Innovation Procurement Empowerment Centre

A Connected Places Catapult Initiative

Using natural language processing to map defence industry supply chains

Andrew D. James, Grigory Pishchulov

Research brief no. 22
Publication Date: September 2026

Executive summary

The last five decades have been characterised by the globalisation of the production of manufactured goods and the emergence of complex global value chains such that complex, technologically dependent products are almost always the result of global value chains (Gerrefi et al., 2005). Moreover, rising geopolitical tensions with China and the growing attention paid to "decoupling", "friend shoring" and "China plus one" strategies have led to greater attention to the mapping of supply chains to identify areas of potential vulnerability (McCarthy et al., 2022).

In the defence industry, detailed knowledge of the supply chain to weapons systems is desirable for several reasons. First, concerns about the potential existence of "single points of failure" – single suppliers, potentially in lower tiers of a supply chain, who are critical to a weapons system and whose performance may have implications for the production of a weapons system. Second, concerns about the existence of suppliers from potentially hostile third countries who may generate import dependence. Third, further concerns about the purchase of supplier companies by foreign entities through mergers and acquisitions. Such issues have become increasingly acute as NATO defence industries have sought to ramp-up production in response to the war in Ukraine, revealing supply chain constraints and vulnerabilities. A practical methodology for mapping defence industry supply chains could support defence procurement, including procurement of innovation, in mitigating such constraints and vulnerabilities.

This research brief reports a pilot project to use Natural Language Processing to map the defence industry supply chain and in particular the supply chains of major weapons systems that are playing an important role in the NATO support to the Ukraine war. The methodology used in the study aims at extracting information about supply-chain relationships from a selection of news articles. The outcome proves the ability of the methodology to produce a supply chain map even with a very limited training and utilizing a small sample of articles.

1. Introduction

Supply chain mapping is the process of documenting the organisations engaged in a company's supply chain. Whether in the civilian or defence industries, the mapping of supply chains presents a number of practical problems. First, the size and complexity of many supply chains presents major challenges in trying to capture data on suppliers. Gholz and James (2023), using data from the US DoD Sector by Sector, Tier by Tier study of the US defence industry, show how the supply chain to major weapons systems can involve thousands of suppliers and in some cases seven tiers of the supply chain.

Second, commercial confidentiality is an important constraint on the sharing of information about suppliers – a company's supply chain is an important source of its competitive advantage. Prime contractors may be cautious about revealing their suppliers to their government customer although the customer has some leverage and political considerations may mean that prime contractors may wish to publicise the geographical breadth of their suppliers. Suppliers to prime contractors and in low tier buyer-supplier relationships are even less likely to reveal their suppliers.

Third, the nature of supply chain relationships and contracts combined with commercial confidentiality means that prime contractors have limited visibility of sub-tier suppliers and their suppliers may be reluctant to share data with the prime contractor due to the risk of disclosing valuable competitive information (i.e. a prime contractor will know its suppliers but rarely knows the suppliers to those suppliers).

Fourth, the resources and time required for primary data collection are considerable and the task is made more difficult because commercial confidentiality can make access by analysts challenging. Finally, whilst curated databases of buyer-supplier relationships exist (e.g. Bloomberg) they are very expensive to purchase.

Accordingly, the visibility of defence industry supply chains to Ministries of Defence, defence procurement organisations and other government agencies can be limited. Attempts at comprehensive mapping of supply chains like the DoD's Sector-by-Sector, Tier-by-Tier survey in the US are immensely expensive, resource intensive and obtaining extensive data from suppliers was only possible because of the particular legal powers vested in the US Department of Commerce to oblige disclosure.

Further, any database is only as good as the frequency with which information is updated and the cost of repeating the exercise was deemed prohibitive. The cost of participation by companies was significant as was the effort required by DoD to administer the survey and the initiative was eventually scrapped by the Trump administration. In the UK, the MOD's Defence Equipment & Support organisation announced in 2023 an initiative called SCRIPT to use software to map the Defence supply chain to identify risks and areas of fragility in defence procurement using data from a range of external sources, suppliers and internal information (DE&S, 2023).

2. Methodology

Our aim is to identify relationships between buyers and suppliers in complex, multi-tier supply chains. Drawing on growing capabilities in Artificial Intelligence and data science, we use Natural Language Processing (NLP). NLP is a rapidly developing sub-field of Artificial Intelligence (AI) that specialises in the extraction and manipulation of natural language text or speech based on statistical relationships learnt from text documents during a computationally intensive self-supervised and semi-supervised training process.

There have been some attempts to use Natural Language Processing to map supply chains. Wichmann et al. (2020) used NLP to extract supply chain maps from natural language text in articles published by the news agency Reuters. They noted issues of partial recall (the model only identified some supply relationships) and selection bias (the results were influenced by the nature of the Reuters database). Kaloskampis et al. (2023) analysed global supply chains using Wichmann et al.'s

data base plus text from the BBC and found that they could detect buyer-supplier relationships with good accuracy.

Following Kaloskampis et al. (2023), we employ an NLP pipeline utilising spaCy — an open-source library for advanced NLP. We employ a pretrained language model as part of the pipeline to produce contextual word embeddings. This pipeline has been implemented by means of a relationship extraction template offered by Explosion, the maker of spaCy. We employed the pipeline to analyse textual data from the Factiva database (Dow Jones, 2025). Factiva is a comprehensive news resource providing international/regional news, trade and industry publications, investment analysis, stock exchange feeds & international news feeds dating back 10 years.

We focus on generating supply chain maps for a number of leading defence contractors (Lockheed Martin, BAE Systems and Rheinmetall) and a number of major weapons systems, commonly referred to as programmes, that are playing an important role in the NATO support to the Ukraine war (HIMARS rocket launcher system, M777 howitzer and Rheinmetall's 155 mm shell).

Figure 1 illustrates key steps in our NLP approach which are explained in the following in greater detail.

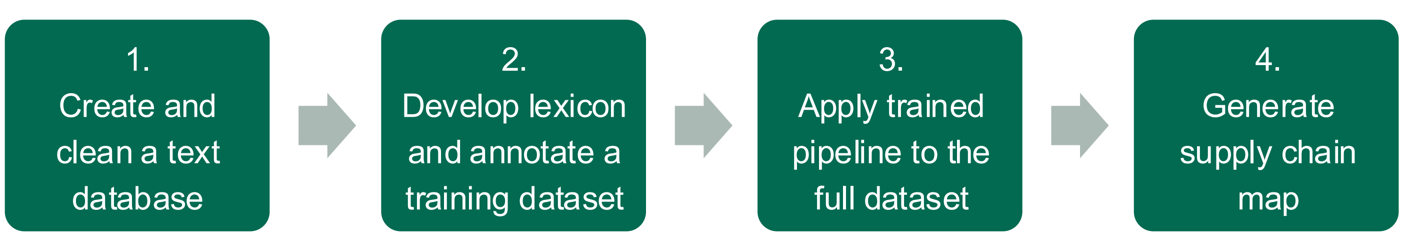

Figure 1. Key steps in the NLP approach used in the study.

Step 1: Create and clean a text database

For the pilot run of the study, an initial sample of articles has been retrieved from the Factiva database by conducting advanced searches for business news articles using a search string that includes the name of a defence contractor and further terms such as supplier, vendor, contract, order, procure, and related ones.

Search results underwent a basic cleaning in terms of removal of line breaks, adding missing spacing, and correcting misspellings caused by character encoding.

Step 2: Develop lexicon and annotate a training dataset

The text database prepared in Step 1 is then used for extracting information about supply-chain relationships between various entities for the overall purpose of supply chain mapping. To this end, we distinguish between two types of entities: Organisation and Programme, and five types of relationships between them: Supplier to an organisation, Ownership of a programme, Supplier for a programme, Collaboration, and Part of an organisation.

Further, we selected a sample of data for annotation — which is the process of identifying and marking occurrences of the specified types of entities and relationships in the text. The annotated sample of data, referred to as the training sample, was then used for training the NLP pipeline — which is the process through which the pipeline learns the patterns of occurrence of entities and relationships in the text by example.

In the pilot run of the study, annotations have been conducted in 30 news articles by two annotators with a research track record relating to the defence sector and supply chains. The annotators have discussed each article to jointly reach a conclusion on the type of entities and relationships observed in the text. Annotations have been conducted using Prodigy — a specialised software by Explosion, the maker of spaCy.

In the second iteration of the study, the scope of annotations has been considerably expanded by involving two junior researchers as annotators whose annotations have been reviewed by the original team of annotators. The outcomes of this iteration of the study will be reported in the follow-up research paper.

Step 3: Apply trained pipeline to the full dataset

The trained pipeline can then be used to comb through unseen data and extract supply chain relationships between entities in automated fashion. Due to the statistical nature of the tasks comprising the NLP pipeline, inferences made about the existence of relationships between entities mentioned in the text are probabilistic in nature. Therefore, we have applied an acceptance threshold of 66% to such inferences, which is a moderately conservative value. Lowering it would result in an over-optimistic acceptance of false positives, whereas raising it would potentially result in omission of some true positives. The exact choice of the threshold requires further experimentation.

In the pilot run of the study, the unseen data comprised 270 news articles obtained in Step 1 due to licensing restrictions. The second iteration of the study, which is currently ongoing, utilises a Factiva API (Application Programming Interface) with access to a much broader data set comprising 5 mln articles. The results below are reported for the pilot run of the study.

Step 4: Generate supply chain map

In order to be included in the supply chain map, the relationships between the entities inferred in Step 3 require additional processing due to variations in the spelling of the entity names in news articles (such as BAE Systems and BAE) which should therefore undergo unification in order to represent the same entity on the supply chain map.

Next, following Kaloskampis et al. (2023), we used R packages 'network' and 'ggnetwork' to visually map supply–buyer relationships between organisations in a network diagram which is presented in Figure 2. This outcome proves the ability of the methodology to produce a supply chain map even with a very limited training and utilizing a small sample of articles. However, the limited amount of training comes at a cost of the errors which become apparent upon a closer examination of the network diagram, where some programmes, such as HIMARS, as well as some non-entities, such as "Martin Aeronautics Elite Supplier Award", have been incorrectly identified as organisations with supplier–buyer relationships. The insufficient amount of training of the pilot run of the study is being addressed in its second iteration, as explained above in Step 3, which should considerably reduce or eliminate such inaccuracies.

3. Conclusion

Supply chain mapping is instrumental in identifying potential constraints and vulnerabilities in complex supply chains characterised by limited visibility into their composition, yet it presents a number of practical problems. The methodology explored in the present study permits automated processing of a vast amount of textual data and generation of supply chain maps through extraction of supply chain relationships from the text, which considerably reduces the cost and time required to map supply chain relationships. An additional strength of the methodology is its applicability to different languages, which aligns with the global character of today's supply chains.

At the same time, application of the methodology poses a number of challenges. First, incomplete, imprecise and ambiguous data may lead to low-quality outcomes. Second, information about supply chain relationships may become outdated, which more recent news articles may fail to reflect, and which renders the supply chain map partly outdated. Third, the need for annotating the data to train the model requires an upfront investment which scales up if the application is featuring multiple languages. Fourth, the inherent error of statistical learning may result in some relationships not being uncovered and some being uncovered wrongly. This consideration requires further exploration using a more extensive training and a more extensive data set, which is the focus of the second, ongoing iteration of the study.

While the study primarily focuses on defence supply chains, this choice is driven by the availability of a unique and comprehensive dataset of defence supply chains, which allows for calibrating and validating our methodology against real-world data. However, the ultimate goal of this research extends far beyond the defence industry.

Once refined and validated within the defence context, this methodology can be adapted and applied to map intricate supply networks in other critical areas such as transport, energy, healthcare, and advanced manufacturing.

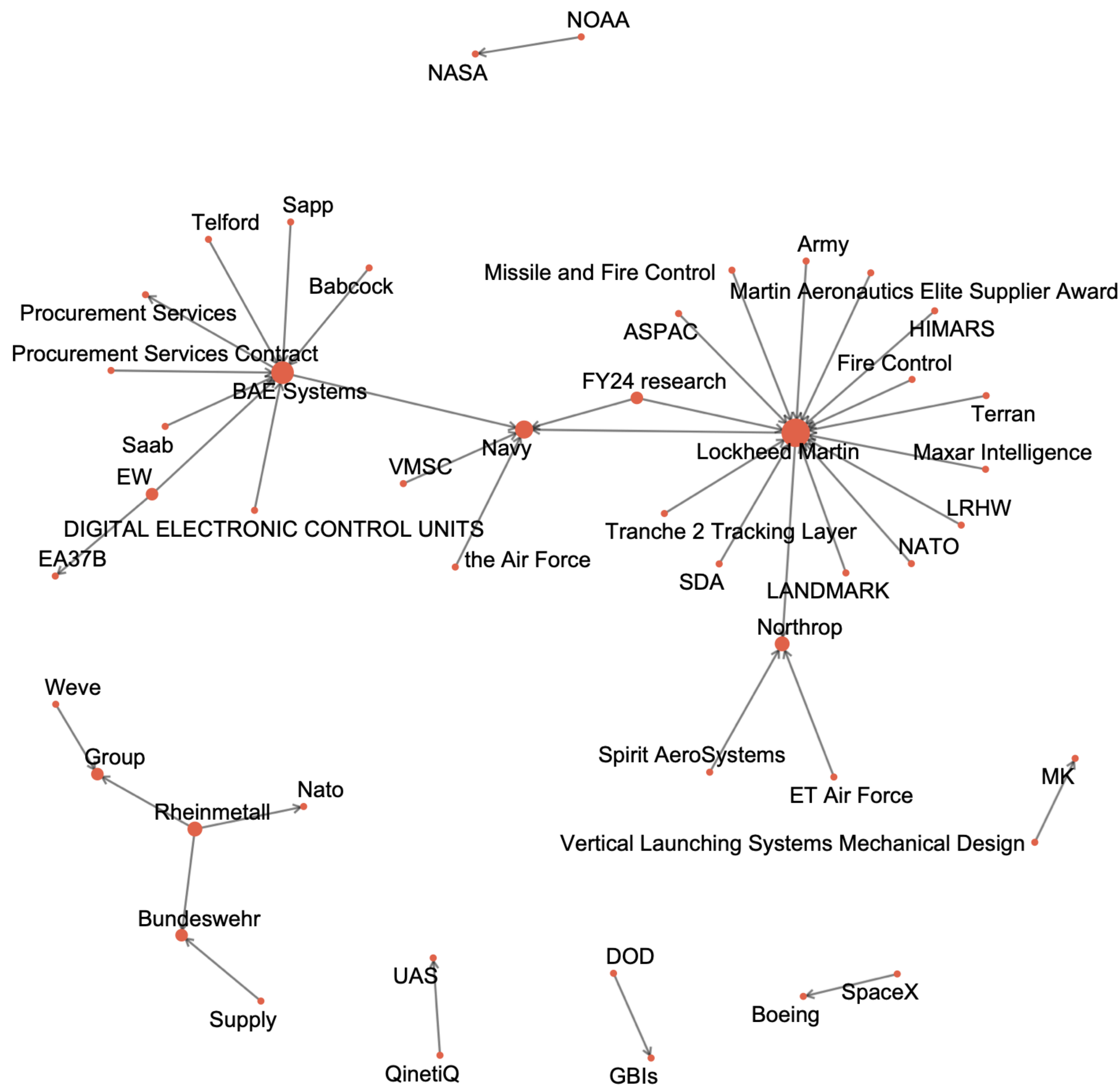

Figure 2. Supply chain map produced in the pilot run of the study.
Note: node sizes correspond to the degrees of the nodes.

4. Implications for policy

Beyond the technical findings, the presented approach has clear implications for policy. NLP-based supply chain mapping offers a cost-effective way to improve visibility and resilience in defence procurement, including procurement of innovation (Selviaridis & Uyarra, 2025). Specifically, the NLP approach reduces reliance on resource-intensive surveys and enables early identification of single points of failure and foreign ownership risks, thereby supporting national security objectives. Once validated, the methodology can be extended to other critical sectors such as energy, healthcare, and transport, aligning with broader resilience and friend-shoring strategies. To realise these benefits, policymakers should consider:

- Funding targeted pilot projects in defence and other critical sectors to test and refine the approach.

- Creating secure data-sharing frameworks that allow contractors and government to exchange supply chain information without compromising commercial confidentiality.

- Integrating NLP-based mapping into procurement risk assessment processes, ensuring that supply chain resilience becomes a standard part of defence and industrial strategy.

References

DE&S (2023). New Defence supply chain mapping tool provides vital visibility, insight and intelligence. Defence Equipment & Support, Ministry of Defence, UK, 29 March 2023, https://des.mod.uk/new-defence-supply-chain-mapping-tool/. Accessed 29 November 2025.

Dow Jones (2025). Dow Jones Factiva - News, Data & Media Monitoring Solutions. https://www.dowjones.com/business-intelligence/factiva/. Accessed 29 November 2025.

Gereffi, G., Humphrey, J., & Sturgeon, T. (2005). The governance of global value chains. Review of International Political Economy, 12(1), 78–104, https://doi.org/10.1080/09692290500049805.

Gholz, E., & James, A. (2023). Assessing the Link between Dependence and Vulnerability: Globalization and China in the U.S. Defense Supply Chain. Mimeo.

Kaloskampis, I., Bradnum, D., McGowan, C., Hunter, P., & Bui, M. (2023). Using Natural Language Processing for the analysis of global supply chains. Data Science Campus, 11 May 2023, https://datasciencecampus.ons.gov.uk/using-natural-language-processing-for-the-analysis-of-global-supply-chains/. Accessed 29 November 2025.

MacCarthy, B. L., Ahmed, W. A. H., & Demirel, G. (2022). Mapping the supply chain: Why, what and how? International Journal of Production Economics, 250, 108688, https://doi.org/10.1016/j.ijpe.2022.108688.

Selviaridis, K., & Uyarra, E. (2025). How intermediaries manage knowledge to support public procurement of innovation: The case of UK defence. Research Policy, 54, 105335, https://doi.org/10.1016/j.respol.2025.105335.

Wichmann, P., Brintrup, A., Baker, S., Woodall, P., & McFarlane, D. (2020). Extracting supply chain maps from news articles using deep neural networks. International Journal of Production Research, 58(17), 5320–5336, https://doi.org/10.1080/00207543.2020.1720925.

.

Innovation Procurement Empowerment Centre

Andrew D. James
Professor of Innovation Management & Policy
Manchester Institute of Innovation Research
University of Manchester
E-mail: Andrew.James@manchester.ac.uk

Grigory Pishchulov
Senior Lecturer in Information Systems and Supply Chain Management
Alliance Manchester Business School
University of Manchester
E-mail: grigory.pishchulov@manchester.ac.uk

IPEC website: https://www.ipec.org.uk/

IPEC Linkedin: https://www.linkedin.com/company/93121184

IPEC email: contact@ipec.org.uk