

# Caramel: Um Framework para Ecosistema de Big Social Data

Paulo Freitas Silva Júnior<sup>1</sup>, Tiago França (coorientador)<sup>2</sup>, Jonice Oliveira (orientadora)<sup>1</sup>

<sup>1</sup>PPGI, Universidade Federal do Rio de Janeiro (UFRJ)  
Térreo, Bloco E,CCMN/NCE, Cidade Universitária – Rio de Janeiro, RJ – Brasil

<sup>2</sup>DECOMP, Universidade Federal Rural do Rio de Janeiro (UFRRJ)  
Av. Gov. Roberto Silveira, S/N – Seropédica, RJ – Brasil

paulofreitas@macae.ufrj.br, tcruz.franca@gmail.com, jonice@ic.ufrj.br

**Abstract.** *Mainstream social media analysis tools create silos that hinder data management and interdisciplinary collaboration. This work proposes CARAMEL, a reference architecture based on microservices and asynchronous messaging to orchestrate distributed Big Social Data ecosystems. The framework automates metadata curation, ensuring traceability and information provenance throughout its entire lifecycle, while allowing the customization of analysis workflows. Validated with massive data volumes during the 2022 and 2024 elections, the infrastructure integrates different actors, reduces technical barriers, and facilitates technological support, artifact reuse, and active collaboration in sociotechnical research.*

**Resumo.** *As principais ferramentas de análise de mídias sociais geram silos que dificultam a gestão de dados e a colaboração interdisciplinar. Este trabalho propõe o CARAMEL, uma arquitetura de referência baseada em microsserviços e mensageria assíncrona para orquestrar ecossistemas distribuídos de Big Social Data. O framework automatiza a curadoria de metadados, garantindo a rastreabilidade e a proveniência da informação ao longo de todo o seu ciclo de vida, além de permitir a personalização dos fluxos de análise. Validada com volumes massivos nas eleições de 2022 e 2024, a infraestrutura integra diferentes atores, reduz barreiras técnicas e facilita o suporte tecnológico, o reuso de artefatos e a colaboração ativa em pesquisas sociotécnicas.*

## 1. Introdução

A necessidade de compreender comportamentos sociais em larga escala evidencia as limitações das ferramentas monolíticas, demandando a adoção de ecossistemas digitais colaborativos focados em *Big Social Data* (BSD) [SILVA JÚNIOR et al. 2023a, DEMIRBAGA 2023]. Interações online produzem grandes volumes de dados semanticamente ricos, servindo de matéria-prima para abordagens de BSD em domínios críticos, como prevenção ao suicídio e resposta a emergências [REHEM et al. 2016, FRANÇA 2019, LIMA FILHO et al. 2020, HARGREAVES et al. 2020]. Para a pesquisa prosperar, é essencial gerenciar dados heterogêneos de maneira distribuída e eficiente [FRANÇA et al. 2014, GROSSI et al. 2018, LIMA et al. 2022].

Embora o BSD seja caracterizado por volume e velocidade massivos [BELLO-ORGAZ et al. 2016, OLSHANNIKOVA et al. 2017], o maior desafio em sistemas colaborativos é o isolamento sociotécnico. A maioria das soluções atuais opera como “silos” baseados em arquiteturas monolíticas, impedindo a interoperabilidade, a escalabilidade independente e o compartilhamento de recursos [FRANÇA 2019, SILVA JÚNIOR et al. 2023a]. Conseqüentemente, artefatos computacionais são raramente reutilizados, gerando desperdício e excluindo especialistas de domínio sem capacitação técnica avançada [ROSSETTI et al. 2017, GROSSI et al. 2018, GROSSI et al. 2021, WANG et al. 2021]. Adicionalmente, a perda crônica da proveniência dos dados compromete a confiança mútua indispensável para a colaboração científica [SILVA JÚNIOR et al. 2023a].

Para superar os desafios desse cenário, a questão central que motiva este trabalho é: como estabelecer uma infraestrutura de *software* que sustente um ecossistema colaborativo de pesquisa em BSD, garantindo o reuso de artefatos, a curadoria de dados e a interoperabilidade tecnológica?

Como solução e objetivo principal, esta pesquisa de mestrado propôs e desenvolveu o CARMEL, uma arquitetura de referência baseada em microsserviços e mensageria assíncrona. Atendendo à demanda por inovação tecnológica, a principal contribuição deste trabalho em Ciência da Computação reside no projeto de uma infraestrutura que utiliza o desacoplamento temporal como mecanismo habilitador da colaboração distribuída. A arquitetura permite que produtores e consumidores operem de forma independente, enquanto um mecanismo automatizado de curadoria assegura a rastreabilidade e a qualidade dos ativos transacionados.

Ao alavancar essas tecnologias, o CARMEL viabiliza o acesso aos dados e padroniza o gerenciamento de BSD, rompendo o isolamento das abordagens monolíticas. A validação do *framework*, conduzida em bases de dados massivas durante as eleições brasileiras de 2022 e 2024, consolida a nossa contribuição, atestando a resiliência técnica em cenários de alta vazão e garantindo o reuso prático por diferentes frentes de pesquisa.

## 2. Metodologia de Pesquisa

Dada a natureza sociotécnica do problema, que envolve a interseção entre artefatos computacionais, processos de gerenciamento de BSD e atores sociais, esta pesquisa adotou o método *Design Science Research* (DSR) [DRESCH et al. 2014]. A abordagem guiou etapas iterativas focadas na criação de uma arquitetura de *software* inovadora, desenhada para atuar como habilitadora da colaboração interdisciplinar e avançar o conhecimento teórico sobre a orquestração de ecossistemas de BSD.

### 2.1. Conscientização do Problema e Fundamentação

O primeiro ciclo iterativo focou na fundamentação teórica por meio de uma Revisão Sistemática da Literatura (RSL) em cinco grandes bibliotecas digitais. A análise de 21 estudos primários selecionados foi crucial para delinear a questão de pesquisa, confirmando que a principal lacuna da área residia na ausência de arquiteturas de referência que suportassem o compartilhamento de recursos computacionais e a gestão colaborativa em ambientes distribuídos (conforme detalhado na Seção 3).

## 2.2. Projeto e Desenvolvimento do Artefato

Fundamentada na RSL, a fase de projeto concebeu o CAMEL como uma plataforma base para um ecossistema. A decisão de *design* de implementar os componentes como microsserviços visou garantir a modularidade e a autonomia das equipes, características essenciais para sistemas colaborativos escaláveis. A infraestrutura foi implementada utilizando Apache Kafka para orquestração assíncrona, garantindo o desacoplamento necessário entre produtores e consumidores de dados, e contêineres Docker para portabilidade. Coletores iniciais (X/Twitter e YouTube) validaram a viabilidade técnica em integrar fontes heterogêneas, com resultados preliminares validados na 19ª edição do Simpósio Brasileiro de Sistemas de Informação (SBSI 2023) [SILVA JÚNIOR et al. 2023a].

## 2.3. Validação Iterativa e Evolução Colaborativa

O processo de validação não se restringiu a testes técnicos isolados, mas avaliou a dinâmica sociotécnica viabilizada pela infraestrutura. Atores externos (alunos de graduação) adaptaram e desenvolveram novos componentes seguindo as diretrizes do CAMEL, o que comprovou a extensibilidade da plataforma e a redução de barreiras técnicas para a interoperabilidade. A validação contínua da proposta ocorreu por meio de avaliações no Workshop de Teses e Dissertações do SBB (WTDBD/SBB 2023) [SILVA JÚNIOR et al. 2023b]. Finalmente, a robustez do modelo frente a requisitos rigorosos de desempenho e escalabilidade foi comprovada pelo reuso massivo de artefatos durante o monitoramento das eleições brasileiras de 2022 e 2024, consolidando o CAMEL como um *framework* resiliente para a pesquisa colaborativa em BSD.

## 3. Trabalhos Relacionados

A análise de grandes volumes de dados sociais demanda soluções que orquestram as interações entre tecnologia e atores [OLSHANNIKOVA et al. 2017]. Neste domínio, definimos um ecossistema de dados como uma rede sociotécnica onde atores interagem com recursos de *software* para promover colaboração e gerar valor científico [OLIVEIRA et al. 2019]. No entanto, a literatura revela uma carência de arquiteturas de referência práticas para sustentar a orquestração dinâmica necessária em ecossistemas de BSD [OLIVEIRA et al. 2019].

Estudos prévios mapearam características e governança de *Big Data* [GROVER and KAR 2017, ANWAR et al. 2021], mas falham em fornecer modelos arquiteturais sobre como orquestrar tecnicamente os componentes para quebrar “silos” tecnológicos. Iniciativas práticas, como o SoBigData [GROSSI et al. 2018, GROSSI et al. 2021], oferecem infraestruturas operacionais, porém centralizadas, dificultando a replicação. Outros *frameworks*, como o DaLiF [SHAH et al. 2021], permanecem no nível conceitual. Adicionalmente, ferramentas de mercado (Gephi, NodeXL, SocioViz) baseiam-se em arquiteturas fortemente acopladas, restringindo a escalabilidade. A Tabela 1 resume o posicionamento do CAMEL, destacando seu diferencial prático como uma infraestrutura leve, baseada em microsserviços e focada na colaboração distribuída.

**Tabela 1. Posicionamento do CAMEL frente aos trabalhos relacionados.**

Trabalho	Foco Principal	Arquitetura	Implementado	Escalabilidade
[OLIVEIRA et al. 2019]	Teoria de Ecossistemas	N/A (Conceitual)	Não	N/A
[GROVER and KAR 2017]	Características de BSD	N/A (Revisão)	Não	N/A
[ANWAR et al. 2021]	Governança de Big Data	N/A (Modelo)	Não	Sim
[GROSSI et al. 2021]	Infraestrutura Colab.	Centralizada	Sim	Sim
[SHAH et al. 2021]	Ciclo de Vida (Gov)	Camadas	Não	N/A
<b>CAMEL</b>	<b>Ecossistema BSD</b>	<b>Microserviços</b>	<b>Sim</b>	<b>Alta</b>

## 4. Arquitetura Proposta do CAMEL

O *framework* CAMEL é uma arquitetura de referência baseada em microserviços, mensageria e contêineres, desenhada para superar as barreiras de ferramentas fortemente acopladas que fragmentam o ciclo de vida da pesquisa [FRANÇA 2019, SILVA JÚNIOR et al. 2023a]. O CAMEL atua como uma plataforma habilitadora, onde a interoperabilidade técnica sustenta a troca de dados e algoritmos entre diferentes domínios.

### 4.1. Microserviços e Contêineres como Habilitadores

Para evitar a formação de “silos”, a arquitetura decompõe as etapas de BSD (coleta, processamento, armazenamento) em serviços autônomos. Cada microserviço possui o seu próprio ciclo de vida, permitindo que diferentes equipes desenvolvam e escalem componentes isoladamente. O empacotamento em contêineres Docker encapsula as dependências técnicas, garantindo uma implantação consistente e portátil. Ao promover contratos de dados padronizados, o *framework* viabiliza a interoperabilidade técnica e a construção de fluxos coerentes de pesquisa.

### 4.2. Desacoplamento Temporal e Curadoria de Dados

O desacoplamento entre os componentes é garantido pela orquestração assíncrona mediada por um *message broker*. Esse modelo de interação orientado a eventos (*Publish/Subscribe*) desacopla temporalmente a coleta de dados do seu consumo analítico. A separação dos fluxos assegura baixa latência, resiliência contra picos de carga e permite a entrada ágil de novos atores no ecossistema.

Nesse processo contínuo, a curadoria do ativo compartilhado é fundamental. Em vez de depender de uma catalogação manual e centralizada, a arquitetura estabelece que cada artefato de *software* gere e gerencie seus próprios metadados, facilitando o rastreamento da proveniência dos dados de forma distribuída. O dado bruto coletado passa por um refinamento onde é injetado com informações cruciais de origem (ex: `caramel.metadata`) e enriquecido com resultados de análises cognitivas (ex: `nlp.analysis`), conforme ilustrado na Fig. 1. O resultado é um objeto JSON padronizado e rastreável, pronto para indexação eficiente e consumo com confiança semântica.

### 4.3. Fluxo Operacional e Extensibilidade

A Fig. 2 ilustra a orquestração prática do ecossistema. O fluxo inicia com microserviços de coleta (ex: X/Twitter e YouTube) atuando como fornecedores de matéria-prima. Esses dados convergem para o Sistema de Mensageria (o *backbone* da arquitetura) e são subsequentemente consumidos por processadores especializados e persistidos em um *Data Lake*. Por fim, os dados processados alimentam o Catálogo de Dados para consumo final.

## Modelo de Transformação de Dados (Schema Evolution) no CAMEL

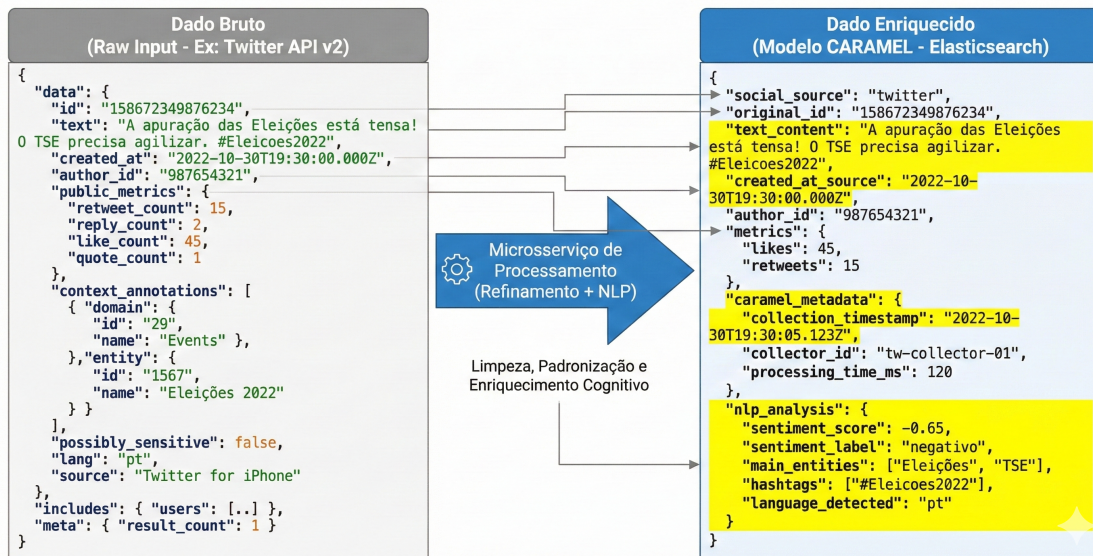


Figura 1. Modelo de Transformação de Dados: do dado bruto ao ativo enriquecido com metadados do ecossistema.

A propriedade de extensibilidade do sistema é destacada no diagrama pelo “Novo Microserviço Customizado”. Essa modularidade exemplifica como pesquisadores e novos parceiros podem acoplar serviços inovadores para consumir fluxos existentes sem a necessidade de refatorar a plataforma central. Esse ambiente de gestão distribuída promove ativamente o reuso, a colaboração interdisciplinar e a rastreabilidade necessárias para uma pesquisa sociotécnica robusta.

## 5. Avaliação e Resultados Obtidos

Esta seção detalha a avaliação prática do CAMEL sob a perspectiva de sua operação como plataforma base para ecossistemas colaborativos. O objetivo foi verificar a viabilidade técnica e a capacidade do *framework* de orquestrar interações complexas em BSD, promovendo reuso de *software* e escalabilidade. A avaliação atuou em demandas reais de pesquisa em larga escala para validar a resiliência sociotécnica do modelo.

### 5.1. Orquestração e Desacoplamento Temporal via Kafka

A característica central que define a escalabilidade do CAMEL é a separação temporal na execução de tarefas de processamento e curadoria. A Fig. 3 apresenta um Diagrama de Sequência ilustrando como a oferta de dados e a geração de conhecimento ocorrem em ritmos independentes, sem bloquear a plataforma.

Como mostrado no diagrama, a orquestração assíncrona ocorre em duas fases de processamento distintas, separadas por um “Gap Temporal”. Na fase de produção, o microserviço Coletor busca ativos brutos de uma API externa e disponibiliza imediatamente o recurso no tópico Kafka. Com a confirmação de recebimento (*ack*), o Coletor é liberado. Na fase de consumo, que ocorre sob demanda da capacidade computacional, o Processador solicita os dados, realiza a agregação analítica (ex: NLP) e confirma a conclusão. Este mecanismo garantiu que picos de demanda no mundo real não desestabilizassem a rede colaborativa.

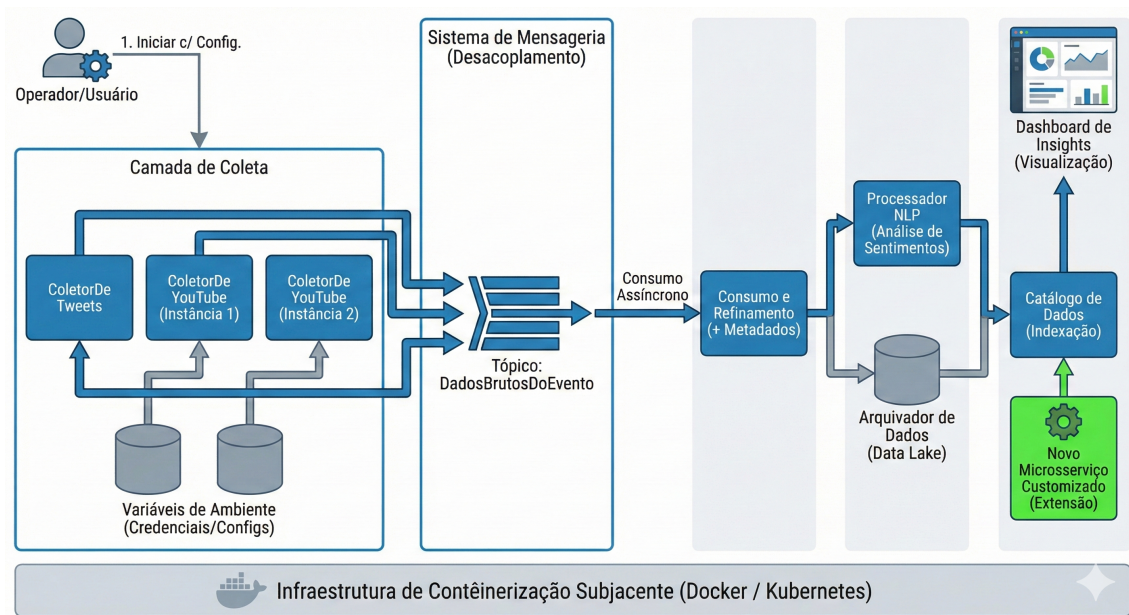


Figura 2. Diagrama de fluxo arquitetural dos microsserviços do CAMEL.

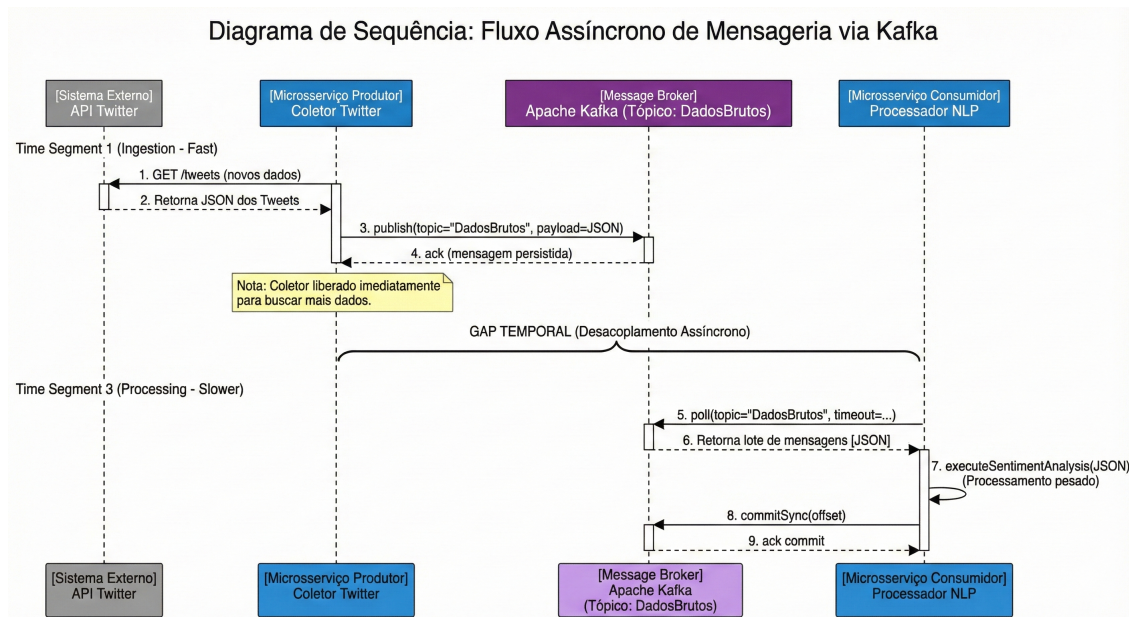
## 5.2. Avaliação em Larga Escala e Bases de Dados Geradas

Para instanciar o ecossistema nos cenários de estresse, a infraestrutura foi implantada na Oracle Cloud utilizando Docker Compose para orquestração de contêineres e Quarkus para microsserviços de baixo consumo, com o Kafka atuando como *backbone*. O ciclo de vida da informação foi validado ponta a ponta, permitindo que pesquisadores configurassem as coletas e consumissem visões analíticas estruturadas diretamente via Kibana e Elasticsearch.

O primeiro cenário de avaliação focou nas eleições presidenciais brasileiras de 2022. Este evento serviu como teste de estresse contínuo e resultou na consolidação de uma base de dados pública massiva de interações sociais. Ao longo de 127 dias ininterruptos, a plataforma manteve estabilidade realizando aproximadamente 1,46 milhão de transações de coleta. O *pipeline* funcionou como uma esteira contínua: os dados fluíam pelos tópicos e eram enriquecidos por análises de sentimentos em tempo real, transformando dados brutos em ativos estruturados de forma ágil.

A segunda avaliação focou na complexidade da integração multicanal durante as eleições municipais de 2024, validando a capacidade de incorporar novas fontes (Instagram e YouTube) sem refatoração da plataforma central. Este experimento empregou o padrão de integração *“fan-out”*, onde um evento de processamento disparava múltiplos fluxos simultâneos de coleta e curadoria (filtragem de ruídos/tokenização). O CAMEL orquestrou a criação de 505 conjuntos de dados distintos, totalizando 1,3 GB de ativos brutos persistidos na base de dados.

A evolução técnica e a capacidade de entrega da plataforma nestes dois eventos críticos estão sumarizadas na Tabela 2.



**Figura 3. Diagrama de sequência ilustrando o desacoplamento temporal entre coleta e processamento via Apache Kafka.**

### 5.3. Validação da Portabilidade do Modelo

Complementarmente, a utilidade e a portabilidade do ecossistema foram validadas em experimentos com alunos da UFRRJ, atuando como novos colaboradores da rede. Os alunos aplicaram a arquitetura para gerar suas próprias coleções de artefatos locais, migrando de soluções em silos para microsserviços desacoplados. Esta fase confirmou a baixa barreira de entrada da infraestrutura, atestando a viabilidade de replicação do *framework* em diferentes contextos de pesquisa interdisciplinar.

## 6. Considerações Finais

Esta pesquisa investigou os problemas estruturais que impedem a consolidação de redes sociotécnicas e ecossistemas colaborativos sustentáveis para a pesquisa em análise de mídias sociais. A complexidade deste domínio, agravada pela rigidez das ferramentas e pela fragmentação do conhecimento em silos tecnológicos, exigia uma abordagem arquitetural que fosse além das soluções fortemente acopladas. Em resposta, propusemos e validamos o CAMEL, não apenas como uma ferramenta técnica, mas como uma plataforma de *software* projetada para orquestrar ecossistemas distribuídos de BSD. Ao promover a interoperabilidade e o reuso de ativos de *software* e dados, o CAMEL avança o estado da arte em Sistemas Colaborativos e Gestão de Dados, oferecendo um modelo de referência baseado em microsserviços que habilita a evolução contínua e compartilhamento de artefatos de pesquisa.

O *framework* consolida uma infraestrutura computacional robusta para sistemas colaborativos, permitindo que equipes distribuídas compartilhem rotinas complexas de BSD. A arquitetura do CAMEL atua, assim, como um elemento habilitador que reduz barreiras de entrada, viabiliza o reuso de artefatos de *software* e fomenta a inovação científica aberta na gestão de dados sociais.

**Tabela 2. Desempenho Operacional e Capacidade de Entrega: Comparação entre Cenários de Avaliação 2022 e 2024.**

Métrica	Cenário 1: Eleições Presidenciais (2022)	Cenário 2: Eleições Municipais (2024)
Estratégia de Coleta	Monitoramento Contínuo de Alto Volume	Coleta Heterogênea Multicanal
Janela de Operação	127 dias contínuos (16 Ago - 20 Dez)	Janelas de coleta baseadas em eventos
Fornecedores de Dados	API do Twitter (X)	YouTube (Vídeos, Transcrições, Chats), Instagram
Padrão de Integração	Processamento em Fluxo (Stream)	Coleta Paralela <i>Fan-out</i> (Micro-lote)
Escalabilidade de Atores	8 instâncias simultâneas de coletores	Instâncias dinâmicas por capital monitorada
Vazão de Serviço	≈ 1 req/min por coletor (1.440/dia)	Ingestão paralela massiva de múltiplos IDs
Transações Totais	≈ 1,46 milhão de requisições de serviço	N/A (Foco na variedade da extração)
Volume de Ativos	> 146 milhões de registros potenciais	<b>505 datasets distintos</b> (1,3 GB dados persistentes)
Serviços Analíticos	NLP em Tempo Real e Indexação	Mineração, Curadoria e Tokenização

### 6.1. Desafios à Validade do Ecossistema Colaborativo

Apesar dos resultados promissores na orquestração técnica e na validação em cenários reais (como as eleições de 2022 e 2024), este estudo apresenta limitações concernentes à dinâmica de gestão do ecossistema proposto:

**Validação da Dinâmica Colaborativa:** Embora a plataforma tenha sido testada tecnicamente com sucesso, os experimentos de extensão ocorreram em ambientes acadêmicos controlados. A ausência de um teste prolongado com uma comunidade externa massiva impediu a identificação de gargalos de governança distribuída em larga escala.

**Interoperabilidade com Ecossistemas Externos (FAIR):** Existe uma limitação na integração direta com o ciclo de vida da pesquisa científica aberta. A versão atual priorizou a robustez da gestão distribuída em detrimento da catalogação automatizada em repositórios institucionais (como o Dataverse), limitando a adesão plena aos princípios FAIR.

**Sustentabilidade da Infraestrutura:** A validação foi restringida pela falta de recursos permanentes de nuvem, impedindo a manutenção do CARMEL como um serviço acadêmico ininterrupto de longo prazo, fator crítico para a consolidação da confiança dos atores na rede.

**Barreiras de Entrada Técnicas:** O desenvolvimento de novos componentes coletores ainda requer um perfil computacional avançado. A ausência de interfaces de alto nível cria uma barreira para especialistas de domínio (como sociólogos e comunicadores), limitando a diversidade interdisciplinar dos atores.

### 6.2. Trabalhos Futuros e Possibilidades de Impacto

Para mitigar as limitações identificadas e expandir o impacto científico, tecnológico e social do CARMEL, sugerimos as seguintes direções para trabalhos futuros:

**Integração com o Ciclo de Vida da Pesquisa Aberta:** Evoluir a camada de curadoria para incluir conectores automáticos com repositórios científicos, transformando o

ecossistema em um fornecedor direto de bases de dados qualificadas para a comunidade.

**Validação Interinstitucional:** Disponibilizar o CAMEL como uma instância permanente para parceiros universitários, avaliando a aceitação tecnológica e os modelos de governança necessários para sustentar a colaboração interdisciplinar em rede.

**Simplificação via Low-Code:** Reduzir drasticamente a barreira de entrada por meio do desenvolvimento de interfaces gráficas intuitivas, permitindo que processos de gerenciamento de BSD sejam configurados visualmente por especialistas não técnicos, sem a necessidade de codificação profunda.

**Agentes Cognitivos no Ecossistema:** Incorporar *Large Language Models* (LLMs) não apenas como ferramentas passivas, mas como atores autônomos integrados aos tópicos de mensageria, capazes de consumir dados brutos e executar rotinas complexas de BSD automatizadamente, elevando a capacidade de geração de valor da rede.

## Agradecimentos

Os autores agradecem ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) — em especial pelo programa PROEXT-PG — e à Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ), processo nº. E-26/210.936/2024.

## Referências

- ANWAR, M. J., GILL, A. Q., HUSSAIN, F. K., and IMRAN, M. (2021). Secure big data ecosystem architecture: challenges and solutions. *EURASIP Journal on Wireless Communications and Networking*, 2021:130.
- BELLO-ORGAZ, G., JUNG, J. J., and CAMACHO, D. (2016). Social big data: Recent achievements and new challenges. *Information Fusion*, 28:45–59.
- DEMIRBAGA, U. (2023). HTwitt: a hadoop-based platform for analysis and visualization of streaming Twitter data. *Neural Computing and Applications*, 35(33):23893–23908.
- DRESCH, A., Lacerda, D. P., and ANTUNES JR, J. A. V. (2014). Design science research. In *Design science research: A method for science and technology advancement*, pages 67–102. Springer.
- FRANÇA, T. C. (2019). *Andare: Um Framework Para Inclusão Da Análise De Dados De Mídias Sociais No Contexto Da Preparação E Resposta À Emergência Em Situações De Manifestações De Massa*. Tese (doutorado), Universidade Federal do Rio de Janeiro.
- FRANÇA, T. C., FARIA, F. F., RANGEL, F., FARIAS, C. M., and OLIVEIRA, J. (2014). Big Social Data: Princípios sobre Coleta, Tratamento e Análise de Dados Sociais. In *XXIX Simpósio Brasileiro de Banco de Dados*, volume 14, page 38. Ed. Porto Alegre.
- GROSSI, V., GIANNOTTI, F., PEDRESCHI, D., MANGHI, P., PAGANO, P., and ASSANTE, M. (2021). Data science: a game changer for science and innovation. *International Journal of Data Science and Analytics*, 11(4):263–278.

- GROSSI, V., RAPISARDA, B., GIANNOTTI, F., and PEDRESCHI, D. (2018). Data science at sobigdata: the european research infrastructure for social mining and big data analytics. *International Journal of Data Science and Analytics*, 6:205–216.
- GROVER, P. and KAR, A. K. (2017). Big Data Analytics: A Review on Theoretical Contributions and Tools Used in Literature. *Global Journal of Flexible Systems Management*, 18(3):203–229.
- HARGREAVES, E., MANGABEIRA, E. F., OLIVEIRA, J., FRANCA, T. C., and MC-NASCHE, D. S. (2020). Facebook news feed personalization filter: a case study during the brazilian elections. In *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 615–618, Netherlands. IEEE, IEEE.
- LIMA, G. d. F. B., OLIVEIRA, M. I. S., and LÓSCIO, B. F. (2022). FASED: A Framework for Data Ecosystems Health Evaluation. *Journal of Information and Data Management*, 13(3).
- LIMA FILHO, S. P., OLIVEIRA, J., and DA SILVA, M. F. (2020). Detection of depression symptoms using social media data. *Simpósio Brasileiro de Banco de Dados (SBB)*, 2020:3–8.
- OLIVEIRA, M. I. S., LIMA, G. d. F. B., and LÓSCIO, B. F. (2019). Investigations into Data Ecosystems: a systematic mapping study. *Knowledge and Information Systems*, 61(2):589–630.
- OLSHANNIKOVA, E., OLSSON, T., HUHTAMÄKI, J., and KÄRKKÄINEN, H. (2017). Conceptualizing Big Social Data. *Journal of Big Data*, 4(1):3.
- REHEM, D., OLIVEIRA, J., FRANÇA, T., BRITO, W., and MOTTA, C. (2016). News recommendation based on tweets for understanding of opinion variation and events. In *Proceedings of the 31st Annual ACM Symposium on Applied Computing*, pages 1182–1185, New York, NY, USA. ACM.
- ROSSETTI, G., MILLI, L., RINZIVILLO, S., SIRBU, A., PEDRESCHI, D., and GIANNOTTI, F. (2017). Ndlb: Studying network diffusion dynamics. In *2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pages 155–164.
- SHAH, S. I. H., PERISTERAS, V., and MAGNISALIS, I. (2021). DaLiF: a data lifecycle framework for data-driven governments. *Journal of Big Data*, 8(1):89.
- SILVA JÚNIOR, P. F., FRANÇA, T. C., and OLIVEIRA, J. (2023a). Caramel: Ecosystem for big social data. In *Proceedings of the XIX Brazilian Symposium on Information Systems*, volume 1, pages 136–142. ACM.
- SILVA JÚNIOR, P. F., FRANÇA, T. C., and OLIVEIRA, J. (2023b). Caramel: Um framework para ecossistema de big social data. In *Anais Estendidos do XXXVIII Simpósio Brasileiro de Banco de Dados (SBB Estendido 2023)*, pages 167–173. Sociedade Brasileira de Computação - SBC.
- WANG, X., DUAN, Q., and LIANG, M. (2021). Understanding the process of data reuse: An extensive review. *Journal of the Association for Information Science and Technology*, 72(9):1161–1182.