

Innovation work in the Data Science Campus: Data Science and AI

Osama Rahman

Director, Data Science Campus, ONS

20 January 2025



Creating the ONS of the Future

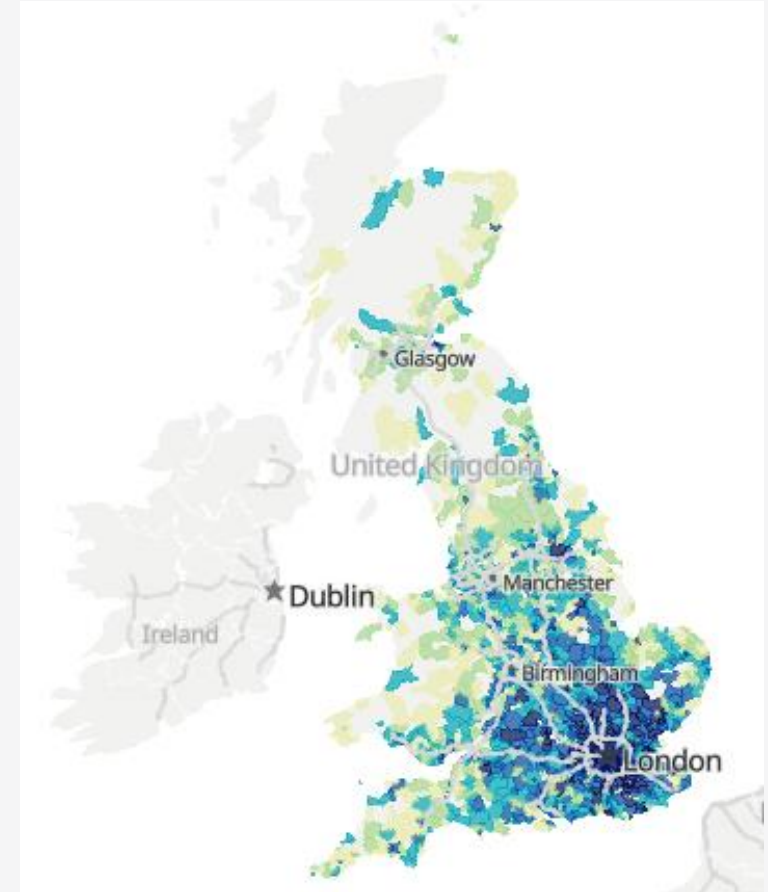
The Three Pillars

Recent applications

Financial Transactions data

- ONS and **Visa** are in a partnership to develop anonymised and aggregated card spending big data into **new economic data** and to support **official statistics production**
- Innovative data published so far include [spending flows across the UK](#), [tourism trends](#) in small areas, estimations of [local consumption zones](#)
- Ongoing projects to explore direct and indirect integration of card data into the statistical production of Household Final Consumption Expenditure (HHFCE), Tourism, Digital Trade, Regional HHFCE, Regional CPI
- Also in partnership with **Pay.UK / Vocalink** to develop interbank payment systems data into **real-time business supply chain** information, similar to input-output tables
- Jan 2025 publication: [monthly payment flows between SIC5 industries](#), capturing over 3 million UK businesses and organisations

*Merchants located in **Covent Garden**, London: where their customers came from, Q3-2023*



Source: Visa and ONS

Publication: [Consumer card spending, flow of spending across the UK: 2019 to 2023](#)

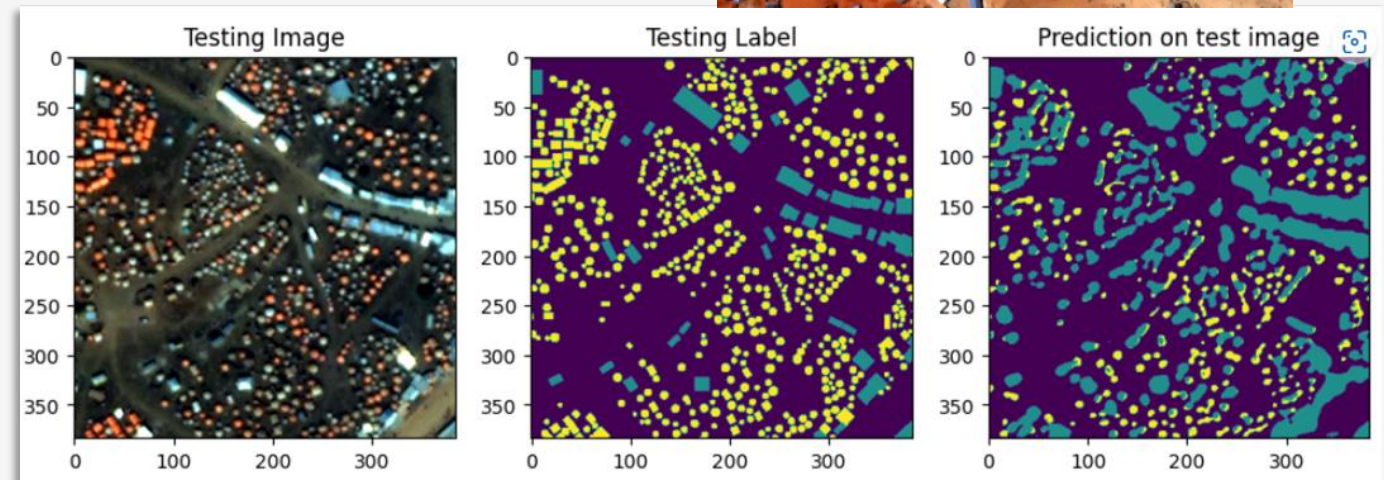
Somalia census support / IDP camp mapping

- Supporting United Nations Population Fund (UNFPA) Somali National Bureau of Statistics (SNBS) and Foreign, Commonwealth & Development Office (FCDO)
- Establishment of enumeration areas to better account for Internally Displaced People (IDP) camps
- ML tool using satellite imagery to predict camp extents



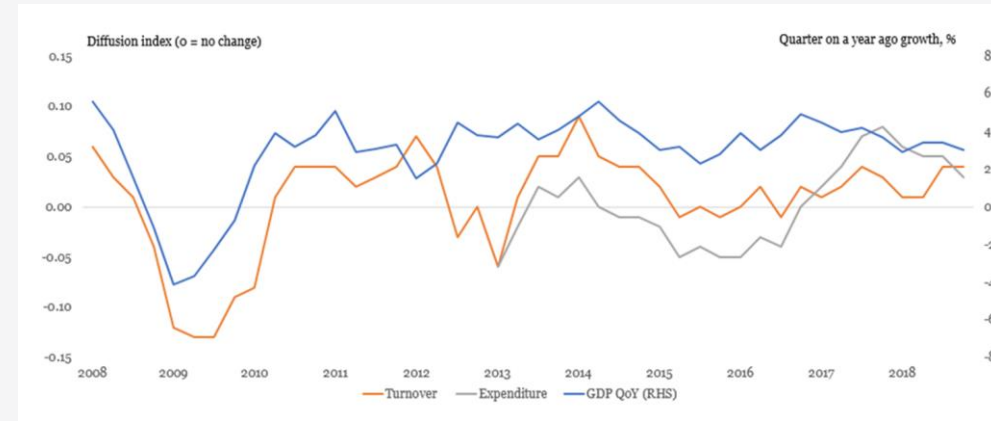
- **Outputs / impacts**

- Tool used to deliver camp footprints to UNFPA
- Development of end user application for application to wider use cases
- Support representation of IDP in censuses



Value Added Tax (VAT) Data

- VAT data can help improve quality and temporal granularity of GDP estimate and give real time insights on economy.
- VAT returns give insights into the turnover and expenditure of over 250,000 UK firms.
- Data is available within a month of reporting - making it useful as an early indicator of economic activity.
- Existing work provides *diffusion indices* – measuring the proportion of firms reporting an increase in activity compared to the preceding period.
- These generally track well with GDP indices, and can be used as an early indicator of economy and to improve GDP estimates.
- Current work is looking to extend to numerical indices – explicitly using turnover returns.



$$VAT\ DI_t = \frac{N_t^{up} - N_t^{down}}{N_t}$$

Data processing: classification

Can the latest AI tools improve coding of labour market survey text data to SIC/SOC?

- Developed a proof-of-concept system, using an enterprise grade LLM.
- Applied this to anonymised survey response data
- Evaluated against clerical coding and a benchmark model.

[Blog](#) and [Code](#) on our website

LLM assisted SIC Coding

What does the organisation mainly make or do?

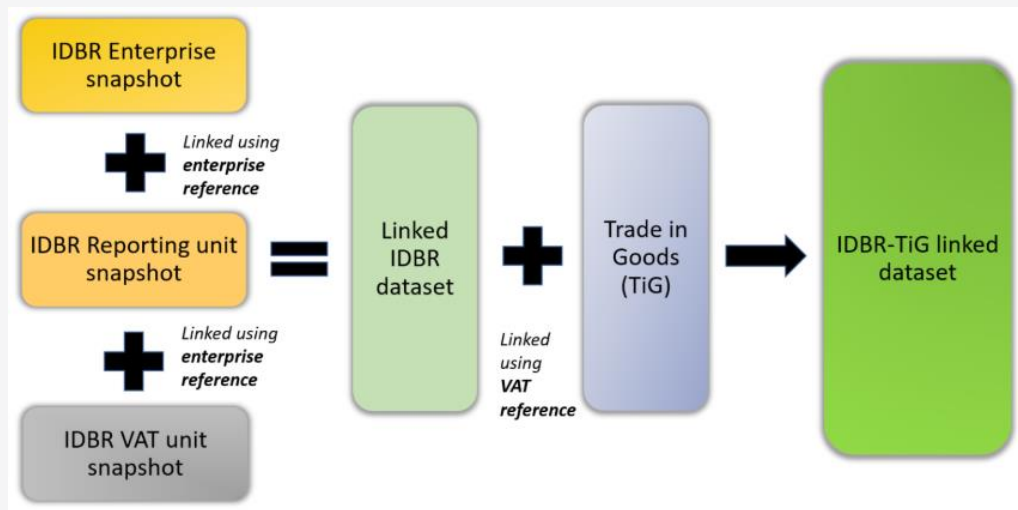
newspaper and website

```
"sic_code" : "5813"
"sic_descriptive" : "Publishing of newspapers"
  "sic_candidates" : [
    0 : {
      "sic_code" : "5813"
      "sic_descriptive" : "Publishing of newspapers"
      "likelihood" : 0.7
    }
  ]
"reasoning" :
"The company's main activity is described as newspaper and website. This suggests that the company is involved in publishing, specifically newspaper publishing (SIC code 5813) and possibly also operating web portals (SIC code 6312)."
```

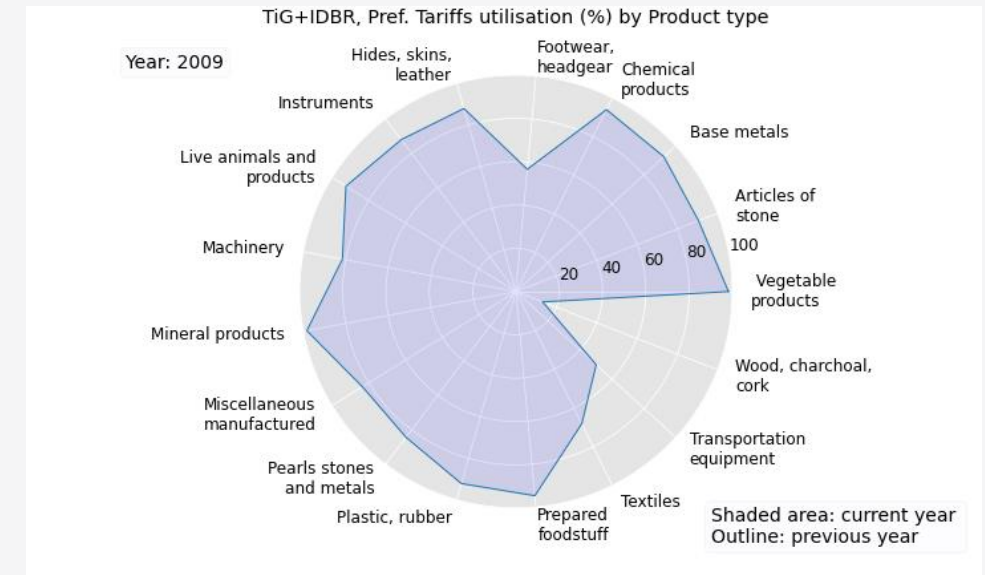

Preferential Tariff Utilisation Rates

Understanding the rate at which UK businesses make use of preferential tariff arrangements when trading with particular countries around the world

The linking of our InterDepartmental Business Register (structures and characteristics of UK businesses) with Trade in Goods data (what imports/exports businesses undertake in the UK)



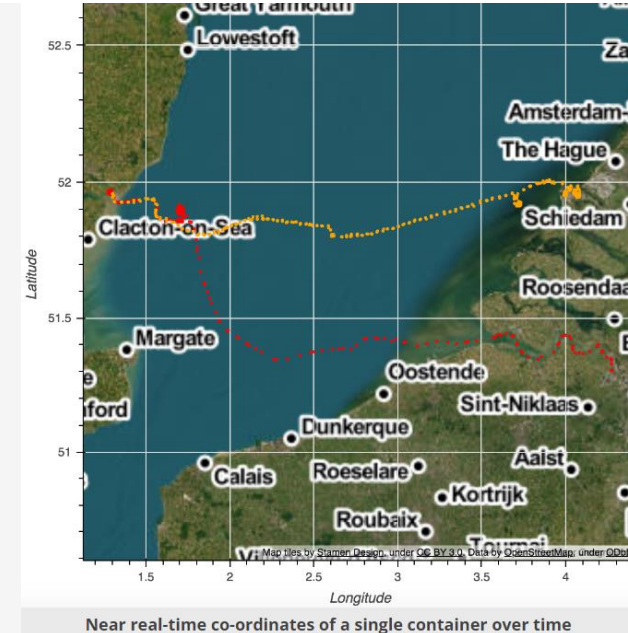
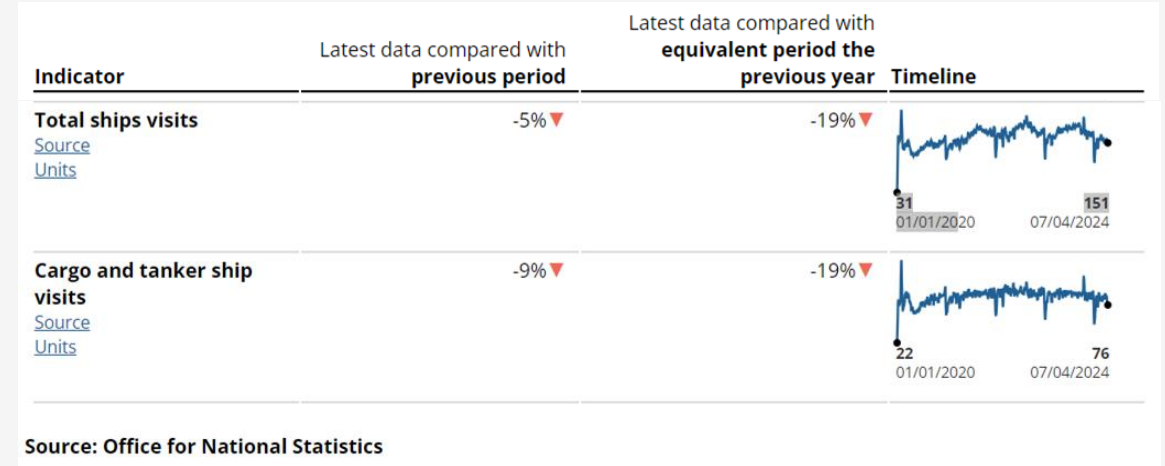
Assessment of which categories of goods are imported using available preferential tariffs



This analysis allows UK Government to better understand which types of businesses across the economy are not benefitting from existing trade deals, and could be additionally supported to improve economic performance

Real Time Shipping Indicators

- AIS provides near real-time information regarding the position of ships, enabling precise monitoring of their movements
- This data has enabled experimental statistics that measure the volume of ship traffic through global maritime passages.
- Data for UK ports enables early indicators for trade statistics – several weeks before central estimates are available.
- Current work is leveraging this further by linking AIS data to shipping instructions – allowing us to track the flow of specific goods.



UK Trade Shipping

Shipping
Instructions
(cargo container contents)



Automated
Identification
System data
(ship transponders and GPS
locations)



See also: UK PortWatch AIS

UK imports, by volume and routes between origin/destination ports



We are helping UK Government (DBT, DfT, CO) understand trade dependencies on critical passages throughout the world ocean.

Not only which critical passages cargo ships navigate through but also their cargo contents, making it possible to gain insights into specific categories of physical goods according to their importance to the UK economy

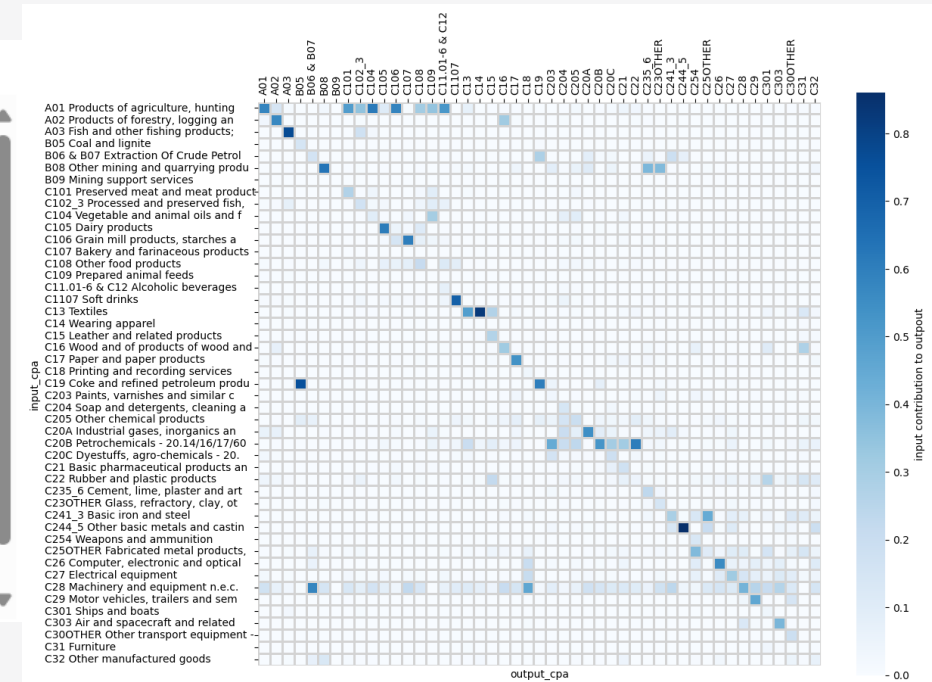
Global Supply Chains with LLMs

Trialling the use of Large Language Models to generate supply chains data across all categories of physical goods, to help identify gaps in existing data sources

LLM suggested values of inputs required to create
Electronic integrated circuits

An input-output table for whole physical economy, based purely on
LLM's intrinsic "knowledge" of each product category

output:	8542 (6) Electronic integrated circuits		
inputs	hscode	description	value
	0 381800	Chemical elements; doped for use in electronics, in the form of discs, wafers or similar forms; chemical compounds doped for use in electronics	30
	1 848640	Machines and apparatus of a kind used solely or principally for the manufacture or repair of masks and reticles, assembling semiconductor devices or electronic integrated circuits, or for lifting, handling, loading or unloading items of heading 8486	15
	2 382499	Chemical products, mixtures and preparations; n.e.c. heading 3824	15
	3 281129	Inorganic oxygen compounds; of non-metals, n.e.c. in item no. 2811.2	10
	4 711299	Waste and scrap of precious metals; waste and scrap of precious metals including metal clad with precious metals, other than that of gold and platinum and excluding ash which contains precious metal or precious metal compounds	10
	5 848620	Machines and apparatus of a kind used solely or principally for the manufacture of semiconductor devices or of	20



By helping comprehensively identify products used across the entire supply chain, we can compare with existing data sources used for both official statistics purposes and for global supply chain intelligence in the Department for Business and Trade more efficiently

Thank you

Any questions?

20 January 2025

