ESCOLA POLITÉCNICA DA UNIVERSIDADE DE SÃO PAULO
ENGENHARIA DE PRODUÇÃO

JOÃO PEDRO MIRANDA MUCCIOLO

Open Finance as a Research Shock: Evidence from Papers with Code in Europe and Brazil

São Paulo
2025

ESCOLA POLITÉCNICA DA UNIVERSIDADE DE SÃO PAULO
ENGENHARIA DE PRODUÇÃO

JOÃO PEDRO MIRANDA MUCCIOLO

Open Finance as a Research Shock: Evidence from Papers with Code in Europe and Brazil

Trabalho de Formatura apresentado à Escola Politécnica da Universidade de São Paulo para a obtenção do diploma de Engenheiro de Produção.

Orientador: Prof. Dr. Tiago Fonseca Albuquerque Cavalcanti Sigahi

São Paulo
2025

FICHA CATALOGRÁFICA

MUCCIOLO, João Pedro Miranda

Open Finance as a Research Shock: Evidence from Papers with Code in Europe and Brazil - São Paulo, 2025. 78 p.

Trabalho de Formatura - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Produção.

I. Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Produção.

Nome: MUCCIOLO, João Pedro Miranda

Título: Open Finance as a Research Shock: Evidence from Papers with Code in Europe and Brazil

Aprovado em:

Banca Examinadora

Prof. Dr. ____

Instituição: ____

Julgamento: ____

Prof. Dr. ____

Instituição: ____

Julgamento: ____

Prof. Dr. ____

Instituição: ____

Julgamento: ____

To my family for always supporting my dreams.

AGRADECIMENTOS

Registro meus mais sinceros agradecimentos a todas as pessoas e instituições que, direta ou indiretamente, tornaram possível este trabalho de formatura.

Sou especialmente grato às duas instituições acadêmicas que moldaram esta jornada: Politecnico di Milano e a Escola Politécnica da Universidade de São Paulo, e ao extraordinário Programa de Duplo Diploma que as une. Agradeço, de modo particular, aos meus professores orientadores, Prof. Stefano Horst Baruffaldi (Politecnico di Milano), Prof. Daniel Fernando De Souza (Politecnico di Milano) e Prof. Tiago Fonseca Albuquerque Cavalcanti Sigahi (Escola Politécnica da Universidade de São Paulo), pela orientação técnica rigorosa, pelas críticas construtivas e pelo estímulo à excelência.

Estendo meu reconhecimento a todo o corpo docente e aos profissionais que asseguram o funcionamento diário das universidades. Sua competência e dedicação criaram um ambiente seguro e propício ao aprendizado, tanto no Brasil quanto na Itália, e foram fundamentais para cada etapa desta formação.

No plano pessoal, agradeço a Deus pela oportunidade de estudar em uma das melhores universidades do Brasil e concluir o mestrado em uma das melhores da Europa por meio do Duplo Diploma.}

À minha família, meu porto seguro, deixo um “Muito Obrigado!” que carrega a dimensão desta caminhada. Aos meus pais, João e Mara, e ao meu irmão, João Victor, devo a formação de valores, a coragem para perseguir objetivos ambiciosos e a serenidade para atravessar os obstáculos. Obrigado por serem meus guias, meus fãs incondicionais e meus investidores-anjos; por estarem ao meu lado inclusive nos dias mais difíceis. Esta conquista é nossa. Amo vocês hoje e sempre.

Aos amigos que se tornaram família, em cada aula, noite de estudo, prova, festa e conversas que me moldaram como pessoa, deixo minha gratidão. As memórias, risadas, debates e a solidariedade nos momentos mais desafiadores fizeram toda a diferença para que eu chegasse ao fim desta jornada. Em especial, agradeço a Victor Lee, Lucas Padial, Gabriel Moraes, Pedro Vanzan, Rodrigo Moreira, Gustavo Kairalla, Ana Henkin, Isabella Messias, Eduardo Anbar, Luciana Lima, Guilherme Amigo, João Madeira, Maria Eduarda Sizini, Larissa Rosa, Cecília Meireles e Tales Lima, pela parceria leal e pela amizade que pretendo levar para a vida.

A todos que, de algum modo, contribuíram para que esta graduação se realizasse, o meu muito obrigado. Que eu seja capaz de retribuir, com compromisso e seriedade, parte do que recebi ao longo deste caminho.

Resumo

Este trabalho de formatura mapeia e quantifica a evolução de dados e pesquisa em finanças na plataforma *Papers with Code*, examinando se marcos regulatórios de Open Finance (PSD2 na Europa/Itália em 2018 e Open Finance no Brasil em 2021) se associam a mudanças no volume e na participação de outputs científicos na área. As bases oficiais foram coletadas, integradas e higienizadas com padronização de chaves, expansão de campos aninhados e controles de qualidade; para suprir ausências de *introduced_date* em *datasets*, empregou-se um resolvedor multifonte com ordem de precedência explícita. A seleção temática utiliza vocabulário ponderado com precedência de *n*-gramas e limiar mínimo com exigência de pelo menos um termo óbvio, resultando em 212 *datasets* (1.4%) e 13,064 artigos (2.6%. A partir desse recorte, construíram-se séries anuais e painéis de contagens e participações, ancorando os marcos como pontos de interrupção em séries globais. As comparações pré/pós e as regressões segmentadas indicam aumentos estatisticamente significativos após 2018 (Europa) e 2021 (Brasil); as participações também se elevam, com saltos de nível e inclinações mais íngremes no pós-marco. As trajetórias temporais sugerem que choques de disponibilidade de dados precedem e viabilizam a produção científica. As contribuições incluem um *pipeline* reprodutível de integração e garantia de qualidade, um imputador transparente do ano de introdução e um classificador temático auditável, além de evidência quantitativa de resposta do ecossistema a marcos de Open Finance.

Palavras Chaves: Open Finance; PSD2; Brasil; Papers with Code; séries temporais; regressão segmentada; ciência de dados.

Abstract

This thesis maps and quantifies the evolution of finance-related data and research in the *Papers with Code* repository and tests whether Open Finance regulatory milestones (PSD2 in Europe at 2018 and Brazil's Open Finance at 2021) are associated with changes in the volume and share of scientific outputs in finance. Official platform datasets were collected, integrated, and cleaned with key standardization, nested-field expansion, and quality controls; missing *introduced_date* values for datasets were resolved via a priority-ordered, multi-source procedure. The finance subset is selected with a weighted vocabulary that gives precedence to longer *n*-grams and enforces a minimum threshold with at least one "obvious" term, yielding 212 datasets (1.4%) and 13,064 papers (2.6%). Annual series and panels of counts and shares treat the milestones as temporal anchors in global time series. Pre/post comparisons and segmented regressions indicate statistically significant increases after 2018 (Europe) and 2021 (Brazil); shares also rise, with level jumps and steeper post-milestone slopes. Temporal patterns suggest that shocks in data availability precede and enable downstream research outputs. The contributions comprise a reproducible data-integration and quality-assurance pipeline, a transparent introduction-year imputer, and an auditable thematic classifier, together with quantitative evidence that the ecosystem responds to Open Finance milestones.

Keywords: Open Finance; PSD2; Brazil; Papers with Code; time-series; segmented regression; data science.

List of Figures

List of Tables

Contents

Executive Summary

Background and Research Problem

Digital transformation in financial services has shifted competitive advantage from product-centric strategies to data-driven capabilities, orchestration of ecosystems, and API-enabled modularity. Within this context, open finance policies have redefined secure data sharing, interoperability, and third-party access, with potential downstream effects on research and innovation. This thesis asks whether these regulatory milestones are temporally associated with observable changes in finance-related scientific artifacts indexed by Papers with Code (PwC).

Objectives

The study pursues four objectives: (i) to map the evolution of finance-related datasets and papers in *Papers with Code* ; (ii) to construct annual global series (2010–2024) for finance counts and shares; (iii) to test for level and trend changes around two external anchors: PSD2 (2018) and Brazil's Open Finance go-live (2021); and (iv) to interpret temporal patterns in light of platformization, open innovation in services, and regulation-as-catalyst perspectives. The analysis treats the milestones strictly as temporal anchors in global series rather than as region-specific treatments, because the source corpus lacks reliable jurisdictional tagging.

Data and Coverage

The empirical corpus comprises *Papers with Code* public JSON archives. After normalization of keys, expansion of nested fields, and conservative deduplication, the descriptive snapshot reports 15,008 datasets in total; of these, 212 ($\approx$ 1.4%) are classified as finance under the thematic procedure detailed below. For papers, the descriptive summary of the papers-with-abstracts index indicates 509,455 total items.

Data Engineering and Quality Assurance

1. Integration and linkage: The workflow normalizes URLs/keys, expands modalities/languages/tasks, and links datasets and papers through evaluation tables and repository pointers.

2. Imputation of introduction year: For datasets with missing `introduced_date`, the thesis implements a priority-ordered, multi-source imputation: ARXIV (official/usage) → WAYBACK → TEXT → URL. This resolves 2,143 of 4,835 missing cases ($\approx$ 44.3%), with provenance recorded and pre-existing values never overwritten; 71% of resolved cases are via arXiv, 23% via text cues, 5.6% via Wayback, and 0.4% via URL.

Finance Thematic Subset

To classify finance-related artifacts, a transparent weighted-vocabulary approach is used. "Obvious" terms (e.g., *open banking*, *finance/financial*, *fintech*) carry higher weights; "indirect" terms (e.g., *credit*, *loan*, *insurance*, *payment*, *risk*, *trading*) carry lower weights. The algorithm prefers longest n-grams to mitigate false positives (*open banking* eclipses *banking*), accumulates multiple occurrences, and admits a record only if it meets a global threshold and includes at least one obvious term.

Design and Methods

Annual global series (2010–2024) are constructed for: finance dataset counts, finance paper counts, and their shares relative to annual totals. Two institutional dates (2018 and 2021) serve as external temporal anchors for interpretation. Distributional pre/post comparisons use Shapiro–Wilk for normality, Welch's t-test where appropriate, and Mann–Whitney U otherwise; magnitudes are summarized by Cohen's d. Interrupted Time Series (ITS) regressions of the form

$$Y_t = \beta_0 + \beta_1 t + \beta_2 \mathrm{post}_\tau + \beta_3 (t \times \mathrm{post}_\tau) + \varepsilon_t$$

estimate level and slope changes at $\tau \in \{2018, 2021\}$.

Main Findings

1. Finance activity expands in absolute and proportional terms. Temporal plots show a near-monotonic rise in total papers and datasets, with finance tracking this expansion on a smaller base. The proportion of finance papers trends up and approaches $\approx$ 1.7–1.8% by 2024. Finance datasets' share peaks near $\approx 2.6\%$ in 2022 and stabilizes around 1.3–1.5% thereafter.

2. Break-type evidence aligns with policy anchors. Pre/post comparisons indicate statistically significant increases in finance counts and shares after 2018 and after 2021. ITS models detect positive post-period level changes and steeper post-2018 slopes for finance papers, with a pronounced dataset jump around 2021–2022 followed by a higher plateau.

3. Cross-metric associations are positive. The correlation matrix reports positive associations across finance metrics (absolute and proportional), including a moderate co-movement between finance datasets and finance papers, consistent with complementary mechanisms rather than a single driver.

Interpretation

The time ordering suggests a plausible sequential mechanism: regulatory and market initiatives expand accessible data and standardize APIs, which is first reflected in dataset availability and later in publications that reuse those artifacts. Platform curation dynamics in *Papers with Code* and small-base effects can amplify year-on-year variation without overturning the medium-run upward trend. The thesis maintains a conservative stance: anchors are used for timing; results are descriptive, not causal.

Contributions

- A reproducible data-integration and quality-assurance pipeline for *Papers with Code* JSON archives, including documented linkage across datasets, papers, tasks, and repositories.
- A transparent, provenance-preserving imputation of dataset introduction year that resolves $\approx 44.3\%$ of missing cases, prioritizing arXiv and recording sources.
- An auditable thematic classifier for finance (weighted vocabulary with longest-n-gram precedence and a minimum-threshold rule) applicable to both datasets and papers.
- Quantitative evidence (via pre/post contrasts and segmented regressions) that finance-related research activity in *Papers with Code* rises after the 2018 and 2021 anchors.

Implications

For research. Consolidating domain benchmarks (tasks, metrics, reference datasets) can convert sporadic dataset spikes into sustained research pipelines; improved documentation/licensing lowers replication and audit costs. **For industry.** API-first strategies, high-quality data governance, and security-by-design become preconditions for effective participation in open finance ecosystems; investment in standard interfaces and conformance testing is complementary to product innovation. **For policy.** Governance choices (centralized, federated, or market-led) affect interoperability, adoption, and systemic risk; aligning technical standards with supervisory goals can propagate data-access shocks into measurable innovation outcomes without compromising privacy or resilience.

Avenues for Further Work

Short-run extensions include HAC-robust ITS with placebo windows and explicit lead diagnostics, as well as descriptive DiD using non-finance series as a proxy comparator. Medium-run extensions include subdomain analyses by prominent financial tasks and the consolidation of open benchmarks to strengthen reproducibility and diffusion.

1 | Introduction

1.1. Context

Digital transformation denotes a socio–technical process through which digital technologies reshape organizational strategies, processes, and value-creation mechanisms, going beyond mere digitization to reconfigure roles, capabilities, and outcomes (1, 2, 3). In finance specifically, this concept unfolds through successive waves of general-purpose and domain-specific technologies (cloud computing, data analytics, machine learning/AI, standardized application programming interfaces), in tandem with institutional and regulatory changes that open data and infrastructure to new complementarities among actors (1, 4).

At the organizational level, digital transformation shifts competitive advantage from product portfolios toward data capabilities, personalization, and ecosystem orchestration, requiring sustained investments in security, data governance, and analytics (2, 5). For society, digital transformation can expand access to payments, savings, and credit, and reduce transactional frictions; yet it also elevates risks related to cybersecurity, privacy, and digital exclusion, which in turn call for appropriate institutional design and regulatory safeguards (6, 7).

1.2. Core concepts

Digital transformation in the financial sector materializes when financial services are modularized and exposed as interoperable components via APIs and lawful data sharing, under multi-actor governance involving banks, fintechs, bigtechs, infrastructure providers, and regulators. Platform and ecosystem theories supply the analytical language for this transition, explaining how vertically integrated firms evolve into orchestrators that manage complementarities between a core and third-party complements (8, 9, 10, 11).

In empirical terms, digital transformation in the financial sector manifests in Banking-as-a-Service (BaaS) and related models that expose regulatory licenses, payment rails, and data services to third parties, allowing non-banks to embed financial functionality without building proprietary infrastructures (12, 13). This reconfiguration intensifies competition in customer experience layers, lowers entry barriers for specialized providers, and redistributes value capture across incumbents and new entrants, while raising the bar for resilience and security (1, 7).

"Platformization" organizes innovation around multi-sided architectures: platform sponsors set interfaces, rules, and incentives; complementors develop compatible services; modularity and governance mechanisms balance variety with interoperability (8, 9, 11). In banking, consulting and scholarly work often portray incumbents as ecosystem orchestrators that provide compliance and payments capabilities, while third parties compete at the front end; BaaS and API-first strategies are emblematic (10, 14, 15).

Open Innovation models innovation as managed flows of internal and external knowledge (outside-in sourcing and inside-out commercialization) (16, 17). In services, this type of innovation helps explain bank–fintech partnerships, accelerators, and infrastructure collaborations that reduce time-to-market and expand product variety in digital transformation for financial services (18, 19).

Regulation acts as a catalytic variable, defining data-access regimes, security baselines, and secure communication obligations (via PSD2 and the RTS on Strong Customer Authentication and Common and Secure Communication) thereby regulating access for AISPs and PISPs under customer consent (20, 21, 22). Implementation has exposed challenges of technical harmonization, supervisory clarity, and security alignment, underscoring the importance of governance design across centralized, federated, and market-led models (7, 23, 24).

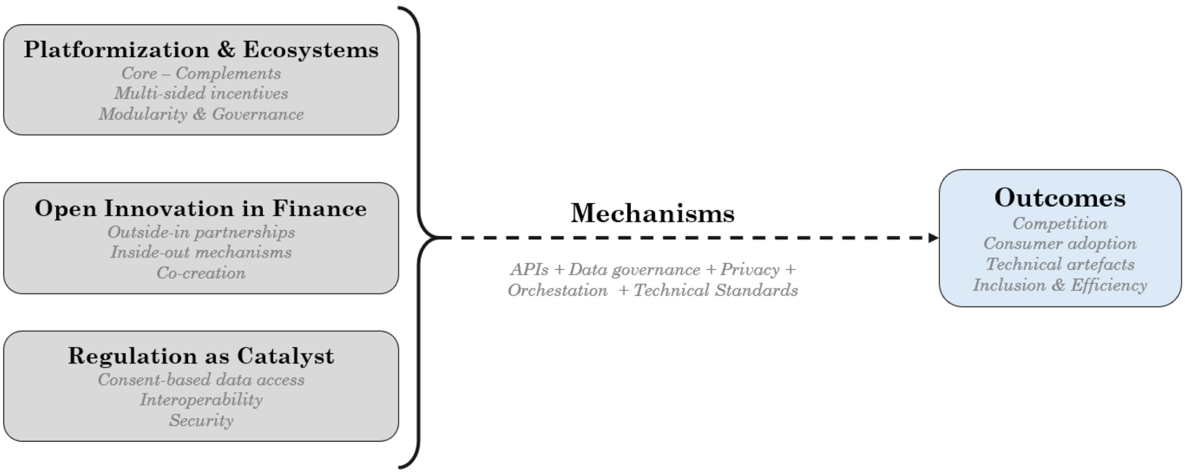

Figure 1.1: Conceptual framework for digital transformation in financial services

1.3. Research relevance

This thesis is motivated by the public relevance and comparative value of studying Open Finance through the combined lenses of platformization, Open Innovation in services, and regulation. For individuals, open financial infrastructures change how people pay, save, and access credit and data; for organizations, they demand API-first capabilities, secure data governance, and new partnering logics; for the ecosystem and financial system, they restructure competition and infrastructure layering; for public policy, they provide levers for inclusion, innovation, and contestability; and for society, they promise lower frictions and costs, conditional on robust privacy and security safeguards (25, 26, 27, 6).

In Brazil, the practical integration between Open Finance and instant payments is made explicit in the Pix Management Report (section “Serviço de iniciação: o elo que conecta o Open Finance e o Pix”); the DREX pilot, currently underway, sets out guidelines for tokenization and programmability (28, 29). In Europe, the PSD2 trajectory shows trade-offs among standardization, competition, and innovation speed, and documents harmonization challenges and obstacles in dedicated interfaces (e.g., fallback, performance, testing), as reflected in EBA Opinions and WG on APIs, and in recent evaluations (22, 30, 20, 31).

1.4. Research gaps

The literature maps the rise of Open Banking/Open Finance, discusses API security and business model shifts, and documents regulatory outcomes (competition, innovation) alongside implementation frictions (32, 33, 31, 7). Brazilian studies describe phased implementation and organizational adjustments required by Open Finance (34).

Two gaps persist. First, there is limited quantitative evidence linking specific regulatory milestones (e.g., PSD2 in 2018; Brazil’s Open Finance phases in 2021) to time-structured changes in the production of technical scientific artifacts (datasets, papers, code) in finance. Second, existing accounts seldom integrate, in one empirical design, the three lenses of platformization, open innovation in services, and regulation to explain cross-jurisdictional variation (EU vs. Brazil) and temporal patterns. The thesis addresses both gaps by leveraging *Papers with Code* based series and aligned policy timelines.

1.5. Research objective

This thesis investigates whether global finance-related artifacts in *Papers with Code* display temporal inflections around two exogenous regulatory anchors: PSD2 (EU, 2018) and Brazil’s Open Finance (2021). In the absence of reliable jurisdictional tags in the source corpus, the empirical series are not stratified by region. The milestones serve strictly as temporal reference points for interpretation rather than as region-specific treatments.

- Map the concepts and regulatory milestones most relevant to digital transformation in financial services (EU and Brazil), emphasizing interoperability, security, and governance obligations (21, 35).

- Construct annual time series for finance-related artifacts in *Papers with Code* , using a transparent “finance” filter.
- Estimate pre and post contrasts and Interrupted Time Series (ITS) models to explore level and slope changes aligned with PSD2 (2018) and Brazil’s Open Finance phases (2021).
- Interpret temporal patterns around two external temporal anchors (EU (2018) and Brazil (2021)) in a global *Papers with Code* panel; no jurisdictional stratification is attempted.

1.6. Research contributions

The study integrates platform/ecosystem scholarship, open innovation in services, and the regulatory dimension as a catalytic variable into an anchor-based global design. Rather than contrasting jurisdictions, we quantify the dynamics of finance-related artifacts in *Papers with Code* and link them to two exogenous milestones used as temporal anchors: PSD2 (2018) and Brazil’s Open Finance (2021). This framing informs banks and fintechs (API/product prioritization, partnering, security/compliance) and policymakers (technical standards, governance choices, timing), while keeping claims strictly descriptive.

2| Literature review

2.1. Digital Transformation and the Evolution of Financial Systems

The twenty-first century has witnessed an unprecedented transformation in the global financial services sector, fundamentally altering the relationship between financial institutions, technology, and consumer expectations (3). The transformation extends beyond mere technological adoption to encompass comprehensive reorganization of business models, operational processes, and value creation mechanisms within financial ecosystems (36).

The emergence of breakthrough innovative technologies, including artificial intelligence, robotics, blockchain technologies, the Internet of Things, Big Data, and cryptocurrency, has fundamentally disrupted traditional systems in the financial services market, leading to the formation of complex ecosystems that blur the boundaries between financial and non-financial activities (1). This transformation has catalyzed the replacement of traditional bank-based and market-based models with symbiotic ecosystems, creating conditions for non-bank financial intermediation models to flourish (1).

Digital transformation in financial services represents a multi-dimensional phenomenon driven by four principal factors: technological advancement, regulatory evolution, evolving consumer behavior, and competitive pressures (2). The rapid advancement of technology, combined with growing demand for digital financial services, has given rise to a thriving fintech ecosystem where startups armed with cutting-edge technologies such as artificial intelligence, blockchain, and big data analytics have disrupted traditional banking models by offering streamlined and customer-centric solutions (37).

The integration of emerging technologies has created a dynamic ecosystem where traditional financial institutions must compete and collaborate with innovative startups, bigtechs, and fintechs (13). This cross-sector fintech phenomenon operates at the intersection of financial services and information technology, disrupting existing business models while creating novel ecosystem dynamics that require new theoretical frameworks for understanding (38). The transformation has been particularly accelerated by COVID-19-induced lockdowns, which expedited the digitalization of financial services and forced rapid adoption of digital channels across all market segments

(36).

The shift toward digitalization has fundamentally altered the competitive landscape, with the key competitive advantage now residing in the use of innovative technologies for remote access and personalization of services (1). Financial institutions have been compelled to transition from product-centric to client-centric business strategies, necessitating significant technological investments and organizational restructuring (2). This transformation encompasses not only technological capabilities but also strategic business objectives aimed at creating new forms of value for all stakeholders within the financial ecosystem (2).

The evolution has been characterized by the emergence of sophisticated technology-enabled financial services that have swiftly altered the traditional financial services space, with global adoption of fintech rapidly expanding across diverse markets and regulatory environments (4). Research indicates that this transformation is creating a resilient financial ecosystem capable of adapting to emerging threats, evolving regulatory landscapes, and changing consumer expectations while maintaining security and operational efficiency (5).

2.1.1. Platformization and Open Innovation

The literature on platforms shows that digital transformation is often organized around platformization models: multi-sided structures that orchestrate complementarities between core providers (platform sponsors) and external complementors, reducing transaction costs and increasing ecosystem generativity. Core concepts as modularity, gatekeeping, ecosystem governance, and coordination mechanisms between "core" and "complements", have been extensively developed in the literature on platforms and ecosystems. These approaches offer useful analytical tools to understand how financial institutions transform from vertically integrated firms into orchestrators of digital markets (8, 9, 10, 11).

In the banking context, the platform logic is manifested in the emergence of models such as Banking-as-a-Service (BaaS) and Banking-as-a-Platform (BaaP), where incumbent institutions or new providers make regulatory capabilities, payment infrastructure, and data services available to third parties via APIs. This strategic shift allows banks to provide the regulatory and liquidity core plumbing, while third parties (fintechs, retailers, bigtechs) compete in the front-end and customer experience layer, a rearrangement that both academic and consulting literature describe as part of the platformization of the banking industry (14, 15, 19).

In parallel, the Open Innovation paradigm (Chesbrough) proposes that firms generate innovation through managed flows of external and internal knowledge, that is, by combining outside-in strategies (incorporating external ideas) and inside-out strategies (commercializing internal ideas through third parties). Applied to services, the concept of Open Services Innovation emphasizes co-creation and the integration of external capabilities, which is precisely what we observe when banks collaborate with fintechs, incubators, and infrastructure providers. These practices reduce the cost and time to develop new financial services and expand the diversity of offerings available to consumers (16, 17, 18).

Regulatory action functions here as a catalyst: data access regimes and technical standards (for example, PSD2 in Europe and the Open Finance architecture in Brazil) navigate the trade-off between consumer protection and market entry facilitation, defining interoperability obligations (standardized APIs, security requirements) that lower entry barriers for external providers. Such public interventions not only encourage the adoption of platform-based models but also shape the research and technological development agenda by stimulating investments in APIs, data governance, and open business models (39, 19).

By combining the above theoretical lenses: platformization and ecosystem studies, Open Innovation theory applied to services, and the regulatory dimension as a catalytic variable, one obtains a robust analytical framework to examine the impact of Open Finance on scientific production, regulatory design, and technological trajectories. This framework will be used in the comparative analysis between the European and Brazilian cases, allowing the articulation of empirical evidence (investments, API adoption, incumbent–fintech partnerships) with theoretical explanations on institutional change, value capture, and ecosystem governance (10, 8, 16, 18).

2.2. Regulatory Frameworks and the Emergence of Open Banking

2.2.1. The European Foundation: PSD2 and Its Implementation Challenges

The regulatory response to digital transformation has been instrumental in shaping the current landscape of open banking and open finance initiatives worldwide. In Europe, the Second Payment Services Directive (PSD2), officially designated as Directive (EU) 2015/2366, established the legal foundation for open banking by mandating that account servicing payment service providers grant third-party providers access to customer account information and payment initiation services (40). The directive aims to enhance competition and innovation in the financial services industry while maintaining consumer protection and data security standards, representing a paradigmatic shift from traditional closed banking systems to open, API-driven ecosystems (41).

In addition to the legislative text, the European Banking Authority (EBA) has issued multiple Regulatory Technical Standards (RTS) and Opinions providing detailed guidance on strong customer authentication (SCA), secure communication, and API standardization, which have been crucial for harmonizing implementation across Member States (21, 42). In the Italian context, the Banca d'Italia has complemented the European framework by publishing specific communications and supervisory reports addressing operational resilience, cybersecurity compliance, and the integration of PSD2 requirements into national banking infrastructures (43, 44).

The implementation of PSD2 has demonstrated the effectiveness of regulatory intervention in creating open technical infrastructure within payment services markets, significantly lowering entry barriers for new service providers (31). Empirical analysis reveals that PSD2 has successfully facilitated the development of innovative financial services and enhanced competition in the European payments landscape, with third-party providers leveraging API access to offer value-added services to consumers (33). However, the directive's implementation has also revealed significant challenges related to inconsistent regulatory requirements for non-bank financial companies, imperfect development of API-based infrastructure, and issues with transaction processing that have impeded the full realization of open banking's potential (31).

A comprehensive analysis of PSD2 compliance reveals a discernible lack of harmonization and clarity concerning the technical security specifications required for effective implementation across European member states (7). This lacuna substantiates the challenges banks face in fully grasping the extensive spectrum of compliance obligations mandated by PSD2, particularly regarding the integration of cybersecurity frameworks such as NIS2, the Cybersecurity Act, GDPR, ISO 27001:2022, and PCI DSS (7). The study demonstrates that while the open banking paradigm brings enhanced consumer choice and market competition, it concurrently exposes the financial ecosystem to potential security vulnerabilities and privacy risks that require sophisticated technical and regulatory responses (7).

2.2.2. The United Kingdom's Pioneering Approach

The European experience with PSD2 has influenced open banking initiatives globally, with countries such as the United Kingdom developing their own regulatory frameworks that extend beyond the basic requirements of the European directive (45). In the United Kingdom, the concept of open banking was initially developed to foster competition by enabling the sharing of client data among competitors with customer consent, representing one of the most comprehensive implementations of open banking globally (46).

The UK's approach has been characterized by industry coordination and technical standards development, aimed at avoiding market fragmentation while building a robust ecosystem that serves all consumers and small-to-medium enterprises (47). The Competition and Markets Authority (CMA) implemented requirements that went further than the EU's PSD2, including a co-regulatory obligation for the nine largest retail and small business banks to agree on a common technical interface (API) and standards for security, user experience, and other areas identified as important to customers (48). This regulatory approach evolved from ineffective portability requirements (Midata) to in-depth interoperability obligations, which have enabled hundreds of firms to create a thriving UK "fintech" market of complementary financial services (48).

However, recent analysis suggests that the UK's open banking initiative has faced implementation challenges, with incumbent banks successfully slowing the reform process through strategic acquisitions and partnerships with fintech firms (49). This research reveals the "quiet politics" of open banking reform, demonstrating how established financial institutions have used their market position and regulatory influence to moderate the disruptive potential of open banking while maintaining their competitive advantages (49). Despite regulatory mandates, the study shows that direct competition with incumbent banks has been limited, though the reforms have successfully fostered innovation in complementary financial services (48).

2.2.3. Global Regulatory Convergence and Divergence

The global adoption of open banking initiatives reflects diverse regulatory approaches and market conditions that influence implementation outcomes, with significant variations in consumer adoption rates across different jurisdictions (50). Research on global open banking implementation reveals that countries implementing PSD2-regulated initiatives, market-driven approaches, and other regulatory frameworks show different levels of success in promoting formal saving and digital remittance behaviors (50). Specifically, consumers in PSD2-regulated countries demonstrate higher marginal utilities from digital remittance services compared to countries without open banking initiatives, indicating the effectiveness of regulatory mandates in driving adoption (50).

Comparative analysis of regulatory frameworks reveals fundamental differences in approach across jurisdictions. The Indonesian Financial Services Authority (OJK) Regulation adopts a 'bank-centric' model, granting financial institutions discretionary power over third-party provider (TPP) access through bilateral partnerships (51). This contrasts with jurisdictions like the EU, UK, and Australia, which employ legislative tools that obligate banks to provide TPPs access to customer data upon explicit consent (51). The research demonstrates that the 'bank-centric' model could reduce consumer choice and create conflicts of interest, with banks potentially favoring their own TPPs and undermining fairness in the ecosystem (51).

2.2.4. Regulatory Technology and Innovation Sandboxes

The development of regulatory frameworks for open banking has been accompanied by the emergence of regulatory technology (RegTech) solutions and innovation sandboxes that facilitate controlled experimentation with new financial technologies (52). The UK's regulatory sandbox for financial technologies represents a pioneering approach to balancing innovation with consumer protection, though research reveals challenges in attracting stakeholder participation from the fintech sector (52). The study finds that fintech firms have a range of practical and reputational considerations that influence their decisions to participate in regulatory sandboxes, highlighting the importance of regulator reputation and stakeholder engagement strategies (52).

The integration of artificial intelligence and machine learning technologies into financial services regulation has prompted the development of new regulatory frameworks that address algorithmic governance and automated compliance monitoring (53). UK financial regulators are experimenting with the conversion of rulebook content into machine-readable and executable code, driven by the belief that algorithmic approaches can eliminate the need for human interpretation as a deliberative process (53). However, this trend raises important questions about the role of judgment and discretion in financial regulation, particularly in complex and rapidly evolving technological environments (53).

2.2.5. Consumer Trust and Behavioral Adoption

The success of open banking regulatory frameworks ultimately depends on consumer adoption and trust, which are influenced by regulatory design and implementation strategies (54). Empirical research analyzing consumer behavior in Spain, a country with established open banking regulation under PSD2, reveals that initial trust and social influence are essential determinants of behavioral intention to adopt open banking services (54). The study demonstrates that perceived usefulness, social influence, and initial trust account for 85% of the variance in behavioral intention to embrace open banking, while perceived ease of use plays a minor role (54).

These findings have important implications for both private sector participants and policymakers in designing effective open banking ecosystems. Private agents should focus on highlighting the benefits of open banking services and building consumer trust through transparent communication and robust security measures (54). Simultaneously, policymakers should work on developing regulatory frameworks that increase clients' initial confidence in open banking systems through clear consumer protection measures, standardized security requirements, and effective dispute resolution mechanisms (54).

The analysis of consumer behavior also reveals the importance of perceived risk as a determinant of propensity to adopt Account Information Services under PSD2 (55). Understanding these behavioral factors is critical for regulators seeking to design frameworks that maximize the benefits of open banking while ensuring adequate consumer protection and market stability (55).

2.3. Brazil's Open Finance Model and Comparative Development

Brazil's approach to open financial systems stands out for its comprehensive scope, regulatory innovation, and integration of digital public infrastructure. Unlike the European model, which is primarily rooted in the banking sector and focused on payment services, Brazil's Open Finance initiative extends beyond traditional banking to include insurance, investments, foreign exchange, and pension products, reflecting a broader vision of financial ecosystem integration (56).

2.3.1. Regulatory Evolution and Implementation Phases

The regulatory journey began with the Central Bank of Brazil's BC# Agenda, which articulated a vision for a more competitive, inclusive, and efficient financial system through digital transformation and open data policies (24, 56). The phased implementation of Open Finance in Brazil was designed to gradually expand the scope of data sharing and service interoperability (57, 35):

- **Phase 1** (2021): Focused on the sharing of standardized public data on financial institutions and their products.
- **Phase 2**: Enabled sharing of customer registration and transactional data, with explicit customer consent.
- **Phase 3**: Introduced payment initiation and forwarding of credit proposals through third-party providers.
- **Phase 4**: Expanded to include data from insurance, foreign exchange, investments, and pension products.

This regulatory architecture was accompanied by robust governance mechanisms, including the establishment of the Open Finance Governance Structure, which oversees technical standards, security protocols, and market coordination (34, 35).

2.3.2. Integration with Digital Public Infrastructure

A distinctive feature of the Brazilian model is its integration with digital public infrastructure, notably the PIX instant payment system and the planned DREX central bank digital currency (CBDC). PIX, launched in 2020, has rapidly become a cornerstone of Brazil's digital payments landscape, enabling real-time, low-cost transactions and supporting the operationalization of Open Finance use cases (28, 56). The ongoing development of DREX aims to further enhance interoperability, programmability, and financial inclusion by providing a secure platform for digital assets and smart contracts within the Open Finance ecosystem (29, 56).

2.3.3. Technological and Operational Impacts

Empirical research highlights that the implementation of Open Banking and Open Finance in Brazil has required significant technological investments from financial institutions, particularly in API development, cybersecurity, and data governance (34). The transition from a product-centric to a client-centric paradigm has necessitated organizational restructuring and the adoption of agile, data-driven business models. Despite the regulatory support and a favorable environment for innovation, many banks have faced challenges in meeting the Central Bank's ambitious implementation timelines, often due to the complexity of legacy systems and the need for cultural change within organizations (34).

2.3.4. Market Dynamics and Ecosystem Effects

The Brazilian Open Finance model has fostered a dynamic ecosystem in which fintechs, incumbent banks, and new entrants compete and collaborate to deliver personalized financial services. The regulatory framework encourages interoperability and competition while maintaining a strong focus on consumer protection and data privacy. As a result, Brazil has seen a proliferation of innovative financial products, increased consumer empowerment, and improved access to financial services for previously underserved populations (56).

Comparative analyses indicate that, while inspired by the European PSD2 framework, Brazil's approach is distinguished by its broader scope, integration with digital public infrastructure, and phased regulatory design. These features position Brazil as a global reference in the development of open financial ecosystems, offering valuable lessons for other jurisdictions seeking to advance financial innovation and inclusion (56, 34).

2.3.5. Challenges and Future Directions

Despite notable progress, the Brazilian Open Finance initiative faces ongoing challenges, including ensuring interoperability across diverse financial products, maintaining robust security standards, and fostering widespread consumer trust and engagement. Continued collaboration between regulators, industry stakeholders, and technology providers will be essential to address these challenges and to realize the full potential of Open Finance in promoting financial inclusion, innovation, and systemic resilience (56, 58).

2.4. Technological Architecture and Innovation Dynamics

The technological architecture of open banking and open finance systems is grounded in the use of standardized Application Programming Interfaces (APIs), which enable secure, real-time data exchange and service interoperability among financial institutions and third-party providers (41). The adoption of APIs has been pivotal in transforming traditional banking infrastructures into modular, interconnected ecosystems, fostering innovation and competition by lowering barriers to entry for fintechs and new market participants (41, 33).

2.4.1. API Standardization and Security

API standardization is a critical element for ensuring interoperability, scalability, and security within open finance environments. Research by Almehrej et al. highlights the importance of robust protocol design, including the use of RESTful APIs, OAuth 2.0 for secure authorization, and strong authentication mechanisms to prevent unauthorized access to sensitive financial data (41). Modesti et al. provide a detailed security analysis of the Open Banking Account and Transaction API, identifying potential vulnerabilities such as improper token management and recommending best practices for mitigating risks through formal verification and continuous monitoring (33).

The European experience, as analyzed by Gounari et al., demonstrates that API-based frameworks must be integrated with comprehensive cybersecurity standards, including alignment with regulations such as GDPR, NIS2, and ISO 27001:2022, to address privacy, data protection, and operational resilience (7). The lack of harmonization in technical and security specifications across jurisdictions remains a challenge, underscoring the need for ongoing collaboration between regulators and industry stakeholders (7).

2.4.2. Semantic Technologies and Data Integration

As open banking expands to include a wider array of financial products and services, the complexity of data integration increases. Paneque et al. emphasize the role of semantic technologies in managing heterogeneous data sources, proposing semantic models that facilitate the interoperability of diverse datasets and enhance the capacity for advanced analytics and personalized financial services (59). These models leverage ontologies and knowledge graphs to create a shared understanding of financial concepts, enabling more efficient data sharing and supporting the development of intelligent, context-aware applications (59).

2.4.3. Innovation, Modularity, and Ecosystem Effects

The modular architecture enabled by APIs supports the rapid development and deployment of innovative financial products, such as Banking-as-a-Service (BaaS), Payment Initiation Services (PIS), and Account Information Services (AIS).

According to (60, 12), BaaS refers to the provisioning of core banking infrastructure, regulatory licenses, and payment rails as modular services to third parties via APIs. This model enables non-banks (e.g., retailers, bigtechs, fintechs) to integrate banking functionality into their offerings without the need to develop proprietary infrastructures.

PIS, as defined under PSD2 and explored by (61), allows authorized providers to initiate payments directly from a customer's bank account, bypassing traditional card networks. AIS, in turn, enables the aggregation of account and transaction data from multiple financial institutions, providing customers and service providers with a consolidated financial overview (62). Both PIS and AIS have been key catalysts for the creation of data-driven financial services, particularly in personal finance management, credit scoring, and tailored product recommendations.

This modularity allows financial institutions and fintechs to combine and recombine services, fostering a dynamic ecosystem where new business models can emerge and evolve quickly (41, 33). The flexibility of API-driven systems also facilitates regulatory compliance, as updates to security protocols and data governance policies can be implemented more efficiently across the ecosystem (7).

2.4.4. Governance Models in Open Banking and Open Finance

The governance of open finance ecosystems is a decisive factor influencing adoption rates, innovation outcomes, and market stability. The literature distinguishes three primary governance archetypes:

- **Centralized governance**, where a single regulatory or industry body defines and enforces API standards, security requirements, and compliance rules. The UK Open Banking Implementation Entity (OBIE) is a paradigmatic example (23).

- **Federated governance**, where multiple stakeholders (regulators, banks, fintechs) collaborate in standard-setting and enforcement, typically under a public-private partnership model. The Berlin Group initiative in the EU exemplifies this approach (62).

- **Market-led governance**, where standards emerge organically through competitive forces and bilateral agreements, with minimal regulatory intervention. The U.S. open finance ecosystem, as discussed by (63), illustrates this model.

Each governance model presents trade-offs between innovation speed, interoperability, and systemic risk. Centralized models tend to ensure higher interoperability and consumer protection but may slow down innovation due to rigid compliance requirements. Market-led approaches promote rapid service innovation but risk fragmentation and inconsistent security practices. Federated models attempt to balance both dynamics, though they require complex coordination mechanisms among diverse actors (23, 62).

2.4.5. Challenges and Future Directions

Despite significant progress, several challenges persist in the technological architecture of open finance. These include ensuring the scalability and reliability of API infrastructures, maintaining robust security in the face of evolving cyber threats, and achieving true interoperability across national and international frameworks (7, 59). Moreover, as (60) note, the long-term sustainability of open finance models will depend on clear governance structures, viable monetization strategies for API providers, and alignment between regulatory objectives and market incentives.

Ongoing research and collaboration between regulators, industry consortia, and technology providers will be essential to address these challenges and to realize the full potential of open banking and open finance as platforms for innovation, inclusion, and systemic resilience.

2.5. Global Perspectives and Comparative Analysis

The global adoption of open banking demonstrates substantial heterogeneity, shaped by regulatory frameworks, market maturity, and technological infrastructure across regions. Comparative research highlights that while Europe has led the way with regulatory-driven models such as PSD2, other regions have pursued market-driven or hybrid approaches, each yielding distinct outcomes in consumer adoption and innovation (64, 50).

2.5.1. Diversity of Regulatory Models and Adoption Rates

A comprehensive review of open banking adoption reveals that PSD2-regulated countries in Europe exhibit higher consumer engagement with digital remittance and savings products, attributable to strong regulatory mandates and standardized API frameworks (50). The United Kingdom stands out, with more than half of consumers having used open banking products at least once, reflecting both regulatory impetus and effective market coordination (64, 65). In contrast, market-driven models, such as those in the United States, face challenges from fragmented regulatory oversight, resulting in slower and less uniform adoption (66).

Emerging economies, including India and China, have demonstrated rapid growth in digital payments and open banking adoption, propelled by government-led digital infrastructure initiatives and high mobile penetration (67, 64). For example, India's Unified Payments Interface (UPI) has transformed the payments landscape, driving financial inclusion and accelerating consumer uptake of fintech services (67). However, these successes are often accompanied by ongoing challenges in achieving full financial inclusion and addressing the needs of unbanked populations (67).

2.5.2. Innovation, Ecosystem Effects, and Global Trends

Open banking is recognized as a catalyst for innovation, enabling the commercialization of APIs, the emergence of new value-added services, and the creation of platform-based business models (32). The collaborative nature of open banking fosters partnerships between incumbents, fintechs, and technology providers, leading to a proliferation of services that enhance customer experience and expand financial product diversity (64, 32).

Bibliometric analyses show a sharp increase in open banking research since 2018, with the United Kingdom, India, and China emerging as leaders in both academic output and practical implementation (64). Despite this growth, persistent gaps remain in areas such as API security, risk management, and the harmonization of technical standards, underscoring the need for ongoing research and international collaboration (64, 65).

2.5.3. Opportunities and Challenges in Global Adoption

The global landscape of open banking presents both opportunities and challenges. Opportunities include enhanced competition, improved financial inclusion, and the democratization of financial data, which collectively drive economic growth and innovation (66, 67). However, challenges such as regulatory fragmentation, cybersecurity risks, and consumer trust remain significant barriers to universal adoption and the realization of open banking's full potential (65, 66).

Comparative studies emphasize the importance of cohesive regulatory policies, secure interoperability, and the adoption of global best practices to ensure sustainable and inclusive growth of open banking worldwide (66, 32). As open banking continues to evolve, cross-border collaboration and the alignment of standards will be critical to addressing emerging risks and maximizing the benefits for all stakeholders.

2.5.4. Socioeconomic Impacts and Financial Inclusion

Beyond technological and regulatory considerations, open banking and its broader evolution into open finance have the potential to deliver significant socioeconomic benefits, particularly in emerging markets such as Brazil. Studies by the World Bank, the Inter-American Development

Bank (IDB), and the Consultative Group to Assist the Poor (CGAP) highlight the role of interoperable financial infrastructures in expanding access to credit, enabling alternative credit scoring models, and reducing the cost of financial services for underserved populations (25, 26, 27).

In the Brazilian context, integration between instant payment systems such as PIX and open finance frameworks has facilitated new credit offerings for small and medium-sized enterprises (SMEs) and microentrepreneurs, segments historically constrained by limited access to working capital (68). This aligns with government and central bank policies aiming to use data-sharing frameworks to foster entrepreneurship and local economic development.

2.5.5. Security, Privacy, and Ethical Considerations

While regulatory frameworks such as GDPR, NIS2, and ISO 27001 provide a solid foundation for data protection, there is growing scholarly debate on the ethical dimensions of financial data sharing. Concepts such as data sovereignty, consumer empowerment, and privacy-preserving computation are increasingly central to discussions of sustainable open finance adoption (69, 70).

Emerging literature proposes the application of trust-by-design principles, embedding transparency, accountability, and user control into the architecture of financial platforms, to strengthen consumer trust and regulatory compliance (71). Such approaches not only mitigate cybersecurity and privacy risks but also enhance public perception of open finance ecosystems as secure and equitable.

2.5.6. Future Trends and Cross-Border Interoperability

Looking ahead, the concept of Open Finance 2.0 (integrating financial services into non-financial platforms via embedded finance) is gaining traction in both European and Brazilian policy discussions (72, 73). This evolution is expected to blur the boundaries between banking, commerce, and digital ecosystems, enabling seamless, context-aware financial interactions.

Internationally, cross-border interoperability is emerging as a strategic priority. Initiatives exploring integrations between Brazil's PIX, India's UPI, and European instant payment systems illustrate the potential for a globally interconnected payments and data-sharing infrastructure (74, 75). In Latin America, regional collaboration could accelerate financial integration, reduce remittance costs, and enhance trade settlement efficiency.

Incorporating these trends into policy and infrastructure design will be essential to ensuring that open finance develops as an inclusive, secure, and innovation-driven global framework.

2.6. Open science as an engine of innovation

2.6.1. Theoretical foundations of open science and cumulative innovation

Across the economics of science, a central insight is that the institutional logic of open science (with priority-based rewards, disclosure norms, and reputational incentives) aligns individual motives with collective knowledge accumulation. In their seminal contribution, David formalizes how priority races, disclosure, and the public-good nature of scientific knowledge support cumulative discovery (76), while he articulates the economic logic by which open regimes maximize the growth of the stock of reliable knowledge (77). Complementing these accounts, Paula Stephan synthesizes how scientific careers, tools, and collaborations create incentives to produce and share technoscientific instruments (data, code, methods) (78) (79). In this framework, openness reduces replication and search costs, expands the audience for validation, and increases the reusability of artifacts, thereby accelerating cumulative and recombinant innovation.

Two additional mechanisms are salient for technology-intensive domains such as finance. First, openness codifies tacit knowledge: datasets with detailed documentation, executable code, and clear protocols transform know-how into shareable capabilities that can be recombined in downstream applications. Second, openness standardizes interfaces for collaboration: common file formats, shared software stacks, and explicit data schemas reduce integration frictions across organizations and disciplines. The relevance for financial technologies is direct. Open datasets, models, and code lower barriers to experimentation and shorten the distance between research breakthroughs and production-grade solutions in payments, risk, and compliance, while simultaneously improving auditability and external scrutiny.

2.6.2. Open data as technological input and public policy

Recent instruments from the Organisation for Economic Co-operation and Development codify a pragmatic continuum of openness: from open data to non-discriminatory and conditioned access arrangements, with legal and technical safeguards to maximize social value while minimizing risk (80, 81). This moves beyond a binary view of "open/closed" and offers implementation guidance for domains, including finance, where privacy, IP, and security constraints bind. Classic open government data work provides operational frameworks to design and evaluate initiatives (82) and to compare policies along context, content, metrics, and public value dimensions (83). Empirically, open data availability is associated with measurable reuse and downstream impact (for example, a robust citation advantage and persistent third-party reuse in biomedicine (84)) while community standards and repositories improve preservation and reproducibility (85).

For Open Finance, these principles translate into tiered access regimes with clear governance:

public product and price-level information exposed via standardized endpoints; consented customer-level data shared under purpose limitation and strong authentication; and controlled research or supervisory access mediated through secure data environments. Effective execution requires attention to data management details that determine real-world usability: rich metadata and persistent identifiers; explicit licensing; versioning and changelogs; high-quality documentation and examples; and stable, well-monitored APIs with test suites and conformance criteria.

2.6.3. Open datasets as technoscientific tools: the case of AI

Open, labeled datasets reconfigure research agendas and lower the cost of learning, benchmarking, and diffusion. In AI, open labeled datasets catalyzed deep learning development, not merely by scale but via shared tasks and evaluation norms (86). Historically, ImageNet (87) and MS-COCO (88) illustrate how widely accessible data, clear task definitions, and standardized metrics orchestrate community effort, spur method innovation, and speed transfer to applications.

The analogy for finance is straightforward. Open, well-curated datasets on payments, fraud patterns, SME finance, market microstructure, or supervisory text can coordinate innovation by creating common tasks and benchmarks (for example, payment initiation failure prediction, merchant risk profiling, or document classification for compliance). Benchmarks discipline empirical claims, reduce redundant effort, and channel competition into problem-solving rather than bespoke data wrangling. At the same time, dataset stewardship must address sector-specific concerns: representativeness and drift, fairness and disparate impact, and confidentiality.

2.6.4. Open-source software as innovation infrastructure

The economics of open-source software (OSS) highlights motivations, governance, and competitive implications (89) and frames open development as a private-collective model in which free revealing can maximize sectoral innovation under broad conditions (90). Firm-level evidence indicates that contributing, rather than merely consuming, yields learning benefits and higher value capture from shared technologies (91). At the ecosystem level, OSS participation is associated with higher rates and distinctive directions of entrepreneurial activity across countries (92).

In financial services, OSS libraries and platforms for data engineering, cryptography, explainability constitute general-purpose infrastructure that underpins faster iteration and regulator-friendly transparency in production systems. Reference implementations reduce integration costs and serve as de facto standards; readable source code improves inspectability for model risk and compliance; and community maintenance distributes the burden of security patches and performance improvements across a broader contributor base. These properties mitigate vendor lock-in and strengthen organizational capability-building, which is critical where systems must evolve with regulation and market structure.

2.6.5. Open standards, APIs, and technical governance: from data policy to sectoral ecosystems

Platform openness involves trade-offs: "how open is open enough?" depends on governance choices that balance generativity and control (93). In standard-setting, rules shape diffusion, litigation risk, and market entry (94). A recent Research Policy special issue reviews the evolving standards–innovation nexus, with direct relevance to data and API economies (95). Practically, open banking infrastructures show how standard APIs, shared security profiles, and reference implementations lower integration costs and enable third-party complements; open platforms such as the Open Bank Project exemplify an OSS path to PSD2/Open Banking compliance and ecosystem experimentation (96).

For Open Finance, technical governance is the layer where policy becomes programmable interoperability. Mature ecosystems define canonical data models, authentication and consent flows, eventing semantics, and nonfunctional requirements (availability, latency, rate limits), together with certification, conformance testing, and lifecycle management for breaking changes. Observability and incident reporting complete the loop between operators and regulators. When combined with the open science mechanisms outlined above, these governance choices translate openness into measurable changes in innovation outcomes.

3 | Methodology

This chapter describes, in order, the data sources and analytical universe; the procedures for collection, integration, and cleaning; the variables (qualitative and quantitative) used in the analysis; the finance thematic subset; and the analytical strategy (time series, pre/post tests, and segmented regressions).

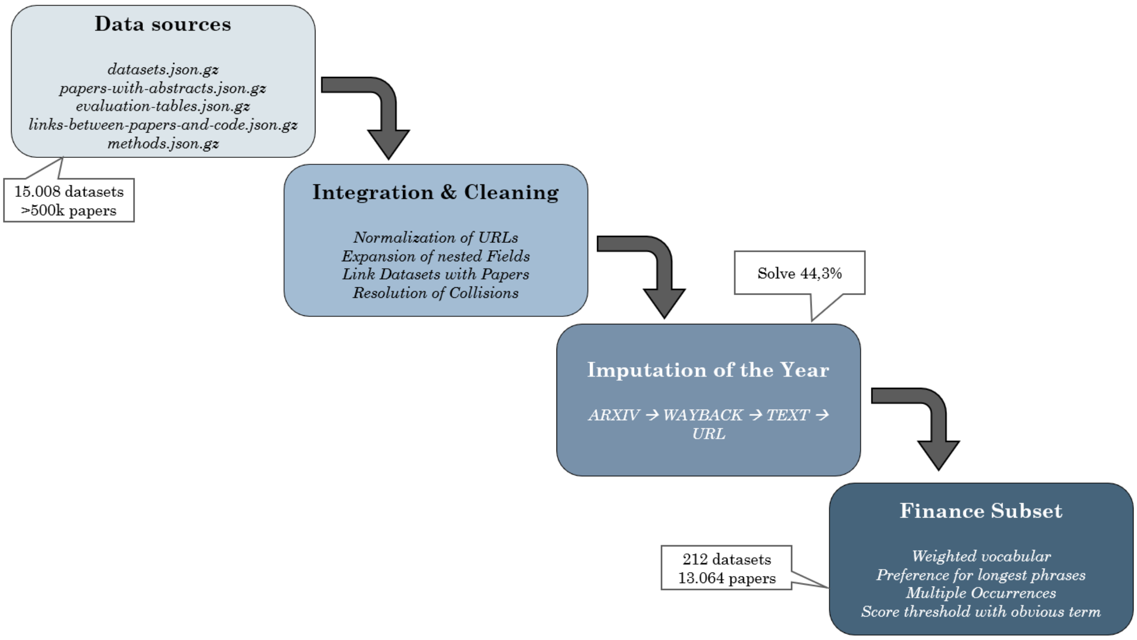

Figure 3.1: Overview of the data pipeline: *Papers with Code* → cleaning/integration → year imputation → thematic subset → series/panels.

3.1. Data sources and analytical universe

The public *Papers with Code* datasets were obtained through the official repository as compressed JSON files. The analytical core comprises:

- ***datasets.json.gz***: a table in which each row represents a dataset. Main columns:
  - `url`: unique locator of the dataset within *Papers with Code* , used as a logical primary key in joins.

- `name`: abbreviated name, useful for display and quick disambiguation.
- `full_name`: standardized full name of the dataset.
- `homepage`: URL of the dataset's original source, when available.
- `description` and `short_description`: long and short textual descriptions, respectively, employed for thematic profiles and the search for temporal cues.
- `parent_dataset`: reference to a parent dataset when the record represents a variation or subset.
- `image`, `audio`, `thumbnail`,
- `paper`: structured citation (title + URL) of the base article that describes the dataset.
- `introduced_date`: date the dataset was introduced on the platform, used for time series.
- `license_name` and `license_url`: licensing metadata, when provided.
- `warning`: curated alert messages.
- `modalities`: data modalities (image, text, audio, graphs, time series, etc.), useful for characterizing the type of information processed.
- `tasks`: list of tasks to which the dataset is linked; used for linkage with the evaluation universe.
- `languages`: covered languages (when applicable).
- `variants`: synonyms/abbreviations by which the dataset is referred to in the literature and repositories.
- `num_papers`: number of papers in *Papers with Code* that use or cite the dataset.
- `data_loaders`: pointers to repositories and code snippets for loading the dataset (frameworks and URLs).

• ***papers-with-abstracts.json.gz***: a table in which each row represents a paper. Main columns:

- `paper_url`: locator of the paper within *Papers with Code* , used as a linking key.
- `arxiv_id`: identifier of the paper on *arXiv*, essential for dating publications and tracking versions.
- `title` and `abstract`: title and structured abstract, core fields for the thematic selection.
- `url_abs` and `url_pdf`: links to the page/file on *arXiv* or the conference.
- `proceeding`, `conference`, `conference_url_abs`, `conference_url_pdf`: dissemination metadata (when available).

- `authors`: list of authors as indexed by *Papers with Code* .
- `tasks`: set of tasks covered in the paper.
- `date`: publication/deposit date indexed in *Papers with Code* , used as a temporal marker for the series.
- `reproduces_papers`: count of replicated studies (when recorded).
- `methods`: structured list of methods employed, with summarized metadata.

- ***evaluation-tables.json.gz***: structure of evaluation tasks/benchmarks. Main columns:
  - `task`: named identifier of the task, a central axis of organization.
  - `description`: objective and context of the task.
  - `synonyms`: alternative terms for standardizing search/linkage.
  - `subtasks`: hierarchical breakdown into subtasks.
  - `categories`: higher-level thematic groupings.
  - `source_link`: external official source, when available.
  - `datasets`: list of datasets relevant to the benchmark (basis for dataset↔paper links).

- ***links-between-papers-and-code.json.gz***: links between papers and repositories. Main columns:
  - `paper_url`, `paper_title`, `paper_arxiv_id`, `paper_url_abs`, `paper_url_pdf`: metadata of the paper.
  - `repo_url`: address of the associated code repository.
  - `is_official`, `mentioned_in_paper`, `mentioned_in_github`: indicators of official status and mentions.
  - `framework`: main framework used in the repository.

- ***methods.json.gz***: catalog of methods. Main columns:
  - `url`, `name`, `full_name`: identifiers and names of the method.
  - `description`: brief technical description.
  - `paper`, `source_url`, `source_title`: reference to the article that introduces the method.
  - `introduced_year`: year the method was introduced according to the index.
  - `code_snippet_url`: implementation pointer/snippet.

- `num_papers`: count of associated papers.
- `collections`: collection/area of classification.

Papers with Code does not provide consistent jurisdiction fields that would allow a reliable EU versus Brazil stratification. Deriving geography from affiliations or language tokens would introduce measurement error and selection biases. Accordingly, the analysis uses global series and treats the EU (2018) and Brazil (2021) milestones as temporal anchors for interpretation rather than within-sample treatments.

3.2. Descriptive summary

To characterize the *Papers with Code* universe and the finance subset in a compact, reproducible way, this section reports two summary tables: one for datasets and one for papers. Each line corresponds to a descriptive indicator and is computed consistently across the full universe and the finance subset.

Table 3.1: **Descriptive summary — Datasets**

Indicator	**Result**
Number of Datasets (total)	15,008
Number of Datasets (finance)	212
Metadata coverage	Tasks: 31.6% Modalities: 27.4% Date: 68.8%
Number of Tasks per dataset (mean)	1.38
Number of Modalities per dataset (mean)	0.75
Modality distribution (Top-5)	Text: 38.5% Time series: 12.1% Tabular: 12.1% Financial: 11.0% Audio: 6.0%
Multi-task (%)	31.7%
Multi-modal (%)	15.8%
Papers per dataset (mean)	26.8
Datasets without associated paper	15.9%

Notes: "Finance" selection follows the weighted keyword model (requires ≥1 obvious term).

Table 3.2: Descriptive summary — Papers

Indicator	Result
Number of Papers (total)	509,455
Number of Papers (finance) and share of total	13,064 (2.56%)
Metadata coverage	Authors: 2.4% Abstract: 2.8% Tasks: 23.9%
Authors per paper (mean)	4.26
Methods per paper (mean)	1.28
Tasks per paper (mean)	1.99

Notes: "Finance" selection follows the weighted keyword model (requires ≥1 obvious term).

3.3. Collection, integration, and cleaning

Collection and storage: The `*.json.gz` files were downloaded directly from the official repository for efficient reading.

```
def download_all_datasets():
    datasets = {
        "datasets": "https://production-media.paperswithcode.com/about/datasets.json.gz",
        "papers-with-abstracts": "https://production-media.paperswithcode.com/about/papers-with-abstracts.json.gz",
        "evaluation-tables": "https://production-media.paperswithcode.com/about/evaluation-tables.json.gz",
        "links-between-papers-and-code": "https://production-media.paperswithcode.com/about/links-between-papers-and-code.json.gz",
        "methods": "https://production-media.paperswithcode.com/about/methods.json.gz"
    }
```

Listing 3.1: Programmatic download of *Papers with Code* archives (`.json.gz`)

Integration across tables: The procedure performs:

- normalization of `url` and keys such as `paper_url`, ensuring uniqueness;
- expansion of nested fields (`tasks`, `modalities`, `languages`) into normalized tabular structures;
- Dataset↔Papers linkage from `evaluation-tables.json.gz` (extraction of names in the `datasets` column and merging with the `datasets` and `papers` bases);
- conservative resolution of collisions by `full_name`/`variants` with threshold-controlled string matching

Imputation of the dataset introduction year: For cases without `introduced_date`, a multi-source pipeline with fixed priority was employed:

- **ARXIV**: the year of the first paper that introduces the dataset is sought; in its absence, the earliest use detected via EVALUATION links and via repositories in `data_loaders` is adopted.
- **WAYBACK**: the earliest captured year for the canonical `homepage` is considered, preferring years explicitly indicated on the archived page.
- **TEXT**: scanning of `verified_name`, `full_name`, `description`, `short_description` for `YYYY` sequences with contexts such as "introduced in" / "released".
- **URL**: extraction of a year from patterns in the `homepage` or simple repository metadata.

The algorithm stops at the first credible source, records provenance, and stores detailed candidates. Preexisting values are not overwritten. In the universe of 15,008 datasets, 2,143 of the 4,835 originally missing cases were resolved ($\approx$ 44.3%): 71% via arXiv, 23% via text, 5.6% via Wayback, and 0.4% via URL, leaving 2,692 unresolved.

```
# ------------------------------
# Unified ARXIV (official + usage)
# ------------------------------
def arxiv_official_candidates(row) -> Tuple[List[int], str]:
    ids, titles = set(), set()
    ids1, tit1 = _parse_paper_cell(row.get("verified_paper"))
    ids2, tit2 = _parse_paper_cell(row.get("paper"))
    ids |= ids1 | ids2
    titles |= tit1 | tit2
    years = set()
    err = []
```

```
    if ids:
        any_year = False
        for aid in ids:
            y = arxiv_year_map.get(aid)
            if y:
                years.add(int(y))
                any_year = True
        if not any_year:
            err.append("ID_NOT_IN_ARXIV_MAP")
    else:
        err.append("NO_ARXIV_ID")
    if titles:
        any_title_year = False
        for tit in titles:
            key = _norm_title(tit)
            y = title_year_map.get(key)
            if y:
                years.add(int(y))
                any_title_year = True
        if not any_title_year:
            err.append("TITLE_NOT_IN_TITLE_MAP")
    else:
        err.append("NO_TITLE")
    if years:
        return sorted(set(int(y) for y in years if YEAR_MIN <= int(y) <= YEAR_MAX)), ""
    return [], "+".join(sorted(set(err))) if err else "UNKNOWN"
# ------------------------------
# TEXT
# ------------------------------
pend_text = base_mask  # process for all eligible base
res = df.loc[pend_text].apply(find_year_in_texts, axis=1, result_type="expand")
df.loc[pend_text, "year_text"] = clamp_years(res[0])
df.loc[pend_text, "err_text"] = res[1].mask(res[1] == "", pd.NA)

# ------------------------------
# URL
# ------------------------------
pend_url = base_mask
res = df.loc[pend_url].apply(find_year_in_urls, axis=1, result_type="expand")
df.loc[pend_url, "year_url"] = clamp_years(res[0])
df.loc[pend_url, "err_url"] = res[1].mask(res[1] == "", pd.NA)

# ------------------------------
```

```python
# WAYBACK -- use external CSV (url -> first year)
# ------------------------------
WAYBACK_CSV_PATH = "wayback_years_checkpoint.csv"
for c in ["year_wayback", "err_wayback"]:
    if c not in df.columns:
        df[c] = pd.NA
pend_wayback = base_mask

def _load_wayback_csv(path: str) -> Optional[pd.DataFrame]:
    if not os.path.exists(path):
        return None
    wb = pd.read_csv(path)
    # Minimal normalization
    if "url" not in wb.columns or "year" not in wb.columns:
        raise RuntimeError("wayback_years_checkpoint.csv must have columns 'url' and 'year'.")
    wb["url"] = wb["url"].astype(str).str.strip().str.rstrip("/")
    wb["year"] = pd.to_numeric(wb["year"], errors="coerce")
    # Keep the first year per URL (min value)
    wb = wb.dropna(subset=["url"])
    wb_min = (
        wb.groupby("url", as_index=False)["year"]
          .min()
          .rename(columns={"year": "wb_year"})
    )
    wb_min["wb_year"] = wb_min["wb_year"].where(
        (wb_min["wb_year"] >= YEAR_MIN) & (wb_min["wb_year"] <= YEAR_MAX)
    )
    try:
        wb_min["wb_year"] = wb_min["wb_year"].astype("Int64")
    except Exception:
        pass
    return wb_min
```

Listing 3.2: Unified imputation of dataset introduction year: ARXIV → WAYBACK → TEXT/URL heuristics

Quality best practices: Duplicates were removed, task/method names were standardized, temporal consistency was verified (truncation of years outside a plausible range), and missingness reports by field were issued.

3.4. Study variables

Primary variables

- *From `datasets.json.gz`*: `url` (dataset key in *Papers with Code*), `name/full_name` (short and full names), `homepage` (source URL), `description/short_description` (descriptions), `tasks` (associated tasks), `languages` (languages), `variants` (synonyms), `num_papers` (number of associated papers), `introduced_date` (introduction date), `introduced_date_method` (winning imputation source), and `data_loaders` (code pointers for loading).

- *From `papers-with-abstracts.json.gz`*: `paper_url` (paper key in *Papers with Code*), `arxiv_id` (identifier on *arXiv*), `title/abstract` (title and abstract), `date` (indexed date), `tasks` (tasks), and `methods` (methods).

Auxiliary variables

- `finance_score`: continuous score computed by a weighted vocabulary that differentiates obvious and indirect terms; used to classify relevance to finance.

- `finance_flag`: binary indicator marking records above the threshold and with at least one obvious term detected.

- `year_intro`: year derived from `introduced_date` (datasets), after safe conversion.

- `year_pub`: year derived from `date` (papers).

Aggregated variables

- D_t^{fin} and P_t^{fin}: annual counts of finance datasets and finance papers, respectively.

- D_t and P_t: total annual counts.

- $s_t^D = D_t^{fin}/D_t$ and $s_t^P = P_t^{fin}/P_t$: annual shares of finance in datasets and papers.

- `post_EU2018`: indicator (dummy) that takes the value 1 for $t \geq 2018$ (alignment with the PSD2 implementation window)

- `post_BR2021`: indicator for $t \geq 2021$ (Open Finance Brazil go-live window).

3.5. Finance Thematic Subset

The thematic selection was structured by a **term-weight scoring model**:

- **Weighted vocabulary**: a group of obvious terms with weight 5; and a group of indirect terms with weight 1.
- **Counting rules**: preference for the longest n-gram when overlaps exist; multiple occurrences accumulate points.
- **Selection rule**: a record enters the subset if it reaches the score threshold and contains at least one obvious term.
- **Validation**: manual validation on a sample of 50 items from the subset

Applying the model to the *Papers with Code* JSONs resulted in **212** datasets ($\approx 1.4\%$) and **13,064** papers ($\approx 2.56\%$) classified as finance.

The selection threshold was fixed ex ante based on face-validity after a manual check. A full metric-based validation (ROC/PR, out-of-sample) is left for subsequent work.

To avoid divergence across documents, the detailed lists of obvious and indirect terms are maintained as a single source of truth in the companion code artifacts and reproduced in Appendix E. Ambiguous tokens such as "bank" (e.g., “memory bank” in computer vision) are documented explicitly. The screening mitigates such ambiguities by (i) applying longest-n-gram precedence (so “open banking” eclipses “banking”/“bank”), and (ii) enforcing the global threshold rule. No retuning or reclassification is introduced in this thesis; it simply documents the vocabulary and its caveats for transparency and reproducibility.

```

    # Definition of keywords/phrases
    KEYWORDS = {
        "obvious": [
            r"\bfinanc\w*",            # finance, financial, finances, financing
            r"\bfintech\w*",
            r"\bopen[\s\-]*bank(?:ing)?",   # open banking / open-bank / openbanking
            r"\bopen[\s\-]*financ\w*",      # open finance / open-finance
        ],
        "indirect": [
            r"\bbank(?:ing|s)?\b",     # bank, banking, banks (avoids 'treebank')
            r"\bloan\w*",
            r"\bcredit\w*",
            r"\binsur\w*",             # insurance, insure, insurtech
            r"\btrading\b|\btrader(?:s)?\b|\btrade\b",
            r"\bstock(?:s)?\b",
```

```
            r"\bmarket(?!ing)\w*\b",  # avoids 'marketing', but allows marketplace, markets
            r"\binvest\w*",
            r"\bpayment\w*",          # payment, payments
            r"\brisk\w*",
        ]
    }

    # Summary parameters
    HIGH_SCORE_THRESHOLD = 3     # "high score" (adjust as needed)
    HIGH_SCORE_TOP_PCT = 0.10    # top 10\% as alternative "high" (also adjustable)

    # === PROCESS ============================================================

    # Minimal checks
    _required_cols = {"full_name", "description", "modalities", "tasks"}
    missing = _required_cols - set(df_datasets.columns)
    if missing:
        raise ValueError(f"Missing columns in df_datasets: {missing}")

    compiled = _compile_patterns(KEYWORDS)

    # Normalized text per column (avoids recomputing in loops)
    texts_by_col = {
        "full_name": df_datasets["full_name"].apply(_to_text),
        "description": df_datasets["description"].apply(_to_text),
        "modalities": df_datasets["modalities"].apply(_to_text),
        "tasks": df_datasets["tasks"].apply(_to_text),
    }

    # Explainability columns (lists of tokens per group/column)
    for group in ("obvious", "indirect"):
        for col in ("full_name", "description", "modalities", "tasks"):
            df_scored[f"hits_{group}_{col}"] = None  # filled below
        df_scored[f"hits_{group}_all"] = None  # aggregation of all columns

    # Vectorized loop by index (keeps clarity and allows fine audit)
    for idx in df_scored.index:
        row_texts = {col: texts_by_col[col].iat[idx] for col in texts_by_col}
        matches = _find_matches_by_col(row_texts, compiled)
        scores = _score_from_matches(matches)

        # Fill explainability per column
        for group in ("obvious", "indirect"):
```

```python
            all_tokens = []
            for col in ("full_name", "description", "modalities", "tasks"):
                tokens = matches[(group, col)]
                df_scored.at[idx, f"hits_{group}_{col}"] = tokens if tokens else []
                all_tokens.extend(tokens)
            # Aggregation (may keep repetition if COUNT_MULTIPLE_OCCURRENCES=True)
            df_scored.at[idx, f"hits_{group}_all"] = all_tokens if all_tokens else []

        # Fill scores
        df_scored.at[idx, "score_obvious"] = scores["score_obvious"]
        df_scored.at[idx, "score_indirect"] = scores["score_indirect"]
        df_scored.at[idx, "score_total"] = scores["score_total"]
```

Listing 3.3: Datasets: weighted keyword scoring (obvious vs. indirect)

3.6. Data analysis procedures

Construction of series and panels: Annual series were generated for 2010–2024 with four outcomes: D_t^{fin}, P_t^{fin}, s_t^D, and s_t^P. The series were aligned with two institutional milestones: 2018 (Europe) and 2021 (Brazil).

```python
try:
    total_datasets = len(df_datasets)
    total_finance_datasets = len(df_datasets_finance)
    pct_finance_datasets = (total_finance_datasets / total_datasets * 100) if
    ↪ total_datasets else 0.0

    total_papers = len(df_papers)
    total_finance_papers = len(df_papers_finance)
    pct_finance_papers = (total_finance_papers / total_papers * 100) if total_papers
    ↪ else 0.0

    df_overview_datasets = pd.DataFrame(
        [{
            "Total Datasets": total_datasets,
            "Finance Datasets": total_finance_datasets,
            "Finance \%": round(pct_finance_datasets, 2)
        }]
    )
    df_overview_papers = pd.DataFrame(
        [{
            "Total Papers": total_papers,
            "Finance Papers": total_finance_papers,
```

```python
            "Finance \%": round(pct_finance_papers, 2)
        }]
    )

def create_temporal_table():
    # Reference years for Open Finance
    OPEN_FINANCE_EUROPE = 2018  # PSD2 implemented
    OPEN_FINANCE_BRAZIL = 2021  # Open Finance Brazil implemented

    # Prepare dataset data
    df_datasets_temp = df_datasets.copy()
    df_datasets_temp['date'] = pd.to_datetime(df_datasets_temp['introduced_date'], errors='coerce')
    df_datasets_temp['year'] = df_datasets_temp['date'].dt.year

    # Prepare paper data
    df_papers_temp = df_papers.copy()
    df_papers_temp['date'] = pd.to_datetime(df_papers_temp['date'], errors='coerce')
    df_papers_temp['year'] = df_papers_temp['date'].dt.year

    # Prepare financial dataset data
    df_datasets_finance_temp = df_datasets_finance.copy()
    df_datasets_finance_temp['date'] = pd.to_datetime(df_datasets_finance_temp['introduced_date'], errors='coerce')
    df_datasets_finance_temp['year'] = df_datasets_finance_temp['date'].dt.year

    # Prepare financial paper data
    df_papers_finance_temp = df_papers_finance.copy()
    df_papers_finance_temp['date'] = pd.to_datetime(df_papers_finance_temp['date'], errors='coerce')
    df_papers_finance_temp['year'] = df_papers_finance_temp['date'].dt.year

    # Define year range
    min_year = 2010
    max_year = 2024
    years = list(range(min_year, max_year + 1))

    # Create consolidated table
    temporal_data = []

    for year in years:
        # Count total datasets
        datasets_total = len(df_datasets_temp[df_datasets_temp['year'] == year])

```

```python
        # Count total papers
        papers_total = len(df_papers_temp[df_papers_temp['year'] == year])

        # Count financial datasets
        datasets_finance = len(df_datasets_finance_temp[df_datasets_finance_temp['year'] == year])

        # Count financial papers
        papers_finance = len(df_papers_finance_temp[df_papers_finance_temp['year'] == year])

        # Calculate percentages
        datasets_finance_pct = (datasets_finance / datasets_total * 100) if datasets_total > 0 else 0
        papers_finance_pct = (papers_finance / papers_total * 100) if papers_total > 0 else 0

        temporal_data.append({
            'year': year,
            'datasets_total': datasets_total,
            'papers_total': papers_total,
            'datasets_finance': datasets_finance,
            'papers_finance': papers_finance,
            'datasets_finance_pct': round(datasets_finance_pct, 2),
            'papers_finance_pct': round(papers_finance_pct, 2),
            'open_finance_europe': year >= OPEN_FINANCE_EUROPE,
            'open_finance_brazil': year >= OPEN_FINANCE_BRAZIL
        })

    df_temporal = pd.DataFrame(temporal_data)
    return df_temporal, OPEN_FINANCE_EUROPE, OPEN_FINANCE_BRAZIL
```

Listing 3.4: Construct annual counts and finance shares (2010–2024) for datasets and papers

Pre/post comparisons: For each metric and each breakpoint:

- The distributional shape is assessed using the normality test by Shapiro (97);
- When normality is plausible in both periods, the *t*-test of mean differences is applied (98, 99);
- Otherwise, the nonparametric Mann–Whitney *U* test is used (100);
- Magnitudes are quantified by Cohen's effect size (*d*) (101).

Segmented regressions (ITS): A time-series model with changes in level and slope is estimated:

$$Y_t = \beta_0 + \beta_1\, t + \beta_2\, \texttt{post} + \beta_3\, (t \times \texttt{post}) + \varepsilon_t,$$

in which β_2 captures the immediate change (level) and β_3 the trend change (slope). The strategy follows the Interrupted Time Series literature (102, 103); when necessary, errors are evaluated for autocorrelation and, in extensions, robust estimators of the Newey–West type are considered (104). Fit is reported by R^2 and visual diagnostics of residuals.

Design choices and inference limits: Series are annual and the post-2021 window is short, which limits statistical power for detecting slope changes. Reported p-values should be read as descriptive; no adjustment for multiple comparisons across outcomes and anchors is applied. Robust (HAC/Newey–West) standard errors are not implemented in the reported fits; serial correlation would widen confidence intervals. Finally, no explicit control group is used; jurisdictional counterfactuals are out of scope given the global design.

3.7. Ethical considerations, limitations, and transparency

The inferences reflect the coverage and biases of the *Papers with Code* ecosystem. Temporal imputation adopts a principle of minimal intervention (do not overwrite existing values; record provenance). The association of regulatory milestones with the series is employed for temporal alignment and does not imply strict causality. Possible concurrent breaks and autocorrelation are addressed by ITS inspections and, where appropriate, by robust errors.

This version reports a 50-item manual sanity check and does not publish precision/recall or ROC/PR curves. A full metric-based validation based on a stratified gold set is planned as post-deposit work. Screening results should therefore be read as transparent and conservative rather than fully optimized.

4 | Results

4.1. Data Scope

This study draws on the public databases exposed by the *Papers with Code* platform, which links machine learning papers to code, datasets, tasks, and benchmark methodologies. Following the data preparation protocol described in the Methodology chapter, we implemented a conservative keyword-based screening across titles, abstracts, descriptions, modalities and task fields.

Three patterns support the validity of this screening. First, independent *Papers with Code* tables (papers, datasets, and evaluation tables) converge on a consistently minor, yet stable, financial footprint when measured as a share of the total corpus. Second, absolute counts and within-domain shares move in the same direction over time, suggesting that the filter captures domain-level dynamics rather than noise. Third, results are robust to inclusion of description fields beyond titles/abstracts in order to mitigate false negatives.

4.2. Temporal evolution

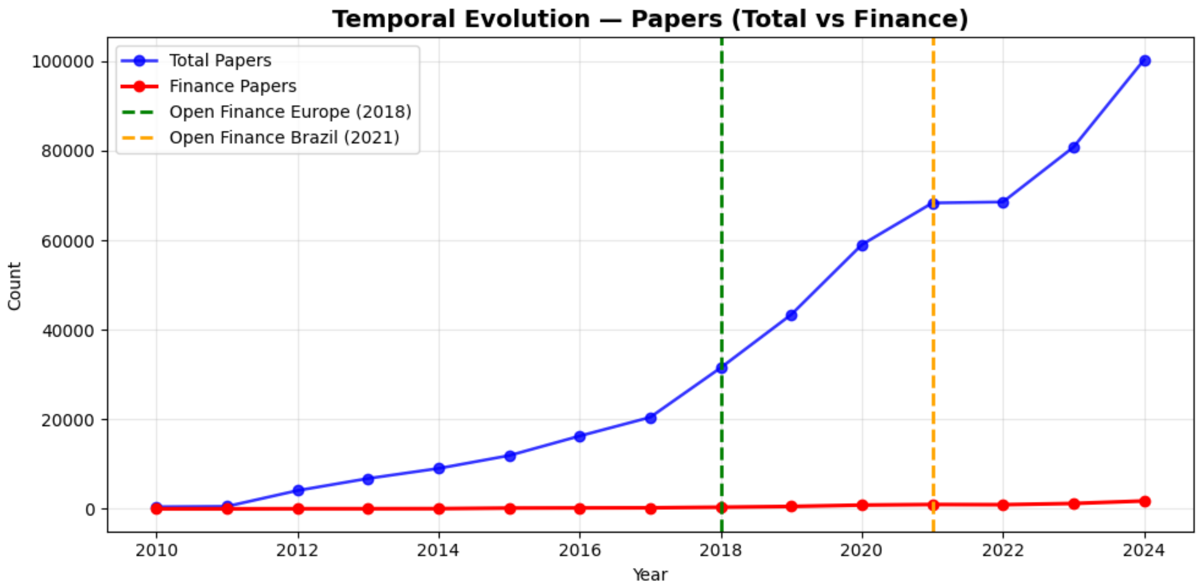

Figure 4.1: Temporal evolution of papers (2010–2024): total vs. finance. ***Note: global Papers with Code series; vertical dashed lines mark PSD2 (EU, 2018) and Open Finance Brazil (2021) as temporal anchors, not region-specific treatments.***

The figure 4.1 shows a near-monotonic expansion in the total volume of papers tracked by *Papers with Code* , with clear accelerations around 2018 and again after 2020. Financial papers follow the same broad trajectory, albeit at a much smaller scale. When expressed as a share of the total, the financial fraction remains close to the 1–1.5% range for most of the period and trends upward in the most recent years, approaching roughly 1.7–1.8% in 2024 (see the bottom-right panel of Figure 4.2). This pattern indicates that finance grows not only in absolute terms but also in relative terms within the broader machine learning literature captured by *Papers with Code* .

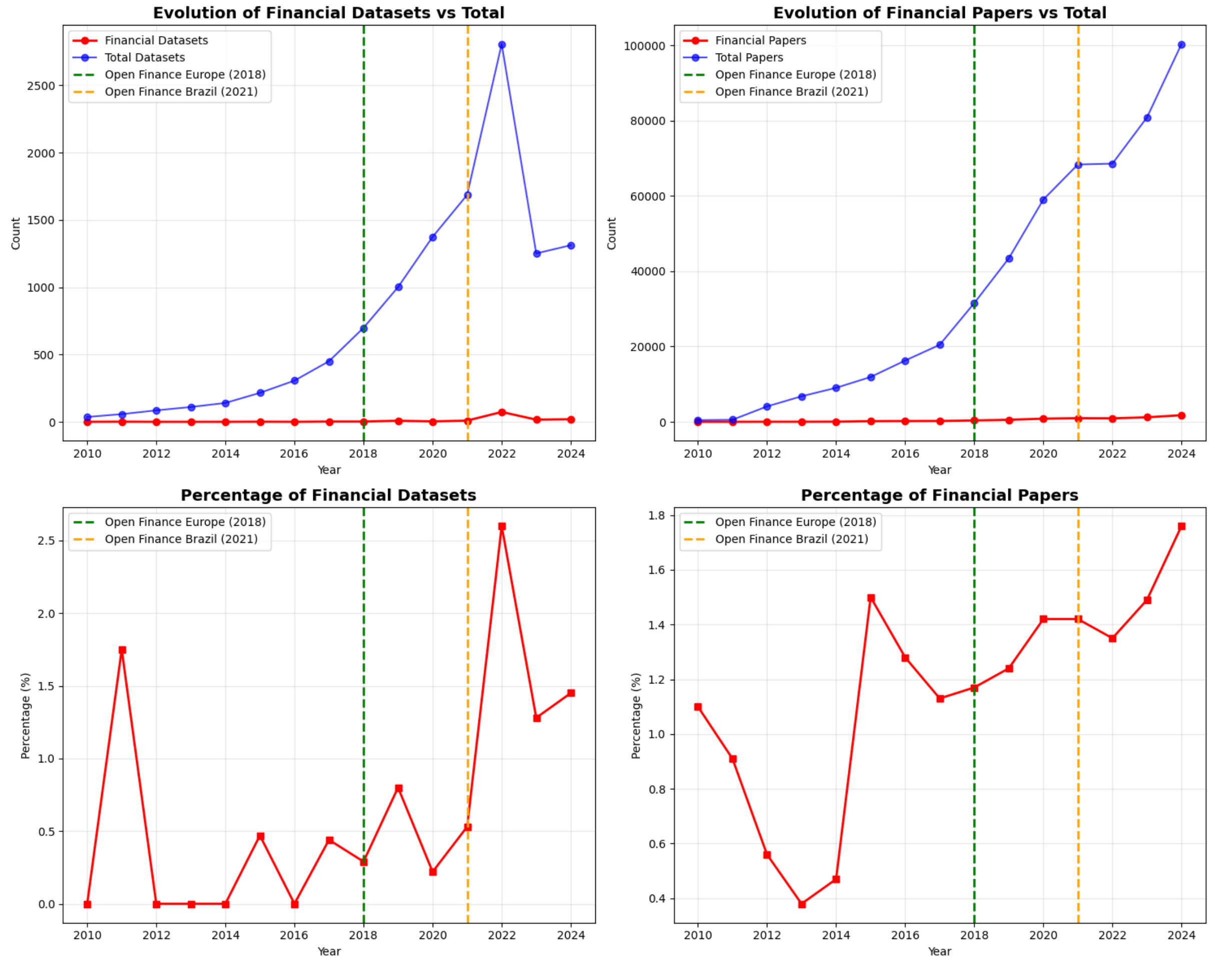

Figure 4.2: Four-panel view: (top) absolute counts for finance vs. totals (datasets and papers); (bottom) finance shares over time. ***Note: global Papers with Code series; vertical dashed lines mark PSD2 (EU, 2018) and Open Finance Brazil (2021) as external temporal anchors.***

Figure 4.3 documents a gradual increase in the total number of datasets up to 2021, followed by a pronounced one-year jump in 2022 and a reversion to a higher plateau in 2023–2024. Financial datasets mirror the overall expansion, with a visible surge around 2021–2022. As a share of all datasets, finance peaks near 2.6% in 2022 and stabilizes around 1.3–1.5% thereafter (bottom-left panel of Figure 4.2). Taken together with the smoother and more persistent rise in the proportion of financial papers (bottom-right panel), this staggered pattern is consistent with data-availability shocks preceding downstream research outputs.

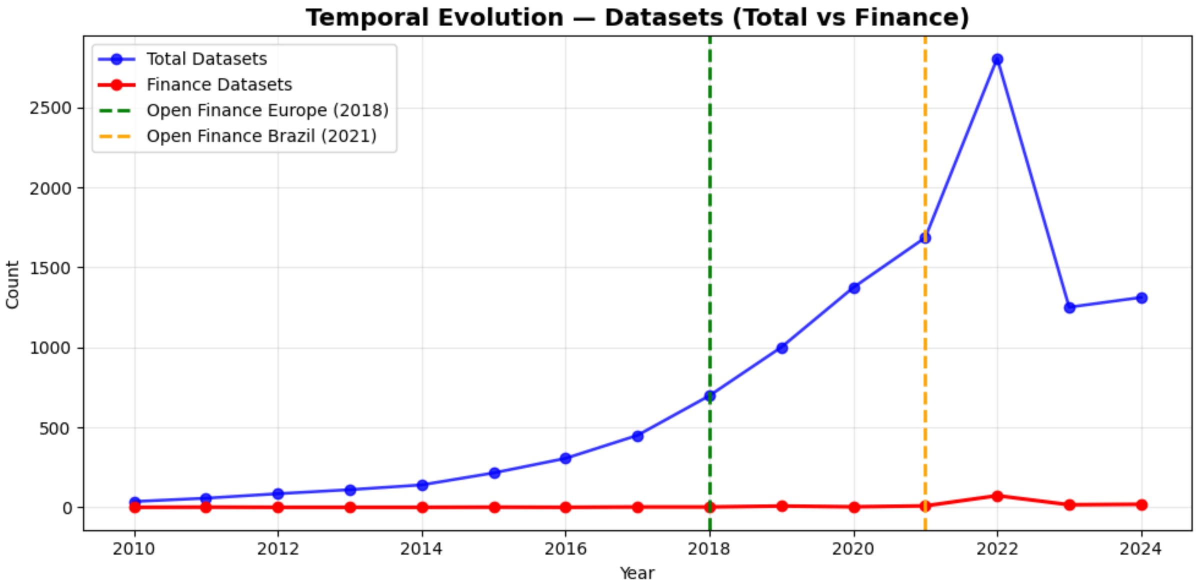

Figure 4.3: Temporal evolution of datasets (2010–2024): total vs. finance. ***Note: global Papers with Code series; PSD2 (2018) and Open Finance Brazil (2021) are used as temporal anchors, not treatments.***

Across papers and datasets, the temporal figures suggest two complementary facts: (i) the field of finance is a small fraction of the *Papers with Code* universe, and (ii) its footprint is non-negligible and has increased in recent years both absolutely and proportionally. Importantly, the timing is not identical across outputs: the jump in datasets is sharper and earlier (2021–2022), while the proportional rise in papers is smoother and more persistent (post-2018 through 2024). This staggered pattern is consistent with the idea that data availability shocks precede, and possibly enable, downstream research outputs.

4.3. Descriptive distributional evidence

To probe whether the temporal breaks visible in Figures 4.1–4.2 coincide with meaningful distributional changes, the summary of finance-related outputs is based on before-and-after comparisons around two reference dates: 2018 (Europe) and 2021 (Brazil). The boxplots in Figure 4.4 reveal upward shifts in both the median and the upper tail for financial papers and datasets after each milestone. The shift is more pronounced for datasets in the post-2021 period than for papers, which display a steadier rise. In proportional terms, the within-domain shares also move up after both cutoffs, with the post-2021 distributions stochastically dominating their pre-2021 counterparts in datasets.

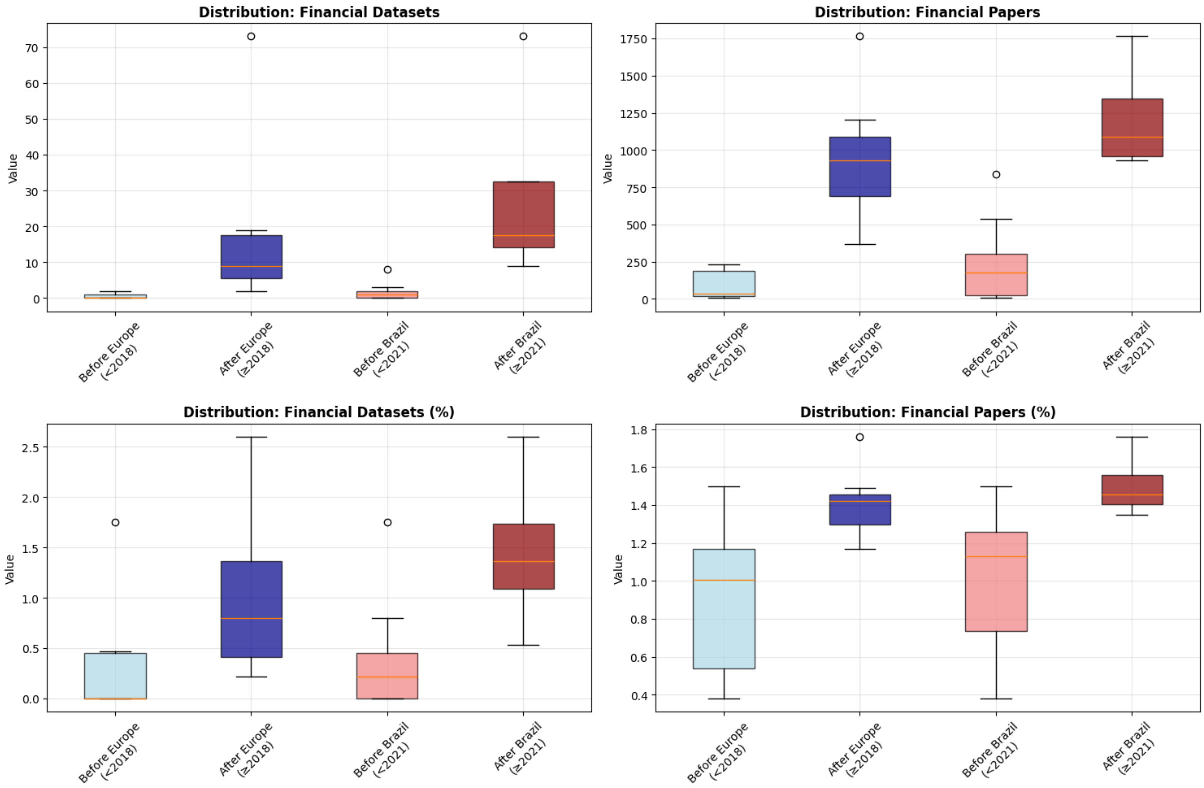

Figure 4.4: Distributional comparisons before and after the 2018 (Europe) and 2021 (Brazil) milestones.

4.4. Associations among financial indicators

The correlation matrix in Figure 4.5 summarizes linear associations across the four finance metrics used in this study (absolute and proportional counts for papers and datasets). Correlations are positive throughout. Notably, absolute and proportional measures within the same output type are strongly associated, and the cross-output association between financial datasets and financial papers is also positive. While correlations should not be interpreted causally, this pattern is consistent with the notion that increases in data supply co-occur with increases in domain-specific research activity.

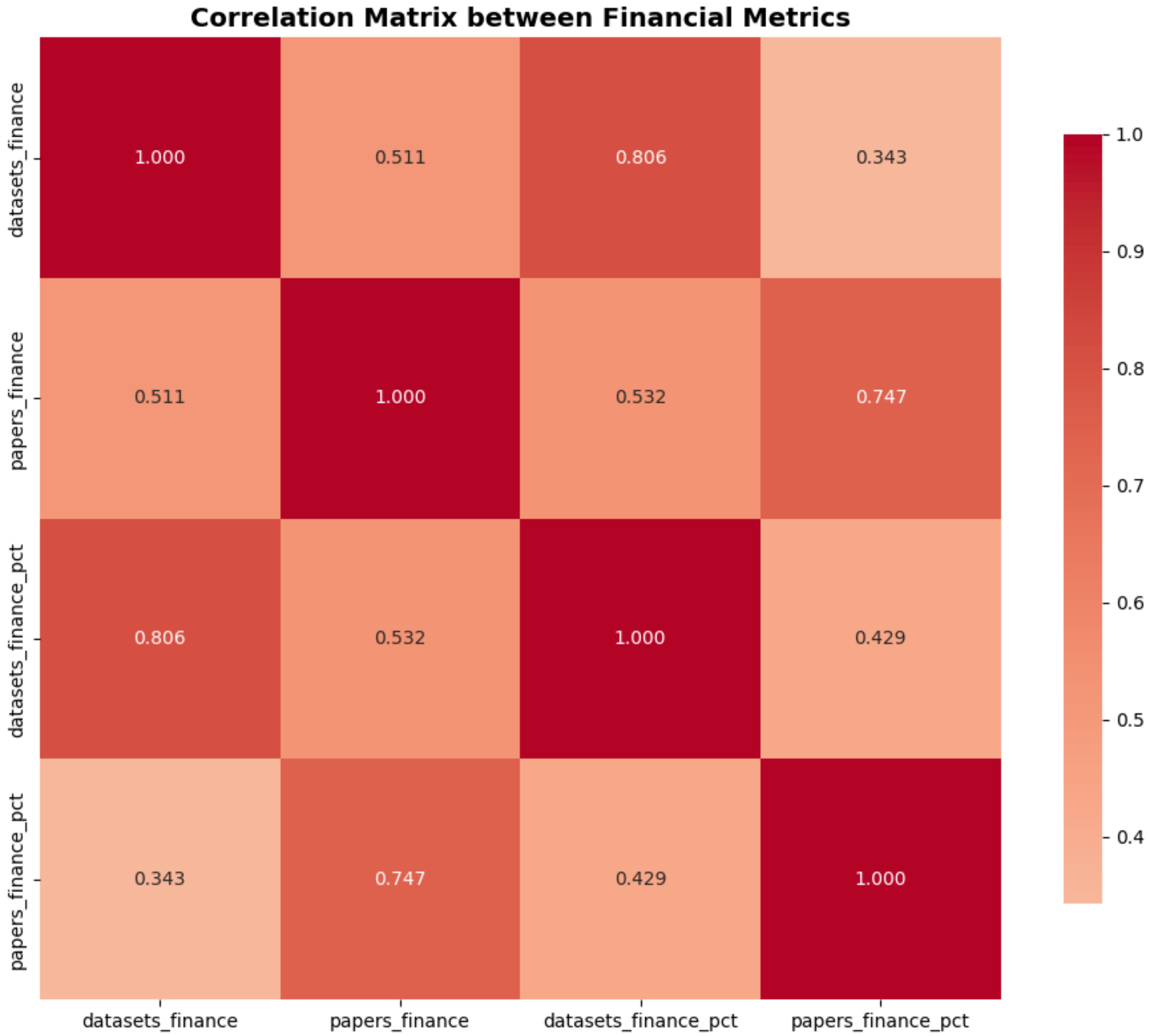

Figure 4.5: Correlation matrix across finance metrics (absolute and proportional counts for datasets and papers).

5 | Discussion

5.1. Linking objectives, methods, and results

(O1) Map and quantify the evolution of research and data in finance within *Papers with Code* . The time–series built from the *Papers with Code* corpus maps both the absolute and proportional evolution of finance-related outputs. Finance remains a small slice of the overall corpus, yet its levels and shares rise over time. The construction of annual series for datasets (by introduction year) and papers (by publication year) enables a granular reading of these trajectories within a common observation window.

(O2) Assess whether major regulatory initiatives align with observable changes in finance-related outputs. Anchoring the series to European and Brazilian milestones indicates discrete upward shifts around PSD2 implementation (2018) and Brazil's Open Finance go-live (2021). These shifts are most pronounced for datasets after 2021 and for the proportional share of finance papers post-2018. While the design is not causal, the pre/post comparisons and segmented trends are consistent with the view that regulatory standardization expands data access and, through that channel, stimulates scientific and engineering activity in finance.

(O3) Relate data availability to scientific production in the domain. The positive association between finance datasets and finance papers, combined with the sharper, earlier jump in datasets, supports a sequential mechanism in which improved data access and API standardization reduce entry costs for research and applications, with papers materializing after a lag. The sparser prevalence of finance-specific benchmarks in evaluation tables suggests that the consolidation of standardized tasks remains a bottleneck for diffusion and reuse.

5.2. Interpretation and plausible mechanisms

Taken together, the empirical patterns admit a coherent interpretation grounded in the institutional and technical architecture of open finance.

(i) Data-supply shocks with lagged research response. Open finance reforms explicitly target access, interoperability, and security through APIs and consent flows (e.g., PSD2 and associated RTS/SCA in the EU; phased data sharing, payment initiation, and scope extension in Brazil). In innovation theory, this is the "outside-in" channel of open services innovation (16, 17), combined with platformization effects in which sponsors expose stable interfaces that invite complementary entry (8, 9, 10, 11). Our finding of a dataset spike in 2021–2022, followed by smoother, persistent increases in finance papers, aligns with this mechanism: infrastructure and data arrive first; research and applications follow with a delay as capabilities are absorbed and recombined.

(ii) Platform curation and measurement dynamics. Coverage and ingestion policies of aggregators such as *Papers with Code* can shift over time, especially for datasets. Bibliometric work on open banking shows uneven standardization and harmonization across providers and jurisdictions (7), and policy process studies of the UK emphasize how co-regulatory governance accelerated some layers (common APIs) while incumbents strategically slowed others through partnerships and acquisitions (48, 49). These dynamics can produce one-off spikes without altering the medium-run trend, a pattern we observe for datasets.

(iii) Small-base effects. Because finance accounts for a modest share of the *Papers with Code* corpus, proportional measures are sensitive to moderate absolute changes. Reading levels and shares jointly mitigates misinterpretation due to base effects, a standard concern in platform and ecosystem measurement (10, 11).

Comparative alignment with regulatory models

European PSD2 created a legal right of access and an API-secured channel for AIS/PIS, but implementation frictions (heterogeneous security profiles, inconsistent technical specs, and gaps at the interface with GDPR/NIS2) are widely documented (7). The UK's OBIE model added centralized technical coordination and certification, catalyzing a large complementor ecosystem even as "quiet politics" tempered head-on competition with incumbents (48, 49). Brazil's phased Open Finance, tightly integrated with digital public infrastructure (PIX and, prospectively, DREX), broadened scope beyond payments to credit, insurance, investments, and pensions (34, 28, 56, 29). The sharper post-2021 dataset jump fits a setting where interoperable rails (PIX) and governance bodies lower the fixed costs of data publishing and API exposure.

Across governance archetypes, the literature anticipates different trade-offs between speed, interoperability, and risk (23, 62, 63). The empirical patterns here are most consistent with contexts that combined legal access mandates and operational standardization, echoing findings on higher adoption and innovation where mandates and coordination co-exist (64, 50).

Open science mechanisms as transmission channels

Economics of science predicts that open data and code reduce search and replication costs, expand validation audiences, and increase reuse, thereby accelerating cumulative innovation (76, 77, 79). In AI, widely available, well-specified datasets and shared evaluation norms orchestrated community effort and method progress (86, 87, 88). Our results mirror this pattern: increases in finance datasets precede increases in finance papers, but the relative scarcity of finance, specific benchmarks likely slows diffusion. The policy corollary is that documentation, persistent identifiers, licensing clarity, versioning, and conformance tests are not peripheral niceties; they are the levers that convert legal openness into scientific productivity and safer deployment (82, 83, 84, 85).

Behavioral and trust frictions

Consumer adoption and thus the business case for API exposure depends on perceived usefulness, social influence, and initial trust; perceived risk dampens uptake even when access is legally guaranteed (54, 55). These behavioral channels help explain why the research response is smoother and lagged: firms and users must traverse learning curves, and credible use cases accumulate over time, not instantly at legal go-live.

5.3. Implications for research, practice, and policy

This section articulates how the evidence changes how scholars investigate the topic, how firms build and operate, and how public authorities design rulebooks and technical governance.

5.3.1. Implications for researchers

- **Invest in benchmark consolidation:** The literature shows that shared tasks and metrics concentrate effort and accelerate diffusion (86, 87, 88). Our findings point to finance benchmarks (e.g., payment-failure prediction, consent-flow anomaly detection, SME credit using consented data) as a high-leverage public good. Curate datasets with rich metadata, licenses, and baselines; publish reproducible leaderboards.

- **Treat APIs and schemas as research objects:** Platformization and open services innovation suggest that interfaces shape feasible inquiry (8, 17). Documenting API stability, rate limits, and event semantics enables robust experimental design and external validity.

- **Adopt open-science stewardship:** Release code, data loaders, and model cards; use persistent identifiers; track dataset versioning; and report data lineage (85). Where privacy binds, employ tiered access (secure data environments) and privacy-preserving methods while maintaining reproducibility.

- **Design identification around institutional timing:** Our descriptive anchors (2018 EU; 2021 Brazil) motivate event-study designs with leads and placebo windows that leverage institutional calendars rather than ad hoc breaks.

5.3.2. Implications for practice (industry and managers)

- **View Open Finance as capability infrastructure:** The dataset jump preceding paper growth suggests that organizations that professionalize API management (observability, SLAs, deprecation policy) and data governance will capture learning and partnership advantages (8, 10).

- **Shorten the path from data to product:** Establish internal "research-to-production" patterns to reduce compliance friction, consistent with secure, API-first architectures (41, 33).

- **Co-produce open assets:** Contributing to OSS and sector benchmarks yields learning spillovers and reputational benefits (89, 90, 91, 92). In finance, reference connectors, consent UX patterns, and auditable feature pipelines are particularly valuable.

5.3.3. Implications for policy and technical governance

- **Couple legal mandates with operational standardization:** The strongest post-milestone responses occur where access rights co-exist with coordinated interfaces (e.g., OBIE; Brazil's governance layers) (23, 62, 34). Invest in certification, conformance, and change-management regimes that keep APIs usable in practice.

- **Build trust by design:** Explicit security profiles, transparent dispute resolution, and machine-readable rules reduce perceived risk and increase adoption (7, 54, 53). Support secure data environments for research access to high-sensitivity data.

- **Leverage digital public infrastructure:** Brazil's integration with PIX (and DREX in prospect) illustrates how public rails can amplify private innovation (28, 56, 29). Where feasible, align open-finance APIs with instant-payment and identity systems to reduce onboarding costs.

- **Promote cross-border interoperability:** Aligning schemas and consent semantics across jurisdictions compounds innovation and inclusion benefits (66, 32, 75).

5.4. Limitations and threats to validity

- **No jurisdictional stratification:** The *Papers with Code* corpus lacks reliable geography; milestones are used strictly as temporal anchors rather than for EU vs. Brazil splits.
- **Annual aggregation and short post-2021 window:** Power to detect slope changes is limited, particularly after 2021.
- **Multiple testing:** Several outcomes across two anchors were examined without FWER/FDR control; significance is exploratory.
- **Autocorrelation:** HAC/Newey–West corrections were not applied to the reported estimates; serial dependence would widen confidence intervals.
- **Absence of systematic placebos and leads/lags:** Placebo windows and lead/lag diagnostics are not yet included; these are important to sharpen timing claims.
- **Screening metrics and validation:** ROC/PR and precision/recall are not reported; the 50-item check is a sanity check, not a full validation. A larger, year-stratified gold set is planned post-deposit.
- ***Papers with Code* as a measurement lens:** Coverage and ingestion dynamics at *Papers with Code* may bias levels and year-to-year variation; we treat medium-run patterns as more reliable than single-year spikes.

5.5. Avenues for further analysis

1. **Interrupted time series (ITS) with diagnostics:** Extend segmented regressions with explicit lead/lag windows, placebo milestones, and HAC/Newey–West standard errors. Where break dates are uncertain, use grid searches or information-criterion selection over candidate windows.
2. **Descriptive difference-in-differences (DiD):** Use non-finance *Papers with Code* series as a proxy comparison group to benchmark post-break dynamics; treat estimates as descriptive given global coverage and potential composition shifts.
3. **Subdomain granularity:** Disaggregate into salient financial tasks (payments/fraud, SME credit, supervisory text, risk management) to identify where benchmark consolidation would yield the highest marginal returns and to test for heterogeneous treatment timing.

6 | Conclusion

This thesis examined whether the emergence of open-finance regulation operates as a research shock in the global machine-learning ecosystem, using the *Papers with Code* platform as an observable lens on data and research artifacts. Building on a conceptual integration of platformization, open services innovation, and regulation-as-catalyst, the study framed two exogenous regulatory anchors and asked whether finance-related datasets and papers tracked by *Papers with Code* display systematic inflections around those dates (PSD2 in Europe at 2018 and Brazil's Open Finance at 2021).

6.1. Synthesis of findings

Across annual series for 2010–2024, finance remains a small fraction of the *Papers with Code* corpus but increases both in absolute terms and as a share of totals. Visual evidence and pre/post contrasts show upward shifts aligned with the 2018 and 2021 anchors. A pronounced one-year rise in finance datasets occurs around 2021–2022, followed by a higher plateau; the proportional share of finance papers rises more smoothly and persistently from the late 2010s. The correlation structure is positive across metrics, and timing patterns are consistent with a mechanism in which shocks to data availability precede and enable downstream research outputs. These conclusions are descriptive: the anchors are used strictly for temporal alignment in a global series rather than for jurisdiction-specific treatment effects.

6.2. Methodological contributions

The thesis contributes a reproducible pipeline for integrating and auditing *Papers with Code* public JSON archives: (i) normalization and linkage across datasets, papers, tasks, methods, and code; (ii) a transparent, priority-ordered imputation of dataset introduction years that preserves provenance (ARXIV → WAYBACK → TEXT/URL); and (iii) an auditable thematic selector for finance based on a weighted vocabulary that privileges longer n-grams and requires at least one "obvious" term. These components reduce measurement error in temporal series, make classification choices explicit, and support independent replication and extension.

6.3. Substantive implications

Read jointly with the literature on digital transformation and open finance, the results have three practical implications:

First, for researchers and standard-setting communities, the staggered pattern, it means datasets peaking before papers, underscores the catalytic role of high-quality, well-documented, and stable data access. Consolidating finance-specific benchmarks (tasks, metrics, and reference implementations) can convert episodic data spikes into durable research pipelines.

Second, for firms, API-first capability, robust data governance, and contribution to open-source tooling emerge as complementary investments. Clear licensing, conformance suites, and documentation reduce integration frictions and accelerate translation from research to production.

Third, for policymakers, the alignment between regulatory milestones and observable activity in the research-data ecosystem points to the salience of technical governance choices (security profiles, consent and authentication flows, observability and incident reporting, certification regimes). Where trustworthy access and interoperability are programmatically enforced, the ecosystem appears more likely to generate reusable artifacts and measurable scientific outputs.

6.4. Limitations

The design privileges transparency over strong identification and has several limitations. (i) *Papers with Code* offers no reliable, comprehensive geography; hence the analysis uses global series and treats the EU and Brazil dates as external temporal anchors rather than as region-specific interventions. (ii) Annual aggregation and the short post-2021 window limit power to detect slope changes and to separate trend from transitory effects. (iii) Multiple outcomes are inspected around two anchors without family-wise error control; p-values should be read as descriptive. (iv) Serial correlation is not explicitly modeled in reported fits; robust (HAC/Newey–West) errors would widen intervals. (v) The imputation of introduction years, while conservative and provenance-preserving, remains a source of measurement error. (vi) The thematic screen favors precision through a conservative threshold; a fully validated classifier with gold-standard labeling is left to future work. These constraints are documented in the Discussion and Methodology chapters to delimit external validity and guide interpretation.

6.5. Avenues for further research

Three extensions are immediate and feasible using the current data assets. (i) Enrich the interrupted time-series framework with placebo windows and HAC-consistent errors; extend visualization with local projections to sharpen timing. (ii) Construct descriptive difference-in-differences using non-finance *Papers with Code* segments as a proxy comparison group; treat estimates as suggestive given common shocks to the platform. (iii) Increase task-level granularity (e.g., payments initiation reliability, fraud detection, credit scoring, supervisory text mining) to identify where benchmark consolidation would deliver the largest marginal returns. Beyond these, two structural extensions merit attention: (iv) micro–macro linkage that matches specific dataset releases (or API standard updates) to subsequent paper production at fine lags; and (v) triangulation with external repositories and bibliometric sources to assess coverage biases in *Papers with Code* .

Concluding remarks Within the limits of an anchor-based global design, the evidence indicates that open-finance milestones are associated with measurable shifts in the production of domain-relevant datasets and research, as captured by a widely used open-science platform. The most parsimonious reading is sequential: policy and market arrangements that increase lawful access to data and standardize interfaces act first on the supply of reusable artifacts, which then propagate to research outputs with lags. While causality is not claimed, the patterns are robust across outcomes, consistent with theory, and supported by complementary statistical checks reported alongside the figures and tests. Taken together, these results reinforce the view that open finance, when paired with credible technical governance, can operate as an engine of cumulative, auditable, and socially valuable innovation.

Bibliography

1 FIRSANOVA, O.; BROVCHA, S. Features of the financial services market and consumer behavior on it in the context of digital transformation. *Scientific Journal of Polonia University*, v. 49, n. 6, p. 177–187, 2021.

2 FLORENCE, S. et al. Systematic review of digital transformation strategies in legacy banking and payments infrastructure. *International Journal of Multidisciplinary Research*, v. 6, n. 4, p. 1–18, 2024.

3 THOTTOLI, M. M. et al. Embracing digital transformation in financial services: From past to future. *SAGE Open*, v. 13, n. 4, p. 1–24, 2023.

4 ANIFA, M. et al. Fintech innovations in the financial service industry. *Journal of Risk and Financial Management*, v. 15, n. 7, p. 287, 2022.

5 BHATNAGAR, S.; MAHANT, R. Strengthening financial services through secure computing: Challenges, solutions, and future directions. *International Journal of Advanced Research in Science, Communication and Technology*, v. 4, n. 7, p. 234–251, 2024.

6 ZHANG, Y.; SILVA, M.; JOHNSON, R. Digital transformation as a driver of the financial sector sustainable development: An impact on financial inclusion and operational efficiency. *Sustainability*, v. 15, n. 1, p. 207, 2022.

7 GOUNARI, M. et al. Harmonizing open banking in the european union: an analysis of psd2 compliance and interrelation with cybersecurity frameworks and standards. *International Cybersecurity Law Review*, v. 5, p. 79–120, January 2024. Disponível em: <https://link.springer.com/article/10.1365/s43439-023-00108-8>.

8 PARKER, G. G.; ALSTYNE, M. W. V.; CHOUDARY, S. P. *Platform Revolution: How Networked Markets Are Transforming the Economy and How to Make Them Work for You*. New York: W. W. Norton & Company, 2016.

9 GAWER, A.; CUSUMANO, M. A. Industry platforms and ecosystem innovation. *Journal of Product Innovation Management*, v. 31, n. 3, p. 417–433, 2014.

10 JACOBIDES, M. G.; CENNAMO, C.; GAWER, A. Towards a theory of ecosystems. *Strategic Management Journal*, v. 39, n. 8, p. 2255–2276, 2018.

11 CENNAMO, C.; SANTALó, J. Generativity tension and value creation in platform ecosystems. *Organization Science*, v. 30, n. 3, p. 617–641, 2019.

12 McKinsey & Company. *Banking-as-a-Service: Rethinking value creation in banking*. [S.l.], 2021. Disponível em: <https://www.mckinsey.com/industries/financial-services/our-insights/banking-as-a-service-rethinking-value-creation-in-banking>.

13 BETHLENDI, A. How the fintech ecosystem changes with the entry of big tech companies. *Investment Management and Financial Innovations*, v. 19, n. 3, p. 22–31, 2022.

14 DINIZ, B. *The Platformization of the Banking Industry*. 2021. Article / column (translated) — FDATA / Exame (PDF). Originally published in *Exame*; available at: <https://fdata.global/wp-content/uploads/2021/07/Exame-The-Platformization-of-the-Banking-Industry-Bruno-Diniz_10_01_21.pdf>.

15 CARSON, B. et al. *Building the AI bank of the future*. [S.l.], 2021. Global Banking & Securities; report/compendium. URL: <https://www.mckinsey.com/~/media/mckinsey/industries/financial%20services/our%20insights/building%20the%20ai%20bank%20of%20the%20future/building-the-ai-bank-of-the-future.pdf>.

16 CHESBROUGH, H. W. *Open Innovation: The New Imperative for Creating and Profiting from Technology*. Boston, MA: Harvard Business School Press, 2003.

17 CHESBROUGH, H. W. *Open Services Innovation: Rethinking Your Business to Grow and Compete in a New Era*. San Francisco, CA: Jossey-Bass / Wiley, 2011.

18 FRANCESCONI, A.; TANDA, A. *Open innovation in banking: a bibliometric study*. [S.l.], 2024. Available at:: <https://economiaemanagement.dip.unipv.it/sites/dip10/files/2024-11/DEMWP0224.pdf>.

19 STEFANELLI, V.; MANTA, F.; TOMA, P. *Digital financial services and open banking innovation: are banks becoming 'invisible'?* 2022. ArXiv preprint. ArXiv:2210.01109. URL: <https://arxiv.org/abs/2210.01109>.

20 European Commission. *Commission Delegated Regulation (EU) 2018/389 of 27 November 2017 supplementing Directive (EU) 2015/2366 with regard to regulatory technical standards for strong customer authentication and common and secure open standards of communication*. 2018. RTS on SCA and CSC under PSD2. Disponível em: <https://eur-lex.europa.eu/eli/reg_del/2018/389/oj/eng>.

21 European Banking Authority. *Opinion on the elements of strong customer authentication under PSD2*. 2019. Opinion under Directive (EU) 2015/2366 (PSD2). Disponível em: <https://www.eba.europa.eu/regulation-and-policy/payment-services-and-electronic-money>.

22 European Banking Authority. *Opinion on obstacles under Article 32(3) of the RTS on SCA & CSC*. 2020. Clarifies practices that constitute obstacles to TPP services under PSD2 dedicated interfaces. Disponível em: <https://www.eba.europa.eu/sites/default/files/document_library/Publications/Opinions/2020/884569/EBA%20Opinion%20on%20obstacles%20under%20Art.%2032%283%29%20RTS%20on%20SCA%26CSC.pdf>.

23 ZETZSCHE, D. A. et al. The governance of open banking: New risks and new solutions. *Journal of Financial Regulation*, Oxford University Press, v. 6, n. 2, p. 243–278, 2020.

24 Banco Central do Brasil. *Agenda BC#: Competitividade, Inclusão, Transparência e Educação — Diretrizes para Open Banking/Open Finance*. 2020. Pilar de competitividade da Agenda BC# e diretrizes para Open Banking no Brasil. Disponível em: <https://www.bcb.gov.br/acessoinformacao/agendaBC>.

25 BANK, W. *Open Finance: Unlocking Financial Inclusion Through Data Sharing*. [S.l.], 2023. Disponível em: <https://documents.worldbank.org/en/publication/documents-reports/documentdetail/>.

26 BANK, I.-A. D. *Financial Inclusion in Latin America: Opportunities and Challenges for Open Finance*. [S.l.], 2022. Disponível em: <https://publications.iadb.org/>.

27 CGAP. *Interoperability and Open Finance: Building Inclusive Digital Financial Ecosystems*. [S.l.], 2021. Disponível em: <https://www.cgap.org/research/publication/interoperability-and-open-finance>.

28 Banco Central do Brasil. *Relatório de Gestão do Pix: Concepção e primeiros anos de funcionamento 2020–2022*. 2023. See section “Serviço de iniciação: o elo que conecta o Open Finance e o Pix”. Disponível em: <https://www.bcb.gov.br/content/estabilidadefinanceira/pix/relatorio_de_gestao_pix/relatorio_gestao_pix_2023.pdf>.

29 Banco Central do Brasil. *Projeto Drex (Real Digital): Piloto e Relatórios de Acompanhamento*. 2024. Página oficial do projeto, atualizações do piloto e documentação técnica. Disponível em: <https://www.bcb.gov.br/estabilidadefinanceira/drex>.

30 European Banking Authority. *EBA responses to issues XXXII–XXXVIII raised by participants of the EBA Working Group on APIs under PSD2 (Seventh set of issues)*. 2021. Non-binding clarifications on API topics incl. downtime, payment status/rejection reasons, access restrictions. Disponível em: <https://www.eba.europa.eu/sites/default/files/document_library/News%20and%20Press/Press%20Room/Press%20Releases/2021/1022210/Seventh%20set%20of%20issues%20raised%20by%20the%20EBA%20WG%20on%20APIs.pdf>.

31 POLASIK, M. et al. Evaluating the regulatory approach to open banking in europe: An empirical study. *Financial Law Review*, v. 34, n. 2, p. 58–90, November 2024. Disponível em: <https://doi.org/10.4467/22996834FLR.24.007.20612>.

32 AWASTHI, S. Strategies for commercialising open banking: Approaches from around the globe. *Journal of Digital Banking*, v. 8, n. 1, p. 15–29, 2024.

33 MODESTI, P. et al. Security analysis of the open banking account and transaction api protocol. *arXiv preprint*, p. arXiv:2003.12776, 2020.

34 GONçALVES, A. V.; ARAUJO, F. O. d. The implementation of open banking in brazil. *Brazilian Journal of Production Management*, v. 14, n. 3, p. 1–15, 2022.

35 Banco Central do Brasil. *Estrutura de Governança do Open Finance no Brasil*. 2021. Modelagem de governança, papéis e processos do ecossistema Open Finance. Disponível em: <https://www.bcb.gov.br/estabilidadefinanceira/openfinance>.

36 BHATT, A.; JOSHIPURA, M.; JOSHIPURA, N. Decoding the trinity of fintech, digitalization and financial services: An integrated bibliometric analysis and thematic literature review approach. *Cogent Economics & Finance*, v. 10, n. 1, p. 2114160, 2022.

37 SAINI, A. The fintech revolution and future of banking and financial institutions: A quantitative investigation. *Psychology and Education*, v. 60, n. 6, p. 123–140, 2023.

38 ZHANG-ZHANG, Y.; ROHLFER, S.; RAJASEKERA, J. An eco-systematic view of cross-sector fintech: The case of alibaba and tencent. *Sustainability*, v. 12, n. 21, p. 8907, 2020.

39 BABINA, T.; BUCHAK, G.; GORNALL, W. *Customer Data Access and Fintech Entry: Early Evidence from Open Banking*. [S.l.], 2022. Current version April 21, 2022. URL: <https://ssrn.com/abstract=4071214>.

40 European Parliament and Council. *Directive (EU) 2015/2366 on payment services in the internal market (PSD2)*. 2015. Primary legislation.

41 ALMEHREJ, A.; FREITAS, L.; MODESTI, P. Account and transaction protocol of the open banking standard. *IEEE Computer Security Foundations Symposium*, v. 33, p. 158–173, 2020.

42 European Banking Authority. *Guidelines on major incident reporting under the Payment Services Directive 2 (PSD2)*. 2021. Includes requirements relevant to SCA/CSC implementation and interface operation under PSD2. Disponível em: <https://www.eba.europa.eu/regulation-and-policy/payment-services-and-electronic-money>.

43 Banca d'Italia. *Relazione Annuale 2020: Attuazione della PSD2 e sviluppi dei servizi di pagamento*. 2021. Sezioni sulla disciplina PSD2 e implementazione nazionale. Disponível em: <https://www.bancaditalia.it/pubblicazioni/relazione-annuale/index.html>.

44 Banca d'Italia. *FinTech in Italy: surveys, supervisory initiatives and innovation hubs — 2022 overview*. 2022. Documentazione e rapporti sul FinTech e innovazione nei servizi finanziari. Disponível em: <https://www.bancaditalia.it/compiti/vigilanza/innovazione/index.html>.

45 Competition and Markets Authority. *The Retail Banking Market Investigation Order 2017*. 2017. UK primary instrument establishing Open Banking remedies.

46 ARALUZE, G. K. B. D.; PLAZA, N. C. Open banking: A bibliometric analysis-driven definition. *PLOS ONE*, v. 17, n. 10, p. e0275496, 2022.

47 BABIN, R.; SMITH, D. Open banking and regulation: Please advise the government. *SAGE Open*, v. 12, n. 2, p. 20438869221082316, 2022.

48 BROWN, I. The uk's midata and open banking programmes. *Technology and Regulation*, v. 2022, p. 71–93, 2022.

49 CLARKE, C.; MACARTNEY, H. The second financial services revolution? the quiet politics of open banking reform in the uk. *West European Politics*, v. 48, n. 3, p. 1–24, 2025.

50 NIANKARA, I.; HASSAN HASSAN ISMAIL ANS TRAORET, R.; ISLAM, A. R. M. Consumer savings and digital remittance in open banking: Insights from bibliometric and geospatial econometric analysis. *Human Behavior and Emerging Technologies*, v. 2025, p. 9352257, 2025.

51 PATI, U. K.; PRATAMA, A. M. Dasguptadavid1994indonesia's open banking future: Designing effective regulatory approaches. *Jambe Law Journal*, v. 7, n. 2, p. 189–210, 2025.

52 FAHY, L. Regulator reputation and stakeholder participation: A case study of the uk's regulatory sandbox for fintech. *European Journal of Risk Regulation*, v. 12, n. 3, p. 602–629, 2021.

53 GEORGOSOULI, A. Metarules, judgment and the algorithmic future of financial regulation in the uk. *Oxford Journal of Legal Studies*, v. 43, n. 4, p. 1134–1159, 2023.

54 ARALUZE, G. B. d.; PLAZA, N. C. The relevance of initial trust and social influence in the intention to use open banking-based services: An empirical study. *SAGE Open*, v. 13, n. 3, p. 1–15, 2023.

55 ROSATI, P.; FOX GRACE ANDA CUMMINS, M.; LYNN, T. Perceived risk as a determinant of propensity to adopt account information services under the eu payment services directive 2. *Universal Journal of Accounting and Finance*, v. 10, n. 2, p. 425–440, 2022.

56 ALMEIDA, V. d. R. S.; MACADAR, M. M. Pix, open finance, and drex: Cases of digital public infrastructure in brazil. In: *Proceedings of the 2025 International Conference on Digital Government Research*. [S.l.: s.n.], 2025. p. 1–6.

57 Banco Central do Brasil and Conselho Monetário Nacional. *Resolução Conjunta nº 1, de 4 de maio de 2020*. 2020. Estabelece o escopo e as diretrizes para o Open Banking/Open Finance no Brasil.

58 Banco Central do Brasil. *Open Finance no Brasil: avanços, desafios e próximos passos*. 2024. Comunicações oficiais e materiais institucionais do BCB sobre a evolução do Open Finance. Disponível em: <https://www.bcb.gov.br/estabilidadefinanceira/openfinance>.

59 PANEQUE, M.; ROLDáN-GARCíA, M. d. M.; GARCíA-NIETO, J. A semantic model for enhancing data-driven open banking services. *Applied Sciences*, v. 13, n. 3, p. 1447, 2023.

60 PwC. *Seizing the Open Banking Opportunity: How banks can create value in the new ecosystem*. [S.l.], 2022. Disponível em: <https://www.pwc.com/gx/en/industries/financial-services/assets/pdf/open-banking-opportunity.pdf>.

61 European Banking Authority. *EBA Report on the impact of BigTech on payment institutions' and e-money institutions' business models*. [S.l.], 2021. Disponível em: <https://www.eba.europa.eu>.

62 Boston Consulting Group. *The Future of Open Finance: Accelerating innovation and competition*. [S.l.], 2023. Disponível em: <https://www.bcg.com/publications/2023/the-future-of-open-finance>.

63 McKinsey & Company. *Open Finance: The next frontier in financial innovation*. [S.l.], 2022. Disponível em: <https://www.mckinsey.com/industries/financial-services/our-insights/open-finance-the-next-frontier>.

64 SHACHEENDRAN, V. et al. The rise of open banking: A comprehensive analysis of research trends and collaborative networks. *International Journal of Economics and Financial Issues*, v. 14, n. 6, p. 88–104, 2024.

65 ABBASOV, A. The development of open banking and the analysis of third-party providers in azerbaijan. *Asian Financial Journal*, v. 11, n. 1, p. 100–115, 2023.

66 KOPPERAPU, R. Open banking and apis: Revolutionizing financial data access and innovation in the u.s. financial sector. *SSRN Electronic Journal*, 2024.

67 KHAN, R. Digital payments impact on indian economy through banking systems & global comparisons. *International Journal of Scientific Research and Engineering Management*, v. 8, n. 5, p. 45–59, 2024.

68 SILVA, M.; ROCHA, A. The role of open finance and pix in expanding credit for smes in brazil. *Journal of Financial Innovation in Emerging Markets*, v. 9, n. 2, p. 45–68, 2024.

69 KHAN, A.; MüLLER, T. Ethical considerations in financial data sharing: From gdpr to trust-by-design. *International Journal of Data Governance*, v. 6, n. 1, p. 12–34, 2023.

70 SMITH, R.; LOPEZ, C. Privacy-enhancing technologies for open finance: Challenges and opportunities. *Journal of Cybersecurity and Privacy*, v. 3, n. 4, p. 78–95, 2022.

71 LI, W.; GONZáLEZ, C. Trust-by-design in open banking: Building consumer confidence through secure architectures. In: *Proceedings of the International Conference on Financial Technology and Security*. [S.l.: s.n.], 2023. p. 89–104.

72 GROUP, B. C. *Open Finance 2.0: The Next Frontier of Financial Services*. [S.l.], 2023. Disponível em: <https://www.bcg.com/publications/2023/open-finance-2.0>.

73 OECD. *Open Finance 2.0 and the Rise of Embedded Finance*. [S.l.], 2024. Disponível em: <https://www.oecd.org/finance/embedded-finance>.

74 SETTLEMENTS, B. for I.; CEMLA. *Cross-Border Interoperability in Instant Payments: Lessons for Latin America*. [S.l.], 2024. Disponível em: <https://www.bis.org/publ/othp59.htm>.

75 BANK, E. C. *Instant Payments in Europe: Status and Challenges*. [S.l.], 2023. Disponível em: <https://www.ecb.europa.eu/pub/pdf/other/instantpayments2023.en.pdf>.

76 DASGUPTA, P.; DAVID, P. A. Toward a new economics of science. *Research Policy*, v. 23, n. 5, p. 487–521, 1994. Disponível em: <https://www.sciencedirect.com/science/article/pii/0048733394010021>.

77 DAVID, P. A. *The Economic Logic of "Open Science" and the Balance Between Private Property Rights and the Public Domain in Scientific Data and Information: A Primer*. [S.l.], 2003. Disponível em: <https://siepr.stanford.edu/publications/working-paper/economic-logic-open-science-and-balance-between-private-property-rights>.

78 STEPHAN, P. E. The economics of science. *Journal of Economic Literature*, v. 34, n. 3, p. 1199–1235, 1996. Disponível em: <https://ideas.repec.org/a/aea/jeclit/v34y1996i3p1199-1235.html>.

79 STEPHAN, P. E. *How Economics Shapes Science*. Cambridge, MA: Harvard University Press, 2012. ISBN 9780674069148. Disponível em: <https://www.hup.harvard.edu/books/9780674069148>.

80 OECD. *Recommendation of the Council on Enhancing Access to and Sharing of Data (OECD/LEGAL/0463)*. 2024. OECD Legal Instruments. Disponível em: <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0463>.

81 OECD. *Enhancing Access to and Sharing of Data in the Age of Artificial Intelligence: Companion Document to the OECD Recommendation on EASD*. 2025. Disponível em: <https://one.oecd.org/document/COM/DSTI/CDEP/STP/GOV/PGC(2024)1/FINAL/en/pdf>.

82 UBALDI, B. *Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives*. [S.l.], 2013. (OECD Working Papers on Public Governance, 22). Disponível em: <https://www.oecd.org/en/publications/open-government-data_5k46bj4f03s7-en.html>.

83 ZUIDERWIJK, A.; JANSSEN, M. Open data policies, their implementation and impact: A framework for comparison. *Government Information Quarterly*, v. 31, n. 1, p. 17–29, 2014. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0740624X13001202>.

84 PIWOWAR, H. A.; VISION, T. J. Data reuse and the open data citation advantage. *PeerJ*, v. 1, p. e175, 2013. Disponível em: <https://peerj.com/articles/175/>.

85 National Academies of Sciences, Engineering, and Medicine. *Open Science by Design: Realizing a Vision for 21st Century Research*. Washington, DC: National Academies Press, 2018. Disponível em: <https://nap.nationalacademies.org/catalog/25116/open-science-by-design>.

86 SOUZA, D. F. de; GEUNA, A.; RODRÍGUEZ, Á. How small is big enough? open labeled datasets and the development of deep learning. *arXiv*, 2024. Disponível em: <https://arxiv.org/abs/2408.10359>.

87 RUSSAKOVSKY, O. et al. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, v. 115, n. 3, p. 211–252, 2015. Disponível em: <https://link.springer.com/article/10.1007/s11263-015-0816-y>.

88 LIN, T.-Y. et al. Microsoft coco: Common objects in context. In: *European Conference on Computer Vision (ECCV)*. Springer, 2014. p. 740–755. Disponível em: <https://link.springer.com/chapter/10.1007/978-3-319-10602-1_48>.

89 LERNER, J.; TIROLE, J. Some simple economics of open source. *Journal of Industrial Economics*, v. 50, n. 2, p. 197–234, 2002. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1111/1467-6451.00174>.

90 HIPPEL, E. von; KROGH, G. von. Open source software and the “private-collective” innovation model: Issues for organization science. *Organization Science*, v. 14, n. 2, p. 209–223, 2003. Disponível em: <https://pubsonline.informs.org/doi/10.1287/orsc.14.2.209.14992>.

91 NAGLE, F. Learning by contributing: Gaining competitive advantage through contribution to crowdsourced public goods. *Organization Science*, v. 29, n. 4, p. 569–587, 2018. Disponível em: <https://pubsonline.informs.org/doi/10.1287/orsc.2018.1202>.

92 WRIGHT, N. L.; NAGLE, F.; GREENSTEIN, S. Open source software and global entrepreneurship. *Research Policy*, v. 52, n. 9, p. 104846, 2023. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0048733323001300>.

93 WEST, J. How open is open enough? melding proprietary and open source platform strategies. *Research Policy*, v. 32, n. 7, p. 1259–1285, 2003. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0048733303000520>.

94 SIMCOE, T. Open standards and intellectual property rights. In: CHESBROUGH, H.; VANHAVERBEKE, W.; WEST, J. (Ed.). *Open Innovation: Researching a New Paradigm*. Oxford: Oxford University Press, 2006. p. 161–183. Disponível em: <https://people.bu.edu/tsimcoe/documents/published/OpenStandards_IPR.pdf>.

95 BLIND, K. et al. Standards and innovation: A review and introduction to the special issue. *Research Policy*, v. 52, n. 8, p. 104830, 2023. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0048733323001142>.

96 Open Bank Project. *OBP-API: Open Source API Platform for Open Banking*. n.d. GitHub repository. Disponível em: <https://github.com/OpenBankProject/OBP-API>.

97 SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). *Biometrika*, v. 52, n. 3-4, p. 591–611, 1965. Disponível em: <https://doi.org/10.1093/biomet/52.3-4.591>.

98 STUDENT. The probable error of a mean. *Biometrika*, v. 6, n. 1, p. 1–25, 1908. Disponível em: <https://doi.org/10.1093/biomet/6.1.1>.

99 WELCH, B. L. The generalization of ’student’s’ problem when several different population variances are involved. *Biometrika*, v. 34, n. 1-2, p. 28–35, 1947. Disponível em: <https://doi.org/10.1093/biomet/34.1-2.28>.

100 MANN, H. B.; WHITNEY, D. R. On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, v. 18, n. 1, p. 50–60, 1947. Disponível em: <https://doi.org/10.1214/aoms/1177730491>.

101 COHEN, J. *Statistical Power Analysis for the Behavioral Sciences*. 2. ed. Hillsdale, NJ: Lawrence Erlbaum Associates, 1988. ISBN 0-8058-0283-5.

102 MCDOWALL, D. et al. *Interrupted Time Series Analysis*. Beverly Hills, CA: SAGE Publications, 1980. (Quantitative Applications in the Social Sciences, 21). ISBN 0-8039-1493-8.

103 BERNAL, J. L.; CUMMINS, S.; GASPARRINI, A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. *International Journal of Epidemiology*, v. 46, n. 1, p. 348–355, 2017. Disponível em: <https://doi.org/10.1093/ije/dyw098>.

104 NEWEY, W. K.; WEST, K. D. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, v. 55, n. 3, p. 703–708, 1987. Disponível em: <https://doi.org/10.2307/1913610>.