

Introduction to “Global dataset on phosphate mining and beneficiation”

Anna Shchiptsova

February 7, 2024

Our global dataset on phosphate mining and beneficiation covers worldwide industrial production of phosphate rock (PR) disaggregated at the level of company and ore deposit. The main objective was to gather site-specific information on the phosphorus content of mined and beneficiated resource and on the recovery rates of beneficiation process. The global dataset comprises 52 phosphate mining and beneficiation complexes, which represent current worldwide production of PR. Phosphate mining and beneficiation complex refers to a phosphate ore deposit with adjacent mine(s) and beneficiation plant(s).

The data was gathered manually from 106 publicly available sources. Firstly, the data sources consisted of publications of the phosphate mining companies, including their annual reports, technical reports for mineral projects, and other information available on the company websites. Secondly, data was acquired from public industry intelligence reports, government reports at regional and national level, and IFA, USGS, and IFDC publications. Thirdly, the sources included scientific articles and industry reviews on mineralogy and geochemistry of mineral deposits, and on metallurgy and mineral processing technologies.

To ensure high global coverage of the dataset, we selected countries responsible for 99% of total PR value (P_2O_5) produced based on the data from the USGS Minerals Yearbooks 2010-2022 (U.S. Geological Survey 2024). Starting from the country-specific structure of phosphate industry given by USGS, we compiled a list of phosphate mining and beneficiation complexes which includes 49 operational complexes and 3 complexes under development. Our dataset includes one representative data point for China - the Haikou mining complex in Yunnan province. In total, the data collected covers 97% of 167 Mt global PR supply capabilities in 2022 (IFA 2024; excluding China and East Asia).

Relevant information on PR production was divided into three categories. The first category comprises information on the general properties of individual phosphate mining complexes, including their name, geographical location, type of rock (sedimentary or igneous), ownership, production capacity, and operational status. Location is represented by country and sub-national administrative area (e.g., region, state, province). The second category describes mining output and efficiency of beneficiation process. We recorded available data on the P_2O_5 grade of ore, P_2O_5 grade of PR, mass recovery, and mineral recovery. Ore is considered as run of mine (ROM) - ore from mining prior to beneficiation. PR is considered as run of beneficiation plant - ore after beneficiation averaged over processing streams. The mass recovery refers to the ratio of the mass of produced PR to the mass of ore prior beneficiation. The mineral recovery refers to the percent of P_2O_5 recovered in the PR from the ore after beneficiation. The third category provides information on the P_2O_5 grade of the representative phosphate-bearing mineral associated with an ore deposit. All data points are provided on the annual basis and are linked to their respective source documents.

The data processing pipeline consists of data quality assignment, gap-filling, and restructuring.

In case of second-category mining and beneficiation data, high confidence is assigned to the data obtained from company records (or from publications based on company records) which are relevant for year 2013 or later. Medium-confidence data means that one of the following conditions holds: 1) data was obtained from company records (or from publications based on company records) which are relevant for period 2003-2012; 2) data was estimated based on the country-level USGS data; 3) data was obtained from the company records on mineral reserves; 4) data was copied from the site with similar beneficiation process; 5) data was estimated from time series with high variability; 6) data was estimated from more than one data source; 7) data was obtained from scientific publications. Otherwise, it stated that there is low confidence in data. The third-category mineral data is assumed to change little over time. In this case, low confidence means that data was copied from the site with similar deposit and/or data was averaged. Otherwise, the quality of data is defined as standard.

Recovery data is usually only available for either mass recovery or mineral recovery, and the available data is used as a proxy for the missing one. Missing value was estimated from the following relationship:

$$R_{mineral}G_{ore} = R_{mass}G_{PR},$$

where G_{ore} is the ore grade, G_{PR} is the PR grade, R_{mass} is the mass recovery, and $R_{mineral}$ is the mineral recovery.

The dataset is structured into 7 tables. All tables can be accessed using R package *phosphateRock* and using the set of files encoded in comma separated values (CSV) format. The *mining_complexes* table stores the general properties of individual phosphate mining complexes. The *ore*, *PR*, and *mineral* tables hold information on the P_2O_5 grade of ore, P_2O_5 grade of PR, and P_2O_5 grade of mineral. The *recovery_mass* and *recovery_mineral* tables store recovery rates of processing. The *sources* table contains description of data sources, including their identifiers, references to source document(s), data relevance, data quality, and details on the performed data collection.

References:

- [1] U. S. Geological Survey. (2024). Phosphate rock statistics and information: minerals yearbook 2014-2022. Accessed January 2024 from <https://www.usgs.gov/centers/national-minerals-information-center/phosphate-rock-statistics-and-information>
- [2] International Fertilizer Association. (2023). World phosphate rock capability 2023: summary report. Retrieved from <https://www.ifastat.org/supply/Phosphate%20Products/Phosphate%20Rock>