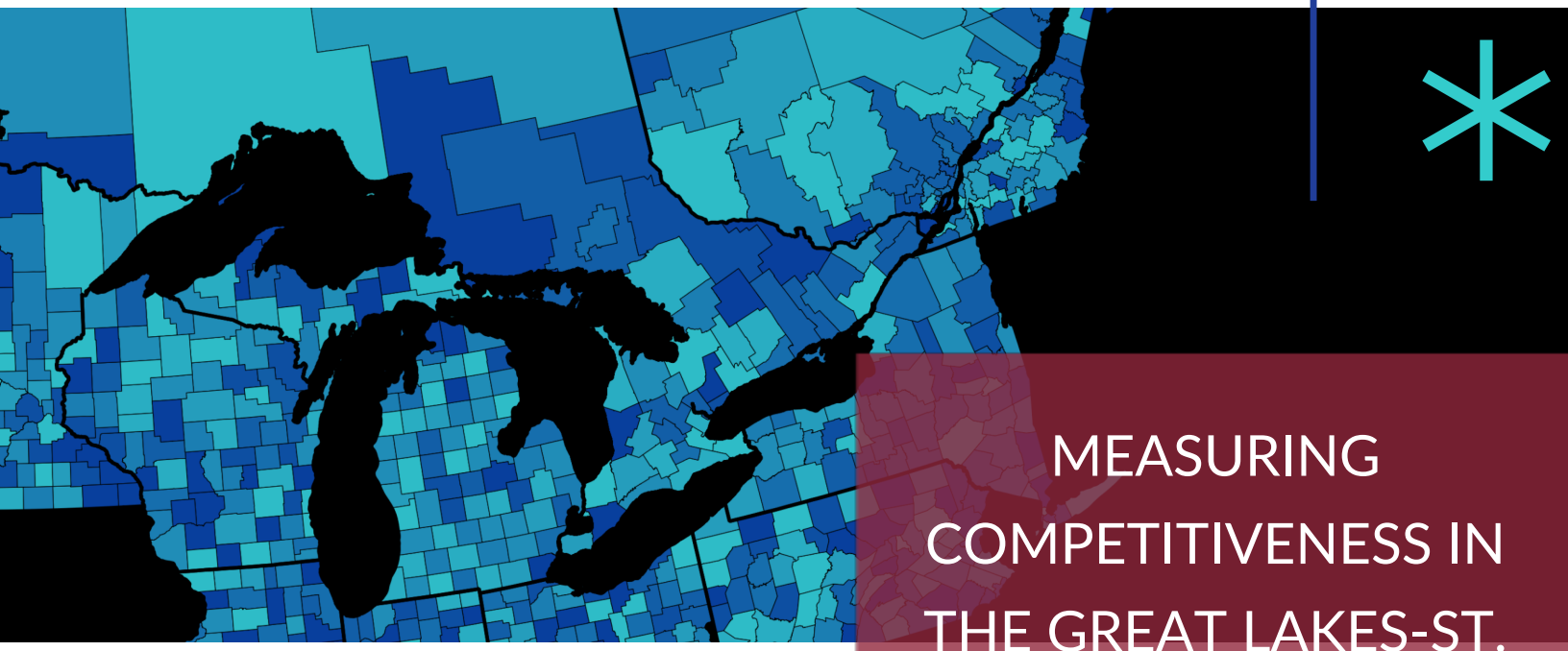
CIRANO
Knowledge into action



MEASURING COMPETITIVENESS IN THE GREAT LAKES-ST. LAWRENCE REGION USING A DIGITAL TWIN: A GEOSPATIAL DATA SCIENCE APPROACH

NATHALIE DE MARCELLIS-WARIN
MARTIN TRÉPANIER
THIERRY WARIN

With the collaboration of Lucien Chaffa, Sarah Elimam, Aïchata Kone, Molivann Panot, Jeremy Schneider, Bilal Siddika, Adam Touré

PR

2024PR-04
FOR REFLECTION

The papers For Reflection... aim to propose, through applied research results or discussion documents, actions to be taken to accelerate recovery, ensure sustainable economic growth, energize Quebec's regions and reduce the future budget deficit while maintaining adequate funding for health and education. These documents are the sole responsibility of the authors.

Les documents Pour Réflexion... visent à proposer, par l'entremise de résultats de recherche appliquée ou de documents de réflexion, des actions à privilégier pour accélérer la reprise, assurer une croissance économique durable, dynamiser les régions du Québec et résorber le déficit budgétaire à venir tout en maintenant un financement adéquat pour la santé et l'éducation. Ces documents sont sous la seule responsabilité des auteurs.

CIRANO is a private non-profit organization incorporated under the Quebec Companies Act. Its infrastructure and research activities are funded through fees paid by member organizations, an infrastructure grant from the government of Quebec, and grants and research mandates obtained by its research teams.

Le CIRANO est un organisme sans but lucratif constitué en vertu de la Loi des compagnies du Québec. Le financement de son infrastructure et de ses activités de recherche provient des cotisations de ses organisations-membres, d'une subvention d'infrastructure du gouvernement du Québec, de même que des subventions et mandats obtenus par ses équipes de recherche.

CIRANO Partners - *Les partenaires du CIRANO*

Corporate Partners *- Partenaires corporatifs*
Autorité des marchés financiers
Banque de développement du Canada
Banque du Canada
Banque nationale du Canada
Bell Canada
BMO Groupe financier
Caisse de dépôt et placement du Québec
Énergir
Hydro-Québec
Innovation, Sciences et Développement économique Canada
Intact Corporation Financière
Investissements PSP
Manuvie Canada
Ministère de l'Économie, de l'Innovation et de l'Énergie
Ministère des finances du Québec
Mouvement Desjardins
Power Corporation du Canada
Ville de Montréal

Academic Partners - *Partenaires universitaires*
École de technologie supérieure
École nationale d'administration publique
HEC Montréal
Institut national de la recherche scientifique
Polytechnique Montréal
Université Concordia
Université de Montréal
Université de Sherbrooke
Université du Québec
Université du Québec à Montréal
Université Laval
Université McGill

CIRANO collaborates with many centers and university research chairs; list available on its website. *Le CIRANO collabore avec de nombreux centres et chaires de recherche universitaires dont on peut consulter la liste sur son site web.*

ISSN 2563-7266 (online version)

Measuring Competitiveness in the Great Lakes-St. Lawrence Region Using a Digital Twin: A Geospatial Data Science Approach

De Marcellis-Warin, N.[1], Trépanier M.[2], Warin T.[3]

with the collaboration of Lucien Chaffa, Sarah Elimam, Aïchata Kone, Molivann Panot, Jeremy Schneider, Bilal Siddika, Adam Touré [4]

Abstract/ Résumé

The study of competitiveness has long been constrained by traditional trade analyses, which focus on inter-industry flows between countries while overlooking the intricate interconnectedness of supply chains. This position paper advocates for the use of digital twin technology to replicate complex economic systems, enabling the modeling of firm-to-firm interactions and uncovering the micro-level impacts of macroeconomic phenomena. We present an integrated analytical framework to analyze the bi-national Great Lakes-St. Lawrence (GLSL) region, spanning Canada and the United States. The creation of a digital twin for this region represents a transformative step in the digitalization of regional economies. This framework provides an integrated analysis of trade, transportation, and environmental systems, enhancing our understanding of regional competitiveness and supporting strategic decision-making. It emphasizes the critical role of multimodal transportation networks, particularly in addressing the challenges posed by climate change, as a key determinant of regional competitiveness.

L'étude de la compétitivité a longtemps été limitée par les analyses traditionnelles des flux commerciaux, centrées sur les flux inter-industries entre pays tout en négligeant l'interconnexion complexe des chaînes d'approvisionnement. Cet article propose l'utilisation d'un jumeau numérique pour répliquer les systèmes économiques complexes, permettant de modéliser les interactions entre entreprises et de révéler les impacts microéconomiques des phénomènes macroéconomiques. Nous présentons un cadre analytique intégré pour analyser la région binationale du Saint-Laurent et des Grands Lacs (SLGL), englobant le Canada et les États-Unis. La création d'un jumeau numérique pour cette région constitue une étape innovante dans la numérisation des économies régionales. Ce cadre offre une analyse intégrée des systèmes de flux commerciaux, de transport et d'environnement, améliorant notre compréhension de la compétitivité régionale et appuyant la prise de décisions stratégiques. Il met en lumière le rôle critique des réseaux de transport multimodaux, notamment face aux défis posés par les changements climatiques, en tant que déterminant clé de la compétitivité régionale.

[1] Professor, Polytechnique Montréal, President and CEO and Fellow, CIRANO

[2] Professor, Polytechnique Montréal, Director, CIRRELT, Co-Pi Pole on Data Science for Trade and Intermodal Transportation – *GVCdtLab,* CIRANO

[3] Professor, HEC Montréal, Researcher and Fellow, CIRANO, Co-Pi Pole on Data Science for Trade and Intermodal Transportation – *GVCdtLab,* Fellow, *Digital, Data and Design (D^3) Institute* at Harvard Business School.

[4] Research Team, Pole on Data Science for Trade and Intermodal Transportation – GVCdtLab, CIRANO

Keywords/ Mots-clés: Competitiveness, Digital Twin, Complexity Economics, Global Value Chains, Machine Learning, Multimodal Transportation, Great Lakes-St. Lawrence region / Compétitivité, Jumeau numérique, Économie de la complexité, Chaînes de valeur mondiales, Apprentissage automatique, Transport multimodal, Région du Saint-Laurent et des Grands Lacs

To quote this document / Pour citer ce document

De Marcellis-Warin, N., Trépanier, M., & Warin, T. (2024). Measuring Competitiveness in the Great Lakes-St. Lawrence Region Using a Digital Twin: A Geospatial Data Science Approach (2024PR-04, Pour réflexion, CIRANO.) https://doi.org/10.54932/DKBC6587

Introduction

If you were to guess which was the third largest economy in the world, you would probably not think of the Great Lakes-St-Lawrence (GLSL) region. Cutting across Canada and the United States, the provinces of Québec and Ontario, and the states of New York, Pennsylvania, Ohio, Michigan, Wisconsin, Indiana, Illinois, and Minnesota together are home to a USD $7.9 trillion economy (Siddika, 2024). In industries such as automotive, aerospace and agri-food within this region, value chains are highly integrated across borders (Fisher et al., 2017). Individual firms collaborate and compete to supply intermediate goods and services, creating value that extends beyond their own industrial and geographic boundaries.

Global trade can be conceptualized as a dynamic aggregation of flows resulting from firm-to-firm relationships that transcend traditional market boundaries. Through the lens of network science, firms and the industries they represent emerge as critical nodes within a complex and interdependent web of regional specialization. These nodes, varying in importance, are interconnected by linkages that reflect the intricate structure of global value chains (GVC). Notably, approximately one-third of global trade flows consist of intra-firm exchanges orchestrated by multinational enterprises operating across borders. This underscores the centrality of firm-level interactions in shaping trade dynamics, emphasizing the need for innovative approaches, such as digital twins, to capture and analyze these intricate networks within the context of regional competitiveness (Lanz & Miroudot, 2011). The rest are flows originating from a relatively small number of the most competitive firms (Bernard et al., 2007).

Importantly, the networked structure of trade is further characterized by firm-to-firm flows that predominantly operate within national boundaries, representing most of the trade. These domestic linkages are integral to understanding the broader economic network, as they underpin the regional specialization and interdependencies that drive competitiveness. By extending the focus beyond cross-border exchanges, this initiative highlights the need to examine the intricate, localized interactions that collectively contribute to the dynamics of global trade. This perspective reinforces the value of deploying advanced analytical tools, such as digital twins, to capture these complex intra-national flows and their implications for regional and global economic systems.

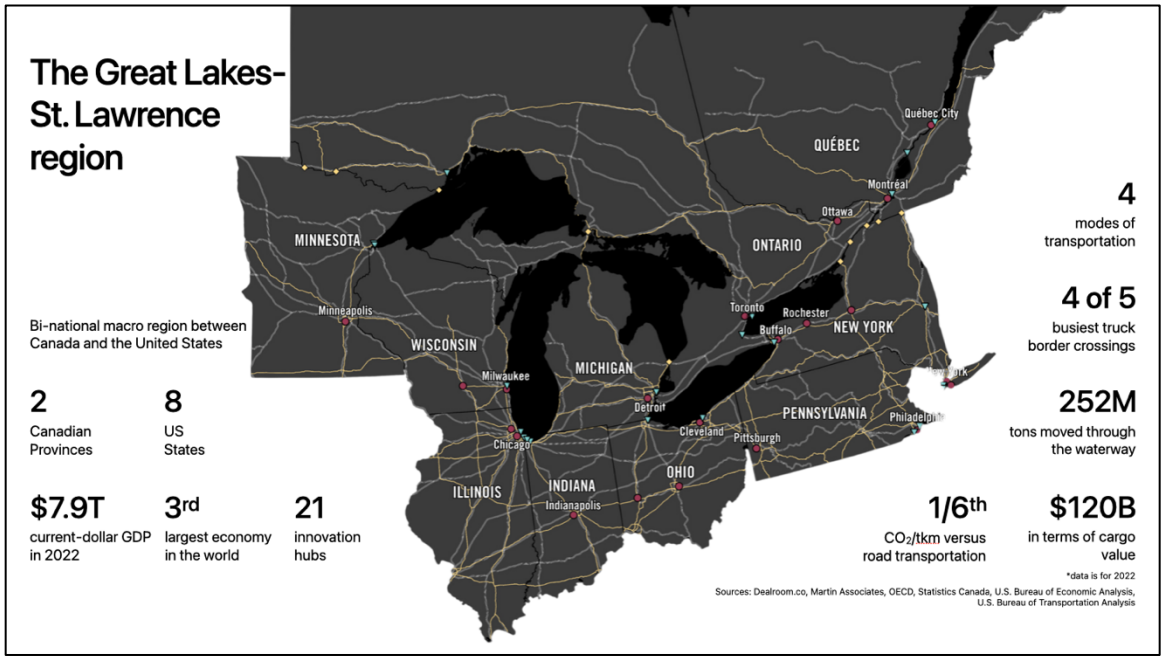
Consequently, it is efficient logistics that facilitates firm-to-firm linkages by reducing the time and cost to trade between domestic and international markets, making it a key determinant of competitiveness. At an aggregate level, the performance of the logistics

sector is directly correlated with a region's level of economic development (Hausman et al., 2013). While trade policy can either incentivize or disincentivize direct linkages with certain partners in strategically important industries, such as semiconductors (Freund et al., 2024), impediments in crucial nodes of the global freight transportation network sends shockwaves through GVCs. Subsequently, a blockade of ships in the Red Sea has a cascading effect on the interconnected global economy (Acemoglu et al., 2012).

Indeed, globalization shifts the spatial emphasis of economies away from hard territorial delineations towards a more abstract relational network between firms (Amin, 2002). Economies, as complex systems, consist of firms that interact and imperfectly optimize their value chains, guided primarily by market forces and the regulatory frameworks within which they operate.

These interactions, shaped by both local and global dynamics, often result in the formation of economic clusters—densely connected networks of activity that transcend political borders. A striking example of this phenomenon is the GLSL region. This bi-national corridor, emblematic of economic interdependence, accounts for approximately one-third of the combined economic output of Canada and the United States. By examining such regions through advanced methodologies, such as a digital twin, we gain a deeper understanding of the spatial and systemic factors that underpin regional competitiveness and global economic integration.

Figure 1: The GLSL region's vibrant economy historically developed around its navigable inland waterway



The GLSL region carries a rich historical legacy, with the St. Lawrence River serving as a vital gateway for European settlers to access the expansive Great Lakes. Over time, this strategic waterway has shaped the development of the region, fostering the growth of some of North America's most prominent cities along its shores. Today, these urban centers stand as testaments to the region's enduring role as a hub of economic and cultural activity (see Figure *1*: The GLSL region's vibrant economy historically developed around its navigable inland waterway), supported by trade flows moving through more than 110 ports along the waterway and the region's roads, railways and airports (Martin Associates, 2023). The intensity of trade in this region creates immense demand on its transportation network, making it home to some of North America's busiest roads, railways, border crossings, ports and airports (National Academies of Sciences, Engineering, and Medicine, 2012).

The region's transportation network faces several challenges due to its unique characteristics. The presence of the international border, along with the lakes, rivers, and topography, creates bottlenecks that add complexity to the system. The competitiveness of many of this region's industries on global markets depends on the capacity and performance of its multi-modal freight transportation network.

Climate change poses both challenges and opportunities for the Great Lakes-St. Lawrence region, including the potential for year-round navigation on the previously ice-bound

Great Lakes and the risk of insufficient water depth in sections of the St. Lawrence River, disrupting shipping routes. These evolving environmental dynamics necessitate a deeper understanding of their impacts on the region's economy, transportation networks, and overall competitiveness. At the same time, the study of competitiveness must advance beyond traditional frameworks to address the complexities of modern economies. How can such intricate challenges be effectively tackled?

This position paper advocates for the adoption of digital twin technology—a cutting-edge tool capable of replicating and dynamically modeling the economic, transportation, and environmental systems of a region. By providing an integrated and real-time perspective, digital twins offer a comprehensive approach to analyzing and enhancing regional competitiveness. To illustrate the transformative potential of this methodology, we detail the development of a digital twin for the bi-national GLSL region. With its highly integrated economy and distinctive multi-modal freight transportation network, the GLSL region presents an ideal case study for exploring how digital twin technology can address the pressing challenges of climate change and economic complexity.

The remainder of this paper is structured as follows. Section 2 describes our application of digital twin technology to study competitiveness, followed by Section 3, which introduces the conceptual framework underpinning our analysis. Section 4 outlines the methodological steps and data integration techniques used. Section 5 assesses the capabilities of digital twins for predictive analytics and scenario simulations, with potential to foster development amid geopolitical uncertainties and climate change. We also consider the implications of our findings for policymakers, business leaders and academics. Finally, Section 6 synthesizes the key contributions and projected impacts of this novel approach.

1. Modelling Complexity in an Economy with Digital Twins

Recognizing the need to advance the study of trade and intermodal transportation (Dudoit et al., 2021), CIRANO launched the GVCdtLab in 2023, a multidisciplinary research lab focused on creating new tools and applying innovative methods to this end. At GVCdtLab, we are developing a digital twin of the GLSL region, leveraging leading-edge innovations in geospatial data science and machine learning to identify and explain why economic activity occurs as it does. The ability to replicate phenomena arising in complex systems and dynamically simulate interactions makes digital twins well suited to address the quantitative challenges of assessing flows in globally networked economies (Warin, 2023).

Our research is guided by the pivotal question: *How can digital twins, through the integration of geospatial data analytics and machine learning techniques, transform our understanding of competitiveness?*

In addressing this question, our findings contribute to the burgeoning field of digital twin technology by establishing a comprehensive blueprint for its application in regional competitiveness analysis. We illuminate its transformative potential in synthesizing and interpreting vast streams of near real-time data, providing a nuanced view of evolving economic dynamics with unprecedented speed, precision and depth. By integrating both aggregated and granular data from diverse disciplines, our digital twin offers a holistic examination of the GLSL region's economy as a complex system that more accurately depicts reality. More importantly, our analysis seeks to address timely questions with a nuanced, bottom-up analytical perspective that prioritizes heterogeneity, granularity and adaptability.

Amidst rising protectionism, resilience concerns and climate change, policymakers are increasingly taking an active role in shaping GVCs. While ongoing research has provided policymakers with a broad understanding of the impacts of tariffs, port blockages and changing water levels, there is a growing need to better understand the micro level outcomes of macro level phenomena. A key challenge is that traditional economic theories and policy tools rest on assumptions that prove deficient in the face of real-life complexities (Kirman, 2016). These models, grounded in predefined hypotheses and tested against empirical data, often fail to capture the evolving and interconnected nature of modern economic systems, which in turn limits our understanding and response capabilities.

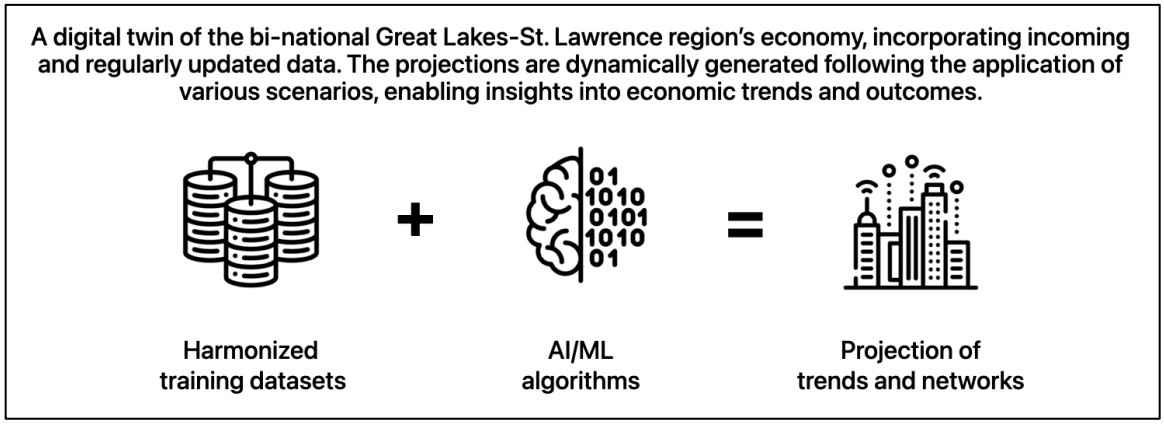
The digital twin, encompassing two Canadian provinces and eight U.S. states within the broader global economy, is accessible through our digital twin platform. Its development follows an iterative and incremental approach, allowing for the continuous addition of features and variables to the digital twin's architecture. This methodology ensures flexibility and ongoing enhancement, with the goal of accurately replicating firm-to-firm dynamics within GVCs. The digital twin traces how firms respond to demand within the constraints of transportation networks and environmental systems, offering unprecedented insights into regional competitiveness.

The digital twin of the GLSL region's economy integrates and harmonizes a wide array of datasets, including firm characteristics, industry input-output tables, transport network infrastructure and operations data, transportation flows by mode, freight traffic, and water levels in the St. Lawrence River. Furthermore, the geographic resolution of certain

datasets is refined to subnational levels, enabling a more granular unit of analysis. This approach enriches the depth of analysis and facilitates an inductive framework, uncovering novel questions and insights into the complex interplay of economic, transportation, and environmental systems.

These data are used to train unsupervised machine learning algorithms to model networks and predict future flow trends based on scenarios involving potential changes in transportation and economic systems, as well as disruptions and other circumstances affecting various components of the digital twin, as shown in Figure *2*. Through a systems perspective and inductive approach, we aim to decipher the drivers of the GLSL region's competitiveness by drawing key elements from the most pertinent economic frameworks.

Figure 2: Developing *a digital twin of a complex system*



In this context, the integration of unsupervised machine learning techniques represents a pivotal advancement. These methodologies enable an inductive approach to economic research, which, unlike its hypothetico-deductive counterpart, does not start with a predetermined hypothesis. Instead, it allows for the data itself to guide the discovery of non-linear patterns, relationships and variables that more accurately reflect the complex interconnected dynamics of modern economic systems (Barbaglia et al., 2021).

Through unsupervised machine learning, we can systematically grasp knowledge that were previously obscured or unattainable, thereby enriching our analytical framework with a depth and flexibility that traditional methods cannot provide.

Adopting an inductive, data-driven approach enables a more authentic investigation of economic phenomena, allowing for the discovery of novel variables and the development of more contextually relevant questions. By harnessing the extensive and diverse data

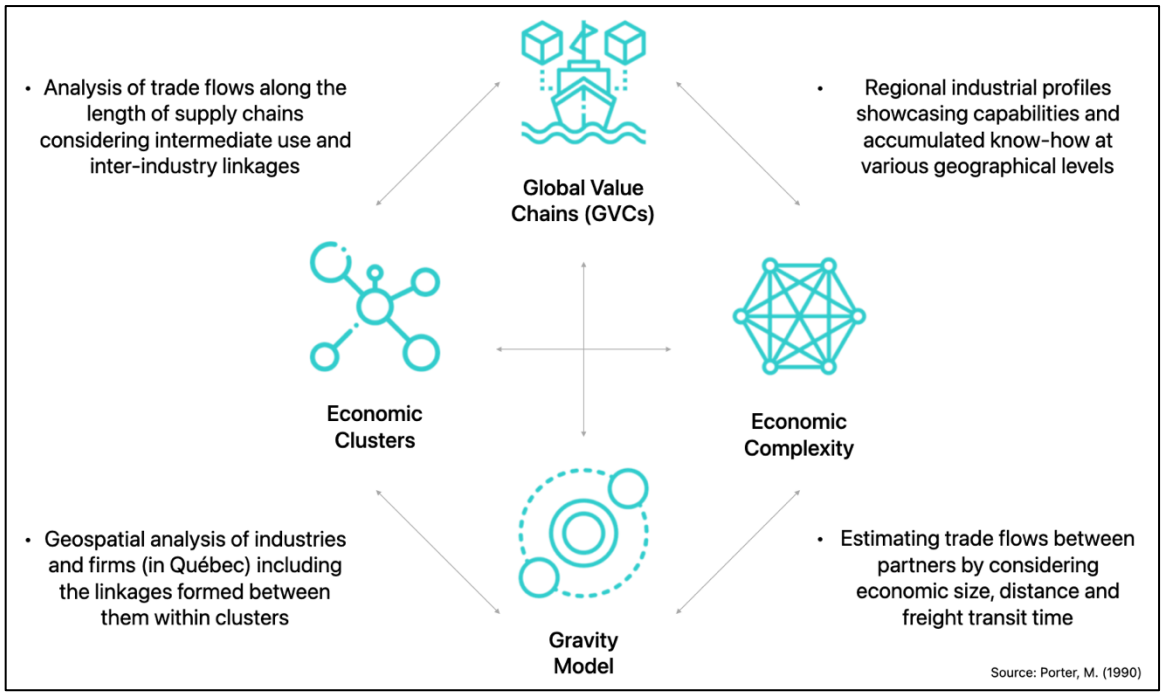
generated by global economic activities, unsupervised machine learning techniques empower researchers to move beyond the constraints of traditional assumptions and rigid classifications. This approach ensures that analyses and outcomes remain adaptable and relevant over time, reflecting the dynamic nature of modern economies. Moreover, it significantly enhances the precision and depth of economic analysis, providing policymakers with the nuanced insights needed to effectively address the complexities and disruptions characteristic of today's interconnected economic landscape.

2. A Holistic Analytical Framework

To analyze and elucidate the complex interactions within the region's economic system and the factors driving its competitiveness, we employ a holistic analytical framework that synthesizes and builds upon key elements from foundational economic theories (ref to framework paper). This framework is inspired by Porter's Diamond Model, which integrates macroeconomic and microeconomic dimensions to characterize the competitive environment in which firms operate.

The concept of regional competitiveness, extensively developed in Michael Porter's seminal work *The Competitive Advantage of Nations* (Porter, 1990), leverages the Diamond Model to explain disparities in economic performance across different geographic contexts. By extending and adapting this foundational model, the holistic analytical framework provides a nuanced tool for understanding the interplay of local and global factors shaping economic outcomes.

Figure 3: Dimensions of our holistic analytical framework



Porter's emphasis on geography, particularly in fostering innovation —a key component of competitive advantage that cannot be effectively physically relocated—makes his model a compelling starting point. He highlights how leading firms in specific industries tend to spatially collocate with firms from related industries, facilitating the exchange of expertise and knowledge (Porter, 1994). While Porter's work rightly identifies the central role of firms in creating wealth by meeting demand within their business environment, in today's interdependent economies, firms are also influenced by factors beyond their own geographic borders. Our augmentation of the Diamond Model addresses this deficiency by considering interactions beyond political borders.

Economic globalization is a complex phenomenon that does not exist in a vacuum. The global integration of economies has created uneven gains and losses, turning local challenges into global issues. While Porter's Diamond posits that localized innovation and upgrading are crucial for spatial competitiveness (Porter, 1990), the interdependencies created by the participation of firms in GVCs is crucially missing. Additionally, considering the trajectory of industry diversification adds value to the study of competitiveness. Analyzing the dynamics of firms without considering the constraints imposed by transportation networks and environmental systems also omits critical real-world complexity. Including these factors makes time to trade, alongside distance, an important variable when evaluating flows.

Building on Porter's contributions, we propose a multidimensional framework for real-time economic analysis of the bi-national GLSL region, leveraging four complementary dimensions to more accurately explain the dynamics of its complex economic system. These dimensions—Global Value Chains, Economic Clusters, Economic Complexity and the Gravity Model—form the cornerstones of our holistic analytical framework (To analyze and elucidate the complex interactions within the region's economic system and the factors driving its competitiveness, we employ a holistic analytical framework that synthesizes and builds upon key elements from foundational economic theories (ref to framework paper). This framework is inspired by Porter's Diamond Model, which integrates macroeconomic and microeconomic dimensions to characterize the competitive environment in which firms operate.

The concept of regional competitiveness, extensively developed in Michael Porter's seminal work The Competitive Advantage of Nations (Porter, 1990), leverages the Diamond Model to explain disparities in economic performance across different geographic contexts. By extending and adapting this foundational model, the holistic analytical framework provides a nuanced tool for understanding the interplay of local and global factors shaping economic outcomes.

Figure *3*). This framework integrates data science methodologies to systematically bridge these established areas of research and contribute to their scholarship with new methods and indicators. Ultimately, our holistic analytical framework will offer new insights and support the development of indicators that deepen our understanding of the micro level outcomes of economic globalization.

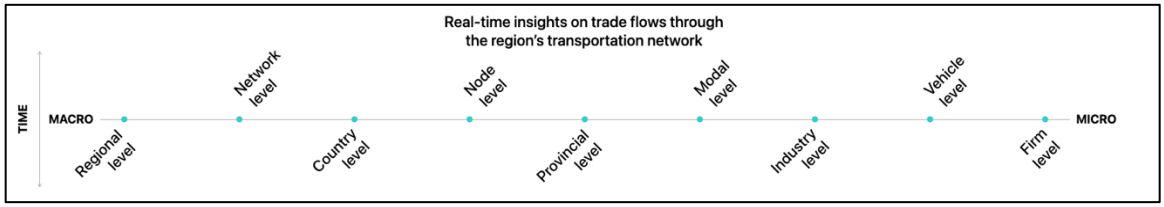
3. Significance and Applications of the Digital Twin

Developing a digital twin of the bi-national GLSL region that authentically reflects real-world interactions is a complex endeavor. This process is challenged by inconsistencies in data collection methodologies, variations in dataset features, observation intervals, classification standards, and issues such as missing data and formatting discrepancies. Addressing these challenges often necessitates data imputation, which involves leveraging information from complementary datasets or making informed assumptions based on related data. This section outlines the initial methodological steps and data integration strategies employed to establish a functional model of the region's multifaceted economic system, enabling preliminary analyses.

The starting point was to create a reliable representation of trade flows within the GLSL region. Recognizing that most of the firm-to-firm trade occurs domestically, relying solely on export data would result in an incomplete picture. Furthermore, country-level export data typically lacks critical details on provenance, value-added contributions, and final use, all of which are essential for mapping industry value chains. To address these gaps, we utilized the OECD's input-output tables (IOTs), which provide inter-industry flow data at the country level. Building on this foundation, we identified complementary datasets to construct regional IOTs at the provincial and state levels, ensuring alignment with the OECD's unit of analysis.

By employing a multi-scalar approach, we reconstructed aggregate trade flows and generated granular estimates to create a comprehensive trade and transportation network for the GLSL region. This approach not only enhances the granularity of the analysis but also provides a robust framework for understanding the intricate dynamics of regional trade and its integration within broader global economic systems, as shown in Figure *4*.

Figure 4: Varying spatial scales to create a bottom-up perspective



Constructing our IOTs required the use of sub-national IOTs from Canada and the US, as well as freight analysis frameworks from both countries to base off estimates for missing values. With this integration, we can analyze industry-level trade flows from Quebec, for instance, not only to Ontario and New York but also to Germany. This spatial and industrial granularity enables us to capture a more nuanced picture of global trade dynamics, revealing intricate patterns of economic interdependence. At this stage, we can add value to the already extensive list of indicators produced by the OECD through their country-level IOTs1[1]. Using machine learning, such low-frequency data can be

[1]See OECD's Trade in Value-Added (TiVA) indicators: https://www.oecd.org/en/topics/sub-issues/trade-in-value-added.html

forecasted to provide a current estimate of the state of the economy. This opens a world of possibilities for developing the networked structure of the economy in our digital twin.

Linking broader longitudinal statistics on trade and transportation with near real-time geospatial data on firms and individual vessels of transport unlocks a highly localized and dynamic perspective of globalization. For instance, high-frequency granular data, such as custom records and business establishment information, can be combined with the data from our IOTs to reconstruct aggregate cross-border trade. Parallelly, a full-scale model of the multimodal transportation network including roads, ports, airports and railroad is integrated so freight traffic and capacity constraints on transportation infrastructure can be estimated, including the effect of weather on water and ice levels in the St. Lawrence River, which impacts the movement of ships through the waterway. The digital twin can then be used to model real-time wealth creation at the firm level or to study the micro effects of shocks in the global economy.

Following this, the next step was to pool and integrate the identified datasets (see Annex 1), together with those used to construct the IOTs, into a relational database, which would then feed into the modelling and analysis on our digital twin platform. This database can then be used to pull data for creating new variables and indicators or for direct use in modelling and analysis. The datasets that have so far been integrated into our digital twin are listed in Table 1. Ultimately, high-quality data that previously existed in siloes becomes interoperable, adding complexity and depth to the analysis by serving as additional variables. Over time, this will lead to a systems perspective, allowing to address current economic challenges using data science methods through an inductive approach.

The digital twin serves as a dynamic testing ground for the application of our holistic analytical framework, providing iterative opportunities to refine both the framework's dimensions and our understanding of the GLSL region. By integrating machine learning techniques, the digital twin unlocks the potential of the underlying database, facilitating the identification and exploration of complex patterns while enabling continuous model training. This iterative process not only enhances the analytical capacity of the digital twin but also informs the development of hypotheses grounded in data-driven insights.

Machine learning, as a predictive and analytical tool, excels at processing vast datasets and approximating complex, unknown models through flexible functional forms. It effectively mitigates the risk of overfitting by employing regularization techniques, ensuring robust and reliable outcomes. A key strength of machine learning lies in its minimal reliance on a priori assumptions about data structure or model shape, allowing

for a more exploratory approach to data analysis. This flexibility makes it an indispensable component of the digital twin, enabling a nuanced understanding of the region's economic, transportation, and environmental dynamics while driving innovation in regional analysis (Stevanovic, 2021).

Coulombe et al. (2022) examine the key features of machine learning modelling that enhance macroeconomic prediction. Machine learning models have demonstrated significant advantages over traditional linear macro-econometric methods by effectively capturing key nonlinearities associated with uncertainty and financial frictions. This capability enhances forecasting accuracy and allows for the identification of new variables and indicators that emerge organically from the data. Such a characteristic aligns with the inductive research approach, which prioritizes discovering relationships and structures within data without relying on predefined hypotheses. Moreover, the predictive power of machine learning facilitates scenario modeling, such as analyzing the sectoral impacts of supply chain disruptions, thereby offering detailed insights into complex economic phenomena.

These tools enable the exploration of critical metrics such as centrality, density, and betweenness, as well as the study of themes like the spatial and organizational clustering of firms. Researchers can integrate their existing theories and models into the platform, transitioning from static to dynamic frameworks. This shift allows theoretical constructs to evolve continuously in response to real-time data, ensuring that academic models remain empirically relevant and reflective of the ever-changing global landscape.

At the core of the digital twin is digital data, which underpins the data science-driven inductive research model. However, the availability of current and high-quality datasets often presents a bottleneck in the development of the digital twin. By addressing this challenge, the platform ensures that users can engage with robust, harmonized data, fostering transformative advancements in economic analysis, policymaking, and academic research. Dudoit et al. (2021) identify a gap in trade data for marine and road modes with Transport Canada, which reduces the traceability for many goods. Additionally, differing time lags in data reporting and gaps in data collection further complicate analysis.

Economic data lags the speed of economic change, unlike financial data. We address this gap by using proxies and machine learning to extrapolate data, offering a near real-time window into the economy. Details on the assumptions we make will be made available on the digital twin.

The initiative to create a digital twin of the GLSL region embodies a significant stride towards a comprehensive understanding and enhanced management of this critical economic corridor. By mimicking the characteristics and behaviors of the real-life economy of the GLSL region onto the virtual realm, we enable a dynamic study of the underlying systemic interactions of actors connected through economic activities. With a comprehensive and real-time view of the economy, stakeholders will gain insights into the performance, efficiency and challenges facing the economy, helping them better decide on the future course of their actions. It not only presents an unparalleled opportunity for cross-disciplinary collaboration but also sets a new benchmark in the application of data science to regional economic analysis.

4. Spurring a Paradigm Shift in Research with Data Science

Big data presents a host of opportunities and challenges, marking the start of a new industrial revolution (Warin et al., 2014). Innovation and economic performance increasingly hinge on the quantity and quality of data. The insights data provides leads to the development of personalized products and services, drives productivity and results in resource efficiency gains across the economy (Dudoit et al., 2021). It promises to completely change the way governments, businesses and organizations work, allowing them to tackle societal, environmental and climate-related challenges.

Machine Learning: Unlocking Predictive Power for Economic Analysis

The integration of data science with real-time, geospatial, structured, and unstructured data, combined with machine learning techniques, signals a transformative evolution in economic research, particularly for the GLSL region. This innovative methodology offers a more detailed and nuanced understanding of economic dynamics, addressing limitations inherent in traditional approaches.

Real-time data adds immediacy to economic analysis, enabling researchers and policymakers to monitor economic activities as they occur. In the context of a rapidly evolving global economy, where conditions and opportunities can shift swiftly, this immediacy is critical. Real-time access to data empowers stakeholders to make timely, informed decisions, enhancing their ability to respond effectively to emerging trends and challenges.

Geospatial data introduces a vital spatial dimension, allowing for the examination of economic activities within their geographical context. This is especially pertinent for the

GLSL region, where geographic factors significantly shape economic interactions. By spatially mapping trade flows, transportation networks, and economic activities, stakeholders can identify regional strengths and vulnerabilities, informing targeted strategies for intervention and development.

The combination of structured and unstructured data enriches the analytical process by providing a more comprehensive dataset. Structured data, organized for straightforward analysis, complements unstructured data—such as text, images, and social media—which provides qualitative insights. Together, they create a multidimensional view of the economic landscape, merging quantitative precision with contextual depth.

Machine learning techniques further elevate this approach by enabling the development of predictive models that reveal patterns and relationships within the data, often inaccessible through conventional analysis. These algorithms handle vast datasets, identify trends, and produce highly accurate forecasts. This predictive capability is indispensable for planning and decision-making, equipping stakeholders to anticipate and adapt to future changes. As a result, we will now be able to analyze value flows holistically, rather than viewing transportation network volumes and financial flows between jurisdictions as separate entities.

By leveraging these advanced data science methodologies, the project enhances the precision and scope of economic research while democratizing access to insights. This inclusivity enables a broader range of stakeholders to participate in and benefit from the findings. The shift to a data-driven, inductive research model marks a significant break from traditional methods, unlocking deeper understandings of economic complexities and fostering informed decision-making across the GLSL region. This paradigm shift, underpinned by state-of-the-art technology and methodologies, redefines the boundaries of economic research and policymaking, offering a dynamic and effective framework for managing and developing economic systems in an increasingly complex world.

The digital twin fosters a structured dialogue among a diverse range of stakeholders, including policymakers, academia, and the business community, promoting collaboration that transcends traditional boundaries. This collaborative framework is essential for integrating varied perspectives and expertise, leading to enriched outcomes and a more holistic approach to regional development. Notably, the grouping of stakeholders into these categories does not imply mutually exclusive roles; for instance, a research institute may contribute both to academic inquiry and policy development.

For policymakers, the digital twin provides a dynamic tool to simulate and evaluate the real-time impacts of policy decisions, thereby enhancing the effectiveness of strategies aimed at fostering economic growth, competitiveness, and resilience. This capability is especially valuable for addressing the complexities of cross-border trade and transportation, enabling data-driven and well-informed decision-making processes.

The academic community gains unprecedented access to real-time data and sophisticated analytical tools through the digital twin. By offering a large, harmonized database and deploying machine learning methodologies, the platform encourages a shift from traditional, hypothesis-driven research to a more inductive, data-centric approach. This transition enables the development of dynamic theories and models that are continuously updated, reflecting the rapidly evolving economic landscape and fostering new avenues of economic inquiry.

Businesses and private consulting firms find the digital twin to be an indispensable resource for understanding the GLSL region's trade and transportation dynamics. The platform's extensive data repository and advanced visualization capabilities provide comprehensive insights into supply chains, industrial clusters, and economic complexities. These tools empower businesses to optimize supply chain configurations, enhance competitiveness, and make strategic decisions rooted in robust data analysis.

The creation of a digital twin for the GLSL region's economy, built on a multidisciplinary foundation, offers significant opportunities for partnerships and collaboration across various disciplines. By establishing a platform for structured dialogue, the initiative facilitates an exchange of expertise among stakeholders, fostering the much-needed cross-border cooperation between government entities and the private sector. This collaboration has the potential to redefine economic planning and management practices, creating a more integrated and resilient economic framework.

Furthermore, the digital twin presents academics with unique opportunities to delve into research on the GLSL region and experiment with data science as a means of analyzing and understanding complex economic systems. By bridging academic, policy, and business perspectives, this initiative serves as a catalyst for transformative change in how regional economies are studied, managed, and developed.

The Digital Twin: A Collaborative Framework for Stakeholder Engagement

Policymakers

This category encompasses governmental departments, agencies, nonprofit corporations, and think tanks, which play a pivotal role in shaping the economic landscape of the GLSL

region. The region presents a unique opportunity to transform latent potential into tangible drivers of economic growth and competitiveness for its provinces and states. Given the shared characteristics and interdependencies between Canada and the United States, treating the GLSL region as a singular economic entity allows policymakers to capitalize on low-hanging fruits, fostering synergies with neighboring partners. A coordinated strategy to establish a complementary business environment can facilitate cross-border business interactions, leading to inclusive growth and enhanced employment opportunities.

The lingering effects of the COVID-19 pandemic, coupled with escalating geopolitical challenges in regions such as Taiwan, Ukraine, and the Middle East, have prompted businesses to reassess their globally dispersed supply chains. In this context, the digital twin offers policymakers detailed insights into the geographic footprint of supply chains, supported by analyses of clustering patterns and economic complexity. Viewing trade within the region through the lens of its multi-modal transportation networks provides invaluable perspectives, particularly as stakeholders adapt to the long-term implications of climate change.

Real-time access to trade data enables policymakers to gain a refined understanding of the systemic movements underlying globalization. Leveraging AI tools integrated within the digital twin allows decisions to be data-driven, supporting the formulation of policies spanning industrial, innovation, transportation, cluster, and supply chain domains. These insights are critical for policymakers striving to boost economic competitiveness while identifying areas for improvement and assessing the potential impacts of their strategies.

As Gereffi et al. (2021) observe, trade restrictions designed to limit the flow of goods across borders can yield unintended consequences. The intricate production and sourcing networks within global value chains (GVCs) mean that such distortions can exacerbate uncertainty in the global economy. Firms have often responded to these disruptions by altering production locations, markets, and suppliers, or by upgrading their value chain activities. Policymakers must therefore develop a more nuanced understanding of the sectoral and temporal effects of trade restrictions to mitigate unintended outcomes.

In this regard, the digital twin becomes an invaluable tool for policymakers when designing initiatives and responding to external policy measures. It facilitates an objective evaluation of public investments and non-monetary interventions in key industrial clusters. Additionally, when trade restrictions are employed in response to geopolitical challenges, such as through tariffs, the digital twin offers a comprehensive perspective on the far-reaching effects of these measures within GVCs. Moreover, investments in

transport infrastructure, supported by policies like the Inflation Reduction Act, can be optimized by assessing their broader benefits across the transportation network. This ensures that each dollar invested delivers maximum efficiency and impact, contributing to the region's resilience and competitiveness.

Academia

This category encompasses academics, academic institutions, and research institutes dedicated to exploring economic and scientific dimensions of the GLSL region.

For researchers and research centers, the digital twin provides an unparalleled resource, offering a wealth of trade and transportation data specific to the bi-national GLSL region. By harmonizing existing reliable data sources on a continuous basis and making them accessible through a real-time API, the digital twin enables users to engage with an expansive dataset. Moreover, the integration of advanced machine learning techniques alongside traditional econometric tools allows researchers to transcend conventional methodologies, fostering a shift from the hypothetico-deductive model—prone to biases from predefined hypotheses—to a more exploratory, inductive approach.

Research institutions will gain unprecedented opportunities to generate profound insights into the complexity of the surrounding economic landscape. Machine learning models, with their ability to identify intricate patterns in large datasets, offer capabilities far surpassing those of traditional econometric approaches. These models effectively handle the idiosyncrasies and intricacies of economic data, facilitating nuanced analyses that can underpin the development of more robust, data-driven theories.

The availability of real-time data further enhances this capacity, allowing researchers to continuously update their theories and models in alignment with the evolving economic landscape. This adaptability contrasts sharply with the limitations of traditional research reliant on historical data, which quickly becomes outdated upon publication. The ability of machine learning models to adapt and learn from emerging data patterns ensures that research remains relevant and reflective of real-world dynamics, making the digital twin a transformative asset for academia.

Several research institutions within the GLSL region already focus on trade and transportation, with an increasing interest in utilizing AI to optimize supply chains. Notably, Quebec province alone hosts three dedicated research institutes and a nonprofit organization focused on GLSL seaway research. The digital twin enhances the capabilities of such institutions, presenting an innovative platform that bridges advanced analytics

with real-time insights, ultimately revolutionizing the scope and impact of economic research in the GLSL region.

Businesses

The private sector stands to gain significantly from the digital twin platform, which serves as a comprehensive repository of trade and transportation data for the GLSL region. Businesses and consulting firms can leverage this resource for in-depth research and strategic planning. The platform's advanced capabilities, including the visualization of global supply chains and the analysis of industrial clusters and economic complexity, offer business leaders a dynamic and actionable perspective on trends within their industries.

By enabling the visualization of the full extent of industry and product supply chains, broken down by value added at each geographic stage, the digital twin empowers analysts to identify inefficiencies, opportunities, and strategic points of intervention. This granular insight allows businesses to explore and optimize alternative configurations of their supply chains, tailoring their strategies to maximize efficiency and competitiveness. By integrating geographically dispersed economic activities into a coherent framework, the digital twin provides a powerful tool for informed decision-making, enhancing the adaptability and resilience of private sector stakeholders in an increasingly interconnected and dynamic global economy.

5. Conclusion

The development of the digital twin represents a transformative advancement in understanding the competitiveness of the GLSL region through the innovative application of digital twin technology. By replicating economic activity in a way that mirrors real-world interactions, the digital twin reveals micro-level outcomes of macro-level phenomena. By integrating diverse strands of economic theory and practice—specifically global value chains, economic clusters, economic complexity, and the gravity model—into a cohesive analytical framework, this initiative transcends traditional methods to provide a comprehensive perspective on trade and transportation within this pivotal bi-national economic corridor.

This multidisciplinary approach, powered by cutting-edge data science, facilitates an unprecedented cross-disciplinary understanding of the geospatial and organizational behavior of firms and industries. This methodological innovation bridges gaps between fragmented economic data sources while pioneering new ways to visualize firm-to-firm interactions. By doing so, it enriches the academic landscape and sets a new benchmark for the dynamic representation and analysis of economic activities.

The digital twin fosters an inductive research paradigm, leveraging the wealth of data available to uncover new trade patterns and economic relationships within the GLSL region. It challenges conventional notions of economic geography by transcending traditional political boundaries and employing machine learning techniques to develop theories and models that evolve in real-time, keeping pace with the complexity and dynamism of the global economy.

The impact of the digital twin extends far beyond academic circles, offering policymakers and business leaders a sophisticated analytical framework for navigating the complexities of today's economic landscape. With global supply chains increasingly threatened by geopolitical tensions and climate-related challenges, the need for such a robust tool has never been more critical. The digital twin provides a collaborative platform where stakeholders from various sectors and jurisdictions can engage in meaningful dialogue, share expertise, and craft strategies to mitigate risks and capitalize on opportunities effectively.

This initiative exemplifies the transformative potential of digital twin technology in economic analysis and strategic decision-making. It stands as a forward-looking tool, capable of adapting and responding to the dynamic forces shaping the global economy. Through this groundbreaking effort, the digital twin redefines the boundaries of economic research and practice, paving the way toward a digitally enhanced future.

Annex 2 : A sample of datasets integrated into our digital twin [December 2024][2]

Dataset	Description	Domain	Coverage	Time period	Unit	Frequency	Source
Trade in Embodied CO_2 ($TeCO_2$)	Estimates on embodied carbon in final demand emitted anywhere in the world along global production chains	Climate	Global	1995-2018	metric tons of CO_2	Annual	OECD
Canadian International Merchandise Trade Web Application	Trade between Canada (at the provincial level) and the rest of the world - Also includes trade within Canada	Economic	Global	1988-Present	Dollar values	Monthly	Statistics Canada
Inter-Country Input-Output (ICIO)	IOTs for 76 countries (and the rest of the world)	Economic	Global	1995-2020	Dollar values	Annual	OECD
TradeStats Express	Trade between US (at the state level) and the rest of the world	Economic	Global	2009-2023	Dollar values	Annual	U.S. Census Bureau
Registre des entreprises	Information on active and inactive enterprises in Québec	Economic	Québec	19XX-Present	Firm	Bi-weekly	Gouvernement du Québec
StateIO/stateior	IOTs for states in the US	Economic	US	2012-2020	Dollar values	Annual	U.S. Environmental Protection Agency
Freight Analysis Framework	Data on freight flows by commodity and mode of transport within the US	Transport	US	Estimates based on CFS 2017	Volume and dollar values	Annual	Bureau of Transportation Statistics
Canadian Freight Analysis Framework	Data on freight flows by commodity and mode of transport within Canada	Transport	Canada	2011-2017	Volume and dollar values	Annual	Statistics Canada
Grain Supply Chain Dashboard	Movement of grain by rail in Canada at the station and corridor level	Transport	Canada	2016-Present	Volume	Daily	Statistics Canada

[2] Find the latest list here.

Air Carriers: T-100 Segment (US Carriers Only)	Time taken by planes between O/D pairs	Transport	Canada/US	1990-Present	Time	Monthly	Bureau of Transportation Statistics
Border Crossing Entry Data	Number of trucks, trains and containers entering the US through border entry points with Canada	Transport	Canada/US	1996-Present	Volume	Daily	Bureau of Transportation Statistics
Great Lakes St. Lawrence Seaway System Intermodal Map	Major ports in the St. Lawrence - Great Lakes Seaway	Transport	Canada/US	NA	Port	NA	Great Lakes St. Lawrence Seaway System
Global Shipping Lanes	Shipping routes geo-referenced from CIA's Map of the World Oceans	Transport	Global	NA	NA	NA	Benden, P. (2022)
Spillover Simulator	Maritime capacity at risk of facing delays due to port disruptions affecting outgoing vessel movement	Transport	Global	2022 (base year)	Dollar values and time	NA	IMF PortWatch
Trade-and-Transport Dataset	Cost of transportation by mode and commodity between O/D pairs	Transport	Global	2016-2021	Dollar values	Annual	UNCTAD, World Bank
Interstate Truck Trips by Origin and Destination	Provides the annual number of interstate trips undertaken by freight carrying trucks in the US	Transport	US	2020-2022	Trips	Annual	Bureau of Transportation Statistics

References

Acemoglu, D., Carvalho, V. M., Ozdaglar, A., & Tahbaz-Salehi, A. (2012). The Network Origins of Aggregate Fluctuations. *Econometrica*, *80*(5), 1977–2016. https://doi.org/10.3982/ECTA9623

Amin, A. (2002). Spatialities of Globalisation. *Environment and Planning A: Economy and Space*, *34*(3), 385–399. https://doi.org/10.1068/a3439

Barbaglia, L., Consoli, S., Manzan, S., Reforgiato Recupero, D., Saisana, M., & Tiozzo Pezzoli, L. (2021). Data Science Technologies in Economics and Finance: A Gentle Walk-In. In S. Consoli, D. Reforgiato Recupero, & M. Saisana (Eds.), *Data Science for Economics and Finance: Methodologies and Applications* (pp. 1–17). Springer International Publishing. https://doi.org/10.1007/978-3-030-66891-4_1

Bernard, A. B., Jensen, J. B., Redding, S. J., & Schott, P. K. (2007). Firms in International Trade. *Journal of Economic Perspectives*, *21*(3), 105–130. https://doi.org/10.1257/jep.21.3.105

Coulombe, P. G., Leroux, M., Stevanovic, D., & Surprenant, S. (2022). How is machine learning useful for macroeconomic forecasting? *Journal of Applied Econometrics*, *37*(5), 920–964. https://doi.org/10.1002/jae.2910

Dudoit, A., Panot, M., & Warin, T. (2021). *Towards a multi-stakeholder Intermodal Trade-Transportation Data-Sharing and Knowledge Exchange Network* (2021RP-28). CIRANO. https://doi.org/10.54932/MVNE7282

Fisher, M., Phillips, J., & Shuster-Leibner, J. (2017, August 11). International trade and the Great Lakes-St. Lawrence region. *Policy Options*. https://policyoptions.irpp.org/magazines/august-2017/international-trade-great-lakes-st-lawrence-region/

Freund, C., Mattoo, A., Mulabdic, A., & Ruta, M. (2024). Is US trade policy reshaping global supply chains? *Journal of International Economics*, *152*, 104011. https://doi.org/10.1016/j.jinteco.2024.104011

Hausman, W. H., Lee, H. L., & Subramanian, U. (2013). The Impact of Logistics Performance on Trade. *Production and Operations Management*, *22*(2), 236–252. https://doi.org/10.1111/j.1937-5956.2011.01312.x

Kirman, A. (2016). Complexity and Economic Policy: A Paradigm Shift or a Change in Perspective? A Review Essay on David Colander and Roland Kupers's Complexity and the Art of Public Policy. *Journal of Economic Literature*, *54*(2), 534–572. https://doi.org/10.1257/jel.54.2.534

Lanz, R., & Miroudot, S. (2011). *Intra-Firm Trade*. *114*. https://doi.org/10.1787/5kg9p39lrwnn-en

Martin Associates. (2023). *Economic Impacts of Maritime Shipping in the Great Lakes-St. Lawrence Region*. Martin Associates. https://greatlakes-seaway.com/wp-content/uploads/2023/07/eco_impact_full_2023_en.pdf

National Academies of Sciences, Engineering, and Medicine. (2012). *Multimodal Freight Transportation Within the Great Lakes–Saint Lawrence Basin* (R. Stewart, Ed.). The National Academies Press. https://doi.org/10.17226/22742

Porter, M. E. (1990). The Competitive Advantage of Nations. *Harvard Business Review*, *68*(2), 73–93.

Porter, M. E. (1994). The Role of Location in Competition. *International Journal of the Economics of Business*, *1*(1), 35–40. https://doi.org/10.1080/758540496

Siddika, B. (2024, January 31). *Montreal: A city in Canada or a node in the world economy?* GVC.dtLab. https://gvcdtlab.com/en/montreal-a-city-in-canada-or-a-node-in-the-world-economy/

Stevanovic, D. (2021). Prévision macroéconomique dans l'ère des données massives et de l'apprentissage automatique. In *Le Québec Économique 9* (pp. 315–351). CIRANO. https://cirano.qc.ca/files/publications/2020CH-12.pdf

Warin, T. (2023). Leveraging Digital Twins to Attain a Competitive Edge. *California Management Review*. May 8, https://cmr.berkeley.edu/2023/05/leveraging-digital-twins-to-attain-a-competitive-edge/

Warin, T., de Marcellis-Warin, N., Troadec, A., Sanger, W., & Nembot, B. (2014). *Un état des lieux sur les données massives* (2014RB-01). CIRANO. https://cirano.qc.ca/files/publications/2014RB-01.pdf