant information on mining was divided into two categories. The first category comprises information on the general characteristics of individual mines, including name, location, mined materials, and other time-independent information. The second category comprises data that can be attributed to a certain time period, such as information on material extraction, mineral production, reserves, resources, and economic data.

In the online research process conducted from March 2019 to September 2021, data was gathered from more than 1900 source documents, primarily consisting of publications by mining companies, such as annual, quarterly, and sustainability reports. When the latest company reports were not available, different additional sources were screened. The database is fully transparent, with all points being linked to their source documents. After manual screening and entry of the data, automatic cleaning, harmonisation, and data checking was conducted to generate a comprehensive and consistent database.

The greatest obstacle of online research that relies on publicly available company reports is the missing of information, often due to reports not being made public by mining companies. While missing information presents the largest obstacle to construct a consistent mining database with high global coverage, the information published in company reports was also often incomplete and ambiguous. Regarding ambiguous reporting, three dimensions were identified. First, arbitrary aggregation of elements – if present, these were explicitly documented. Second, arbitrary aggregation of data for several mines – a two-level accounting structure of facilities allows to record such aggregated data in a consistent manner. Third, ambiguous unit specifications – we applied computational data checks to correct erroneous unit entries.

After the construction of the accounting framework, and its population with data from mining company reports, the obtained spreadsheet contained manually entered, raw data. This spreadsheet served as the input file for the data processing pipeline, at the end of which a consistent, final data product resulted. The data processing pipeline consists of harmonisation, conversion, aggregation, gap-filling, geographic referencing, and restructuring. Additionally, data checks were applied during post-processing. All processing steps were conducted as R scripts and can be reproduced with the code available at www.github.com/fineprint-global/compilation_mining_data.



Geographic coverage

The database includes 2413 facilities, of which 1435 are actual facilities and 978 are sub-sites. Of the 1435 facilities, 1323 have one or more coordinate points specifying their exact location, while 112 do not have coordinates, but are only mapped to GADM regions [6] for spatial determination. The 2413 facilities and sub-sites are composed of 2066 mines, 96 smelters, 67 mineral refineries, and 184 entities of other forms. Facilities are located in 80 different countries. Countries with the most recorded facilities are the USA (233), Australia (213), Russia (100), South Africa (100), China (95) and Brazil (89).

Figure 1 shows the location of mines and processing facilities, with mines categorised by their primary commodity production. Spatial agglomeration of mining in certain areas, such as along the Andes Mountain range, in South Africa, in Kalimantan in Indonesia, or in Western Australia is visible. However, some important mining areas are underrepresented. Particularly for China, the largest producer of many mining commodities, our open database lacks entries.



Figure 1: Global map of facilities covered in the open database on global coal and metal mine production. Mines are categorised by their primary commodity.

Production coverage

Our database includes extraction and production values for 80 different materials of which 6 refer to coal types, 21 to metal ores, 12 to metal concentrates, and 19 to metals. All materials included in the database are accessible on Zenodo. To assess the extent of total mining production covered in the open database on global coal and metal mine production, its reported production was compared to national production figures, using data from the UNEP IRP Global Material Flows Database [3].

Figure 2 depicts the share of global production coverage for nine selected materials in the period 2000–2018. The blue layer depicts coverage across all mining countries and the red layer shows

global coverage excluding all extraction values from China. The reason for this distinction is that while China is the largest producer of many extractive commodities, the share of Chinese extraction covered by the open database for most materials is significantly smaller than for other countries. Coverage for most materials has increased in recent years, a trend that can be explained by an easier accessibility of company reports.



Figure 2: Share of global production covered in the open database on global coal and metal mine production, for the years 2000–2018. The blue layer depicts global coverage, the red layer depicts global coverage excluding China.

Applications of the database

Our database enables a wide spectrum of global, multi-commodity, multi-year analyses on mining activities at the spatially explicit level of individual mines. Applications can include the modelling of global raw material flows, environmental assessments of the mining sector, socioeconomic analyses of the growth and employment impacts in mining regions, and studies on the origins of critical raw materials. If high production coverage in a region of interest is required for an analysis, the database can be easily extended via the data processing framework described in the data descriptor paper. Thus, the database adds value by providing comprehensive, global, and open information on mine-level coal and metal production that is directly usable for analyses, and by providing an expandable data foundation for analyses that require very high production coverage in a specific region.

Citation

Jasansky S., Lieber M., Giljum, S., Maus, V., 2023. An open source database on global coal and metal mine production. FINEPRINT Brief No. 18. Vienna University of Economics and Business (WU). Austria.



- [1] Reichl C, Schatz M. World mining data 2022, vol. 37. Federal Ministry of Agriculture, Regions and Tourism of Republic of Austria. Vienna, 2022.
- [2] OECD. Global material resources outlook to 2060: Economic drivers and environmental consequences. Paris, 2019.
- [3] UNEP IRP. Global Material Flows Database, 2021.
- [4] S&P Global Market Intelligence. S&P Capital IQ Pro, 2022.
- [5] Global Energy Monitor. Global coal mine tracker, 2022.
- [6] Global Administrative Areas. GADM database of global administrative areas version 3.6., 2019.