# Appendix B: About the construction of the ITS dataset

## Technological Import Dependency Ratio (TIDR) as a proxy to logistical vulnerability

The formula applied for constructing the TIDR for mining and agriculture was based on the Cereal Import Dependency Ratio (CIDR), which is a measure created by the FAO to characterize the dependency of a country on cereal imports (FAO, 2023). The formula for the CIDR is as follows:

Following the principle of parsimonia, we will use the CIDR formula as the foundation for the construction of the TIDR in agriculture and mining, but adapted to the available data in the source, which is the full volume of trade operations (imports and exports) of specific technologies in these sectors accross the globe. Therefore, the formula for the TIDR will be as follows:

This approach has the main advantage of being easily replicable, adaptable and interpretable, as it is based on the simple comparison of the volume of technology imports and exports in each sector. On the other hand, it has the disadvantage of not considering production data, which could be relevant to assess emergent development in technology production for agriculture and mining (because if a country develops internal technologies that are not traded internationally, its wouldn’t be visible with the TIDR formula until said country trades that innovation globally).

The trade operations gathered refer to the following technologies in the agriculture and mining sectors, whose HS codes are used to identify them in the COMTRADE database:

| **HS\_Code** | **Description** |
| --- | --- |
| 841710 | Furnaces and ovens for roasting, melting or other heat-treatment of ores, pyrites or of metals |
| 842320 | Weighing machinery (excluding balances of a sensitivity of 5 cg or better), including weight-operated counting or checking machines; weighing machine weights of all kinds |
| 842441 | Agricultural or horticultural sprayers |
| 842482 | Agricultural or horticultural appliances for projecting, dispersing or spraying liquids or powders |
| 8427 | Fork-lift trucks; other works trucks fitted with lifting or handling equipment |
| 8428 | Other lifting, handling, loading or unloading machinery (for example, lifts, escalators, conveyors, teleferics) |
| 8429 | Self-propelled bulldozers, angledozers, graders, levellers, scrapers, mechanical shovels, excavators, shovel loaders, tamping machines and road rollers |
| 843031 | Machinery for soil preparation or cultivation; lawn or sports-ground rollers |
| 843039 | Other machinery for soil preparation or cultivation |
| 8432 | Agricultural, horticultural or forestry machinery for soil preparation or cultivation; lawn or sports-ground rollers |
| 8433 | Harvesting or threshing machinery, including straw or fodder balers; grass or hay mowers; machines for cleaning, sorting or grading eggs, fruit or other agricultural produce, other than machinery of heading 8437 |
| 8434 | Milking machines and dairy machinery |
| 8435 | Presses, crushers and similar machinery used in the manufacture of wine, cider, fruit juices or similar beverages |
| 8436 | Agricultural, horticultural, forestry, poultry-keeping or bee-keeping machinery, including germination plant fitted with mechanical or thermal equipment; poultry incubators and brooders |
| 8437 | Machines for cleaning, sorting or grading seed, grain or dried leguminous vegetables; machinery used in the milling industry or for the working of cereals or dried leguminous vegetables, other than farm-type machinery |
| 847410 | Machinery for sorting, screening, separating, washing, crushing, grinding, mixing or kneading earth, stone, ores or other mineral substances, in solid (including powder or paste) form |
| 847420 | Machinery for agglomerating, shaping or moulding solid mineral fuels, ceramic paste, unhardened cements, plastering materials or other mineral products in powder or paste form; machines for forming foundry moulds of sand |
| 8701 | Tractors (other than tractors of heading 8709) |
| 8704 | Motor vehicles for the transport of goods |
| 8709 | Works trucks, self-propelled, not fitted with lifting or handling equipment, of the type used in factories, warehouses, dock areas or airports for short distance transport of goods |
| 9015 | Surveying (including photogrammetrical surveying), hydrographic, oceanographic, hydrological, meteorological or geophysical instruments and appliances, excluding compasses; rangefinders |

We suggest that the COMTRADE dataset is a powerful tool to set the ground for an empirical analysis of the ITS in general, because it provides enough data to identify the very extremes of technological trade’s dependence: net exporter countries (whose TIDR value is expected to be negative and lower than -100, because they export more technology than they import) and net importer countries (whose TIDR value is expected to be 100 or near, because they import more technology than they export) [[1]](#footnote-1).

TIDR values will be considered as a proxy for a logistical kind of ITS, because it allows us to identify economies whose technified capital (i.e. machinery) is based at least to some extent on its import rather than its production. With this in consideration, the authors suggest that a TIDR value above 50 is a sign of a high dependency on technology imports.

## The patent data as a proxy for an innovation and development subordination

Datasets available through the World International Patents Office (WIPO) and other global stakeholders on the International Patents Systems offer valuable information about intellectual property around the world, but these registers are not directly aggregated or ready to be analyzed in a categorical manner through their web interfaces, which means that the *surface* data isn’t directly suited for analyzing inventive activity with the specific aim of this article (that is, the search for indicators of sustained hierarchies among countries in particular technologies).

To gather the data necessary for the analysis (that is, at least a count of patents granted per country for agricultural and mining technologies through time), a programmatic approach is taken based on *The WIPO Patent Analytics Handbook* (Oldham 2022) applied to the *Google Patents Public Dataset* (Google Cloud 2024). Following the suggestions of the handbook, the data is gathered taking advantage of five columns offered by Google’s dataset: patent\_number, country\_code, publication\_date, family\_id and cpc\_code. Respectivelly, these columns offer the patent identifier number, the country of origin, the publication date, the patent family identifier number and the Cooperative Patent Classification (CPC) code of the patent; the latter being a classification system used to categorize patents by field and/or technology by the European Patent Office and other global stakeholders. Given the interest in looking for patents of global relevance in agriculture and mining, during the data download and processing a column is created to confirm that each observation is a patent associated with a patent family (that is, that ‘family\_id’ isn’t 0 or null).

To identify the patents of interest in agriculture and mining, the CPC codes are used to filter the patents in the dataset according to the general field of appliance of the patented technology. In order to take into account the elements of the 4.0 revolution in the innovation landscape, CPC codes for data process technologies are also considered. Therefore, CPC codes identified through the search engine provided by the European Patents Office (European Patent Office 2024) and used for the purpose of this research are the following:

| **CPC\_Code** | **Description** | **CPC\_broad\_class** |
| --- | --- | --- |
| A01 | Agriculture; forestry; animal husbandry; hunting; trapping; fishing | Agriculture patents and related |
| E21 | Earth or rock drilling; mining | Mining patents and related |
| G05 | Controlling; regulating systems | Data process patents and related |
| G06 | Computing; calculating or counting systems | Data process patents and related |
| G08 | Signalling systems | Data process patents and related |
| G11 | Information storage systems | Data process patents and related |
| G16Y | Information and communication technology specially adapted for the internet of things [IoT] | Data process patents and related |

With this approach to patent data, the authors aim to describe global trends of intellectual property accumulation in the agriculture, mining and data processing sectors, as a way to complement the TIDR analysis and to provide a more comprehensive view of the ITS as global inequality in technified capital development.

## Financial subordination through net FDI labeled as Fixed Gross Capital Formation

As stated above, the stock FDI as percentage of GFCF data per year is a easily available, but rather general metric. Stock of Gross Fixed Capital Formation (GFCF) was selected as a proxy over stock FDI in USD because it is a more direct measure of the technified capital accumulation in a given country and year, for which it better suits the longitudinal perspective of this analysis. For this article, the data was processed in a similar manner to the TIDR data (that is, input - output of an economy):

The resulting value follows a similar logic to the TIDR: positive values indicate a net inflow of FDI (that is, receiving more FDI than sending out), while negative values indicate a net outflow of FDI (that is, sending more FDI than receiving). While a net inflow isn’t necessarily a sign of subordination, and the abscence of a measurement of local fixed capital financed by national investment limits the possibility to identify autarky, this ratio allows for two things: (1) to identify extreme cases of financial subordination or dominance in the GE, and (2) to keep the principles of parsimonia and data accesibility in the construction of the dataset.

## Final dataset constructed for the analysis of the ITS in mining and agriculture globally

To merge all datasets together into a single file that can be used for the analysis and then make it available to the public use, the authors use the years and country names as *main keys* (that is, the columns that will be used to track each case across datasets). The *countrycode* R package developed by Arel-Bundock, Enevoldsen, and Yetman (2023) is used to standardize the country names across datasets, and the *dplyr* R package developed by Wickham et al. (2023) is used to merge and clean the datasets into a single file. The final dataset is then exported to a *.csv* file that can be used by other researchers to replicate or further the analysis and is available along with the necessary documentation in a public repository curated by one of the paper’s authors [[2]](#footnote-2).

The composition of the final dataset is as follows:

| **Column / Variable** | **Description** | **Interpretation** |
| --- | --- | --- |
| year | Year of observation | Year of observation |
| country | Country of observation | Country of observation |
| IDIT | Proxy for logistical subordination (LS) in agriculture and mining | Values closer to 100 indicate a high dependency on technology imports, negative values indicate a net exporter |
| IDIT\_China | LS related to China | Values closer to 100 indicate a high dependency on technology imports, negative values indicate a net exporter |
| IDIT\_USA | LS related to USA | Values closer to 100 indicate a high dependency on technology imports, negative values indicate a net exporter |
| main\_tech\_capital\_source | Main source of technological capital imports | Main source of LS in agriculture and mining technologies |
| agr\_pat | Proxy for intellectual property subordination in agriculture | High values indicate a high relative position in the global intellectual property system |
| dat\_pat | Proxy for intellectual property subordination in mining | High values indicate a high relative position in the global intellectual property system |
| min\_pat | Proxy for intellectual property subordination in data processing | High values indicate a high relative position in the global intellectual property system |
| Net\_GFCF | Net FDI as percentage of GFCF | Positive values indicate a net inflow of FDI as GFCF, negative values indicate a net outflow of FDI as GFCF |
| ECI | Economy Complexity Index (ECI) | A measure of the productive capabilities of a country, higher values may indicate a more diverse (or expanded) economy |
| ECI\_rank | ECI ranking within the available global economy data | Ranking of each country according to ECI indicator. A country in top 10 ranking can be considered as a country that its economy have expanded greatly respective to the rest of the world |

This dataset contains information for 199 countries and 34 years, but not necessarily continuous; for example, a country may have full data for TIDR but not for patents, or partial data of both (for a complete summary of the final dataset composition, see Appendix B). The authors suggest that the dataset is useful as a first approach to creating an ITS set of indicators that can be used to research some GE dynamics. Among other things, the authors suggest that the dataset can be used to identify the relative position of a country in the global economy of the sectors analyzed, and to aggregate (for example, by size in USD or location of the economy) or complement the data with other sources to test hypotheses about the global development of imperialism and subordination in specific countries. In this line, ECI indicators are also added to test the capability of the ITS dataset to understand global division of labour, and to encourage other researchers to incorporate metrics into the data set that will enrich the debate on the measurement of development in a global perspective. Therefore, the dataset wil be used to test the following hypotheses that emerge from the historical analysis in which the ITS framework was built.

# References

Arel-Bundock, Vincent, Nils Enevoldsen, and CJ Yetman. 2023. *Countrycode: Convert Country Names and Country Codes*. <https://CRAN.R-project.org/package=countrycode>.

Food and Agriculture Organization of the United Nations (FAO). 2023. “Cereals Imports Dependency Ratio.” <https://www.fao.org/faostat/en/#data/FS>.

Google Cloud. 2024. “Google Patents Public Data.” <https://console.cloud.google.com/marketplace/product/google_patents_public_datasets/google-patents-public-data?project=global-patents-analysis>.

Oldham, Paul. 2022. *The WIPO Patent Analytics Handbook*. <https://wipo-analytics.github.io/>.

R., Rodrigo B. 2024. “Public ITS Repository.” <https://github.com/RodrigoBR1/Public_ITS>.

1. Using the adapted FAO’s cereal dependency formula interpretation. [↑](#footnote-ref-1)
2. The repository is available at *https://github.com/RodrigoBR1/Public\_ITS.git*, for further information on the data and the analysis. [↑](#footnote-ref-2)