



► **Project Insight**

# Uncovering the complexities of global value chains

July 2026

Project partners:



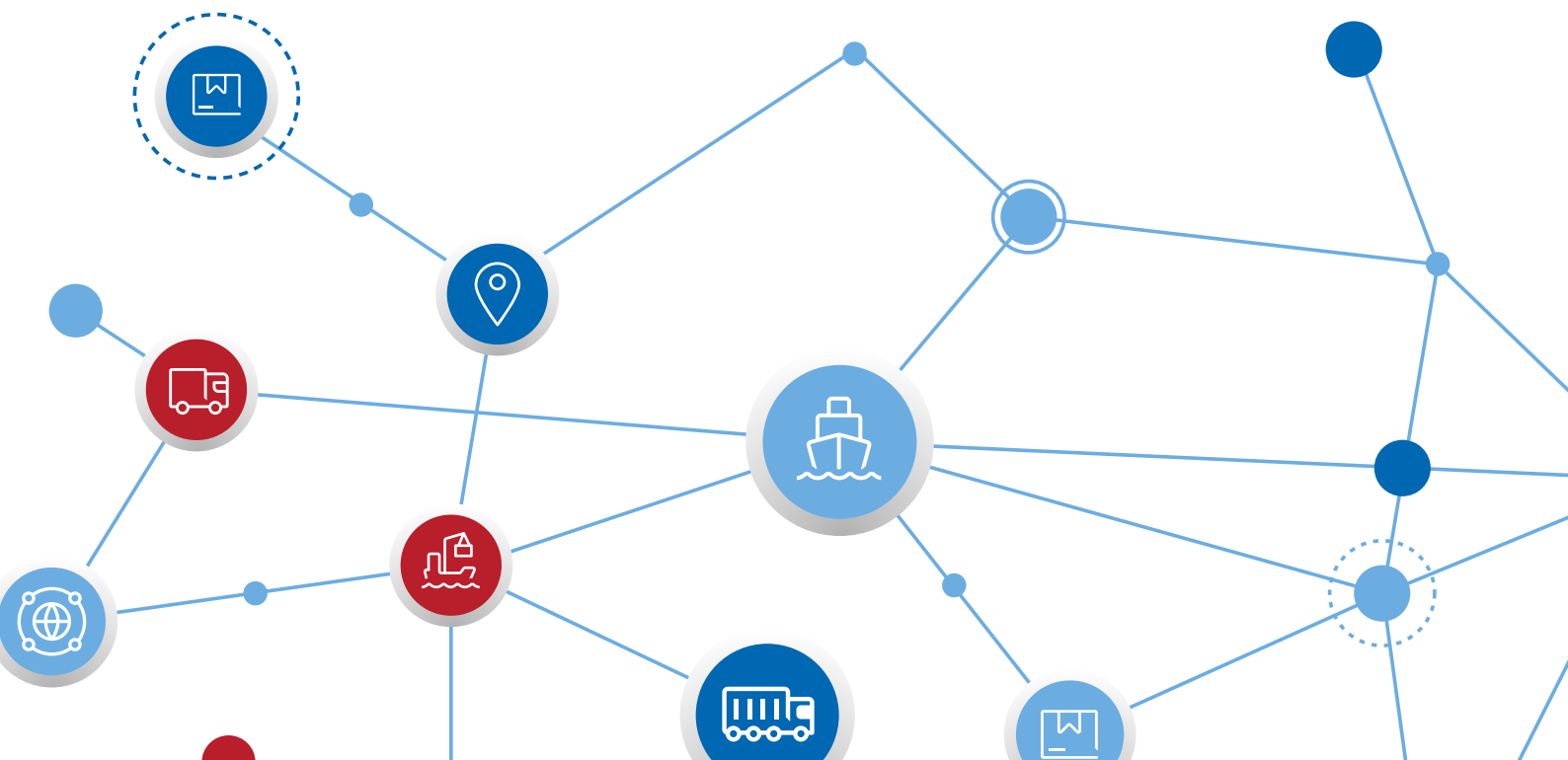
HONG KONG MONETARY AUTHORITY  
香港金融管理局



OECD



DIW BERLIN



# Contents

---

<b>Executive summary</b>	<b>4</b>
<b>1. Introduction</b>	<b>6</b>
<b>2. Current gaps in monitoring global value chains</b>	<b>8</b>
<b>3. Project objectives and scope</b>	<b>11</b>
<b>4. Data selection</b>	<b>13</b>
4.1. Commercial data: maritime shipments, supply chain relationships, entity attributes and vessel tracking	15
4.2. Public data: macroeconomic data and cross-economy input-output tables	17
<b>5. Solution architecture and data preprocessing</b>	<b>19</b>
5.1. Solution architecture	19
5.2. Data preprocessing	20
<b>6. Analytics and visualisation</b>	<b>30</b>
6.1. Overview of global supply chains	31
6.2. Aggregate time series trends	33
6.3. Resilience and bottlenecks	35
6.4. Trade exposure and concentration	38
6.5. Network metrics	41
<b>7. Conclusion</b>	<b>44</b>
<b>References</b>	<b>45</b>
<b>Contributors and acknowledgements</b>	<b>48</b>

## Executive summary

---

Global value chains (GVCs) underpin production and trade with direct implications for the real economy and financial system, yet they remain poorly understood at the granular level. Limited visibility into these complex, firm-level supply chain networks can obscure vulnerabilities, with risks remaining undetected until they materialise and shape macroeconomic outcomes such as growth, inflation and financial stability. Recent shocks such as the Covid-19 pandemic have highlighted the risks of this lack of transparency. For central banks and policymakers, the challenge is clear – aggregate statistics cannot capture the intricate firm-to-firm relationships or the early structural shifts that shape GVC dynamics well before they appear in headline data. This lack of visibility matters because shocks propagate through these networks in ways that can amplify inflationary pressures, mask concentration risks and create correlated balance sheet exposures with spillovers to credit markets and asset prices. In an increasingly interconnected global economy, understanding and monitoring GVC dependencies can therefore support central banks in their core mandate to safeguard monetary and financial stability.

Project Insight attempts to address this gap by developing a proof of concept (PoC) dashboard to enhance the monitoring of GVC developments. The project was conducted in collaboration with project partners – the Hong Kong Monetary Authority (HKMA), the Organisation for Economic Co-operation and Development (OECD) and DIW Berlin. It also brought together a global observing-member stakeholder group of supply chain experts across central banks, international organisations, academia and the private sector. The development was grounded in user needs, drawing on continuous feedback from this international expert group to identify concrete monitoring challenges and data gaps faced by central banks and other policymaking institutions.

Leveraging a snapshot of historical, granular supply chain data from 2018 to 2024, Project Insight integrates various detailed, commercial data sets on firm-to-firm activity with publicly available macroeconomic information. These inputs were transformed into a harmonised suite of foundational GVC metrics, covering cross-sectional and time series trends, trade exposure and concentration risks, resilience and bottlenecks, and network centralities. These metrics spanned multiple layers of granularity of the supply chain network, from entity-to-entity up to economy-to-economy levels.

A custom-designed interactive dashboard was built leveraging open source tools to feature the data and analytics. The dashboard is equipped with interactive filters across the time, product, economy and trading-direction dimensions. This front-end interface was developed using a modern, component-based architecture, enabling modular design and reusable interface elements that can be flexibly extended as new data and metrics are incorporated, allowing this dashboard to be used for other purposes. The dashboard enables potential users to drill down from aggregate patterns to specific sectors, counterparties or relationships, supporting both high level monitoring and more detailed analytical exploration. To support adaptability, this report includes a level of technical detail on data and methodology that may be of particular interest to analytical and research audiences.

The PoC demonstrates how granular firm-to-firm data can support more timely and informed policy decisions. For example, it could improve the assessment of inflationary pressures driven by global supply chain conditions and help to identify hidden concentrations and transmission risks that could impact financial stability. In addition, Project Insight distils a set of practical approaches to data cleaning, core metrics and visualisation that policymakers can adapt and build on using their own data sets and use cases. Rather than addressing a single policy question, the PoC prioritised the development of robust data structures and a standardised technological and analytical foundation that can be adapted to a wide range of policy, research and risk monitoring applications. This enables deeper and more timely visibility into evolving supply chain structures, supports the identification of emerging risks and opportunities, and strengthens the capacity for informed policy analysis.



## 1. Introduction

---

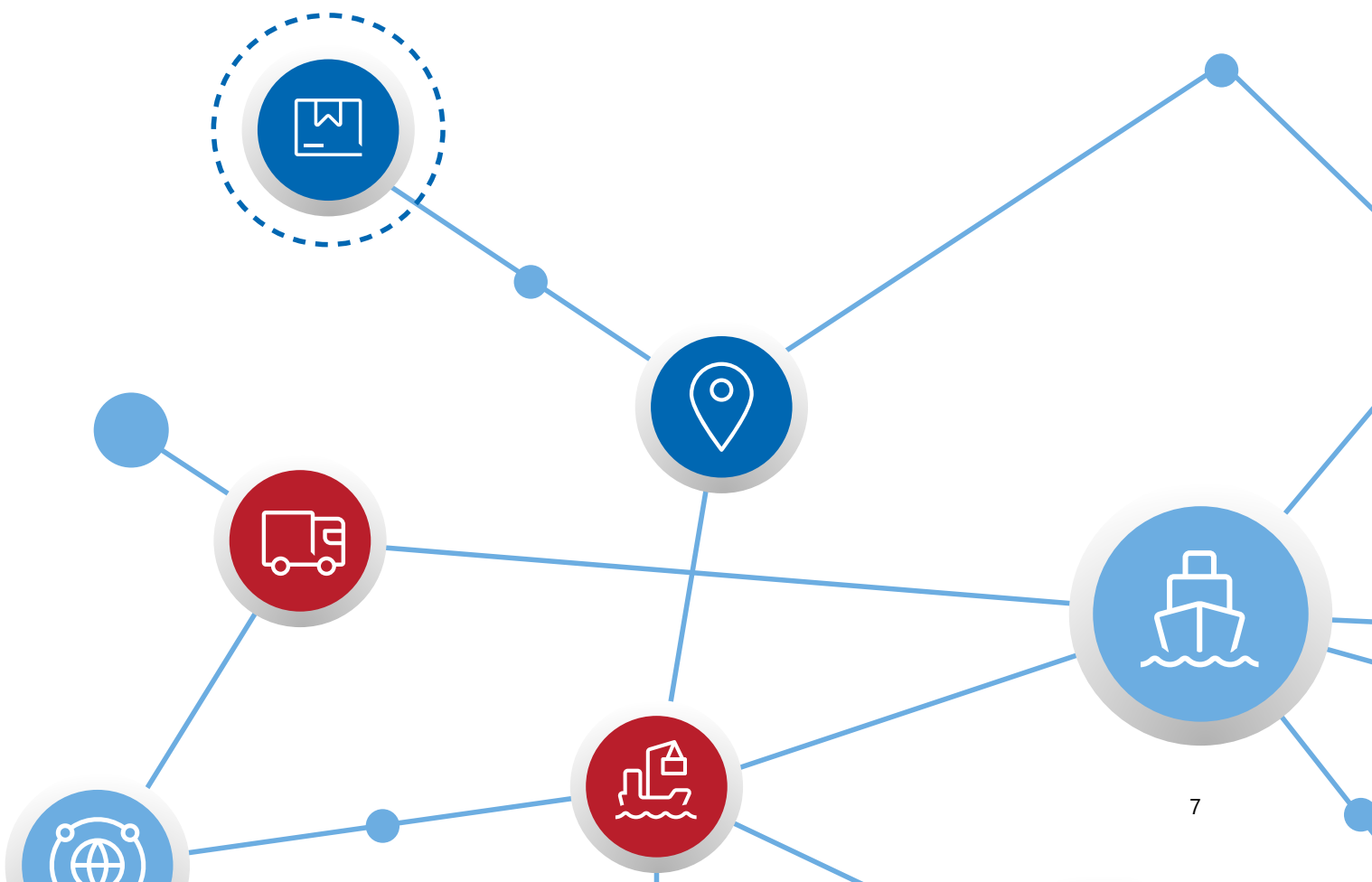
Global value chains (GVCs) are made up of a highly complex web of interlinked firms, balance sheets and production stages spread across multiple economies and industries (Antràs and Chor (2022)). Distributing production processes around the globe has long been accepted as common practice (Mann (2012)), but significant gaps remain in our understanding of the underlying complexities of supply chains. Beyond aggregate trade statistics, only limited information is available at a global and cross-sectoral level to capture the firm-level direct and indirect relationships that drive GVC trends and dynamics.

This presents a major challenge for central banks and policymakers, as these supply chain dependencies can have implications for both financial and monetary stability. Shocks propagate through these granular networks (Acemoglu et al (2012); Magerman et al (2016); Baqaee and Farhi (2019)), and concentration risks and vulnerabilities can easily be obscured in the aggregate data. Firms may be highly exposed to a narrow set of upstream suppliers or downstream customers, giving rise to correlated balance sheet risks and cross-border transmission channels. Furthermore, given the prevalence of foreign subsidiaries in supply chains, disruptions can spread not only through trade but also through ownership structures, blurring the distinction between real and financial shocks (Qiu et al (2025)). As a result, supply chain disruptions can spill over into credit markets, equity valuations and bank balance sheets, particularly when highly central firms in the network are affected. Addressing this lack of visibility is therefore relevant to the financial stability mandate of central banks and supervisory authorities.

GVCs also play an important role in shaping inflation dynamics. International input-output linkages can increase the sensitivity of domestic inflation to global economic slack beyond what conventional, aggregate trade measures would suggest (Auer et al (2017); Auer et al (2019)). Structural shifts in trade patterns also occur at the micro level long before they start to show up in macroeconomic statistics. Indeed, recent shocks to global supply chains such as the Covid-19 pandemic exposed just how little we know about granular supply chain relationships and how they can spread disruptions and affect economic growth, inflation and financial markets (Chen et al (2023); IMF (2024)). It is therefore important for central banks and policymakers to understand granular GVC dependencies and monitor their shifts and trends.

Against this backdrop, Project Insight developed a proof of concept (PoC) for analysing GVCs using a snapshot of historical data. The PoC brought together a modular and scalable data architecture, a suite of standardised GVC analytical tools and an interactive dashboard with the aim of sharing practical insights that policymakers can draw on in their own analytical work and use cases. The project combined highly detailed commercial data sets on shipments, firms, ownership linkages and reported supply chain relationships, with publicly available aggregate information, together comprising billions of data points. It applied data engineering techniques to clean, reconcile and integrate these heterogeneous sources, as well as to construct multilayered supply chain networks across entities, industries and economies. Everything was built on a flexible, cloud-based and collaborative data platform that unified data processing, analytics and visualisation within a single ecosystem. This enabled project stakeholders to transform raw, granular data into actionable intelligence through a structured pipeline for cleaning, harmonisation and integration.

The project also developed a bespoke and interactive dashboard, leveraging data visualisation and network-mapping techniques, to enable exploration of complex and multidimensional supply chain linkages. The dashboard interface was custom-built using an open source front end, designed to support modular and reusable components along with scalable integration with the underlying data architecture. This approach ensures that new metrics, data sets and visualisations can be incorporated flexibly over time without requiring a full redesign of the user interface. It also allows for these modules to be used for other projects.

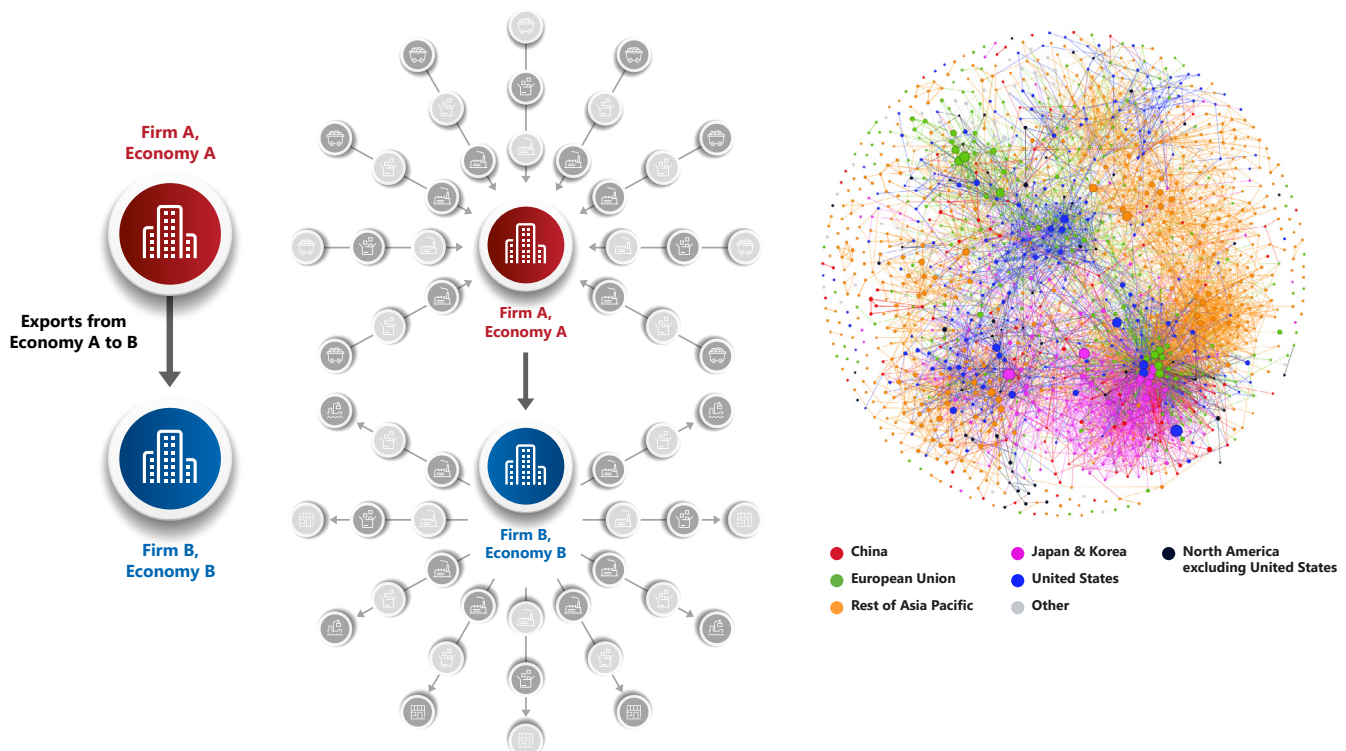


## 2. Current gaps in monitoring global value chains

Despite the importance of supply chains, publicly available data and tools to understand their underlying complexities are limited, especially at a global level and in a timely manner. The main sources of information on international trade linkages are customs-based merchandise trade data, which capture physical goods crossing borders, and balance of payments statistics, which capture transactions for the exchange of goods and services between residents and non-residents (UN (2021)). These two types of data are reported at varying levels of detail and with different publication delays and frequencies across economies. In some cases, these statistics can be provided at a global and highly detailed product level (see eg the UN's Comtrade database (UN (2026))). However, the data cannot shed light on the granularity of the firm or entity, which are the main agents of the supply chain. As firms weave a web of supply chains across the global trading system and as economic activity is increasingly geographically dispersed, new data and measurement systems are needed to track the international footprint of global firms (Avdjiev et al (2018); Qiu et al (2025)).

Aggregate trade data mask significant underlying complexities<sup>1</sup>

Graph 1



1 The network diagram on the right-hand side excludes nodes with less than 10 linkages. Nodes represent firms (sizes of which are proportional to a firm's importance in the overall network, measured by eigenvector centrality), and edges represent an unweighted supplier-customer relationship. Node colour corresponds to the firm's primary location while edges take the colour of the supplier node.

Sources: BIS; H Qiu, H S Shin and L Zhang "Mapping the realignment of global value chains", BIS Bulletin, no 78, October 2023.

Firms engaged in GVCs are typically involved in extensive and intertwined production networks, making it costly and difficult to track their indirect exposures (Pichler et al (2023)). Firms themselves, let alone policymakers, often lack visibility into the linkages they may have further up or down their supply chains beyond the direct relationships (Zhang (2021)). This lack of visibility can mask important details that can have macroeconomic implications, as has been evident during recent episodes of major supply chain disruptions (Bonadio et al (2021); Diem et al (2022)). Take the example shown in Graph 1 of firms A and B (in the red and blue nodes, respectively) located in economies A and B, respectively. Starting from the left-hand side of the graph, if firm A sells goods to firm B, this would typically be reflected in the aggregate official and publicly available statistics as an export of economy A to economy B, as indicated by the downward arrow. While critical to the understanding of aggregate trade dynamics, these statistics cannot capture the string of upstream suppliers and downstream customers that form part of firm A's and firm B's respective supply chains, and who are in turn connected to other strings of supply chain counterparties; this wider ecosystem is illustrated in grey in the centre of the graph. Zooming out even further, supply chains involve millions of entities spread across multiple different industries and economies, as shown by the network diagram of firms (represented by nodes) and their supply chain relationships (represented by edges) on the right-hand side of the graph. Together, these relationships make up a complex web of interlinked firms, sectors and economies.



Various efforts by international organisations and academic institutions, such as the Organisation for Economic Co-operation and Development (OECD), the Asian Development Bank (ADB), the International Monetary Fund (IMF) and the University of Groningen, to capture these complexities have made significant strides in deepening our understanding of supply chains (Guilhoto et al (2022); ADB (2015); Arslanalp et al (2025); Timmer et al (2015)). However, more granular insights down to the level of individual firms or shipments are still lacking due to data limitations. Official statistical, tax and customs agencies collect firm- and shipment-level data, but they are often confidential. These data are also limited to the transactions (whether cross-border or domestic) of the firms in a single jurisdiction or economic area and cannot be merged with the data of other economies, providing only a partial picture of supply chains (eg Bacilieri et al (2025); Magerman et al (2026)). Furthermore, customs data typically omit within-customs-union or domestic trade flows, which may represent critical links in supply chains. Granular supply chain data at a global level can so far be accessed only (i) through commercial data providers who collect such data from firm disclosures and surveys (which tend to be biased toward large and publicly listed firms, excluding smaller entities that may play significant roles in GVCs); (ii) by engaging with the customs agencies of multiple jurisdictions; or (iii) by working directly with shipping carrier companies, among others.

No single data provider has a full view of global supply chain activity or the characteristics of the entities involved in those trading relationships. Capturing this at a more comprehensive level often requires matching multiple different commercial data sets, which comes with its own set of difficulties. Processing raw shipping and customs data presents significant technical and analytical challenges. A substantial portion of these data is manually entered. This often leads to product classification errors, particularly in bills of lading.<sup>1</sup> Further complications arise from inconsistencies in recording standards and trade valuation methods (eg variations between full container loads and less-than-container loads) or differing approaches to international commercial term valuation (eg invoice value; free on board; or cost, insurance and freight). Furthermore, harmonising product and industry classifications is complex and may result in information loss due to non-unique mappings. Firm identities are often represented in various formats (eg company names, commercial identifiers or global reference systems), requiring fuzzy matching or cross-referencing of different reporting standards. Establishing linkages between customs declarations, transport logs and firm-level records at the shipment level is particularly difficult and resource-intensive.

Fine-grained information down to the level of individual firms or shipments at a global level can offer more actionable insights and capture indirect flows across sectors and geographies to better understand GVC dynamics. Furthermore, existing trade and supply chain measures often display low frequency and delayed publication. Official trade data among other relevant statistics often suffer from multi-year reporting lags. More timely data would enable earlier detection of disruptions and trade shifts, which can have important macroeconomic effects.

<sup>1</sup> A bill of lading is a legal document issued by a carrier company to a shipper/exporter that details the type, quantity and destination of goods being transported and acts as a "receipt" for the goods (see DHL (2025)).

### 3. Project objectives and scope

---

Leveraging a snapshot of historical data, Project Insight developed a PoC to better understand global supply chains by combining diverse sources of granular and macro data. It leveraged big data analytical tools, and drew on the expertise of an international network of supply chain experts. The project aimed to deliver a foundational “knowledge base” for central banks and policymakers to better track key developments in GVCs and evaluate their economic and financial implications at a time when understanding supply chains has become vital to economic policy and financial oversight. Project Insight spans data engineering, economic analysis and the creation of interactive dashboards for visualisation. Notably, the project delivered the following three key outputs at various levels of granularity:

- **a standardised set of metrics and analyses** that can provide insight into the structure and evolution of global supply chains and their broader economic and financial implications
- **an interactive dashboard** to monitor global supply chains
- **an end-to-end data architecture** to support the system’s functionality.

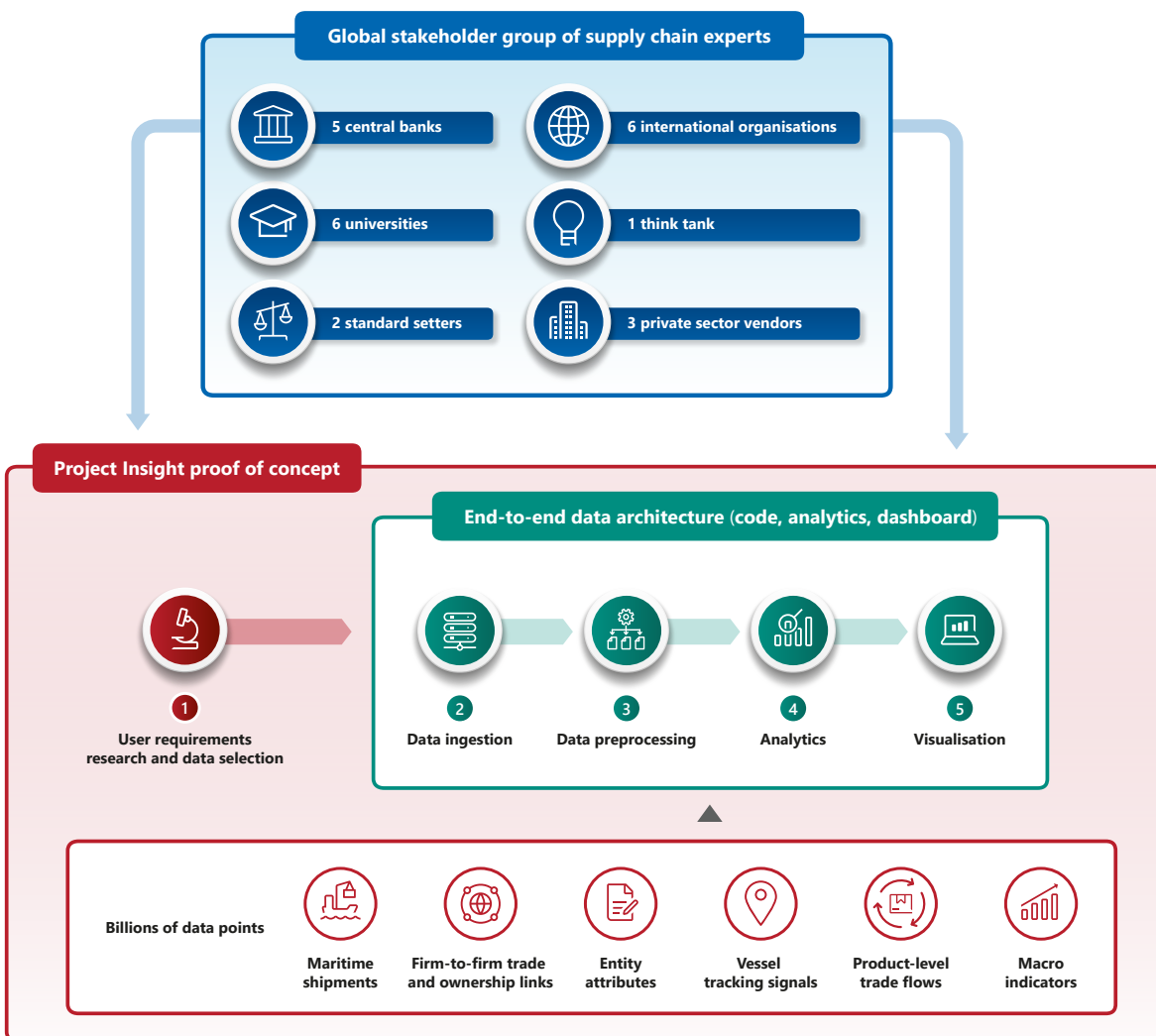
An illustration of the project’s overall structure is provided in Graph 2. Leveraging the BIS Innovation Hub’s role as a focal point for global central bank and policy experts on innovation, the project started by bringing together a group of key stakeholders from around the world. As detailed in the top part of the graph, this included project partners and observing members with supply chain knowledge and experience across five central banks, six international organisations, six universities, one think tank, two standard-setting bodies and three private sector vendors. This ensured that the PoC build was anchored in stakeholder needs and driven by user demand. Stakeholders came from an international network, and they work across various functions within their organisations, including statistical divisions, research departments, innovation hubs, data science teams, policy advisory roles and executive offices. They could therefore provide diverse and rich perspectives.

At the beginning of the project in its first stage, feedback was first collected through systematic research on user requirements. This included a comprehensive questionnaire and several stakeholder-wide virtual meetings. The findings from this research were summarised in the form of a detailed report. After that, feedback on the project’s progress was gathered on a regular basis through bilateral consultations, working groups and other project meetings. This enabled the creation of a structured inventory of challenges faced by stakeholders in monitoring supply chains, their data and analytical needs, user expectations and requirements, best practices and data integration protocols. This inventory then fed into the subsequent data selection process and PoC build. A variety of data sets related to supply chains were selected as detailed along the bottom of Graph 2, covering maritime shipments, firm-to-firm trade and ownership links, entity attributes, vessel tracking signals, product-level trade flows and macroeconomic indicators. Together, these made up billions of observations.

After the completion of this first phase – ie research into user requirements and data selection – the development of the PoC followed four more key phases that together made up the end-to-end data architecture illustrated in the green rectangle of Graph 2. The build process started with one-off ingestion of the various forms of supply-chain-related historical data selected; this was followed by cleaning and preprocessing of the data. This then served as the basis for the creation of a standardised suite of GVC indicators during the analytics phase; and finally the project wrapped up with the development of an interactive and modular PoC dashboard visualisation tool, which showcased some of the cleaned data and the indicators generated during the analytics phase.

Project Insight structure

Graph 2



Sources: BIS.

It is important to note that Project Insight was not intended to be comprehensive in its data selection, analytics or visualisation in monitoring global supply chains. Rather, it aimed to be a starting point that reflects lessons learned from working with historical data and to provide practical references that policymakers may adapt to their own contexts.

## 4. Data selection

---

After rigorous assessment of numerous commercial data sets relevant for measuring GVCs, the project team selected the following types of data with an emphasis on methodological complementarity in GVC analysis:

- detailed package-level maritime shipment data sourced from major global carrier companies
- entity-level<sup>2</sup> supply chain relationships based on entity reporting, surveys and news articles
- entity attributes, such as size, location and industry
- corporate ownership linkages, such as parent-subsidiary relationships
- automated information system (AIS) maritime vessel tracking information.

This was supplemented with relevant data from publicly available channels either provided by project stakeholders or extracted from open sources:

- detailed product-level economy-pair merchandise trade flows
- macroeconomic indicators
- cross-economy input-output tables.

Together, these data points made up more than 15 billion observations. A more detailed overview of each data source is provided in Table 1 and in sections 4.1 and 4.2.

<sup>2</sup> Entities can be firms (defined by a juridical structure or reporting requirements), establishments (separate geographical sub-units of firms) or affiliates (partially or fully owned by a domestic or foreign parent).

Summary of Project Insight data

Table 1

<b>Data set and source</b>	<b>Main features</b>	<b>Frequency and time coverage</b>
<b>Commercial data</b>		
<b>Maritime shipments</b> <i>Source (commercial): Data Vendor 1.</i>	Global seaborne containerised shipments from shipper to consignee entity, collected mainly from major global carrier companies.	<i>Frequency:</i> shipment entry time, up to hourly timestamps. <i>Time period covered:</i> 2018-24.
<b>Supply chain relationships</b> <i>Source (commercial): Data Vendor 2.</i>	Firms' key customers, suppliers, competitors and strategic partners, collected mainly from firm disclosures and publicly available information.	<i>Frequency:</i> annual. <i>Time period covered:</i> since 2003 for North America; since 2011 for Europe; since 2013 for Asia.
<b>Entity attributes</b> <i>Sources (commercial): Data Vendor 1; and Data Vendor 2.</i>	Entity information such as legal identifiers, location, age, industry, size, etc, collected mainly from government registries and publicly available information.	<i>Frequency:</i> annual. <i>Time period covered:</i> two snapshots (1 January 2024 and 1 January 2025) for Data Vendor 1; since 2015 for Data Vendor 2.
<b>Corporate ownership linkages</b> <i>Source (commercial): Data Vendor 1.</i>	Corporate entities, beneficial owners, minority/majority shareholders and ultimate parent companies, collected mainly from firm registries and publicly available information.	<i>Frequency:</i> annual. <i>Time period covered:</i> two snapshots (1 January 2024 and 1 January 2025).
<b>AIS vessel tracking</b> <i>Source (AIS data provided by commercial data vendor): United Nations Global Platform (UNGP).</i>	Automated tracking system for monitoring maritime traffic with information on voyage, vessel and movement-related activity.	<i>Frequency:</i> in-motion vessels send signals every few seconds. <i>Time period covered:</i> historical UNGP-AIS data is available from December 2018.
<b>Public data</b>		
<b>Economy-pair goods product trade flows</b> <i>Source (public): UN Comtrade.</i>	Bilateral trade flows at the detailed product level up to Harmonized System six-digit classification.	<i>Frequency:</i> monthly and annual. <i>Time period covered:</i> 2018-24.
<b>Macroeconomic indicators</b> <i>Sources (public): World Bank; IMF; FRED; CEPII; among others.</i>	Economy macro indicators (GDP, CPI, exchange rates, etc) and economy-pair information (distance, trade agreements, etc).	<i>Frequency:</i> monthly, quarterly and annual, depending on the indicator and source. <i>Time period covered:</i> can go back to the mid-1900s in some cases, depending on the indicator and the source. More recent data usually goes up to the early 2020s.
<b>Cross-economy input-output tables</b> <i>Source (public): ADB; OECD.</i>	Tables that track the flow of goods and services between industries across different economies in terms of value added and final demand flows.	<i>Frequency:</i> annual. <i>Time period covered:</i> 2017-24 for ADB data, 1995-2022 for OECD data.

## 4.1. Commercial data: maritime shipments, supply chain relationships, entity attributes and vessel tracking

The key data underpinning Project Insight's PoC are Data Vendor 1's maritime shipment data set collected from carrier companies (mostly capturing bill of lading information). This data set provides detailed information on over 300 million seaborne containerised packages and shipments<sup>3</sup> between more than one million entities for the period 2018-24 and covers around 50% of global containerised maritime trade in terms of container volume.<sup>4</sup> In particular, these data include the following key line-item details for each shipment:

- **shipment identification (ID)**, as part of a unique identification system for shipments
- **timestamps** for shipment booking entry, vessel departure and arrival, container gate in and gate out, and vessel loading and unloading, all recorded at the Universal Time Code level
- **shipper/exporter and consignee/importer details**, such as their unique identifier, entity name, Standard Industrial Classification (SIC) industry classification,<sup>5</sup> and location
- **port information**, such as departure and destination ports names, locations and United Nations Code for Trade and Transport Locations (UN/LOCODE) identifiers<sup>6</sup>
- **vessel information**, such as name and International Maritime Organization (IMO) number<sup>7</sup>
- **product details**, such as the Harmonized System (HS) code<sup>8</sup> (at the six-digit level in most cases)
- **shipment volume** (in 20-foot equivalent units), **weight** (in kilograms) **and value** (in USD), the latter of which is estimated by Data Vendor 1 using product-specific price indices.

3 A single shipment can be split into multiple packages for operational, safety, compliance and cost reasons and in such cases would be recorded as separate observations in the data set.

4 Maritime transport moves over 80% of goods traded worldwide (see UNCTAD (2025)).

5 The SIC is a four-digit code that identifies a company's type of business or "Industry" (see US Securities and Exchange Commission (2026)).

6 The "UN/LOCODE" is used by most major shipping companies to identify locations such as ports, airports and terminals in international trade and transport (see United Nations Economic Commission for Europe (2025)).

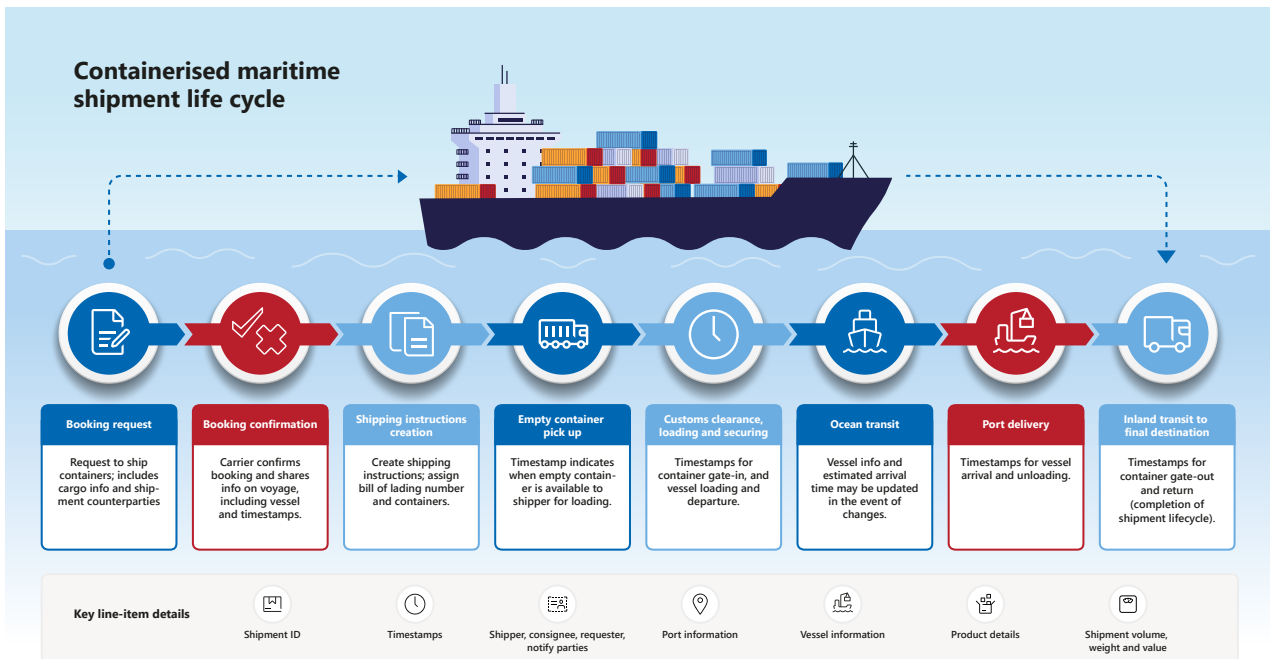
7 The IMO number is a ship identification scheme aimed at assigning a permanent number to each ship to enhance maritime safety (see IMO (2026)).

8 The HS system is an internationally recognised nomenclature created by the World Customs Organization for classifying products traded. It comprises more than 5000 commodity groups identified by six-digit codes, with four- and two-digit aggregations (see WCO (2026)).

Graph 3 illustrates how these details fit into a typical containerised maritime shipment process. The data collection starts from the initial booking request made by a carrier company to move containers from the departure location, to the assignment of a bill of lading number, to the actual movement of containers, to arrival at the final destination.

Maritime shipment data set

Graph 3



Source: BIS.

While the maritime shipment information provides an in-depth snapshot of trade conducted via waterways, it cannot shed light on the supply chain linkages established through other modes of transport, such as between firms in landlocked countries or in the same economy.

To complement the maritime shipment information and capture some of these missing relationships, the project team also selected the Data Vendor 2 supply chain relationships data set, including information on over 400,000 companies and more than three million supply chain relationships identified mainly through reporting (eg firm disclosures, surveys, publicly available information). The data set captures firms' main customers, suppliers, competitors and strategic partners without shedding light on the value of those relationships. Firms and their counterparties are identified primarily by their commercial Data Vendor 2 identifiers.

One key innovation of Project Insight is to combine the maritime shipment data from Data Vendor 1 with the supply chain relationships data from Data Vendor 2 to generate a multidimensional view of global supply chains and relationships between firms. This is because the former has information on the value of supply chain linkages established through maritime shipments, while the latter has information on the extensive margin of trade at both a cross-border and domestic level established through multiple modes of transport (not only maritime, as in the case of Data Vendor 1's data). See Section 5.2.3 for more details on the data-matching process.

For the entities involved in these supply chain linkages, they were also matched to master entity data sets from Data Vendor 1 and Data Vendor 2, each covering around 600 million and 18 million entities, respectively. This helped to capture firm characteristics such as total sales, number of employees, founding date and contact information (eg emails and websites), among others, to better understand the firms making up the global supply chain network.

Furthermore, entities were also connected by ownership relationships using major shareholder information that allowed the identification of immediate and ultimate parent companies. This enabled users to distinguish between foreign subsidiaries and domestic firms in international trade, to capture intrafirm trade among affiliates under the same parent company umbrella and to measure trade from a “nationality” angle (eg see Qiu et al (2025); Alfaro et al (2025)).

Access to detailed, maritime vessel tracking information (AIS) was provided by the UN via its UN Global Platform; the data are sourced from a commercial vendor, and access is limited to international organisations, researchers and statistical offices free of charge to encourage the use of big data for policy applications. AIS is an automated tracking system designed to help sea vessels navigate and avoid collisions. Vessels fitted with AIS equipment send radio messages on their position every few seconds, and different message types contain information related to the voyage (eg location, speed, destination, draught, estimated time of arrival), the vessel (eg name, IMO identifier, type, size) and the position (eg longitude and latitude). While the unit of measurement is at the vessel level (and therefore cannot show the individual shipment or firm), the highly frequent and timely nature of AIS information makes it a useful tool for capturing trade dynamics and supply chain disruptions in real time.<sup>9</sup>

## **4.2. Public data: macroeconomic data and cross-economy input-output tables**

In addition to the core data sets obtained from commercial data providers, the project also leveraged rich data sets made publicly available by international organisations and public institutions, including the UN, IMF, Federal Reserve Bank of St. Louis (St. Louis Fed), World Bank, Centre d'etudes prospectives et d'informations internationales (CEPII), OECD and ADB. For example, the project team used:

- aggregate indicators from the World Bank World Development Indicators for macroeconomic context,
- detailed product-level bilateral trade flows from the UN for cross-checking the shipment-level trends,
- the CEPII Gravity Database for structural trade determinants, and
- exchange rates from the IMF and St. Louis Fed's Federal Reserve Economic Data for the data-cleaning and harmonisation process.

<sup>9</sup> See for example the IMF's PortWatch platform (IMF (2026)) and Bai et al (forthcoming).

Cross-economy input-output (IO) tables from the OECD and ADB were also included to trace production and trade linkages through value added and final demand components across economies and industries at a more aggregate level. Built from harmonised national supply-and-use and IO tables, and international trade statistics, these cross-economy IO tables are a core statistical source for GVC analysis at the economy-industry level. They are also often used to map upstream and downstream industry linkages and to estimate trade in intermediate and final goods and services. While not designed for high-frequency trade monitoring (as these data can be released with a multiyear lag due to the resource-intensive nature of producing such tables) or firm-level analysis, cross-economy IO tables are useful for modelling macroeconomic interdependencies and capturing indirect trade flows.

Together, these historical data sets formed the basis of the PoC build and provided global coverage across geographies and sectors, timely data over a multiyear window, multiple levels of granularity (shipment, entity, product, port, industry and economy) and the ability to build connections across data sets through identifiers and harmonisation protocols. While developed based on bespoke, historical data sets, the data framework was designed to be modular and extensible, enabling users to layer on new data as needed by modifying and building on the existing code. Not all of the aforementioned data sets were included in the PoC dashboard, but they were preprocessed and harmonised for stakeholders to explore in the collaborative sandbox environment and deploy for their own analyses.



## 5. Solution architecture and data preprocessing

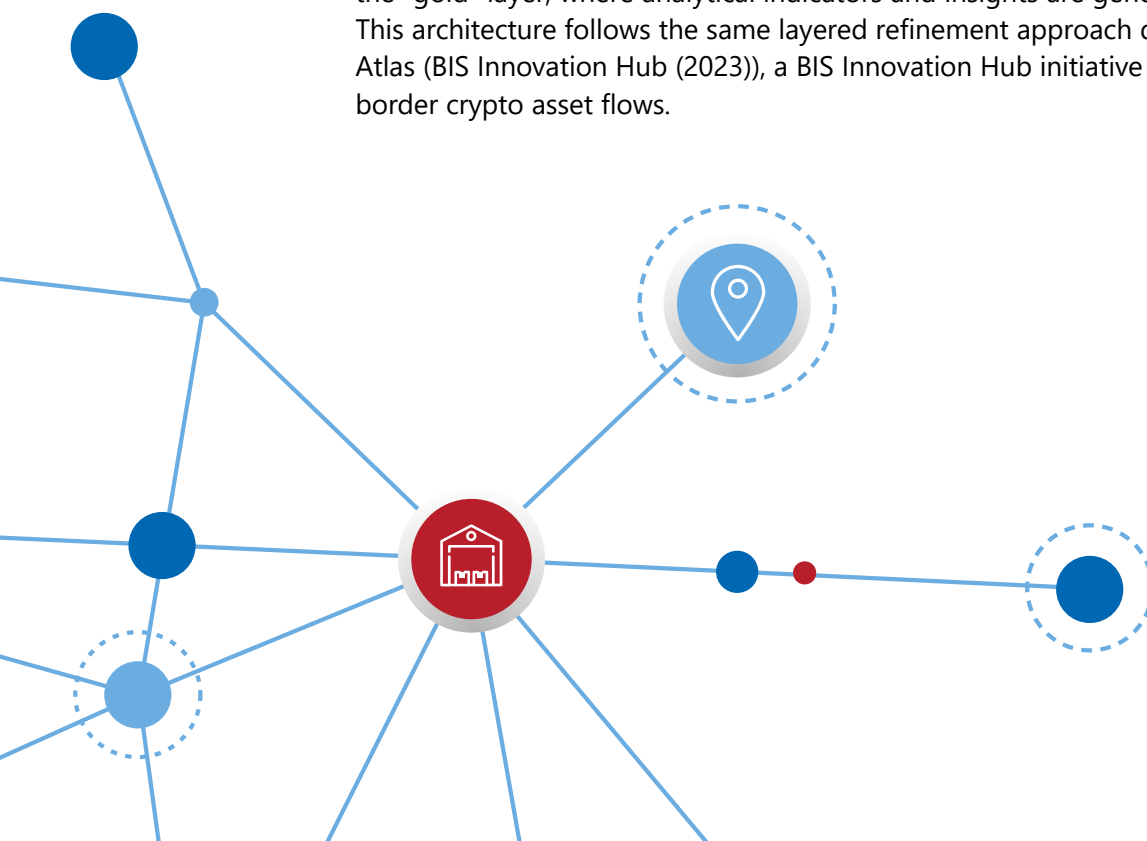
---

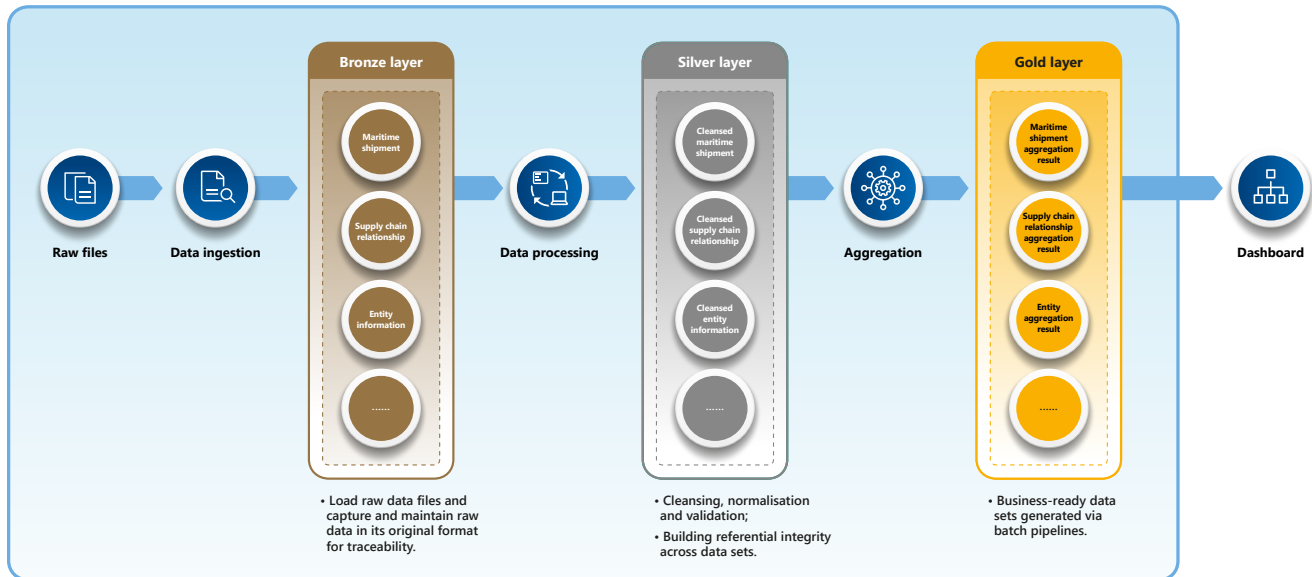
### 5.1. Solution architecture

Turning raw data into actionable intelligence requires a structured data approach built on a flexible data platform to ensure consistency, comparability and analytical usability across the diverse data sets underpinning Project Insight. The project's data platform was deployed in a cloud-based, scalable commercial software solution to flexibly process billions of data points and provide a shared environment for collaboration across the project's multiple stakeholders. The cloud-based solution was chosen for its relatively simple and rapid deployment, scalability, dynamic resource allocation capabilities and seamless integration with diverse data sources.

The project team chose Databricks as the core platform due to its ability to streamline data engineering, data science and business analytics within a single ecosystem. It provided an accessible and secure hosting and processing environment for stakeholders to collectively contribute to the project's analytical outputs, as well as to explore their own use cases. While not an open source platform, Databricks is built on open source technologies such as Apache Spark, which is a powerful distributed computing engine for processing and analysing large data sets. Spark supports a wide range of data engineering, data science and machine learning workflows.

Data processing began with organising Project Insight's historical data sets according to the widely used "medallion" architecture. This divides data into different stages of refinement from raw ingestion to analytics to progressively improve the structure and quality of data as they flow through each layer of the architecture (see Databricks (2026)). Raw data files are ingested into the "bronze" layer, where they are maintained in their original formats for traceability. These raw data sets are then cleaned, normalised and validated in the "silver" layer before being aggregated for the "gold" layer, where analytical indicators and insights are generated (Graph 4). This architecture follows the same layered refinement approach deployed in Project Atlas (BIS Innovation Hub (2023)), a BIS Innovation Hub initiative analysing cross-border crypto asset flows.





Source: BIS.

Project stakeholders conducted data exploration and analytics in Databricks Notebooks. This interactive interface allows analysts to write and execute code in commonly used languages like SQL, Python and R and to automate data workflows through built-in pipeline orchestration tools. Depending on their user permissions, stakeholders could access and query data tables across multiple layers of the medallion architecture, with each data table featuring detailed metadata covering data sources, context, variable names, formats and sample records. The environment supported flexible data engineering and advanced analytics, along with exporting aggregated data for local use.

Custom interactive dashboards were then built leveraging data from the gold layer through integration with front-end dashboard libraries and graph-visualisation components. The Project Insight front end features a custom-developed and interactive web application interface that hosts dynamic dashboards to showcase preprocessed metrics. This offers a user-friendly overview of key indicators and trends developed during the project, along with the ability to drill down into detailed data points. For illustrative purposes, the dashboards were tailored to accommodate various analytical needs, enabling users to filter and examine the data.

## 5.2. Data preprocessing

The project invested significant efforts in the preprocessing and cleaning of the data, which ensured consistency, comparability, integrity and overall readiness for the analytics. This cleaning process included the following:

- standardising identifiers and harmonising classifications (across valuations, products, entities, vessels, ports, industries and economies)

- aligning time periods and frequencies
- defining relevant units of observation and granularity
- deduplicating and handling missing values
- doing data quality checks (eg checking for implausible values and comparing coverage and trends with official statistics).

Where possible, the team linked data sets using globally recognised unique identifiers (eg the Legal Entity Identifier (LEI) and vessel IMO numbers) and correspondence tables (eg those connecting different editions of the HS product codes), or else deployed fuzzy or probabilistic matching methods.

The harmonisation workflow was developed iteratively, with multiple matching and mapping strategies tested before finalising the approach. This standardised process can substantially reduce the time required to onboard and clean similar data sets in future projects. Some examples of the preprocessing steps the project took are provided in further detail in sections 5.2.1 through 5.2.3. These preprocessing outcomes then served as the foundation for the GVC analytics described in section 6.

#### **5.2.1. Maritime shipment data**

Given the large and granular nature of the maritime shipment data, and the fact that information often comes from manually filled bills of lading, several cleaning operations were applied to prepare it for the analytics phase of the project, including but not limited to the following:

- removing observations with:
  - missing or invalid HS product codes
  - missing or negative shipment weight, volume or value
  - inconsistencies in the timestamps of container and vessel movement operations (eg if the vessel leaves the departure port before the container is loaded)
- adding the globally harmonised shipper and consignee International Standard Industrial Classification (ISIC) industry codes by matching entity identifiers across the shipment and entity information data sets to make the data more suitable for comparison with widely used global data sets
- replacing the identifiers of ports not in the list of official UN/LOCODEs with the UN/LOCODE of the nearest port using a combination of port coordinates from the UN and the Azure Map application programming interface.

### 5.2.2. Harmonisation of product codes

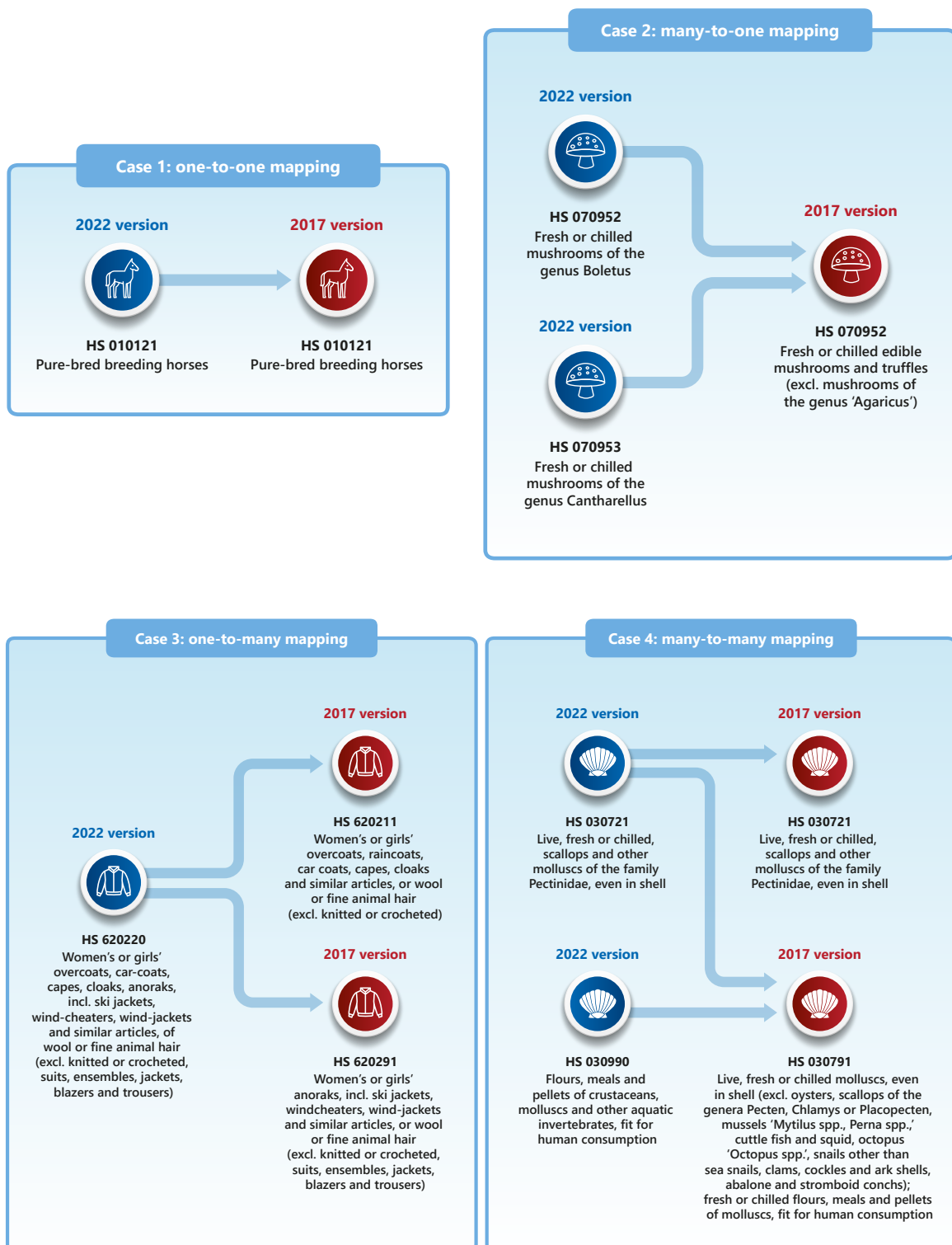
In the maritime shipment data set, each shipment includes information on the contents shipped, which the data provider maps to an HS Code.<sup>10</sup> HS Code versions are updated by the World Customs Organization (WCO) every few years to reflect changes in statistical reporting requirements (eg the arrival of smartphones as a new product), and product code changes can be singular (one-to-one) or non-singular (one-to-many, many-to-one, and many-to-many). As the shipment data cover multiple years of information, they span across both the 2017 and 2022 versions of the HS nomenclature, requiring a harmonisation process to ensure accurate comparability of products over time. Given this, all classifications were harmonised to the 2017 version, using the UN's HS-2022-to-HS-2017 correlation tables.

This conversion was straightforward for one-to-one and many-to-one mappings, where a unique HS 2017 code exists for a single or many HS 2022 products. For example, case 1 in Graph 5 shows that product 010121 of the HS 2022 version, representing pure-bred breeding horses, maps to the same, unique code in the HS 2017 version. Similarly, case 2 in Graph 5 shows that both HS 2022 products 070952 (representing fresh or chilled "Boletus" mushrooms) and 070953 (representing fresh or chilled "Cantharellus" mushrooms) map to the same 070952 code in 2017 version. In both cases, the 2022 product codes could simply be reassigned the matched unique 2017 codes. In the case of one-to-many (eg case 3 in Graph 5) and many-to-many (eg case 4 in Graph 5) mappings, however, the situation becomes more complex. Take the example of case 4 in Graph 5, where product code 030721 (classifying a subgroup of molluscs) in HS version 2022 matches to both product codes 030721 and 030791 in HS version 2017. At the same time, product 030791 in HS version 2017 is also matched to other products in the 2022 version, such as 030990, making this a many-to-many mapping case. In such cases, it is not immediately clear how to best assign the 2017 version of the HS classification.

<sup>10</sup> While this mapping process is straightforward for cases where an HS code was provided in the raw input data, in other cases mapping is determined by applying natural language processing (NLP) methods to the package free-text descriptions. As a result, HS codes in the shipment data can appear at all digit levels (two-, four- and six-digit subheadings), with more aggregated classifications resulting from the NLP mapping process.

## HS 2022 to HS 2017 product mapping example

## Graph 5



Source: BIS; United Nations; product code descriptions from World Trade Organization, "HS Tracker", [hstracker.wto.org/](http://hstracker.wto.org/).

For the more complicated one-to-many and many-to-many scenarios, the project team developed a simple and data-driven mapping methodology by computing “concordance shares” using the shipment values from before and after 2022.<sup>11</sup> Table 2 uses the case 4 example in Graph 5 to illustrate the approach taken, using abstract shipment values (in USD) for simplicity. Suppose that the shipment values are 1,000 and 2,000 for products 030721 and 030990, respectively, after 2022 under the HS 2022 version (column B), and are 400 and 800 for products 030721 and 030791, respectively, before 2022 under the HS 2017 version (column E). Product 030791 of HS 2017 matches both products 030721 and 030990 in HS 2022. Because the data do not observe how much of product 030791’s shipment value is allocated to 030721 or 030990 after 2022, the project team assumed an equal split (ie 400 of 030791’s total 800 shipment value is allocated to each product (column F)). In the case of the 030721-to-030721 match across the two HS versions, the split adjustment value is simply the original pre-2022 value of 030721 (ie 400), given that no other HS 2022 product codes match to 030721 in the HS 2017 version.

The concordance shares (column H) are then derived as the split adjustment value (column F) divided by the total adjusted shipment value for each product before 2022 (column G). These shares are then used to harmonise shipment value, weight and volume across the HS 2022 and 2017 versions. For example, the 1,000 shipment value of product 030721 after 2022 is harmonised to 500 of product 030721 and 500 of product 030791, while the 2,000 shipment value of product 030990 after 2022 is entirely allocated to product 030791 (column I). The allocated shipment values are then aggregated at the HS six-digit level, so that product 030791 takes a final harmonised value of 2,500.

Example of deriving concordance shares

Table 2

A. HS 2022 product code	B. Original value post-2022	C. HS 2017 product code	D. Mapping	E. Value pre-2022	F. Split adjustment value	G. Total adjusted value pre-2022	H. Concordance share (F / G)	I. Harmonised value post-2022 (B * H)
030721	1,000	030721	1:n	400	400	800	0.50	500
030721	1,000	030791	m:n	800	400	800	0.50	500
030990	2,000	030791	m:1	800	400	400	1.00	2,000

<sup>11</sup> This represents one of many potential solutions for HS-revision harmonisation, and the most suitable methodology will depend on the ultimate availability of use cases and data (see eg Butos et al (2026)).

### 5.2.3. Entity resolution across different data providers

Project Insight included two commercial data sets:

- Data Vendor 1 covered data on the existence and value of supply chain linkages at the shipment level between entities, established through maritime transport, along with data on the entities themselves.
- Data Vendor 2 covered data on the existence of supply chain linkages at the entity level without data on the value of the supply chain relationships, established through multiple modes of transport, based on company reporting, along with data on the entities themselves.

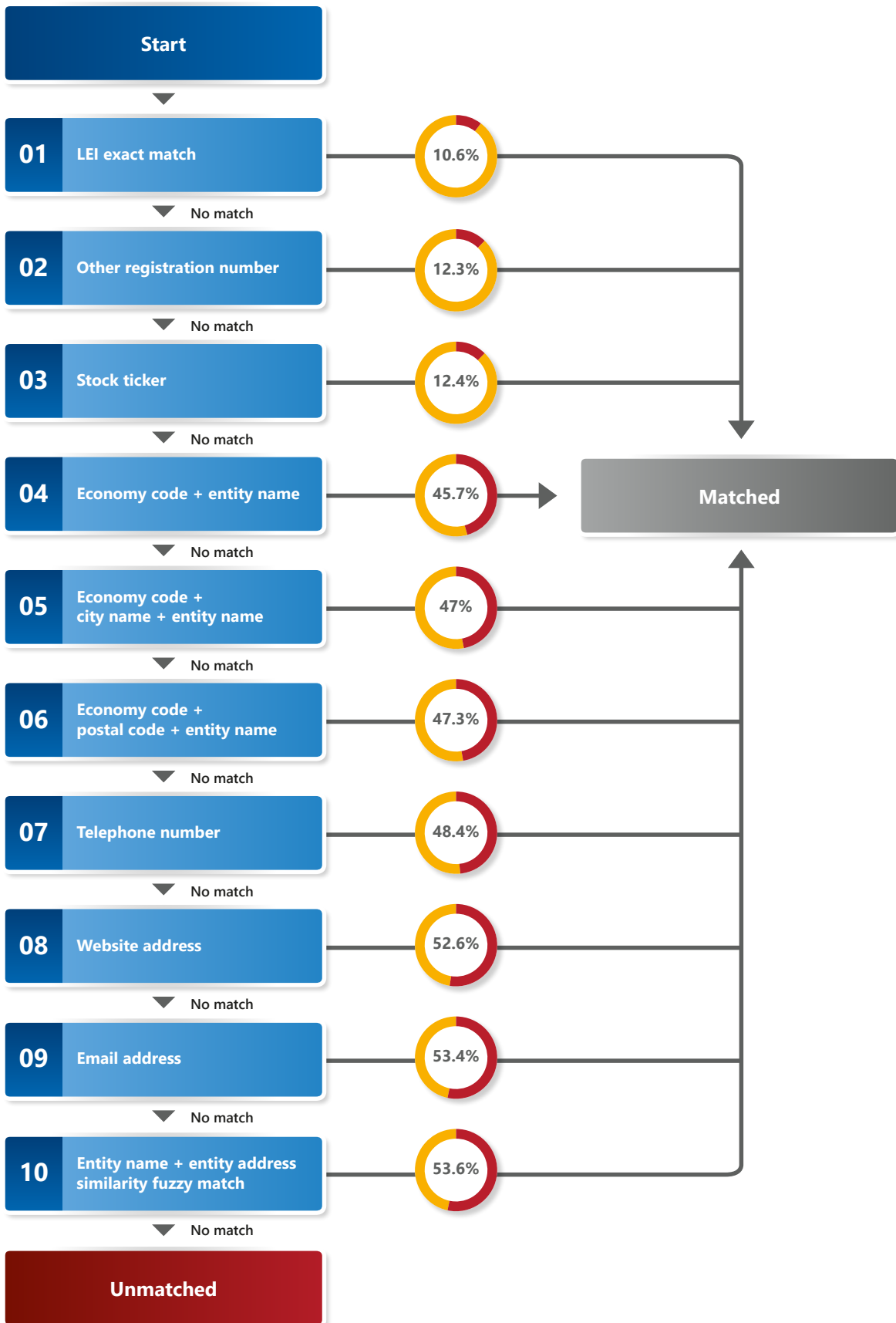
Combined, these two data sets can complement each other by capturing supply chain linkages among entities established through multiple modes of transport and shedding light on both the extensive and intensive margin of trade. This enables more complete supply chain mapping, better measurement of relationship strength and richer analysis compared with using either data set in isolation.

However, each data provider identifies entities based on its own commercial identifiers, and reconciling entities across these heterogeneous sources is not a straightforward task. In general, global standardised corporate identifiers such as the LEI have been introduced to help align entity identification across economies and data providers (see GLEIF (2026)). Adoption is still evolving internationally, and currently only a small share of entities in the Data Vendor 1 and Data Vendor 2 databases are associated with an LEI. Stock tickers are another common identification method, but they are available only for publicly traded entities.

To match entities across the two data sets, the project team adopted a multitiered procedure that leveraged different levels of entity-specific information. The ordering of different attributes used in the matching process was guided by fill-rate patterns of those attributes in the two data sets, with indicators with higher fill rates across the two data sets prioritised over those with more missing information.

Graph 6 outlines the entity resolution process. The first three steps involved looking for exact matches through common global identifiers, such as the LEI (step 1 in Graph 6), economy-specific registration numbers (step 2) and stock tickers (step 3), across the two data sources. If an entity is matched across the two data sets based on the selected criteria, it is placed in the “matched” pool, as shown by the arrows pointing to the grey “matched” rectangle on the right-hand side of Graph 6. Each matched return adds to the cumulative share of the entities in the Data Vendor 2 database being matched with entities in the Data Vendor 1 database.<sup>12</sup>

12 The Data Vendor 2 data set was used as the baseline because its sample of firms was smaller than that of Data Vendor 1.



Source: BIS.

On the other hand, if an entity is not matched, it returns “no match” and is assessed based on the next criteria, as indicated by the arrows pointing from each step to the next on the left-hand side of Graph 6. The first three steps led to a cumulative match rate of 12.4%. Then, entities were matched using combinations of the entity and economy names (step 4), city name (step 5) and postal code (step 6), raising the match rate to a cumulative 47.3%. Finally, they were matched using the telephone number (step 7), website address (step 8) and email address (step 9), bringing the match rate to 53.4%.

For the remainder of entities that lacked well-defined and consistent identifiers or attributes to match, the project team applied fuzzy or probabilistic matching using similarity scoring (step 10 in Graph 6), where entities with the same entity and economy name across the two data sets were matched based on their addresses falling above a threshold “similarity” score. This increased the match rate modestly by 0.2%, due to the stringent threshold set by the project team to avoid false positives.

This multilayered approach was designed to maximise accuracy by ensuring unique matches and by validating candidates through similarity thresholds. The combined use of deterministic and fuzzy techniques successfully matched 53.6% of the 18 million unique entities in the Data Vendor 2 data set to those in the Data Vendor 1 one, and the matched data were then used by project stakeholders for their internal analyses.



#### 5.2.4. Cleaned data structure

After conducting the aforementioned preprocessing steps, the project team applied two additional cleaning measures:

1. It consolidated all observations to the shipment-product level by combining observations with the same shipment ID, product type and timestamp (but different packages) together to focus on shipment-level analyses.
2. It removed shippers and consignees in the transportation industry to focus on trade flows between non-logistics entities.<sup>13</sup>

The data were then aggregated to the month-year level in terms of time frequency and the HS two-digit (HS2) level in terms of product (as some product codes in the shipment data were available only at the two-digit level) to prepare the data for the analytics phase.

Four supply chain network layers were constructed at different levels of aggregation to enable a multitiered view of global supply chain activity:

- **entity to entity:** This layer consisted of data aggregated to an entity-to-entity, HS2 product (where available) and month-year (where available) level (eg combining shipments of the same product shipped from entity A to entity B in the same month-year). Relationships linking entities in this network include maritime shipments, ownership connections and reported binary supply chain relationships.
- **port to port:** This layer consisted of data aggregated to a port-to-port, HS2 product and month-year level (eg combining shipments of the same product shipped from departure Port A to destination Port B in the same month-year). Relationships linking ports in this network include maritime shipments.
- **industry-economy to industry-economy:** This layer consisted of data aggregated to an industry-economy-to-industry-economy, HS2 product (where available) and month-year level (eg combining shipments of the same product shipped from firms in industry A and economy B to firms in industry C and economy D in the same month-year). Relationships linking industry-economy pairs in this network include maritime shipments and reported binary supply chain relationships.

<sup>13</sup> For logistical and operational reasons, it is common for shipper and consignee entities appearing in shipment records to be transportation intermediaries, such as freight forwarders, rather than the ultimate importer or exporter entities. In the maritime shipment data used for Project Insight, transportation intermediaries (classified as entities falling under SIC Industry codes ranging from 40 to 47) represented around 36% of total shipment value across all years. The project team filtered out such shipments to focus on trade flows between entities that were more likely to be the ultimate importer and exporters. A version of the data tables that kept these observations was retained in the silver layer for stakeholders that were interested in exploring the full data set.

- **economy to economy:** This layer consisted of data aggregated to an economy-to-economy, HS2 product (where available) and month-year level (eg combining shipments of the same product departing from economy A to destination economy B in the same month-year). Relationships linking economies in this network include maritime shipments and reported binary supply chain relationships.

These multilayered network representations of GVCs collapsed highly granular data across multiple dimensions into comprehensive, detailed information that was ready to be used by analysts and formed the core of Project Insight's analytical framework. They consisted of nodes – such as entities, vessels, industries, ports and economies – linked through relationships including maritime shipments, reported supply chain relationships, partnerships and ownership structures. All information was harmonised across units, dimensions and over time and served as the input for all gold-layer tables that analysts used for their own downstream analytics, as well as for the dashboard PoC.

The data tables underpinning these aggregated network layers were stored in the gold layer for further analytics. These additional cleaning and aggregation steps reflected strategic decisions the project team made to cater to the specific types of analyses and visualisations conducted in the next phase of the project. However, all cleaned data in their most granular forms were retained in the silver layer for users to explore and aggregate according to their own individual needs.



## 6. Analytics and visualisation

---

The analytical framework of Project Insight transformed integrated entity and trade data into a standard set of global supply chain metrics, which were showcased on an interactive dashboard PoC. Rather than focusing on specific use cases, the project aimed to build robust data structures and a core suite of standardised metrics that offer a solid and flexible foundation for future user-specific applications in policy, research and risk analysis.

Using the cleaned and preprocessed data outlined in the previous section, the project team, together with the help of project stakeholders, developed a suite of foundational GVC metrics grouped into five broad sections: (i) overview; (ii) aggregate time series trends; (iii) trade exposure and concentration; (iv) resilience and bottlenecks; and (v) network metrics. Each group of metrics, in turn, formed a separate section of the dashboard PoC, which focused on clear, simple and interactive visualisations for communicating insights within analytical teams and to senior decision-makers. To ensure maximum flexibility and to support multiple user-driven analyses at varying levels of granularity, each section of the dashboard was equipped with filters across the time, product, economy and trading-direction dimensions, to allow potential users to drill down and slice the data according to their individual needs.

The project team custom-developed the front-end technical stack underpinning the dashboard using the open source JavaScript library React (a popular tool for building user interfaces) and programming language TypeScript (which enhances the structure and robustness of the JavaScript code). This was further supported by open source libraries React Router for webpage navigation and structure, Ant Design for common controls and layouts, and Axios for application programming interface connectivity with the backend.

The dashboard build prioritised a modular design with reusable interface components and a common data-request layer to facilitate the potential addition of new metrics and visualisations without the need to redesign the entire system. Controls, charts, tables and map-based elements could be easily reused across the dashboard's multiple sections while accommodating different metrics and data sets. This approach helped to establish a coherent user experience and minimise duplicative efforts that would otherwise be needed to rebuild interface logic for each section.

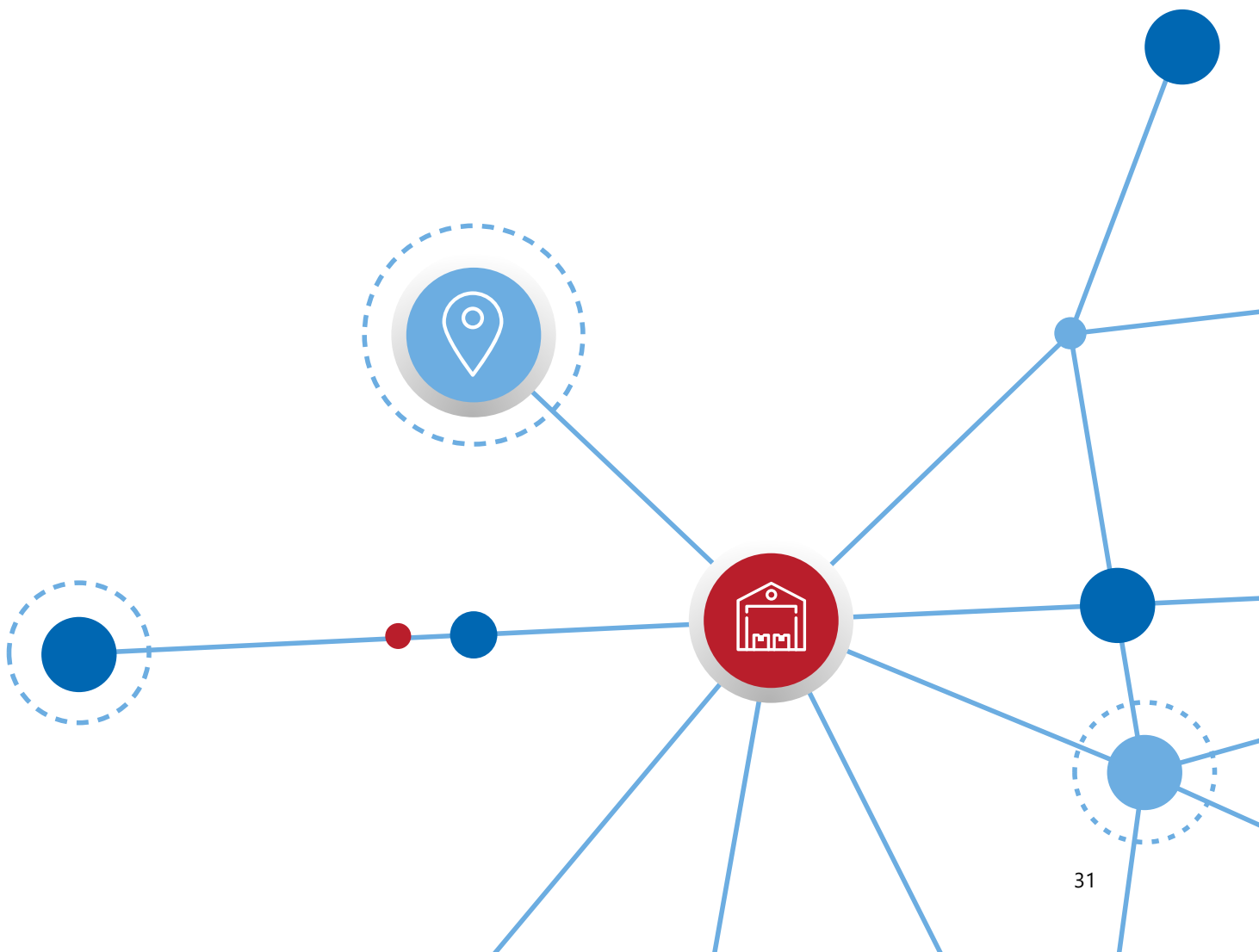
Given that the maritime shipment data can capture both the extensive and intensive margin of supply chain relationships at multiple levels of granularity, the dashboard PoC focused primarily on visualisations of this data set. While the resulting visualisations are specific to the data set used, the dashboard was designed in a modular way to demonstrate how similar tools and analyses could be adapted to other data sources in the future.

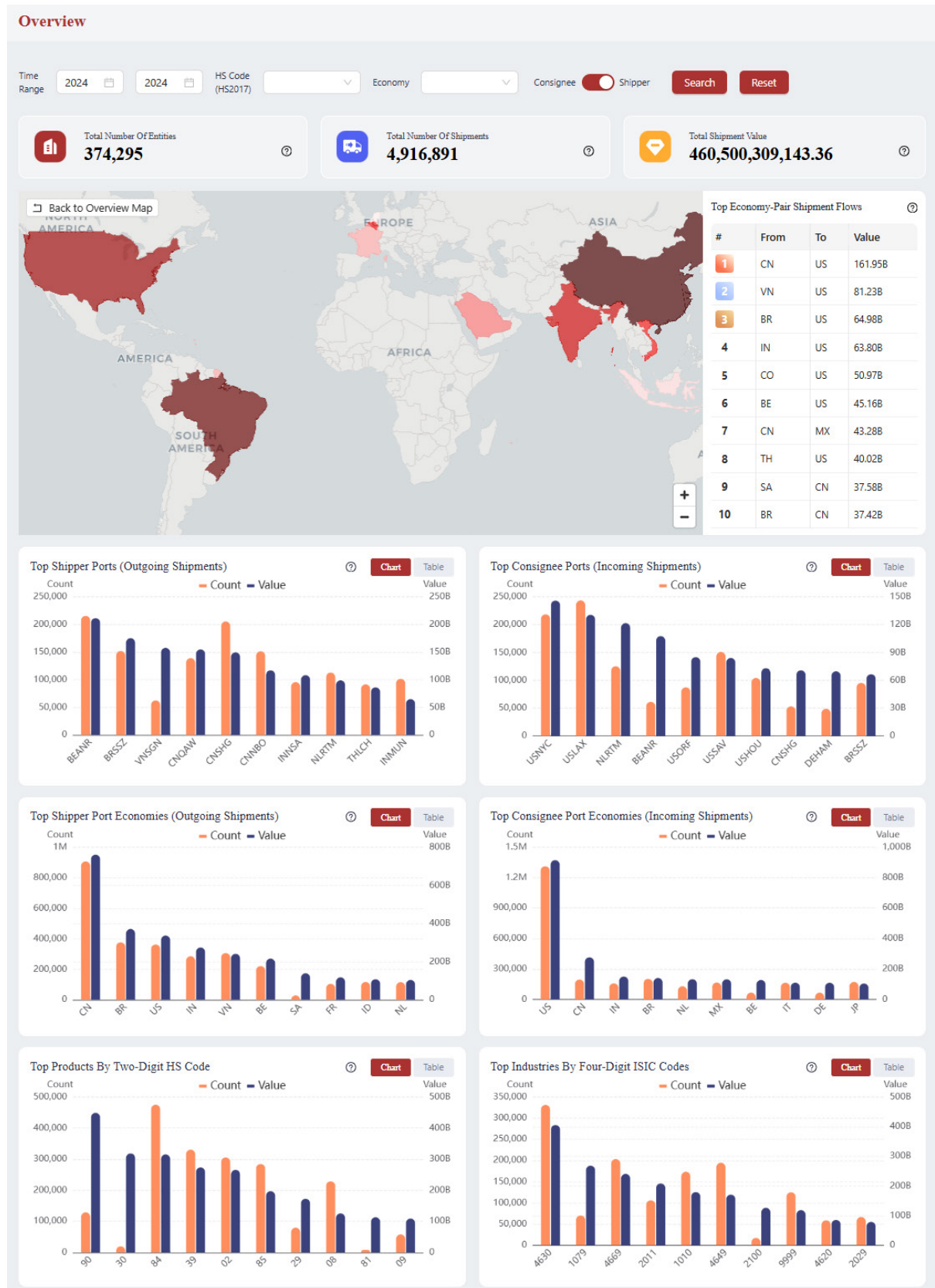
## 6.1. Overview of global supply chains

The “Overview” section provides cross-sectional snapshots of broad global maritime shipment trends. In the dashboard, for a specific year selected at the top of the page, this section includes a tally of the total number of entities and shipments involved in the entire global shipment network and the accompanying total shipment value. It also includes visual insights into the top shipper and consignee ports, economies, products and industries, ranked by both the number and value of shipments, in graph and table formats.

An interactive map enables users to zoom in and out of specific economies and see the top bilateral trading pairs. Filters along the top of the page allow users to filter all visuals by year, HS2 product code, and shipper or consignee economy. While the default shows the latest year of available data (2024) and all product types and economies, a user could select a specific product (Product A) and, for example, a shipper economy (Economy Z), to see the top ports, counterparty economies and industries involving Economy Z shipping Product A.

Each graph, table and interactive feature is equipped with an info-bubble with more detailed information and user guidance. Graph 7 provides a screenshot of the “Overview” section of the dashboard.



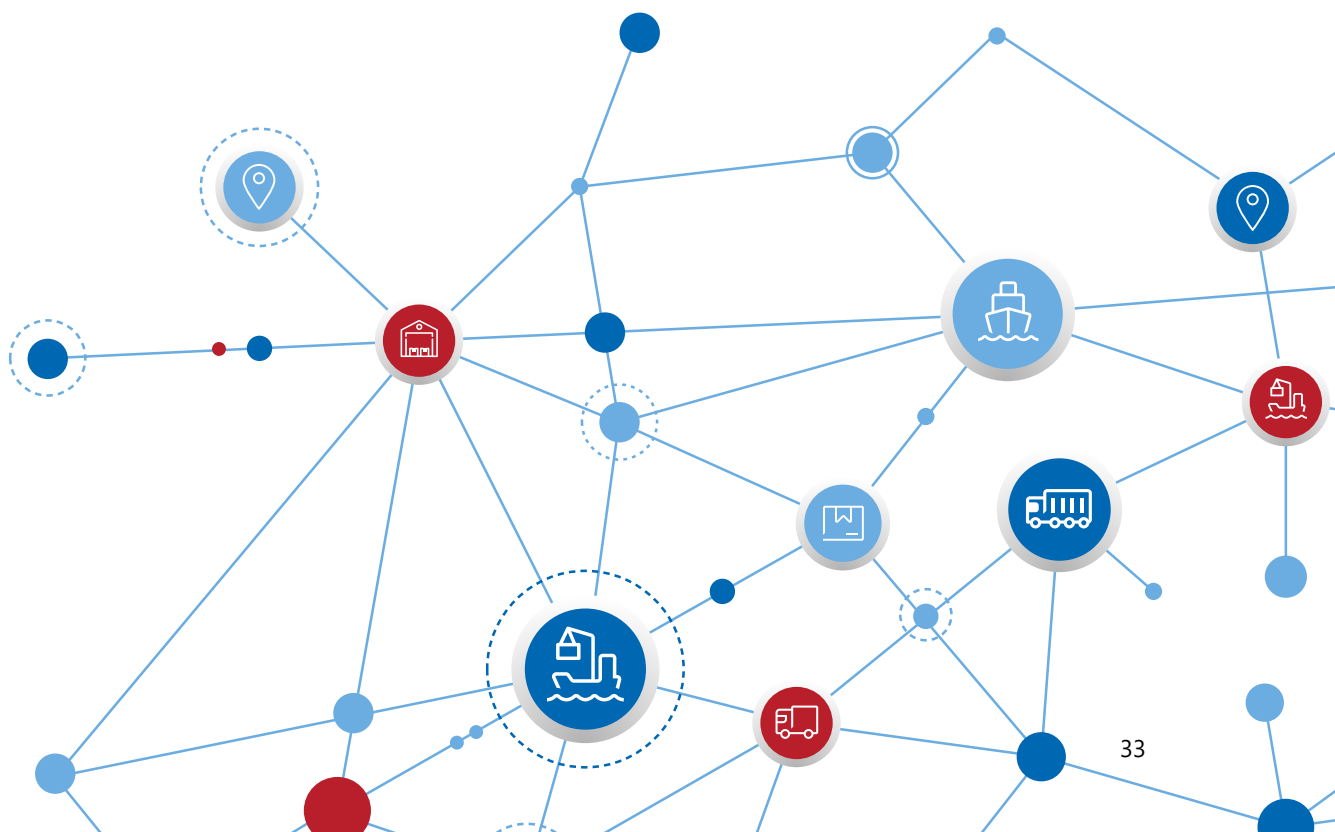


Source: BIS.

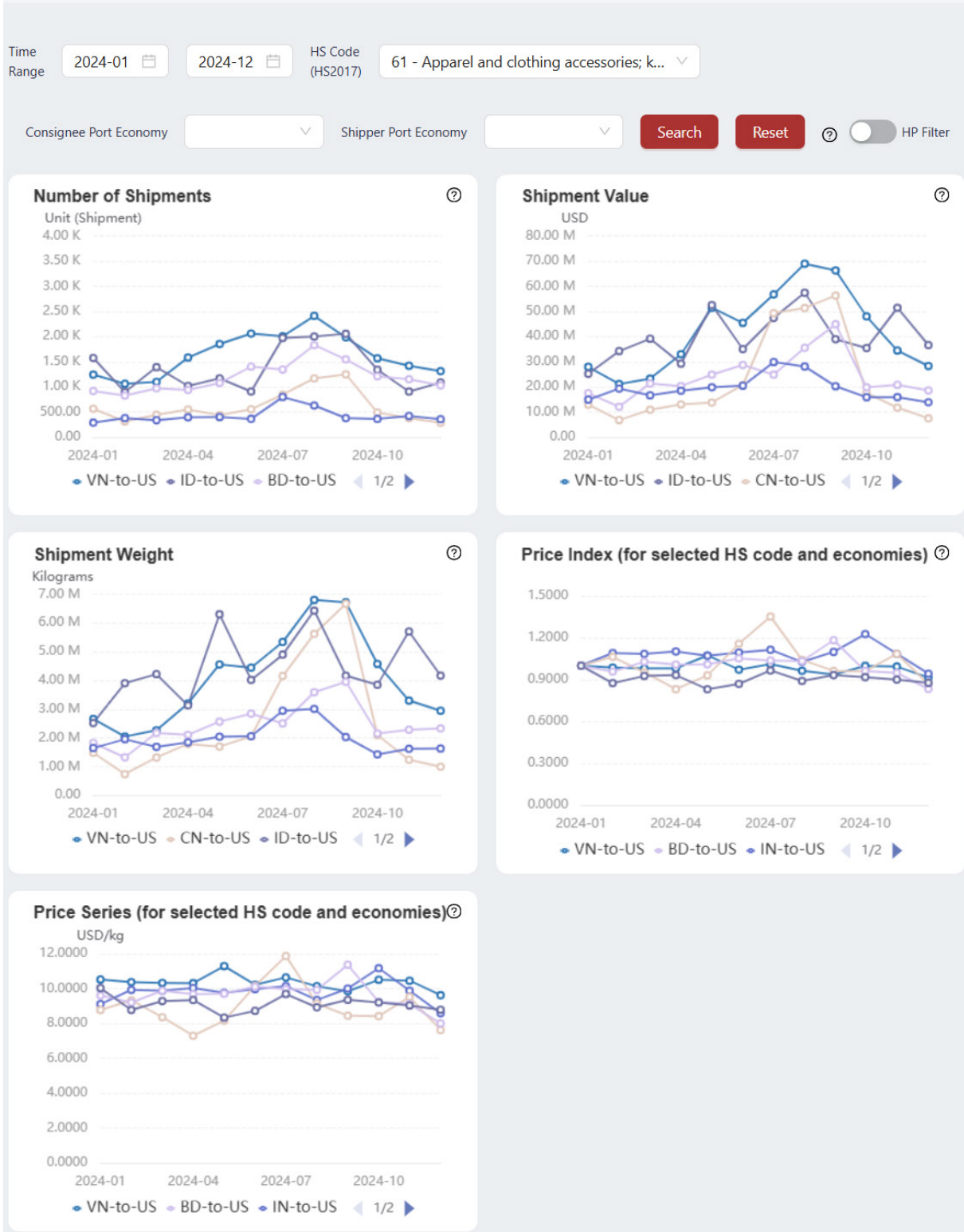
## 6.2. Aggregate time series trends

The “Aggregate time series trends” section provides a monthly time series view of maritime shipment trends at the bilateral economy level along three different dimensions that represent the intensive margin of trade: the number of shipments, shipment value and shipment weight. It also shows the evolution of the prices of products shipped over time (in USD per kilogram and index form), derived from the value and weight of shipments for products at the HS2 level. Price series and indices will show up only if a specific product code is selected in the filter. Because this section provides an overview of how these shipment indicators evolve over time, the product harmonisation conducted during the data preprocessing stage (described in section 5.2.2) was critical to ensure accurate and consistent comparability across different time periods.

By default, the top five economy pairs ranked by bilateral shipment value based on the filters selected at the top of the page will be displayed across all graphs, unless a specific shipper or consignee economy (or both) is selected by the user. The dashboard is also equipped with a Hodrick-Prescott (HP) filter that allows the user to decompose each time series into a trend and cyclical component. The HP filter can be applied to the value, weight and number of shipments for the monthly data at the departure economy or destination economy level, for economy pairs, as well as for the HS2 product dimension. However, as the selection becomes more granular (eg economy-pair and HS2 product level), there may not be enough data points to compute the HP filter, so the function is not available for all filter iterations. Graph 8 provides a screenshot of the “Aggregate time series trends” section of the dashboard.



### Aggregate time series trends



Source: BIS.

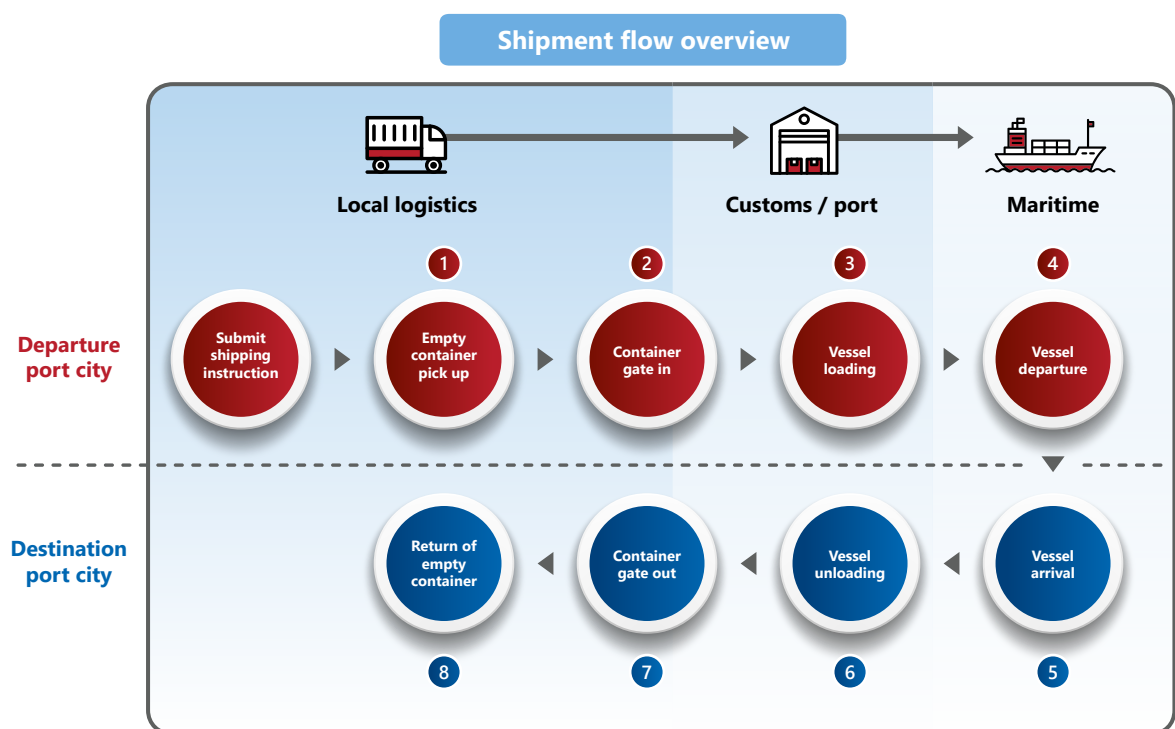
### 6.3. Resilience and bottlenecks

The “Resilience and bottlenecks” section allows users to monitor evolving congestion patterns and systematic vulnerabilities across the global maritime logistics infrastructure by assessing the robustness of supply chain maritime transportation networks and identifying bottlenecks. The project team constructed a suite of metrics for this section using timestamps (at the Universal Time Code level) of various shipment-related activities at ports extracted from the maritime shipment data.

Graph 9 provides an overview of these activities, in the following order: (i) empty container pick up; (ii) container gate in (when the filled container was gated into the departure port), (iii) vessel loading (when the container is loaded onto the vessel at the departure port); (iv) vessel departure (when the vessel departs from the departure port); (v) vessel arrival (when the vessel arrives at the destination port); (vi) vessel unloading (when the container is unloaded from the vessel at the destination port); (vii) container gate out (when the container is gated out from the destination port); (viii) return of the empty container.

Key activities in the container, vessel and port clearing process

Graph 9



Source: BIS.

These detailed timestamps were used to develop the following port-level metrics of efficiency and congestion, which can shed light on the relative resilience of different ports and identify potential bottlenecks<sup>14</sup>.

- **total processing time:** For the departure port, this was computed as the difference between the time of the vessel's departure from the departure port and the time of the container being gated in to the port. For the destination port, this was computed as the difference between the time of the container being gated out of the destination port and the time of arrival of the vessel at the destination port.
- **container clearing time:** For the departure port, this was computed as the difference between the time the container is loaded onto the vessel and the time of the container being gated into the departure port. For the destination port, this was computed as the difference between the time of the container being gated out of the destination port and the time a container is unloaded from the vessel.
- **vessel clearing time:** For the departure port, this was computed as the difference between the time of the vessel's departure from the departure port and the time the container is loaded onto the vessel. For the destination port, this was computed as the difference between the time a container is unloaded from the vessel and the time the vessel arrives in the destination port.
- **vessel congestion rate:** For both departure and destination ports, this was computed as the share of vessel clearing time to total processing time. A higher value represents more congestion (i) for vessels to leave from departure ports because a larger share of time is spent after container loading, or (ii) for vessels to arrive at destination ports because more time is spent before container unloading.

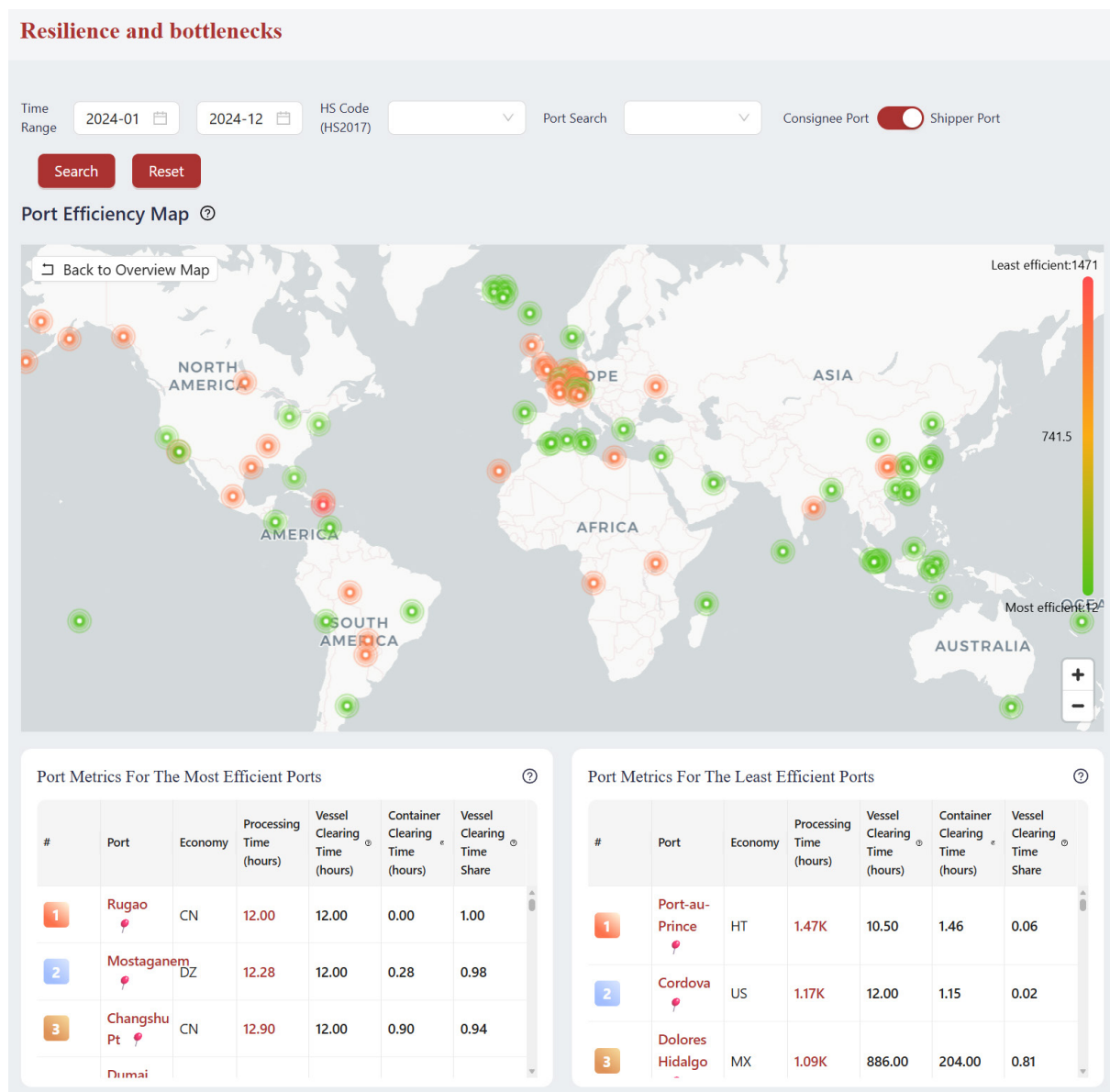
These measures were then averaged at the monthly level and showcased in the dashboard via an interactive map. The map uses different colours to highlight the relative efficiency of different ports based on the selected month-year time range, the product code and the trade direction (eg from the perspective either of outgoing shipments from the shipper and departure port, or of incoming shipments from the consignee and destination port). The different metrics are also provided in table form in the dashboard, and clicking on any individual figure (which provides a snapshot averaged based on the selected time frame) will also show a time series of that metric over the selected time range.

<sup>14</sup> A version of these port efficiency metrics was also calculated at the HS2 product code level to examine possible heterogeneity in how ports handle different types of goods.

In addition to the different efficiency metrics outlined above, correlations of the vessel congestion rates were also computed across departure and destination ports, respectively, to capture a measure of port “redundancy”. These correlations will pop up if a user clicks on any of the port names in the port metrics tables. Ports with more correlated shares of time spent for vessel clearing are more likely to encounter similar congestion issues for departure or arrival of vessels within the same time period. This can help users identify common bottlenecks and alternative landing points in the event of congestion. Graph 10 provides a screenshot of the “Resilience and bottlenecks” section of the dashboard.

Project Insight dashboard proof of concept – resilience and bottlenecks

Graph 10



Source: BIS.

## 6.4. Trade exposure and concentration

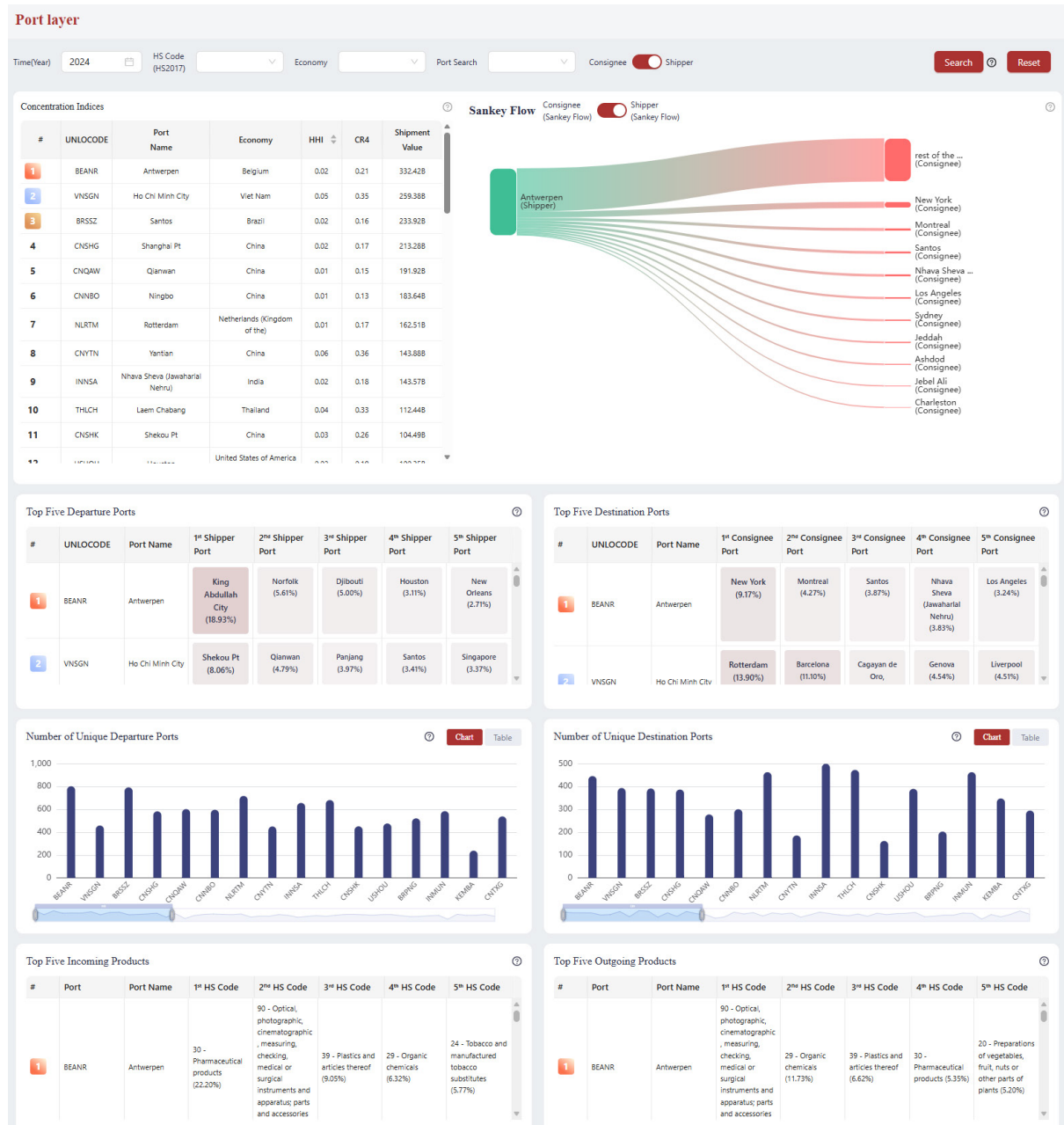
The “Trade exposure and concentration” section quantifies the extent to which different entities, ports and economies depend on only a few trading counterparties or products. This allows users to identify potential vulnerabilities arising from high concentration and greater exposure to disruptions emanating from those counterparties or products (see also IMF (2022); Berthou et al (2024)). To capture these exposures, the project team developed the following trade concentration indicators for the entity, port and economy network layers, with each set of metrics calculated by HS2 product code and by year:

- **number of unique counterparties:** From the shipper angle, this was computed as the count of unique consignees, destination ports or consignee economies. From the consignee angle, this was calculated as the count of unique shippers, departure ports or shipper economies. This can capture how diversified an entity, port or economy is, and it can differ significantly across time, product and shipment direction dimensions.
- **bilateral counterparty and product shares:** Even if a node has many unique connections, exposure may be concentrated among a few connections that make up the majority of trade. To account for this, the team calculated bilateral shares using shipment value, weight and count. From the shipper angle, this was computed as the share of total value, weight and count of shipments shipped to each consignee, destination port or consignee economy. From the consignee angle, this was computed as the share of total value, weight and count of shipments received from each shipper, departure port or shipper economy. This helps users identify the top counterparties that represent relatively higher exposure. To capture concentration specifically of products (regardless of counterparties), the team also computed the shares of HS2 product codes from the shipper and consignee angles to shed light on product specialisation.
- **Herfindahl-Hirschman Index (HHI):** Shares alone, however, do not take into account the number of connections an entity, port or economy has. For instance, a share of 85% for an entity’s top connection implies higher concentration for a firm with only two links compared with a firm with 20 links. The HHI takes this into account by measuring how concentrated shipments are across all counterparties. It is computed as the sum of the squares of the bilateral shipment shares across all of an entity’s, port’s or economy’s counterparties. This can also be calculated for product concentration (ie across an entity’s, port’s or economy’s products shipped or received).
- **four concentration ratio (CR4):** As a simplified and widely used complement to the HHI, the CR4 sheds light on the market power of the top counterparties. It was calculated by summing up the shares of the counterparties that account for the four largest ratios of shipment value. This measure can be adjusted to include more or fewer top counterparties (CR3, CR5, etc).

While these metrics all aim to capture concentration and exposure, they can reveal different and complementary patterns in the data. Each measure is included in the dashboard in a table or chart, with the default showing the counterparty concentration results for the top 50 entities, ports and economies (depending on which network layer is selected) ranked by total shipment value in the year chosen in the time filter.

The dashboard also includes a Sankey flow diagram that allows users to visualise top counterparties in either trading direction by selecting any entity, port or economy in the accompanying tables. Similar to the other dashboard sections, results can be further filtered for specific products and economies. Graph 11 provides a screenshot of the "Trade exposure and concentration" section of the dashboard.





Source: BIS.

## 6.5. Network metrics

The “Network metrics” section includes different measures of the importance or “centrality” of nodes (represented by entities, ports, industries and economies, each making up their own “granularity layer”) in the overall global supply chain network. This allows users to identify key nodes that may more easily and quickly propagate events and disruptions throughout the supply chain, including to seemingly distant nodes. The team also constructed different versions of the networks for each granularity layer for the derivation of these node centrality metrics, including directed and undirected, and weighted and unweighted versions. A directed network considers the direction of edges (eg from shipper to consignee or from departure to destination port), while an undirected network ignores the edge direction. In a weighted network, the importance of a relationship between nodes also takes into account the intensive margin of that linkage, which is captured by the shipment value, weight or count. On the other hand, an unweighted network considers only the existence of the edge, ignoring the magnitude of that relationship.

The project team computed a series of metrics that capture node network importance to analyse the characteristics that make entities, industries, ports or economies disproportionately influential, and also to identify potential vulnerability points within the global trading system. For each granularity layer, the metrics were computed yearly across all HS2 product codes. Each centrality metric and its directed/undirected and weighted/unweighted variants (where relevant) provide alternative notions of the importance of a node in the GVC network.

- First, several network-wide metrics were computed or identified, including the following:
  - **network density:** This captures how sparse an overall network is by taking the fraction of existing linkages, or “edges”, out of the total possible number of edges between nodes.
  - **largest connected component:** This represents the largest set of interlinked nodes connected - by undirected paths in the undirected case and by directed paths in the directed case.
  - **average shortest path distance** in the identified largest connected component: This is computed as the average length of the shortest paths between all nodes. In the weighted version of these metrics, the length between nodes is the minimum of the summed weights across the edges that are part of the path.

- Second, entity, port, industry-economy and economy-specific centrality metrics were computed, including the following<sup>15</sup>.
  - **degree centrality:** This represents the number of edges for each node, computed for both directed and undirected versions of the networks, so that in-degree (the number of incoming linkages) and out-degree (the number of outgoing linkages) centralities are computed separately. Weighted and unweighted versions were also computed, with the former taking into account shipment value, weight and count.
  - **eigenvector centrality:** This accounts not only for how many connections a node has but also for the importance of the nodes it is connected to. It does this by assigning a score to each node that is proportional to the sum of the centrality scores of its neighbours, governed by an eigenvalue of the network's adjacency structure. Eigenvector centrality was calculated for the largest connected component only. This measure was also calculated separately for directed/undirected and weighted/unweighted versions.
  - **betweenness centrality:** This captures how critical a node is in the connectivity of a network by considering how often it is encountered on the shortest path between other nodes. This measure was calculated by summing, over all distinct pairs of nodes in the network, the fraction of each pair's shortest paths that pass through the node in question. This measure was also calculated separately for directed/undirected and weighted/unweighted versions.
  - **closeness centrality:** This captures how close a node is on average to all other nodes, or how easy it is for other nodes to reach the node in question. This measure was calculated by taking the inverse of the total distance from a node to all other nodes in the network, so that a node becomes more central when it can reach every other node using a smaller sum of shortest path distances. This measure was calculated separately for directed/undirected and weighted/unweighted versions.

In the dashboard, overall network density and average shortest path length are displayed at the top of the section alongside the total number of nodes and edges of the network filtered based on the year, network layer, and directed/undirected and weighted/unweighted selections. Below these on the dashboard are tables displaying the four node-specific centrality measures, ranked by highest to lowest centrality (by default for the incoming direction for the directed network). Clicking on any individual node's centrality figure will bring up a time series of its evolution across all years in the data set (2018 to 2024). Graph 12 provides a screenshot of the "Network metrics" section of the dashboard.

<sup>15</sup> For more details of these centrality measures, see MIT OpenCourseWare (2026).

Project Insight dashboard proof of concept – network metrics

Graph 12

### Network metrics

Time (Year)  Shipment Network Layer  Undirected  Directed  Weighted Unweighted  Port Name Search

**Total Number Of Nodes**  
2,485

**Total Number Of Edges**  
76,504

**Network Density**  
0.011

**Average Shortest Path Length**  
2.311

#### Degree Centrality

Node	Economy	Port	In-Degree	Out-Degree
INNSA	India	Nhava Sheva (Jawaharlal Nehru)	500.000	657.000
AEJEA	United Arab Emirates (the)	Jebel Ali	492.000	476.000
THLCH	Thailand	Laem Chabang	473.000	682.000
NLRMT	Netherlands (Kingdom of the)	Rotterdam	463.000	718.000
INMUN	India	Mundra	463.000	585.000
MYPKG	Malaysia	Port Klang (Pelabuhan Klang)	447.000	419.000
BEANR	Belgium	Antwerpen	446.000	803.000
USNYC	United States of America (the)	New York	429.000	396.000
KRPUS	Korea (the Republic of)	Busan	429.000	485.000
ZADUR	South Africa	Durban	414.000	471.000
IDJKT	Indonesia	Jakarta, Java	401.000	488.000
VNSGN	Viet Nam	Ho Chi Minh City	394.000	459.000
SGSIN	Singapore	Singapore	394.000	442.000
BRSSZ	Brazil	Santos	392.000	793.000
USHOU	United States of America (the)	Houston	390.000	477.000

Total 2806 data < 1 2 3 4 5 ... 188 >

#### Eigenvector Centrality

Node	Economy	Port	In-Eigenvector	Out-Eigenvector
AEJEA	United Arab Emirates (the)	Jebel Ali	0.095	0.088
INNSA	India	Nhava Sheva (Jawaharlal Nehru)	0.093	0.093
NLRMT	Netherlands (Kingdom of the)	Rotterdam	0.092	0.098
INMUN	India	Mundra	0.092	0.090
THLCH	Thailand	Laem Chabang	0.092	0.099
BEANR	Belgium	Antwerpen	0.091	0.101
MYPKG	Malaysia	Port Klang (Pelabuhan Klang)	0.089	0.084
KRPUS	Korea (the Republic of)	Busan	0.086	0.089
USNYC	United States of America (the)	New York	0.086	0.080
ZADUR	South Africa	Durban	0.085	0.090
SGSIN	Singapore	Singapore	0.083	0.088
DEHAM	Germany	Hamburg	0.081	0.090
IDJKT	Indonesia	Jakarta, Java	0.081	0.088
USHOU	United States of America (the)	Houston	0.080	0.084
BRSSZ	Brazil	Santos	0.080	0.099

Total 2806 data < 1 2 3 4 5 ... 188 >

#### Betweenness Centrality

Node	Economy	Port	Value
BEANR	Belgium	Antwerpen	0.045
NLRMT	Netherlands (Kingdom of the)	Rotterdam	0.043
BRSSZ	Brazil	Santos	0.042
INNSA	India	Nhava Sheva (Jawaharlal Nehru)	0.041
THLCH	Thailand	Laem Chabang	0.039
INMUN	India	Mundra	0.036
AEJEA	United Arab Emirates (the)	Jebel Ali	0.026
CNSHG	China	Shanghai Pt	0.025
CNNBO	China	Ningbo	0.019
IDJKT	Indonesia	Jakarta, Java	0.019
USSAV	United States of America (the)	Savannah	0.019
MYPKG	Malaysia	Port Klang (Pelabuhan Klang)	0.019
DEHAM	Germany	Hamburg	0.018
CLSAI	Chile	San Antonio	0.017
CNQAW	China	Qianwan	0.017

Total 2806 data < 1 2 3 4 5 ... 188 >

#### Closeness Centrality

Node	Economy	Port	In-Closeness	Out-Closeness
AEJEA	United Arab Emirates (the)	Jebel Ali	0.617	0.615
INNSA	India	Nhava Sheva (Jawaharlal Nehru)	0.614	0.652
NLRMT	Netherlands (Kingdom of the)	Rotterdam	0.609	0.673
THLCH	Thailand	Laem Chabang	0.608	0.662
INMUN	India	Mundra	0.606	0.635
BEANR	Belgium	Antwerpen	0.603	0.689
MYPKG	Malaysia	Port Klang (Pelabuhan Klang)	0.602	0.597
USNYC	United States of America (the)	New York	0.596	0.588
KRPUS	Korea (the Republic of)	Busan	0.594	0.612
ZADUR	South Africa	Durban	0.589	0.612
DEHAM	Germany	Hamburg	0.584	0.624
IDJKT	Indonesia	Jakarta, Java	0.582	0.616
SGSIN	Singapore	Singapore	0.581	0.603
BRSSZ	Brazil	Santos	0.581	0.681
VNSGN	Viet Nam	Ho Chi Minh City	0.580	0.608

Total 2806 data < 1 2 3 4 5 ... 188 >

Source: BIS.

## 7. Conclusion

---

Production has long been globalised, and granular supply chain relationships play an important role in propagating shocks and driving aggregate dynamics. Yet our understanding of the micro-level networks that define GVCs remains limited. For central banks and policymakers, this lack of visibility underscores the need for data and tools. These tools can shed light on the complexities of production networks, inflationary pressures, how disruptions can propagate, where hidden vulnerabilities may reside, and structural shifts that may bring about both risks and opportunities.

Project Insight aimed to address this challenge by developing a PoC that transformed a historical snapshot of highly detailed supply chain information into actionable insights. These included indicators to flag potential supply chain bottlenecks, measures of exposure to concentrated suppliers or markets, and network-based metrics to trace how disruptions could propagate across firms, sectors and economies. The project was grounded in user needs by continuously incorporating feedback and guidance from a global stakeholder group of supply chain experts spanning central banks, international organisations, academic institutions and the private sector. This process started with creating a structured inventory of the challenges project stakeholders face in monitoring supply chains along with their data and analytical needs. This was followed by carefully selecting a set of commercial and public data sets – together making up billions of observations – that can provide complementary and detailed insights into GVC developments.

These data sets were then preprocessed on a flexible and accessible data platform to ensure consistency, comparability and integrity of the subsequent data analytics, which were conducted collaboratively with project stakeholders in the project's cloud-hosted environment. The project then developed a unified analytical framework that converted the cleaned data sets into a foundational suite of GVC measures organised into several broad sections: aggregate cross-sectional and time series overviews of supply chain shipments, resilience and bottlenecks, trade exposure and concentration, and network metrics. These indicators were displayed through an interactive PoC dashboard for easy monitoring and decision-making. Instead of targeting specific use cases, the emphasis was on creating robust data structures and a standardised set of core metrics that could be adapted by users to suit different analytical, policy or research needs.

By unifying commercial firm-level data sets, public sources and advanced analytics within a collaborative platform, the project demonstrated how granular and complex GVC dependencies can be studied more effectively. The resulting data architecture, standardised metrics and dashboard visualisation developed through the PoC demonstrate how such tools can help policymakers to detect emerging vulnerabilities and understand shifting global patterns and can support more informed decision-making. Furthermore, reusable components of the data architecture and preprocessing lay the groundwork for future central bank initiatives that leverage firm-level and supply chain data. This can lower the technical barriers to incorporating these granular data sets into policy analysis.

## References

---

- Acemoglu, D, V Carvalho, A Ozdaglar and A Tahbaz-Salehi (2012): "The network origins of aggregate fluctuations", *Econometrica*, vol 80, no 5.
- Alfaro, L, P Conconi, F Kamal and Z Kroff (2025): "Trade within multinational boundaries", *NBER Working Papers*, no 33887.
- Antràs, P and D Chor (2022): "Global value chains", *Handbook of international economics*, vol 5.
- Arslanalp, S, S M Choi, P Kamali, R Koepke, M McKetty, M Ruta, M Saraiva, A Sozzi and J Verschuur (2025): "Nowcasting global trade from space", *IMF Working Papers*, no 2025/93.
- Asian Development Bank (ADB) (2015): "Key Indicators for Asia and the Pacific 2015", 46th edition.
- Auer, R, Borio, C and A Filardo (2017): "The globalisation of inflation: the growing importance of global value chains", *BIS Working Papers*, no 602.
- Auer, R, A Levchenko and P Sauré (2019): "International inflation spillovers through input linkages", *Review of Economics and Statistics*, vol 101, no 3.
- Avdjiev, S, M Everett, P Lane and H S Shin (2018): "Tracking the international footprints of global firms", *BIS Quarterly Review*, March.
- Bacilieri, A, A Borsos, P Astudlo-Estevez, M Hoefler and F Lafond (2025): "Firm-level production networks: what do we (really) know?", *INET Oxford Working Paper*, no 2025-14.
- Bai, X, J Fernández-Villaverde, Y Li and F Zanetti (forthcoming): "The causal effects of global supply chain disruptions on macroeconomic outcomes: theory and evidence", *American Economic Review*.
- Bank for International Settlements (BIS) Innovation Hub (2023): *Project Atlas: mapping the world of decentralised finance*, October.
- Baqae, D and E Farhi (2019): "The macroeconomic impact of microeconomic shocks: beyond Hulten's theorem", *Econometrica*, vol 87, no 4.
- Berthou, A, A Haramboure and L Samek (2024): "Mapping product-level supply chain vulnerabilities", *OECD Science, Technology and Industry Working Papers*, no 2024/02.
- Bonadio, B, Z Huo, A Levchenko and N Pandalai-Nayar (2021): "Global supply chains in the pandemic", *NBER Working Papers*, no 27224.

Butos, S, E Jackson, D Torun, B Leonard, N Tuzcu, P Lukaszuk, A White, R Hausmann and M Yildirim (2026): "Tackling discrepancies in trade data: the Harvard Growth Lab International Trade Datasets", *Scientific Data*, vol 13, no 170.

Chen, S, E Tsang and L S Y Zhang (2023): "Global supply chain interdependence and shock amplification – evidence from Covid lockdowns", *BIS Working Papers*, no 1123.

Databricks (2026): "What is medallion architecture?", 15 March, [www.databricks.com/glossary/medallion-architecture](https://www.databricks.com/glossary/medallion-architecture).

DHL (2025): "What is the meaning of bill of lading (BoL) in international shipping?", Discover, 18 August, [www.dhl.com/discover/en-sg/logistics-advice/import-export-advice/what-is-bill-of-lading](https://www.dhl.com/discover/en-sg/logistics-advice/import-export-advice/what-is-bill-of-lading).

Diem, C, Pichler, A, Farmer, J and S Thurner (2022): "Quantifying firm-level economic systemic risk from supply network data", *Scientific Reports*, vol 12, article number 7719.

Global Legal Entity Identifier Foundation (GLEIF) (2026): "Legal Entity Identifier (LEI)", 15 March, *Organizational Identity*, [www.gleif.org/en/organizational-identity/introducing-the-legal-entity-identifier-lei](https://www.gleif.org/en/organizational-identity/introducing-the-legal-entity-identifier-lei).

Guilhoto, J, C Webb and N Yamano (2022): "Guide to OECD TiVA Indicators, 2021 edition", *OECD Science, Technology and Industry Working Papers*, no 2022/02.

International Maritime Organization (IMO) (2026): "IMO identification number schemes", 15 March, [www.imo.org/en/ourwork/msas/pages/imo-identification-number-scheme.aspx](https://www.imo.org/en/ourwork/msas/pages/imo-identification-number-scheme.aspx).

International Monetary Fund (IMF) (2022): "Chapter 4: global trade and value chains during the pandemic", *World Economic Outlook*, April.

——— (2024): "Chapter 2: the great tightening: insights from the recent inflation episode", *World Economic Outlook*, October.

——— (2026): *PortWatch*, 15 March, <https://portwatch.imf.org/>.

Magerman, G, E Dhyne, K De Bruyne and J Van Hove (2016): "Heterogeneous firms and the micro origins of aggregate fluctuations", *National Bank of Belgium Working Papers*, no 312.

Magerman, G, A Palazzolo, E Dhyne, A Borsos, D Kulikov, A Linarello, A Paulus, G Romanini and M Saldias (2026): "Granularity in firm-to-firm production networks: a multi-country analysis", working paper.

Mann, C (2012): *"1493: Uncovering the New World Columbus Created"*, Vintage.

MIT OpenCourseWare (2026): "Lecture 4: centrality measures", Introduction to network models, 15 March, [https://ocw.mit.edu/courses/1-022-introduction-to-network-models-fall-2018/resources/mit1\\_022f18\\_lec4/](https://ocw.mit.edu/courses/1-022-introduction-to-network-models-fall-2018/resources/mit1_022f18_lec4/).

Pichler, A, C Diem, J Farmer and S Thurner (2023): "Building an alliance to map global supply networks", *Science*, vol 379, no 6639.

Qiu, H, H S Shin and L Zhang (2023): "Mapping the realignment of global value chains", *BIS Bulletin*, no 78.

——— (2025): "International footprint of global firms: residence and nationality perspectives on maritime shipments", *BIS Bulletin*, no 118.

Timmer, M, E Dietzenbacher, B Los, R Stehrer and G de Vries (2015): "An illustrated user guide to the World Input–Output Database", *Review of International Economics*, vol 23, no 3.

United Nations (UN) (2021): *Handbook on Supply and Use Tables and Input–Output Tables with Extensions and Applications*, Department of Economic and Social Affairs.

——— (2026): "UN comtrade database", 15 March, [comtradeplus.un.org/](https://comtradeplus.un.org/).

United Nations Economic Commission for Europe (2025): *UN/LOCODE*, 15 April, [unece.org/trade/uncefact/unlocode](https://unece.org/trade/uncefact/unlocode).

United Nations Trade and Development (UNCTAD) (2025): "Shipping data: UNCTAD releases new seaborne trade statistics", 23 April, <https://unctad.org/news/shipping-data-unctad-releases-new-seaborne-trade-statistics>.

US Securities and Exchange Commission (2026): "Standard Industrial Classification (SIC) code list", 15 March, [www.sec.gov/search-filings/standard-industrial-classification-sic-code-list](https://www.sec.gov/search-filings/standard-industrial-classification-sic-code-list).

World Customs Organization (WCO) (2026): "What is the harmonized system (HS)?", 15 March, [www.wcoomd.org/en/topics/nomenclature/overview/what-is-the-harmonized-system.aspx](https://www.wcoomd.org/en/topics/nomenclature/overview/what-is-the-harmonized-system.aspx).

World Trade Organization (2026): "HS Tracker: about", 15 March, [hstracker.wto.org/](https://hstracker.wto.org/).

Zhang, S Y (2021): "Using equity market reactions and network analysis to infer global supply chain interdependencies in the context of COVID-19", *Journal of Economics and Business*, vol 115, 105974.

## Contributors and acknowledgements

---

### Bank for International Settlements (BIS)

Bénédicte Nolens (Centre Head, BIS Innovation Hub Hong Kong Centre)

Leanne Si Ying Zhang (Project Lead, Report Author and Adviser, BIS Innovation Hub Hong Kong Centre)

Kevin Tsui (Project Technical Lead and Adviser, BIS Innovation Hub Hong Kong Centre)

Rudolf Biczok (Project Technical Consultant and Azure Cloud Engineer, BIS Innovation Hub)

Han Qiu (Project Analytics Consultant and Economist, BIS Monetary and Economic Department)

Weeris Treeratanajaru (Project Technical Consultant and Adviser, BIS Innovation Hub Hong Kong Centre)

### Hong Kong Monetary Authority (HKMA)

Michael Cheng (Head, Economic Research Division)

Max Kwong (Senior Economist, Economic Research Division)

Chi-sang Tam (Senior Economist, Economic Research Division)

Ken Wong (Economist, Economic Research Division)

Yannson Wang (Economist, Economic Research Division)

Freda Liu (Assistant Economist, Economic Research Division)

Mia Xiao (Assistant Economist, Economic Research Division)

Yifan Zhang (Adviser, Economic Research Division)

### The Organisation for Economic Co-operation and Development (OECD)

Sébastien Miroudot (Senior Trade Policy Analyst, OECD Trade and Agriculture Directorate)

Mattia Cai (Trade Policy Analyst, OECD Trade and Agriculture Directorate)

Charles Cadestin (Smart Data Platform Officer, OECD Trade and Agriculture Directorate)

Ricardo Chiapin Pechansky (Junior Trade Policy Analyst, OECD Trade and Agriculture Directorate)

Xue Han (Trade Modeller, OECD Trade and Agriculture Directorate)

Carmen Zürcher (Statistician, OECD Trade and Agriculture Directorate)

**DIW Berlin**

Sonali Chowdhry (Research Associate, Firms and Markets Department)

**Project observing members and other collaborators (in alphabetical order)**

Asian Development Bank

Bank of Thailand

Global Legal Entity Identifier Foundation

Harvard University, The Center for International Development – Growth Lab

Hong Kong University of Science and Technology, Li & Fung Supply Chain Institute

International Monetary Fund

Peking University, National School of Development

Society for Worldwide Interbank Financial Telecommunication

Swiss National Bank

United Nations

University of Oxford

World Trade Organization

**Acknowledgements**

The BIS authors are grateful to Tommaso Mancini-Griffoli, Miguel Diaz, Karmela Holtgreve, Raphael Auer, Andras Valko, Jon Frost and Haibin Zhu, as well as the HKMA, OECD and DIW Berlin project partners, for reviewing the report and their insightful feedback. We are also grateful to Glenn Magerman, Federico Gallina, Vivi Bong, Anna Nguyen, Peter Sha and Brian Li for their invaluable analytical and technical contributions.



Bank for International Settlements (BIS)

ISBN 978-92-9259-965-2