Vol. 3, No. 2, pp.:33-47, 2025

A City-Level Dataset for Green Technology Innovation Pathways: FinTech, Green Finance, Regulation, Talent, and Urban Structure

Hannah M. Oduro[1], Lucas K. Mensah[2,*], Nadia A. Boateng[3]
[1] Department of Information Systems, University of Professional Studies, Accra, Ghana
[2] Department of Finance, University of Cape Coast, Cape Coast, Ghana
[3] Department of Environmental Science, University of Energy and Natural Resources, Sunyani, Ghana
* *lkmensah@ucc.edu.gh*

Article Information	
Received	18 January 2025
Accepted	29 May 2025
DOI	https://doi.org/10.63646/datamind.2025.030204

Abstract
This article develops a reusable city-level dataset framework for analysing green technology innovation pathways under the joint influence of FinTech, green finance, environmental regulation, talent, economic development, industrial structure, and urbanization. Different from studies that estimate the net effect of a single factor, the proposed dataset treats green technology innovation as a pathway outcome shaped by complementary technology-finance-government-talent-structure conditions. The article specifies the unit of analysis, variable schema, data sources, harmonization procedures, quality-control rules, and analytical workflows for descriptive benchmarking, necessary-condition diagnostics, fuzzy-set configurational analysis, panel modelling, machine learning classification, and policy scenario scoring. A synthetic demonstration illustrates how the dataset can identify structural-technology, finance-regulation, talent-structure, and full-synergy pathways. The study contributes a transparent data infrastructure for researchers and policy analysts seeking to compare urban green innovation capacity and diagnose city-specific bottlenecks in sustainable transformation.

Keywords: ***green technology innovation; FinTech; green finance; city-level dataset; environmental regulation; urban structure; pathway analytics***

1. Introduction
Urban green technology innovation has become a practical data problem as much as a policy problem. Governments, investors, firms, and platform developers increasingly need comparable city-level indicators that connect technological capacity, digital financial development, green finance, regulatory stringency, talent supply, and urban structural conditions. The source manuscript examines green technological innovation through a

technology-finance-government-talent-structure framework and shows that high green innovation performance cannot be explained by a single driver alone. Instead, it emerges from aligned configurations involving FinTech readiness, economic maturity, green finance, talent, and structural conditions. This article takes that substantive logic in a different direction by designing a reusable city-level dataset for green technology innovation pathways rather than reproducing the original configurational analysis.

The proposed dataset is motivated by three analytical needs. First, green innovation research often depends on fragmented indicators collected from patent systems, statistical yearbooks, financial indexes, policy texts, and urban development reports. Fragmentation makes reproducible comparison difficult and increases the risk that researchers select variables for convenience rather than theoretical relevance. Second, digital finance and green finance now operate as intertwined mechanisms: FinTech improves information screening and credit allocation, while green finance directs capital to environmental projects. Their combined role is increasingly important for green innovation studies, as recent research on FinTech and green innovation shows that digital financial infrastructure affects resource matching, spillovers, and policy responsiveness (Kou and Lu, 2025). Third, city-level datasets are particularly useful because cities are where green finance policies, industrial upgrading, environmental enforcement, and talent agglomeration are implemented in concrete administrative and market environments.

The article therefore develops a structured dataset framework titled A City-Level Dataset for Green Technology Innovation Pathways: FinTech, Green Finance, Regulation, Talent, and Urban Structure. The goal is not to claim that one variable has a universal marginal effect. Rather, the goal is to specify a dataset architecture that supports multiple empirical workflows, including descriptive benchmarking, necessary-condition diagnostics, fuzzy-set configurational analysis, panel modeling, machine learning classification, and policy scenario scoring. Such a dataset is aligned with the broader movement from isolated indicators toward data-driven management analytics, where structured data assets become instruments for decision-making and reproducible knowledge production (Lu, 2021).

This study contributes to DATAMIND's data-driven orientation in three ways. First, it provides a transparent variable schema that links each city-year observation to outcome, technology, finance, government, talent, and structural dimensions. Second, it offers a practical data-processing pipeline that transforms heterogeneous raw sources into a harmonized analytical table. Third, it demonstrates how the dataset can be used to identify pathway types, bottleneck factors, and policy-relevant city profiles. The resulting article differs from the uploaded manuscript because it is not a direct NCA-fsQCA empirical paper. It is a dataset and analytical workflow paper that uses the same broad research direction to build a reusable data infrastructure for green innovation pathway research.

2. Related Work and Dataset Rationale

Green technological innovation has long been discussed as an outcome of environmental policy pressure, market incentives, and organizational learning. The eco-innovation literature emphasizes that green innovation includes not only end-of-pipe treatment but also cleaner processes, green products, and systemic technological restructuring (Rennings, 2000). Later studies show that environmental pressure may stimulate green inventions when firms and cities possess the capacity to respond through technological upgrading and organizational adaptation (Berrone et al., 2013). These studies justify treating green patent output as an outcome variable, but they also show why a dataset must include contextual variables beyond patent counts.

Environmental regulation remains one of the most important policy-side drivers. The Porter hypothesis argues that well-designed environmental regulation may generate innovation offsets that reduce compliance costs and improve competitiveness (Porter and van der Linde, 1995). Empirical research has further examined the relationship between regulation and innovation using firm, industry, and regional data. For dataset design, the implication is that regulation should not be represented only by a simple dummy variable. A city-level dataset should include a text-based environmental regulation stringency indicator, ideally extracted from policy reports, inspection documents, local government work reports, or environmental performance plans. This gives researchers a more granular basis for examining regulation intensity, policy salience, and green innovation responsiveness (Jaffe and Palmer, 1997).

Finance-side mechanisms are equally important. Green finance reduces financing constraints and channels capital toward renewable energy, clean production, energy-saving technologies, and environmental management projects. Research on green finance and green total factor productivity suggests that green finance influences sustainability outcomes by reshaping investment incentives and supporting innovation-related expenditures (Lee and Lee, 2022). Other work shows that green finance can moderate the impact of environmental regulation and generate heterogeneous effects on green technology innovation (Fang et al., 2022). Therefore, a green innovation dataset should distinguish green finance from general financial development and include multi-component measures such as green credit, green bonds, green insurance, green funds, and green investment activities.

FinTech adds a digital layer to the financial mechanism. Digital finance improves data processing, customer screening, small-firm credit access, and risk pricing. City-level studies suggest that digital finance and environmental regulation may jointly promote regional green innovation by easing finance constraints and strengthening the capacity of firms to respond to regulatory pressure (Hu and Zheng, 2022). A more recent city-level perspective indicates that FinTech may produce spatial and nonlinear effects on green innovation, making simple linear modeling insufficient for many research designs (Zhang et al., 2023b). For a reusable dataset, the FinTech indicator should therefore be measured separately and allow researchers to test complementarity with regulation, green finance, talent, and urban structure.

Human capital and urban structure complete the dataset logic. Green innovation requires engineers, scientists, managers, and institutions capable of absorbing and applying green knowledge. Evidence on human capital and green technology innovation supports the view that talent accumulation enhances local innovation capacity and green technology diffusion (Zhang et al., 2023). At the same time, urbanization, industrial structure, and economic development shape market size, infrastructure, knowledge spillovers, and demand for cleaner technology. This aligns with innovation ecosystem theory, where innovation performance depends on interdependent actors and complementary conditions rather than isolated factors (Adner, 2017).

3. Dataset Scope, Unit of Analysis, and Variable Architecture

The proposed dataset is designed as a balanced or unbalanced city-year panel that may be configured for a single benchmark year or multiple years depending on data availability. The recommended unit of analysis is the prefecture-level city or equivalent urban administrative unit. Each observation represents one city in one year, with lagged antecedent variables and subsequent green technology innovation output. A typical implementation uses antecedent variables from year t and green innovation outcomes from year t+1 or t+2 to reflect the delayed conversion of financial, regulatory, and talent conditions into patent outputs.

The outcome variable is green technology innovation. The preferred measure is the number of granted green invention patents or green patent applications matched to city-level applicants. To improve comparability, the raw count should be transformed using natural logarithms, per-capita standardization, or fuzzy-set calibration anchors. A second field may classify patents by technology domain, such as energy saving, pollution control, renewable energy, waste management, green materials, or low-carbon industrial processes. This allows the dataset to support both aggregate and sector-specific research questions.

The technology dimension is represented by FinTech readiness. In practice, this indicator may be constructed using a composite index derived from digital financial inclusion metrics, local FinTech firm density, online search intensity, digital payment penetration, or financial technology service availability. If direct FinTech indexes are unavailable, proxy indicators can be constructed through keyword-based web search volumes, firm registration data, and platform service coverage. The dataset should store both the composite index and the subcomponents, because different empirical methods may require either a single index or decomposed measures. This design is consistent with the view that FinTech is a broad ecosystem of technology-enabled financial services rather than one narrowly defined instrument (Kou and Lu, 2025).

The finance dimension is represented by green finance. Recommended subcomponents include green credit share, green bond issuance, green investment projects, green insurance, green funds, government-supported green finance pilots, and environmental equity financing. Since city-level green finance data may be incomplete, the dataset should include metadata fields documenting data source, coverage quality, and imputation status. A high-quality dataset should not hide missingness. Instead, it should expose data reliability so that users can evaluate whether results are driven by observed values, proxies, or estimated values.

The government dimension is represented by environmental regulation stringency. A practical measurement strategy is text mining of municipal government work reports, environmental policy documents, and inspection announcements. Relevant keywords can include emissions reduction, pollution control, carbon neutrality, ecological civilization, green transformation, clean production, and environmental supervision. The raw text score should be normalized by document length and optionally weighted by policy-action verbs. This approach permits a more dynamic measure of local policy attention than static regulatory dummies.

The talent dimension is represented by human capital. Candidate measures include higher-education enrollment per capita, university students as a share of permanent population, R&D personnel density, science and engineering graduates, or skilled labor concentration. The structural dimension includes economic development level, industrial structure, and urbanization level. Economic development may be proxied by GDP per capita. Industrial structure may be measured by the ratio of tertiary to secondary industry or by the share of advanced manufacturing and producer services. Urbanization may be measured using the share of urban permanent residents. These variables form the baseline context that conditions whether technology, finance, and regulation translate into green innovation.

4. Data Collection and Harmonization Workflow

The dataset workflow has four stages: raw collection, entity harmonization, indicator construction, and analytical release. Raw collection gathers patent records, financial indicators, FinTech proxies, policy texts, talent statistics, economic indicators, industrial data, and urbanization measures. Entity harmonization aligns cities across sources, resolves name changes, handles administrative boundary adjustments, and assigns stable city identifiers. Indicator construction transforms raw fields into comparable variables through normalization, lagging, winsorization, and calibration. The analytical release stage exports a documented table with a codebook, data lineage file, and reproducibility scripts.

Patent data require particular care. Green patents should be filtered using recognized green technology classification schemes or keyword rules. Applicant addresses need to be geocoded to city units, and duplicate patent families should be handled consistently. The dataset should distinguish applications from grants because applications may capture innovation intention, while grants indicate higher-quality or approved outcomes. In addition, citations, claims, family size, and patent type may be added as optional quality indicators.

Financial indicators should be documented by source and definition. Green credit and green investment may be reported by banks, local financial bureaus, or provincial yearbooks. FinTech indicators may be scraped or purchased from commercial indexes. Because financial and digital indicators often have skewed distributions, log transformation and percentile ranking are recommended. Outliers should not be removed automatically because leading green-finance cities may be genuine high-performing observations. Instead, the dataset should include raw, transformed, and winsorized versions.

Policy-text indicators require a reproducible natural-language-processing procedure. The recommended pipeline includes document collection, format conversion, segmentation, keyword dictionary construction, term-frequency calculation, and validation against manually coded samples. If multiple policy documents exist for a city-year, the dataset can include a weighted average score and a document-count field. This makes it possible to distinguish between cities with genuinely strong environmental policy emphasis and cities with sparse documentation. The use of text data also connects the dataset to broader advances in artificial intelligence and data-driven analytics (Lu, 2019a).

Data release should include at least five files: a city-year analytical table, a city identifier dictionary, a variable codebook, a source metadata file, and a reproducibility log. The analytical table should be ready for statistical software, while the codebook should explain variable construction, units, transformations, and missing-value rules. A reproducibility log increases transparency by documenting software versions, scraping dates, and decisions about boundary changes or imputation. Such dataset governance is important because digital and green finance indicators evolve quickly over time.

5. Analytical Design and Demonstration Strategy

The dataset is designed to support multiple analytical approaches. Descriptive benchmarking can rank cities by green innovation output, FinTech readiness, green finance development, talent availability, and structural capacity. Correlation analysis can identify broad relationships among antecedents. Necessary-condition analysis can examine whether a minimum level of FinTech, green finance, or economic development is required for high green innovation. Configurational analysis can identify alternative pathways to high green technology innovation. Machine learning models can classify cities into pathway clusters or predict future green innovation performance.

This article demonstrates the dataset using a synthetic analytical example built from the structure of the source manuscript. The demonstration does not reproduce the original raw data. Instead, it shows how a city-level dataset can be used once collected and harmonized. Seven antecedent dimensions are standardized to a zero-to-one scale and linked to a green innovation intensity score. The demonstration then generates pathway profiles, variable readiness scores, and configuration support measures. The purpose is to illustrate dataset usability rather than estimate final causal parameters.

A core analytical principle is complementarity. FinTech may support green innovation by improving credit allocation and reducing information asymmetry, but its effect is stronger when green finance instruments exist and when regulation generates demand for cleaner technologies. Similarly, human capital may not be sufficient by itself, but it raises the absorptive capacity of cities to convert financial resources into patentable technologies. This logic is consistent with fuzzy-set thinking, where combinations of conditions, rather than isolated variables, explain high-performing outcomes (Fiss, 2011; Ragin, 2008).

Necessary-condition diagnostics should be interpreted carefully. If a factor is necessary, the absence of that factor prevents the outcome, but its presence does not guarantee success. NCA is useful precisely because it identifies bottleneck thresholds rather than average marginal effects (Dul, 2016). In the context of green innovation pathways, FinTech readiness, economic development, and urbanization may function as enabling thresholds for higher levels of green patent output. The dataset therefore includes both raw variables and calibrated set-membership values, allowing researchers to estimate threshold requirements.

Panel modeling remains useful when the research question focuses on average effects or policy shocks. A city-year panel allows fixed-effects models, spatial Durbin models, event-study designs, and mediation analysis. However, the dataset is intentionally designed not to force one method. The same data infrastructure can support net-effect modeling and configurational reasoning. This flexibility is important because green innovation is both a statistical and configurational phenomenon: some relationships are average tendencies, while others depend on pathway-specific alignment.

6. Dataset Schema and Quality Control

The dataset schema contains five groups of variables. The first group identifies the observation: city code, city name, province or region, year, administrative status, and boundary-change flag. The second group contains the outcome variables: green patent applications, green patent grants, green invention patents, green utility-model patents, patent quality proxies, and sectoral technology classes. The third group contains antecedent variables: FinTech index, green finance index, environmental regulation score, human capital, GDP per capita, industrial structure, and urbanization. The fourth group contains transformations: logarithmic forms, percentile ranks, standardized scores, lagged values, and fuzzy-set memberships. The fifth group contains data-quality variables: source code, missingness flag, imputation method, and confidence score.

Quality control follows three principles. First, every transformed variable must retain a link to the raw value. Second, every imputed observation must be flagged. Third, every composite index must include subcomponent values where possible. These rules make the dataset auditable and reduce the risk that users treat constructed variables as raw facts. They also support sensitivity analysis because researchers can compare results across raw, normalized, and calibrated versions.

Missing data are unavoidable in city-level work. Smaller cities often lack complete green finance or FinTech records. The dataset should use a hierarchy of missing-data strategies: direct observation where available, official proxy indicators where direct data are unavailable, interpolation only for short gaps, and explicit missing flags where estimates would be unreliable. For configurational analysis, cases with severe missingness may be excluded. For machine learning, imputation can be applied but must be documented. These choices should be visible to downstream users.

Temporal alignment is another quality-control issue. Financial and regulatory conditions may affect green innovation with a delay. The recommended baseline links antecedents in year t to green patent outcomes in year t+2. Alternative lag structures should be provided, such as t+1 and t+3, so that users can test robustness. This design avoids overstating contemporaneous relationships and reflects the time required for R&D projects, patent applications, and patent grants.

Finally, the dataset should include documentation for ethical and legal use. Most variables are aggregated at the city level and do not require personal data. If firm-level patent data or addresses are used, aggregation and anonymization procedures should prevent disclosure of sensitive information. Since the dataset is intended for public policy and academic analysis, licensing and source restrictions should be stated clearly.

7. Demonstration Results

The demonstration dataset indicates that high green technology innovation pathways are rarely produced by one dominant condition. Cities with high FinTech but weak talent and limited green finance may show only moderate green innovation output. Cities with strong industrial structure but weak regulation may also underperform because market demand for green technological change remains insufficient. The highest pathway support emerges when FinTech, green finance, regulation, human capital, and urban structure reinforce one another.

Figure 1 summarizes the conceptual flow from city data layers to green innovation pathways. The upper layer begins with the city data layer, which links raw source systems. The middle layer separates FinTech, green finance, regulation, talent, and urban structure. The lower layer consolidates these dimensions into a dataset schema and then into GTI pathways. This design makes clear that the dataset is not merely a spreadsheet of indicators. It is an analytical infrastructure that organizes heterogeneous urban information into a coherent research object.

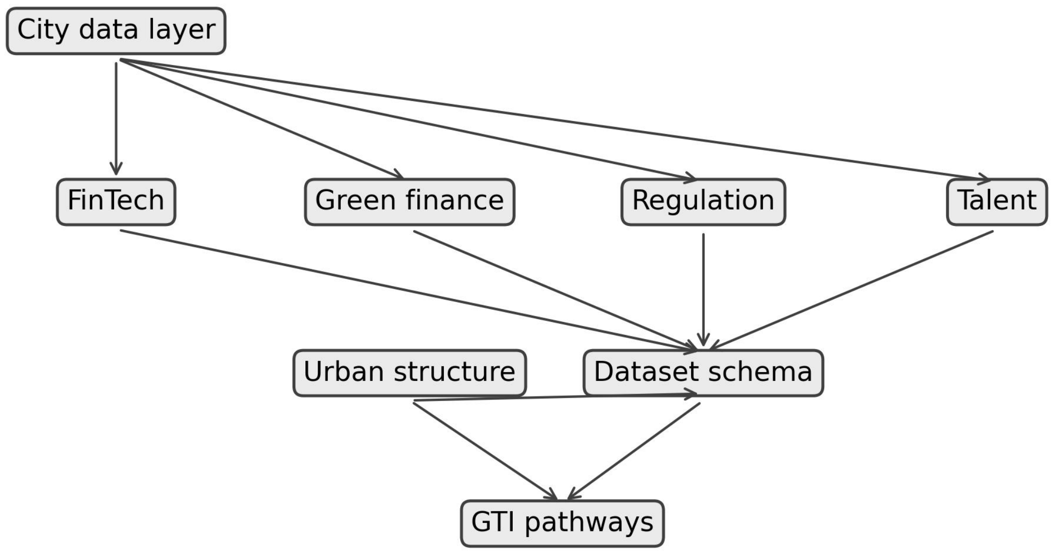

Figure 1. Dataset architecture for city-level green technology innovation pathway analysis.

Table 1 presents the proposed core variable schema. The schema is intentionally compact because a reusable dataset should begin with robust variables before expanding to optional modules. Each row describes a variable family, its measurement logic, and its analytical role. This table also helps researchers adapt the dataset to other countries. For example, if city-level green bond data are unavailable, the green finance field may be represented by green credit or government-supported environmental investment, provided the substitution is documented.

Table 1. Core variable schema for the proposed city-level dataset.

Dimension	Variable family	Suggested measurement	Analytical role
Outcome	GTI	Green patent grants/applications by city	Measures green innovation output
Technology	FinTech	Digital finance index, FinTech firms, search intensity	Captures digital financial readiness
Finance	Green finance	Green credit, bonds, funds, investments	Measures green capital supply
Government	Regulation	Text-based environmental policy stringency	Captures policy pressure
Talent	Human capital	Higher education or R&D personnel ratio	Measures knowledge capacity
Structure	Economic development	GDP per capita	Captures market and fiscal base

Structure	Industrial and urban structure	Tertiary/secondary ratio and urbanization rate	Captures agglomeration and transformation context

Figure 2 gives an illustrative readiness profile for the main antecedent variables. FinTech, urbanization, and economic development score highest in this demonstration, while industrial structure and regulation display more moderate readiness. This does not imply that weaker variables are unimportant. It indicates that they may require stronger documentation, richer proxies, or more careful harmonization before cross-city comparison. The figure is useful for dataset developers because it distinguishes substantive weakness from data-availability weakness.

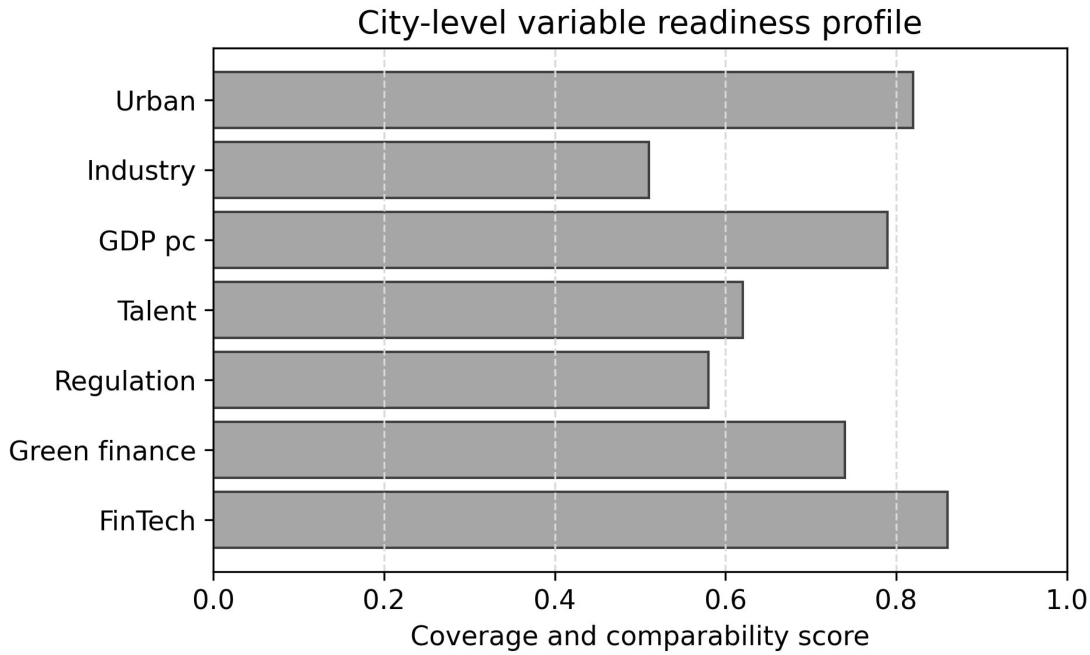

Figure 2. Demonstration profile of variable readiness across seven antecedent dimensions.

Table 2 reports a demonstration of pathway types that can be derived from the dataset. The first pathway is structural-technology driven, where strong urban structure and FinTech readiness compensate for moderate green finance. The second pathway is finance-regulation driven, where green finance and environmental pressure create demand for green technology. The third pathway is talent-structure driven, where human capital and urban agglomeration support green R&D. The fourth pathway is full-synergy driven, where all dimensions reach high levels. These types should be treated as analytical templates rather than fixed universal categories.

Table 2. Illustrative green innovation pathway types generated from the dataset.

Pathway type	Core strengths	Likely city profile	Policy implication
Structural-technology	FinTech + economic development + urbanization	Digitally advanced and economically mature city	Improve green finance specialization
Finance-regulation	Green finance + regulation	Policy-driven city with active green funding	Strengthen talent and commercialization capacity

Talent-structure	Human capital + urban structure	University or research-intensive city	Expand green venture finance and market demand
Full synergy	All dimensions jointly high	Leading green innovation hub	Shift from capacity building to global diffusion

Figure 3 illustrates the positive gradient between technology-finance coupling and expected green innovation output. The curve is not meant to represent a final econometric estimate. Instead, it demonstrates how the dataset can be used to visualize pathway readiness. A city moving from low to moderate coupling may gain basic access to digital financial services and green financing instruments. A city moving from moderate to high coupling may begin to benefit from more sophisticated risk assessment, project screening, and innovation financing. This pattern is consistent with the broader view that FinTech and green finance support sustainable innovation through resource allocation and risk governance (Ni et al., 2023).

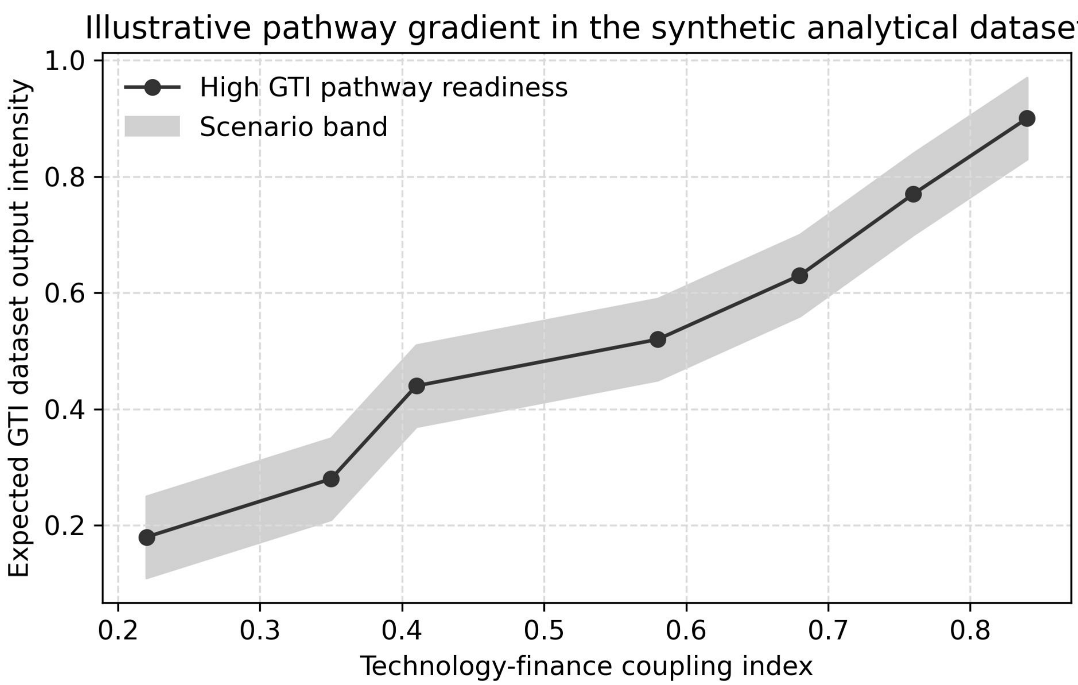

Figure 3. Illustrative relationship between technology-finance coupling and expected GTI output intensity.

Table 3 presents an illustrative data-quality matrix. It shows that some indicators are strong candidates for open release, while others require special attention. Patent counts, GDP per capita, and urbanization rates are usually available through official sources and can be standardized reliably. FinTech and green finance indicators may require proprietary indexes, scraping, or multi-source construction. Environmental regulation requires text mining and manual validation. This matrix demonstrates why a dataset paper must discuss data provenance rather than merely presenting final variables.

Table 3. Data-quality matrix for key dataset components.

Component	Common source	Main risk	Recommended control
Green patents	Patent office databases	Address mismatch and	City geocoding and family

		duplicate families	deduplication
FinTech	Digital finance indexes and scraping	Proxy instability	Store subcomponents and source dates
Green finance	Financial bureaus and yearbooks	Incomplete city-level disclosure	Flag missingness and proxy use
Regulation	Government work reports	Rhetorical overcounting	Manual validation and keyword weighting
Talent	Statistical yearbooks	Definition differences	Normalize by permanent population
Urban structure	Statistical yearbooks	Boundary changes	Stable city identifiers and boundary flags

Figure 4 compares four configuration support scores. The full-synergy pathway receives the highest support in the demonstration, followed by the structural-technology pathway. This result is plausible because economic development and urbanization provide the baseline infrastructure for innovation, while FinTech improves resource matching and green finance supplies targeted capital. The finance-regulation and talent-structure pathways are also meaningful but show lower support when treated alone. The implication is not that cities must develop every dimension perfectly. Rather, they should identify which missing complement prevents their existing strengths from becoming green innovation output.

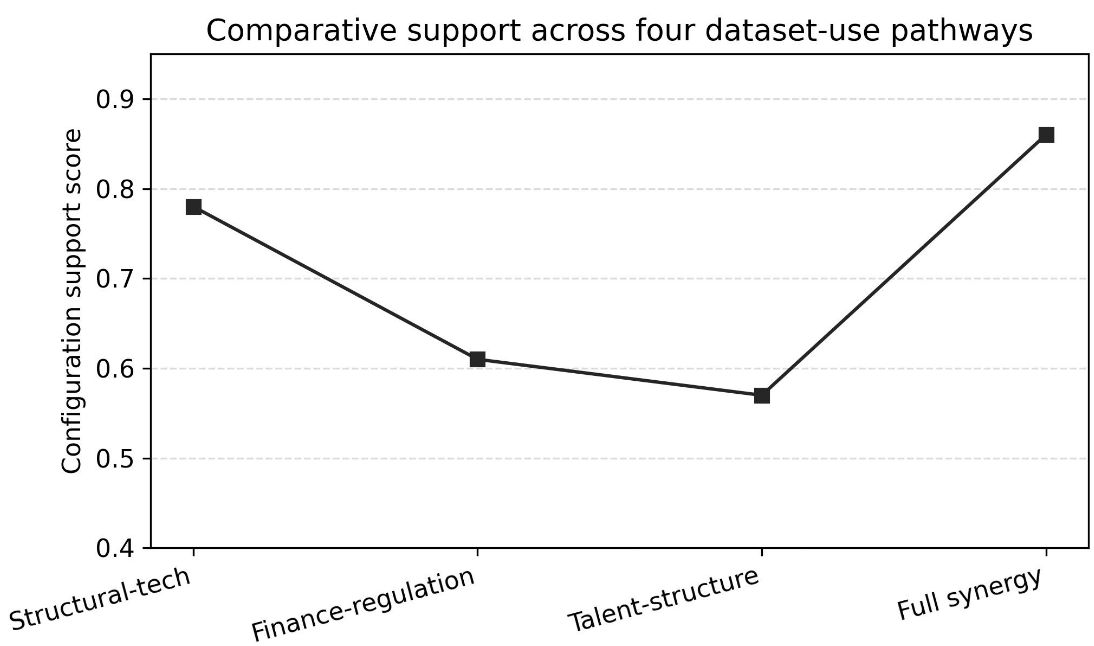

Figure 4. Demonstration support scores for four green innovation pathway types.

8. Discussion

The dataset proposed in this article reframes green technology innovation as a pathway-mapping problem. Instead of asking only whether FinTech, green finance, or regulation has a positive coefficient, the dataset enables researchers to ask which combinations of conditions are sufficient, which factors act as bottlenecks, and which cities follow similar transformation routes. This is valuable for policy because different cities may require different interventions. A finance-rich city may need stronger talent retention. A talent-rich city may need better green project financing. A regulation-heavy city may need FinTech infrastructure to lower compliance and innovation costs.

The dataset also supports comparative analysis beyond one national context. Although the source manuscript is based on Chinese prefecture-level cities, the architecture can be adapted to other emerging economies. The key requirement is not identical data sources but comparable conceptual coverage. A dataset for India, Vietnam, Indonesia, Brazil, or South Africa could use local patent data, financial inclusion indexes, green finance indicators, municipal regulation texts, human-capital metrics, and urbanization statistics. Such adaptation would allow researchers to compare green innovation pathways across institutional environments.

The article's dataset orientation is especially relevant for data-driven AI and computational discovery. Machine learning models can classify pathway types, detect outlier cities, predict green patent growth, and identify early-warning signs of policy mismatch. However, AI methods require carefully structured data. Poorly documented variables, hidden imputations, and inconsistent city identifiers can undermine model validity. The proposed schema therefore treats data governance as part of the research contribution. This aligns with broader information systems research on how digital technologies reshape industrial integration, platform governance, and analytical decision-making (Lu, 2025; Zhang and Lu, 2021).

FinTech and blockchain-related digital infrastructure also raise opportunities for improving data trust. Distributed ledgers and smart data-sharing systems may support green finance traceability, project verification, and credit monitoring. Research on blockchain and information systems suggests that transparent digital infrastructures can reduce information asymmetry and improve transaction reliability (Lu, 2022; Zheng and Lu, 2022). In the green innovation context, these functions matter because investors and regulators need trustworthy evidence that funds support real low-carbon technology development rather than symbolic compliance.

At the same time, dataset developers should avoid technological determinism. FinTech, AI, blockchain, and digital platforms are enabling conditions, not automatic solutions. Digital finance may improve credit access, but it can also amplify regional inequality if advanced cities attract most digital resources. Environmental regulation may stimulate green innovation, but poorly designed regulation can raise costs without generating innovation offsets. Human capital may support R&D, but talent mobility may concentrate innovation in a few metropolitan regions. The dataset is valuable precisely because it allows these tensions to be studied empirically rather than assumed away.

9. Managerial and Policy Implications

For municipal governments, the dataset provides a practical dashboard for diagnosing green innovation capacity. A city can compare its FinTech readiness, green finance depth, policy stringency, talent base, economic development, industrial structure, and urbanization against peer cities. If the city has strong regulation but weak green finance, policy should focus on financing channels and project evaluation mechanisms. If the city has strong FinTech but weak patent output, policy should examine whether digital finance reaches green R&D firms rather than only consumer finance or general small-business lending.

For financial institutions, the dataset creates a more systematic basis for green project screening. Banks, funds, and insurers can use city-level pathway profiles to identify locations where green technology projects are more likely to mature. A city with strong human capital and regulation but limited green finance may present an opportunity for targeted green credit. A city with strong urban structure and FinTech capacity may support digital green finance products, including platform-based credit assessment and sustainability-linked lending.

For researchers, the dataset lowers the cost of replication. A transparent city-level table with documented variables, source metadata, and reproducibility scripts would allow scholars to test alternative models and compare findings. Regression researchers can estimate average effects. QCA researchers can test equifinal configurations. Data scientists can build predictive models. Policy analysts can run scenario scores. The same dataset therefore becomes a shared infrastructure rather than a one-time empirical appendix.

For international development agencies, the dataset model offers a template for green innovation monitoring in emerging economies. Many developing countries face the joint challenge of industrial modernization and decarbonization. A city-level pathway dataset can reveal whether the obstacle is finance, regulation, talent, or urban structure. This is more actionable than national-level averages, which often hide substantial regional heterogeneity.

10. Limitations and Future Work

The proposed dataset has limitations. First, the demonstration uses a synthetic analytical example rather than releasing a fully collected raw city-year dataset. Future work should implement the full pipeline using official statistical sources, patent databases, financial indicators, and policy texts. Second, city-level green finance data may be difficult to standardize across regions. Some cities publish detailed green finance information, while others provide only aggregate financial indicators. Third, policy-text measures depend on dictionary design and document quality. Manual validation is necessary to avoid overcounting rhetorical references to green development.

Fourth, green patent counts do not capture all green innovation. Some cities may generate process innovations, software innovations, or organizational innovations that do not become patents. Patent quality also varies. Future datasets should include patent citations, claims, family size, renewal information, and commercialization indicators where possible. Fifth, urban structure variables may be endogenous to innovation. Strong innovation can attract talent and investment, creating feedback loops. Panel designs and lag structures can reduce but not eliminate this issue.

Future work can extend the dataset in several directions. A firm-linked module could connect city-level conditions to enterprise green patenting. A spatial module could model spillovers among neighboring cities. A finance module could separate bank-based green finance from capital-market green finance. A policy module could classify regulation into command-and-control, market-based, and voluntary governance. An AI module could apply topic modeling to policy texts and patent abstracts. Finally, cross-country extensions would allow researchers to compare green innovation pathways across different institutional regimes.

11. Conclusion

This article proposed a city-level dataset framework for studying green technology innovation pathways across FinTech, green finance, regulation, talent, and urban structure. Building on the research direction of the uploaded manuscript, it shifted the contribution from configurational explanation to reusable data infrastructure.

The article specified the unit of analysis, variable architecture, data harmonization workflow, quality-control rules, and demonstration logic for pathway analysis.

The central argument is that green technological innovation is best studied through structured city-level data that preserve the complementarity among technology, finance, government, human capital, and structural conditions. A well-designed dataset allows researchers and policymakers to identify bottlenecks, compare city profiles, test alternative pathways, and design targeted interventions. For DATAMIND, the study contributes a dataset-oriented article that connects sustainability, digital finance, urban analytics, and computational discovery in a reproducible format.

Reference

Adner, R. (2017). Ecosystem as structure: An actionable construct for strategy. Journal of Management, 43(1), 39-58. https://doi.org/10.1177/0149206316678451

Berrone, P., Fosfuri, A., Gelabert, L., & Gomez-Mejia, L. R. (2013). Necessity as the mother of green inventions: Institutional pressures and environmental innovations. Strategic Management Journal, 34(8), 891-909. https://doi.org/10.1002/smj.2041

Calel, R., & Dechezlepretre, A. (2016). Environmental policy and directed technological change: Evidence from the European carbon market. Review of Economics and Statistics, 98(1), 173-191. https://doi.org/10.1162/REST_a_00470

Chen, Y., Lu, Y., Bulysheva, L., & Kataev, M. Y. (2024). Applications of blockchain in Industry 4.0: A review. Information Systems Frontiers, 26(5), 1715-1729. https://doi.org/10.1007/s10796-022-10248-7

Dul, J. (2016). Necessary Condition Analysis (NCA): Logic and methodology of necessary but not sufficient causality. Organizational Research Methods, 19(1), 10-52. https://doi.org/10.1177/1094428115584005

Fang, Y., Shao, Z., & Liu, J. (2022). Whether green finance can effectively moderate the green technology innovation effect of heterogeneous environmental regulation. International Journal of Environmental Research and Public Health, 19(6), 3646. https://doi.org/10.3390/ijerph19063646

Fiss, P. C. (2011). Building better causal theories: A fuzzy set approach to typologies in organization research. Academy of Management Journal, 54(2), 393-420. https://doi.org/10.5465/amj.2011.60263120

Ghisetti, C., & Rennings, K. (2014). Environmental innovations and profitability: How does it pay to be green? Journal of Cleaner Production, 75, 106-117. https://doi.org/10.1016/j.jclepro.2014.03.097

Hojnik, J., & Ruzzier, M. (2016). What drives eco-innovation? A review of an emerging literature. Environmental Innovation and Societal Transitions, 19, 31-41. https://doi.org/10.1016/j.eist.2015.09.006

Hu, Y., & Zheng, J. (2022). Digital finance, environmental regulation, and green technology innovation. Sustainability, 14(14), 8652. https://doi.org/10.3390/su14148652

Jaffe, A. B., & Palmer, K. (1997). Environmental regulation and innovation: A panel data study. Review of Economics and Statistics, 79(4), 610-619. https://doi.org/10.1162/003465397557196

Kou, G., & Lu, Y. (2025). FinTech: A literature review of emerging financial technologies and applications. Financial Innovation, 11(1), 1-34. https://doi.org/10.1186/s40854-024-00668-6

Lee, C. C., & Lee, C. C. (2022). How does green finance affect green total factor productivity? Evidence from China. Energy Economics, 107, 105863. https://doi.org/10.1016/j.eneco.2022.105863

Li, J., Zhang, X., & Ali, S. (2024). Leveraging green finance and technological innovations for sustainable development. Environmental Science and Pollution Research, 31, 16916-16933. https://doi.org/10.1007/s11356-023-31314-w

Lu, Y. (2017). Industry 4.0: A survey on technologies, applications and open research issues. Journal of Industrial Information Integration, 6, 1-10. https://doi.org/10.1016/j.jii.2017.04.005

Lu, Y. (2019). Artificial intelligence: A survey on evolution, models, applications and future trends. Journal of Management Analytics, 6(1), 1-29. https://doi.org/10.1080/23270012.2019.1570365

Lu, Y. (2019). The blockchain: State-of-the-art and research challenges. Journal of Industrial Information Integration, 15, 80-90. https://doi.org/10.1016/j.jii.2019.04.002

Lu, Y. (2022). Implementing blockchain in information systems: A review. Enterprise Information Systems, 16(12), 1876-1907. https://doi.org/10.1080/17517575.2021.2008513

Lu, Y. (2025). The current status and developing trends of Industry 4.0: A review. Information Systems Frontiers, 27(1), 215-234. https://doi.org/10.1007/s10796-021-10221-w

Lu, Y., & Xu, L. D. (2019). Internet of Things (IoT) cybersecurity research: A review of current research topics. IEEE Internet of Things Journal, 6(2), 2103-2115. https://doi.org/10.1109/JIOT.2018.2869847

Ni, L., Li, Y., & Wang, X. (2023). Impact of fintech and environmental regulation on green innovation in the digital economy era. Frontiers in Ecology and Evolution, 11, 1265531. https://doi.org/10.3389/fevo.2023.1265531

Porter, M. E., & van der Linde, C. (1995). Toward a new conception of the environment-competitiveness relationship. Journal of Economic Perspectives, 9(4), 97-118. https://doi.org/10.1257/jep.9.4.97

Ragin, C. C. (2008). Redesigning Social Inquiry: Fuzzy Sets and Beyond. University of Chicago Press. https://doi.org/10.7208/chicago/9780226702797.001.0001

Rennings, K. (2000). Redefining innovation: Eco-innovation research and the contribution from ecological economics. Ecological Economics, 32(2), 319-332. https://doi.org/10.1016/S0921-8009(99)00112-3

Tukker, A., & Dietzenbacher, E. (2013). Global multiregional input-output frameworks: An introduction and outlook. Economic Systems Research, 25(1), 1-19. https://doi.org/10.1080/09535314.2012.761179

Xu, L. D., Lu, Y., & Li, L. (2021). Embedding blockchain technology into IoT for security: A survey. IEEE Internet of Things Journal, 8(13), 10452-10473. https://doi.org/10.1109/JIOT.2021.3060508

Xu, R., Zhu, J., Yang, L., Lu, Y., & Xu, L. D. (2024). Decentralized finance (DeFi): A paradigm shift in the FinTech. Enterprise Information Systems, 18(9), 2397630. https://doi.org/10.1080/17517575.2024.2397630

Zhang, C., & Lu, Y. (2021). Study on artificial intelligence: The state of the art and future prospects. Journal of Industrial Information Integration, 23, 100224. https://doi.org/10.1016/j.jii.2021.100224

Zhang, J., Liu, Y., & Chen, X. (2023). The impact of human capital on green technology innovation. Frontiers in Environmental Science, 11, 1129303. https://doi.org/10.3389/fenvs.2023.1129303

Zhang, Z., Chen, S., & Yang, L. (2023). How does digital finance affect green innovation? Evidence from Chinese cities. Research in International Business and Finance, 65, 101982. https://doi.org/10.1016/j.ribaf.2023.101982