Location Quotients and Regional Industry Clusters

A Guide to Global Data Sources

May 2026

Prepared for internal use.

Contents

1 Introduction 2

2 Employment-Based Data Sources 2
- 2.1 ILO — ILOSTAT . . . 2
- 2.2 World Bank — World Development Indicators (WDI) . . . 3
- 2.3 OECD — Structural Analysis (STAN) Database . . . 3
- 2.4 Eurostat — Regional Statistics (NUTS Framework) . . . 3

3 Value-Added and GDP-Based Sources 4
- 3.1 OECD — National Accounts and Regional GDP . . . 4
- 3.2 World Bank — GDP by Sector and National Accounts . . . 4
- 3.3 UN National Accounts Statistics . . . 4

4 Trade and Supply Chain Data 4
- 4.1 OECD — Trade in Value Added (TiVA) and ICIO . . . 4
- 4.2 UN Comtrade and WITS . . . 5
- 4.3 WTO Global Trade Data Portal . . . 5

5 Innovation and Knowledge Economy Data 5
- 5.1 WIPO — Global Patent Data . . . 5
- 5.2 OECD — R&D Expenditure (MSTI and ANBERD) . . . 5
- 5.3 Scientific Publications — Scopus, Web of Science, PubMed . . . 6

6 Labour Flow and Firm-Level Data 6
- 6.1 LinkedIn Economic Graph and Similar Platforms . . . 6
- 6.2 National Business Registers and Census Microdata . . . 6

7 Summary: Data Sources at a Glance 6

8 Persistent Limitations of the LQ Framework 7

9 Conclusion 8

1 Introduction

A very common, though limited and misunderstood, means of identifying regional industry clusters is the **Location Quotient** (**LQ**). The LQ is a ratio of employment shares: regional industry i's share of total regional employment over national (or global) industry i's share of total employment in the reference economy. Formally:

$$\mathrm{LQ}_{i,r} = \frac{E_{i,r} \;/\; E_{\cdot,r}}{E_{i,n} \;/\; E_{\cdot,n}}$$

where $E_{i,r}$ is employment in industry i in region r, $E_{\cdot,r}$ is total regional employment, and $E_{i,n}$, $E_{\cdot,n}$ are the corresponding national (or global reference) totals.

- An LQ of **1.0** indicates that the regional economy has the same share of employment in industry i as the reference area as a whole.
- Values **above 1.0** signal relative specialisation — the region is relatively more concentrated in that industry than the reference.
- Values **below 1.0** indicate relative under-representation.

Despite its widespread use in regional economic analysis, economic development practice, and policy planning, the LQ is both analytically limited and heavily dependent on the quality and scope of the underlying data. Extending the LQ framework — or supplementing it with richer alternatives — requires access to a broader range of data sources across geographies, variables, and institutional scales.

This report catalogues the major global and multi-country data sources available for LQ calculation and regional cluster analysis, organised by thematic category, and assesses the strengths and limitations of each.

2 Employment-Based Data Sources

Employment data remain the most widely used input for LQ calculation because they are consistently defined, broadly available, and decomposable by industry at sub-national geographies. The following sources represent the principal options at global and regional scales.

2.1 ILO — ILOSTAT

Institution: International Labour Organization (ILO)

Access: `ilostat.ilo.org` — free and open

ILOSTAT is the world's leading repository of international labour statistics. It provides employment data disaggregated by sector and occupation, covering over 200 countries and territories. The ILO combines nationally reported estimates with statistically modelled values for countries where survey data are incomplete or unavailable, producing consistent regional and global aggregates. Industry classifications follow the International Standard Industrial Classification (**ISIC**), which facilitates cross-country comparability.

Limitations. Modelled estimates carry uncertainty and should not be used for fine-grained country-level comparisons. Sub-national (regional or city-level) breakdowns are generally not available through ILOSTAT; analysts must rely on national statistical offices for sub-country data. Coverage of the informal economy is inconsistent across country survey methodologies.

2.2 World Bank — World Development Indicators (WDI)

Institution: World Bank

Access: `data.worldbank.org` — free and open

The World Development Indicators database republishes ILO modelled employment estimates alongside over 1,500 economic, social, and environmental indicators for more than 200 countries. For cluster analysis, employment by industry as a percentage of total employment provides the raw material for national-level LQ benchmarking across countries. The WDI's API allows bulk download and integration with statistical software.

Limitations. National-level aggregates only; industry classification is broad (agriculture, industry, services) rather than fine-grained NAICS or ISIC 4-digit sectors. Researchers requiring sector-level depth must combine WDI with ILOSTAT or national statistical offices.

2.3 OECD — Structural Analysis (STAN) Database

Institution: Organisation for Economic Co-operation and Development

Access: `stats.oecd.org` — free for registered users

The OECD STAN database provides internationally comparable data on industrial output, value added, employment, labour compensation, investment, and R&D for OECD member countries. Sectoral breakdowns are available at a level of detail comparable to two- and three-digit ISIC, enabling meaningful LQ calculations across industries. The companion **OECD Regional Statistics** database extends some of these measures to sub-national NUTS-equivalent regions.

Limitations. Coverage is limited to OECD member and selected partner economies (approximately 50 countries). Temporal lags of one to three years are common. Sub-national regional data are less complete than national series.

2.4 Eurostat — Regional Statistics (NUTS Framework)

Institution: Statistical Office of the European Union

Access: `ec.europa.eu/eurostat` — free and open

Eurostat publishes the most geographically granular multi-country employment dataset available for any major world region. Data are structured according to the **Nomenclature of Territorial Units for Statistics (NUTS)**, a hierarchical system ranging from NUTS 1 (major socio-economic regions) through NUTS 3 (small regions for specific diagnoses). Employment and economic output are provided at all three levels, disaggregated by NACE Rev. 2 (the European equivalent of ISIC), making Eurostat the gold standard for sub-national LQ analysis within the European Union.

Limitations. Coverage is confined to EU member states and select partner countries. Non-EU analysts must use separate national or international sources. NACE classification does not map perfectly onto NAICS, complicating US–EU comparisons.

3 Value-Added and GDP-Based Sources

Employment-based LQs treat every job as equivalent in productive terms, which can be misleading in industries with wide variation in labour productivity. Value-added or GDP-based LQs correct for this by measuring each industry's contribution to regional output rather than its headcount.

3.1 OECD — National Accounts and Regional GDP

The OECD publishes detailed national accounts, including value added by industry at ISIC disaggregation. The OECD Regional Outlook and Regional Well-Being database extend GDP data to sub-national regions for member countries. Value-added LQs computed from OECD data offer a productivity-adjusted view of regional specialisation that employment counts alone cannot provide.

3.2 World Bank — GDP by Sector and National Accounts

The World Bank's national accounts data, available through the WDI and the **World Integrated Trade Solution (WITS)** platform, include sectoral value-added shares of GDP for over 200 countries. While less granular than OECD STAN, these data enable cross-country benchmarking in a global LQ framework — for example, comparing a country's manufacturing value-added share against a world or regional average.

3.3 UN National Accounts Statistics

The UN Statistics Division publishes national accounts data through the UN Data platform and the **National Accounts Main Aggregates Database**, covering most UN member states. Industry breakdowns follow ISIC and include value added, gross output, and intermediate consumption, enabling more sophisticated LQ variants and input-output extensions in countries not covered by the OECD.

4 Trade and Supply Chain Data

Trade specialisation is an important complement to domestic employment and output measures for identifying clusters with an export orientation. **Revealed Comparative Advantage (RCA)**, itself mathematically analogous to the LQ, is the standard trade-based specialisation index:

$$\mathrm{RCA}_{i,c} = \frac{X_{i,c} \,/\, X_{\cdot,c}}{X_{i,w} \,/\, X_{\cdot,w}}$$

where $X_{i,c}$ is country c's exports of product i and the denominators are the corresponding world totals.

4.1 OECD — Trade in Value Added (TiVA) and ICIO

The OECD **Trade in Value Added (TiVA)** database and the **Inter-Country Input-Output (ICIO)** tables decompose export flows by the domestic value-added content, revealing the true industrial footprint behind trade statistics. For cluster analysts, TiVA exposes supply chain dependencies that a simple employment or output LQ cannot detect — for example, identifying that a region's automotive cluster extends its effective specialisation far upstream into steel and electronics.

4.2 UN Comtrade and WITS

UN Comtrade is the largest global repository of international merchandise trade data, covering over 200 countries at the HS 6-digit product level. The World Bank's WITS platform provides convenient access and analytical tools. Analysts can construct product-level RCA indices as a trade-based analogue to employment LQs, identifying countries or regions whose export baskets are concentrated in specific product clusters.

4.3 WTO Global Trade Data Portal

The WTO's Global Trade Data Portal aggregates indicators from multiple sources — including the WEF, ITC, OECD, and IMF — providing trade competitiveness metrics alongside employment and output data. It includes data on foreign affiliate activity (AMNE database), capturing the multinational dimension of clusters that purely domestic employment statistics miss.

5 Innovation and Knowledge Economy Data

Employment and output LQs capture the current productive structure of a region but are lagging indicators of cluster dynamism. Innovation-focused data sources provide forward-looking signals of cluster strength and technological specialisation.

5.1 WIPO — Global Patent Data

Institution: World Intellectual Property Organization

Access: `patentscope.wipo.int` — free and open

WIPO's **PATENTSCOPE** database contains over 100 million patent documents from national and regional patent offices worldwide. Patent counts geocoded to inventor or applicant location enable construction of patent-based LQs — measuring whether a region's share of patents in technology class t exceeds its share of global patenting. This approach is particularly valuable for identifying emerging innovation clusters in sectors such as biotechnology, semiconductors, and clean energy, which may not yet register in employment or GDP data.

Limitations. Patent propensity varies significantly across industries, firm sizes, and countries. Regions with strong academic research cultures but weak commercialisation may appear as innovation clusters by patent count alone. Name disambiguation — linking patents to unique firms or regions — remains technically challenging.

5.2 OECD — R&D Expenditure (MSTI and ANBERD)

The OECD's **Main Science and Technology Indicators (MSTI)** and **Analytical Business Enterprise R&D (ANBERD)** databases provide R&D expenditure broken down by sector of performance and source of funds for OECD countries. R&D-intensity LQs — comparing a region's sector-specific R&D spend to the national or international benchmark — capture the knowledge investment dimension of cluster analysis and are increasingly used alongside employment LQs in cluster competitiveness assessments.

5.3 Scientific Publications — Scopus, Web of Science, PubMed

For sectors where publication counts are more meaningful than patents — notably biomedical research, basic sciences, and social sciences — geocoded bibliometric databases such as **Scopus** and **Web of Science** allow construction of publication-based LQs. **PubMed**, freely accessible, provides approximately 23 million bibliographic records in the biomedical domain and has been used to identify biomedical innovation clusters without relying on predetermined administrative boundaries.

6 Labour Flow and Firm-Level Data

One of the most significant innovations in cluster analysis in recent years is the use of **labour flow data** — records of workers changing jobs between firms and industries — to identify cluster boundaries from observed economic behaviour rather than from administrative classifications.

6.1 LinkedIn Economic Graph and Similar Platforms

Research using LinkedIn's global employment history data — covering hundreds of millions of individuals and millions of firms — has produced labour flow networks that reveal the hierarchical structure of geo-industrial clusters worldwide. Because workers tend to transition between geographically proximate firms with similar skill requirements, the resulting labour flow networks expose cluster membership and boundaries that employment LQs, anchored to predefined industry codes, cannot reveal. This approach represents the frontier of global cluster identification but requires access to proprietary platform data.

6.2 National Business Registers and Census Microdata

For researchers with access to administrative data, national business registers — such as the UK's Inter-Departmental Business Register, Germany's Establishment History Panel, or the US Longitudinal Employer-Household Dynamics (**LEHD**) database — provide firm- and establishment-level employment with geocoding. These sources enable LQ computation at very fine spatial and sectoral scales (below county level, at 4- to 6-digit industry codes) while also supporting input-output and supply chain analysis through firm-to-firm linkage data.

7 Summary: Data Sources at a Glance

Table 1 summarises the principal data sources discussed in this report, their institutional homes, geographic reach, key variables, and country coverage.

Table 1: Key global data sources for LQ and cluster analysis

Source	Institution	Geography	Key Variables	Coverage
ILOSTAT	ILO	National / regional aggregates	Employment by sector (ISIC)	200+ countries

Table 1 continued…

Source	Institution	Geography	Key Variables	Coverage
World Development Indicators	World Bank	National	Employment by broad sector; GDP	200+ countries
STAN / Regional Statistics	OECD	National + sub-national	Employment, value added, R&D	~50 OECD + partners
Eurostat NUTS	Eurostat	NUTS 1–3 (sub-national)	Employment, GDP, value added	EU + EEA countries
TiVA / ICIO	OECD	National	Trade in value added, supply chains	~65 countries
UN Comtrade / WITS	UN / World Bank	National	Export flows by product (HS)	200+ countries
PATENTSCOPE	WIPO	Geocoded to applicant	Patent counts by technology class	Global
MSTI / ANBERD	OECD	National	R&D expenditure by sector	~45 OECD countries
Scopus / PubMed	Elsevier / NLM	Geocoded to institution	Scientific publications	Global
LinkedIn Economic Graph	LinkedIn / Microsoft	Geocoded to firm	Labour flows, skills, transitions	Global (proprietary)

8 Persistent Limitations of the LQ Framework

Regardless of the data source employed, the LQ carries structural limitations that richer data cannot fully resolve:

1. **Static, single-variable measure.** A high LQ reveals concentration but says nothing about whether firms in the cluster are connected to one another, share a labour pool, or benefit from knowledge spillovers — the defining features of a genuine Marshallian cluster.

2. **Classification-boundary dependence.** Industry classification boundaries (ISIC, NAICS, NACE) are administratively defined, not economically determined. Cluster boundaries routinely cross multiple industry codes, and aggregation choices can create or dissolve apparent specialisations.

3. **Supply-side versus demand-side ambiguity.** The LQ conflates supply-side specialisation with demand-side dependence. A high LQ may reflect a genuine export-oriented cluster or simply a region with a large captive domestic market.

4. **Informal economy exclusion.** LQs based on formal employment miss the informal economy, which is substantial in low- and middle-income countries and which ILOSTAT modelled estimates only partially capture.

5. **Alternative metrics.** The Herfindahl-Hirschman Index (HHI), Ellison-Glaeser concentration indices, Revealed Comparative Advantage (RCA), and input-output network analysis each

address specific LQ limitations but require additional data and methodological investment.

9 Conclusion

The LQ remains a useful first-pass diagnostic for identifying candidate clusters, but its analytical power is heavily constrained by the data source on which it is computed. Employment-only LQs using ILOSTAT or WDI data are appropriate for broad national comparisons. Eurostat NUTS data support the most rigorous sub-national cluster mapping within Europe. Value-added sources (OECD STAN, UN National Accounts) provide a productivity-corrected view of specialisation. Trade-based RCA indices from UN Comtrade expose export-oriented cluster strength. Innovation data from WIPO PATENTSCOPE and OECD MSTI capture technological dynamism that employment counts miss entirely.

The most credible contemporary cluster analyses combine at least two of these dimensions — for example, employment LQs alongside patent LQs and wage data — and supplement quantitative indices with input-output linkage analysis and, where possible, qualitative fieldwork. No single data source, and no single index, can substitute for that multi-layered approach.

Prepared for internal use — May 2026