

Barreiras de interoperabilidade em portais de DGA: uma camada de abstração para acesso unificado

Flavia Bernardini¹, Matheus Erthal¹, Raissa Barcellos²

¹Instituto de Computação
Universidade Federal Fluminense (UFF)
Niterói, RJ

²Instituto de Matemática e Estatística (IME)
Universidade do Estado do Rio de Janeiro (UERJ)
Rio de Janeiro, RJ

fcbernardini@ic.uff.br, matheusamancio@id.uff.br,
raissa.barcellos@ime.uerj.br

Abstract. *Open government data has expanded opportunities for transparency and data-driven public services, but its effective reuse remains limited due to interoperability barriers across heterogeneous portals. This work investigates these barriers through a Rapid Review of the Literature and proposes Hlib, a Python-based programmatic interoperability layer for unified data access. An empirical evaluation involving 14 national portals revealed that only 10 could be successfully integrated, highlighting inconsistencies in APIs, meta-data, and query capabilities. Hlib addresses these issues by abstracting portal-specific differences and enabling unified, asynchronous data retrieval. The results demonstrate that interoperability challenges are persistent and concrete, reinforcing the need for programmatic abstraction to support advanced applications, including AI-based systems.*

Resumo. *A disponibilidade de dados governamentais abertos ampliou as oportunidades de transparência e serviços públicos orientados por dados, mas sua reutilização efetiva permanece limitada devido a barreiras de interoperabilidade entre portais heterogêneos. Este trabalho investiga essas barreiras por meio de uma Revisão Rápida da Literatura e propõe a Hlib, uma camada de interoperabilidade programática baseada em Python para acesso unificado a dados. Uma avaliação empírica envolvendo 14 portais nacionais revelou que apenas 10 puderam ser integrados com sucesso, evidenciando inconsistências em APIs, metadados e capacidades de consulta. A Hlib aborda essas questões abstraindo as diferenças específicas de cada portal e permitindo a recuperação unificada e assíncrona de dados. Os resultados demonstram que os desafios de interoperabilidade são persistentes e concretos, reforçando a necessidade de abstração programática para suportar aplicações avançadas, incluindo sistemas baseados em IA.*

1. Introdução

A crescente disponibilidade de dados governamentais abertos (DGA) tem consolidado portais públicos como infraestrutura fundamental para a promoção da transparência e o

desenvolvimento de serviços orientados a dados [Shah et al. 2021]. No contexto do governo digital orientado por Inteligência Artificial (IA), agentes autônomos e sistemas de *Retrieval-Augmented Generation* (RAG) dependem de acesso programático confiável a essas fontes [Ramlan et al. 2025, Akatkin and Yasinovskaya 2020]. No entanto, os portais de DGA são inerentemente heterogêneos: cada um implementa sua própria API, seu esquema de metadados e sua política de autenticação [Heise and Naumann 2012, Folmer et al. 2020], frequentemente acompanhados de inconsistências semânticas e organizacionais [Neumaier et al. 2016, Margariti et al. 2025]. Como consequência, observa-se um alto esforço de integração, não escalável e que precisa ser reiterado a cada novo portal, comprometendo tanto o reuso humano quanto o consumo automatizado por sistemas de IA [Ciancarini et al. 2024].

Embora a literatura categorize essas barreiras de forma consistente, persiste uma lacuna na disponibilização de artefatos práticos, uma vez que ainda são escassas bibliotecas de propósito geral capazes de abstrair as diferenças entre portais sem demandar alterações em suas infraestruturas. Nesse contexto, este trabalho tem por objetivo investigar as barreiras de interoperabilidade em portais de DGA e propor a Hlib, uma biblioteca Python que implementa uma camada de abstração programática para acesso unificado e assíncrono a portais heterogêneos, com suporte à integração com *frameworks* de agentes de IA.

Nesse sentido, a principal contribuição deste trabalho consiste em demonstrar, de forma empírica e prática, como uma camada de interoperabilidade programática pode mitigar as barreiras identificadas na literatura e viabilizar o uso efetivo de DGA por sistemas inteligentes. Especificamente, este trabalho oferece: (i) uma Revisão Rápida da Literatura que sistematiza barreiras de interoperabilidade nas dimensões técnica, semântica e organizacional; (ii) a Hlib, uma biblioteca Python de código aberto para acesso unificado a portais de DGA; e (iii) uma avaliação empírica envolvendo 14 portais nacionais, com evidências das barreiras observadas na prática. O restante deste trabalho está organizado da seguinte forma: a Seção 2 apresenta o referencial teórico; a Seção 3 descreve a metodologia adotada; a Seção 4 discute os resultados obtidos; e, por fim, a Seção 5 apresenta as conclusões do trabalho.

2. Referencial Teórico

2.1. Uso de IA em Governos Digitais

A adoção de IA no setor público avança em múltiplas frentes: automação de processos internos, entrega de serviços ao cidadão e tomada de decisão orientada a dados [Ciancarini et al. 2024, Shah et al. 2021]. O paradigma de *data-driven government* pressupõe que dados governamentais estejam estruturados e acessíveis como insumo para sistemas de aprendizado de máquina; sem essa infraestrutura, o potencial da IA no setor público permanece subaproveitado [Ramlan et al. 2025, Akatkin and Yasinovskaya 2020]. No domínio de dados abertos, LLMs têm sido explorados para reduzir a barreira cognitiva de acesso à informação pública. [Barcellos et al. 2023, Barcellos et al. 2024] propõem um *framework* e demonstram o uso do ChatGPT para aumentar a interpretabilidade de dados recuperados de portais de DGA; Diniz et al. [Diniz et al. 2025] identificam adoção crescente de assistentes virtuais em sistemas de e-governo. Esses usos pressupõem APIs consistentes e dados bem catalogados, condições frequentemente ausentes nos portais atuais [Neumaier et al. 2016,

Folmer et al. 2020, Musyaffa et al. 2017], o que reforça a necessidade de uma camada de normalização que habilite tanto o consumo humano quanto o processamento automatizado.

2.2. Interoperabilidade em Portais de DGA

Interoperabilidade pode ser entendida como a capacidade de sistemas e conjuntos de dados distintos trocarem informações de forma coordenada, sem perda de significado [Lodato et al. 2018]. No contexto de portais de DGA, essa capacidade se desdobra em três dimensões: *técnica* (compatibilidade de formatos, protocolos e APIs), *semântica* (consistência de vocabulários e metadados) e *organizacional* (processos e acordos institucionais que viabilizam a integração) [Margariti et al. 2025]. Na dimensão técnica, cada portal tende a implementar paginação, filtragem e autenticação de forma própria, obrigando o desenvolvedor a escrever código específico para cada fonte [Heise and Naumann 2012]. A adoção de padrões como *Linked Data* pode mitigar essa fragmentação, mas exige investimento institucional que nem todos os governos sustentam [Folmer et al. 2020]. Na dimensão semântica, a ausência de vocabulários controlados compartilhados, como DCAT, dificulta a integração entre órgãos e esferas de governo [Kato et al. 2018, Toro et al. 2019]; a qualidade dos metadados agrava o problema, com campos faltantes, identificadores ambíguos e codificações inconsistentes recorrentes entre portais [Neumaier et al. 2016, Musyaffa et al. 2017]. A dimensão organizacional, por fim, é condição necessária mas frequentemente negligenciada: sem processos e governança alinhados, a integração técnica e semântica permanece fragmentada [Margariti et al. 2025]. Em conjunto, essas barreiras evidenciam a necessidade de uma camada de abstração programática que normalize o acesso a portais heterogêneos, abstraindo diferenças de API, formato e vocabulário sem impor mudanças na infraestrutura existente.

2.3. Ferramentas de Interoperabilidade Existentes

Os trabalhos mais próximos à proposta deste estudo abordam a integração programática de fontes de dados heterogêneas ou a construção de camadas de normalização sobre portais de OGD. Lehmborg *et al.* [Lehmborg et al. 2017] apresentam o *WInte.r*, biblioteca Java modular com separação clara de responsabilidades entre *schema matching*, resolução de entidades e fusão de dados, que resolve uma camada diferente, da pós-recuperação dos dados. Beluzo [Beluzo 2015] demonstra, com o IDEO, a viabilidade de uma camada ETL para integração de portais nacionais de DGA heterogêneos. Sánchez *et al.* [Sánchez et al. 2023] e Masoumi *et al.* [Masoumi et al. 2020] reforçam que *pipelines* de normalização e contratos semânticos são estratégias eficazes para lidar com fontes estruturalmente distintas.

Contudo, essas soluções apresentam três limitações: (i) acoplamento a domínios específicos (orçamentário ou IoT), restringindo a reutilização; (ii) requisitos de infraestrutura ou expertise semântica que portais de DGA raramente satisfazem; e (iii) ausência de suporte a agentes de IA. A Hlib supera essas restrições operando exclusivamente do lado do consumidor, abstraindo padrões CKAN e APIs REST proprietárias em um modelo unificado (*Dataset*), e expondo funções registráveis como ferramentas de LLMs, sem exigir alterações nos portais fonte.

3. Metodologia

Para atingir os objetivos deste trabalho, foram adotadas duas estratégias metodológicas complementares: (i) uma Revisão Rápida da Literatura (*Rapid Review*) [Dobbins 2017], a fim de identificar abordagens e soluções existentes no estado da arte relacionadas a barreiras de interoperabilidade em portais de dados abertos governamentais; e (ii) o projeto e implementação de uma biblioteca Python, denominada Hlib, como uma camada de interoperabilidade programática que viabiliza o uso efetivo de DGA por sistemas inteligentes e agentes de IA.

3.1. Revisão Rápida da Literatura

A Revisão Rápida da Literatura buscou identificar barreiras de interoperabilidade em portais de DGA e soluções propostas para sua mitigação. A busca incluiu intencionalmente domínios adjacentes - integração de dados, *web services* e semântica - dado que soluções transferíveis frequentemente surgem fora do domínio de OGD. Embora interoperabilidade e integração de dados sejam tratadas como sinônimos na literatura, integração refere-se à combinação de fontes em visão unificada (ETL/virtualização), enquanto interoperabilidade é propriedade mais ampla que abrange dimensões técnicas, semânticas e organizacionais [Margariti et al. 2025]. A pesquisa foi conduzida em três bases científicas: ACM Digital Library, Scopus e IEEE Xplore, com o apoio da ferramenta Parsifal¹, com *strings* adaptadas ao vocabulário de cada base:

- **ACM e IEEE:** ("open data" OR "government data" OR "e-government data") AND (interoperabilit* OR "data integration" OR "system integration" OR "data standard*") AND ("implementation*" OR "solution*" OR "prototype*")
- **Scopus:** TITLE-ABS-KEY(("open government data" OR "government data portal" OR "open data portal") AND (interoperabilit* OR "data integration" OR "semantic standard*" OR "data exchange") AND ("case study" OR "framework" OR "architecture" OR "implementation"))

Em todas as bases, foram aplicados filtros de refinamento: restrição a trabalhos publicados em periódicos (*journals*) e janela temporal a partir de 2010. A *string* original para o Scopus retornou mais de sete mil resultados, exigindo reformulação com termos mais específicos para tornar o corpus analisável.

A seleção dos trabalhos foi guiada por critérios de inclusão e exclusão pre-definidos. Os critérios de inclusão foram: (I1) o trabalho aborda DGA; (I2) o trabalho aborda interoperabilidade de dados; (I3) o trabalho propõe ou avalia uma solução técnica - *framework*, arquitetura, ferramenta ou protótipo; e (I4) o trabalho trata de sistemas ou fontes de dados heterogêneas, mesmo fora do domínio de OGD. Os critérios de exclusão foram: (E1) o trabalho não aborda DGA; (E2) o trabalho não aborda interoperabilidade de dados; (E3) o trabalho é exclusivamente conceitual, político ou normativo, sem contribuição técnica avaliável; (E4) os dados tratados são exclusivamente privados ou comerciais, sem relação com dados abertos; e (E5) o texto completo não está disponível para

¹<https://parsif.al/>

leitura. Ao final, 50 trabalhos foram selecionados como relevantes após a aplicação desses critérios, compondo o corpus analisado neste trabalho. Os resultados da Revisão Rápida influenciaram diretamente o projeto da Hlib, garantindo que a solução proposta abordasse as barreiras de interoperabilidade observadas empiricamente.

3.2. Proposta da Hlib

A contribuição técnica central deste trabalho é a Hlib, biblioteca Python desenvolvida como implementação do módulo de recuperação de dados de um *framework* conceitual proposto pelos próprios autores em trabalho anterior [Barcellos et al. 2023, Barcellos et al. 2024]. Esse *framework*, de autoria dos mesmos autores, estrutura-se em três módulos funcionais: enriquecimento semântico, recuperação de dados e visualização, voltados a aprimorar a interpretabilidade de DGA; a presente implementação cobre exclusivamente o módulo de recuperação.

As decisões de design da Hlib foram diretamente guiadas pelas barreiras sistematizadas na Revisão Rápida: as barreiras técnicas de heterogeneidade de APIs motivaram a abstração `DataAdapter` (camada de transporte); as barreiras semânticas de inconsistência de metadados motivaram o modelo unificado `Dataset`; e as barreiras organizacionais de autenticação variável motivaram o tratamento parametrizado de credenciais na interface `Portal`. A biblioteca está disponível publicamente no ambiente PyPI ².

Arquitetura e Extensibilidade

A arquitetura da Hlib organiza-se em duas camadas de abstração, ilustradas na Figura 1. A `DataAdapter` isola a comunicação HTTP: é responsável exclusivamente pelo transporte, ou seja, gerenciar sessões assíncronas, executar requisições e converter arquivos baixados em `DataFrame`. Por sua vez, a camada `Portal` isola a semântica de cada fonte: conhece os *endpoints*, o esquema de metadados, as regras de autenticação e o mapeamento de campos para o modelo unificado `Dataset`. Ambas são definidas como Classes Base Abstratas (ABCs), com métodos abstratos que cada implementação concreta deve satisfazer. Essas três classes (`DataAdapter`, `Portal` e `Dataset`) compõem os alicerces da arquitetura proposta neste trabalho.

A relação entre as duas camadas é de composição: todo portal concreto instancia e encapsula um adapter adequado ao protocolo do portal. O `Portal` delega ao `DataAdapter` as operações de transporte, mantendo-se agnóstico ao protocolo; o `DataAdapter`, por sua vez, é agnóstico à semântica de qualquer portal. Essa separação garante que uma mudança de protocolo (ex.: migração de REST para CKAN) afete apenas o adapter instanciado, sem alterar a lógica de mapeamento do portal, e vice-versa. A biblioteca fornece dois adapters concretos: `CkanAdapter`, especializado no protocolo CKAN v3, e `ApiAdapter`, para APIs REST com estruturas proprietárias.

O método `fetch_dataset_data()`, definido como concreto na interface `Portal`, orquestra o fluxo completo (busca de metadados, identificação do recurso e delegação do download ao adapter) sem que nenhum portal precise reimplementá-lo. Adicionar um novo portal requer apenas implementar a interface `Portal` com o adapter adequado e registrá-lo no enumerador `PortalType`.

Seleção de Portais

²<https://pypi.org/>

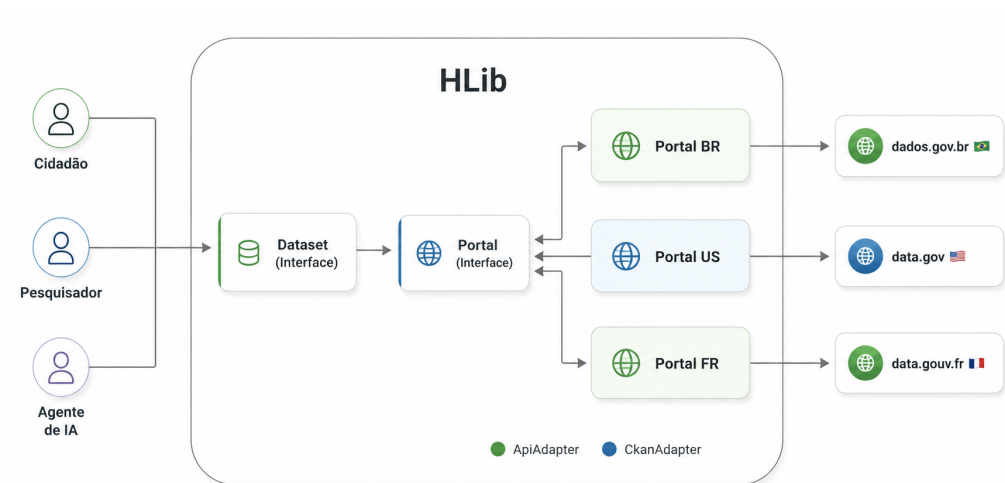


Figure 1. Arquitetura em camadas da Hlib: a camada Portal (semântica) compõe a camada DataAdapter (transporte), expondo ao consumidor uma API unificada independente do protocolo de cada portal.

A seleção dos portais foi orientada por três estudos comparativos de portais de DGA em nível nacional [Máchová and Lněnička 2017, Nikiforova and McBride 2021, Sheoran et al. 2023], cobrindo até 75 países com critérios de qualidade, usabilidade e acessibilidade. Foram priorizados portais com representatividade geográfica diversa, presença recorrente nos três estudos e desempenho positivo em ao menos uma das dimensões avaliadas. A versão atual da Hlib integra dez portais nacionais de dados abertos, abrangendo quatro continentes: Brasil, Estados Unidos, Reino Unido, Suíça, Finlândia, Austrália, França, Espanha, Singapura e Índia. Durante o desenvolvimento, quatro portais adicionais foram investigados e descartados por barreiras técnicas intransponíveis: migração para *Single Page Applications* sem API REST pública (Áustria e Chipre), ausência de *endpoints* acessíveis (Rússia) e restrição de autenticação para operações de busca (Taiwan). Esses casos são documentados como evidência empírica das barreiras de interoperabilidade abordadas neste trabalho.

Integração com IA

A Hlib foi projetada para servir como camada de acesso a dados em *pipelines* que incorporam ferramentas de IA, em particular Grandes Modelos de Linguagem (LLMs). Ao abstrair as diferenças de API, formato e autenticação entre portais heterogêneos, a biblioteca permite que agentes de IA e sistemas RAG (*Retrieval-Augmented Generation*) consultem e recuperem dados governamentais de múltiplas fontes sem necessidade de código específico por portal. Isso viabiliza cenários como a busca semântica em dados abertos governamentais por meio de interfaces em linguagem natural, conforme explorado por Barcellos et al. [Barcellos et al. 2024] em trabalho complementar ao *framework* aqui implementado.

4. Resultados

4.1. Barreiras de Interoperabilidade Identificadas na Literatura

A análise dos 50 trabalhos selecionados na revisão rápida evidencia que as barreiras de interoperabilidade em portais de dados abertos governamentais são recorrentes, multidimensionais e persistentes ao longo da última década [Heise and Naumann 2012, Neumaier et al. 2016, Margariti et al. 2025]. Seguindo a taxonomia de interoperabilidade estabelecida, os resultados podem ser organizados em três dimensões: Barreiras Técnicas, Barreiras Semânticas e Barreiras Organizacionais.

Barreiras Técnicas

A heterogeneidade de formatos de arquivo e de interfaces programáticas é a barreira técnica mais frequentemente apontada. A multiplicidade de formatos (CSV, JSON, XML, RDF, XLSX) exige que consumidores implementem lógica de parsing específica para cada fonte [Heise and Naumann 2012]. A ausência de padronização nas APIs constitui outro obstáculo central, identificado tanto em contextos nacionais [Beluzo 2015] quanto em análises de portais internacionais [Folmer et al. 2020]. A inconsistência na representação de recursos para download (URLs diretas, redirecionamentos, acesso em duas etapas) agrava esse quadro, dificultando a automação do *pipeline* de coleta de dados. Portais baseados em *Single Page Applications* sem API backend exposta figuram como caso-limite: os dados são acessíveis a humanos mas inacessíveis a máquinas [Meyerhoff Nielsen and Jordanoski 2023], o que contraria os princípios fundamentais do dado aberto.

Barreiras Semânticas

A dimensão semântica manifesta-se principalmente na incompletude e inconsistência dos metadados. Neumaier et al. [Neumaier et al. 2016] avaliam automaticamente a qualidade de metadados em mais de 261 portais e demonstram que campos ausentes e inconsistentes comprometem diretamente a descoberta e a integração de conjuntos de dados. Musyaffa et al. [Musyaffa et al. 2017] identificam problemas análogos no domínio de dados fiscais abertos: identificadores ambíguos, codificações inconsistentes e campos faltantes são recorrentes entre portais de diferentes países. A ausência de vocabulários controlados compartilhados, como ontologias de domínio ou perfis DCAT padronizados, dificulta a inferência semântica de equivalência entre datasets nominalmente distintos de diferentes portais [Masoumi et al. 2020, Kato et al. 2018]. O multilinguismo não padronizado representa uma barreira semântica adicional em contextos internacionais, onde a mesma informação pode ser representada em diferentes idiomas com estruturas JSON incompatíveis entre si [Toro et al. 2019].

Barreiras Organizacionais

As barreiras organizacionais incluem a descontinuidade de plataformas sem garantia de *backward compatibility* de APIs, a fragmentação de responsabilidades entre esferas de governo e a adoção de políticas de autenticação que contradizem os princípios do dado aberto. Margariti et al. [Margariti et al. 2025] demonstram que a interoperabilidade técnica e semântica são condições necessárias, mas não suficientes: sem acordos institucionais e modelos de governança alinhados, a integração permanece fragmentada mesmo quando padrões técnicos estão disponíveis. Jetzek [Jetzek 2016] aponta que mesmo em contextos com forte coordenação governamental, como o Programa Básico

de Dados da Dinamarca, a multiplicidade de padrões de publicação entre órgãos distintos gera fricção significativa. No Brasil, a exigência de chave de API para operações de leitura em dados publicamente disponíveis exemplifica como decisões organizacionais podem criar barreiras programáticas incompatíveis com os princípios de dados abertos [Penteado et al. 2022, Beluzo et al. 2021].

4.2. Avaliação da Hlib

A avaliação empírica da Hlib foi conduzida por meio da tentativa sistemática de integração de 14 portais nacionais de dados abertos governamentais, abrangendo diferentes continentes, plataformas e modelos de API. Para cada portal, foram realizadas: (i) descoberta de *endpoints* REST/JSON; (ii) teste de conectividade com requisições HTTP reais; (iii) mapeamento do esquema de metadados para o modelo unificado Dataset; e (iv) implementação e validação com testes automatizados. O critério de integração bem-sucedida exigiu a execução correta das três operações públicas da biblioteca: `search_data()` (busca por palavra-chave), `get_dataset()` (recuperação de metadados por ID) e `fetch_dataset_data()` (download e *parsing* de um recurso em formato tabular).

Portais Avaliados

Dos 14 portais investigados, 10 foram integrados com sucesso e 4 apresentaram barreiras técnicas intransponíveis. Os integrados foram classificados em dois grupos: cinco baseados em CKAN v3 (Estados Unidos, Reino Unido, Suíça, Finlândia e Austrália) e cinco com APIs proprietárias (Brasil, França, Espanha, Singapura e Índia). Os quatro portais descartados ilustram categorias distintas de barreira organizacional e técnica. Chipre e Áustria migraram suas plataformas (de CKAN para Drupal e para SPA Vue.js, respectivamente) sem manter ou documentar um substituto para a API anterior (os dados continuam acessíveis via navegação manual, mas tornaram-se inacessíveis programaticamente de forma silenciosa). A Rússia opera inteiramente como SPA sem API *backend* exposta. Taiwan mantém o acesso individual a datasets via GET, mas restringe a operação de busca/listagem a POST autenticado com chave de API, sem documentação pública desse requisito.

Evidências Empíricas de Interoperabilidade

A integração dos 10 portais bem-sucedidos revelou um conjunto de barreiras técnicas e semânticas que confirmam empiricamente os achados da revisão da literatura.

1. **Heterogeneidade de identificadores:** Nenhum padrão de identificação de datasets é compartilhado entre os portais. Os formatos variam entre UUIDs (portais CKAN), URIs completas (Espanha), identificadores alfanuméricos (França), IDs com prefixo de hash (Singapura) e UUIDs encapsulados em arrays (Índia). Essa ausência de convenção impede referências cruzadas entre portais.
2. **Campos de metadados ausentes:** A análise do mapeamento de campos revelou que nenhum campo do modelo unificado está presente em todos os portais com a mesma representação. Campos críticos como descrição, tags, licença e organização publicadora estão ausentes em um ou mais portais, sendo a licença o campo mais frequentemente omitido (ausente em Espanha, Singapura e Índia).
3. **Inconsistência na capacidade de busca textual:** A funcionalidade de busca por texto livre varia de busca *full-text* via Apache Solr (portais CKAN) a filtro ape-

nas por nome do dataset (Índia), busca sem filtro textual (Espanha) e ausência completa de *endpoint* de busca funcional (Taiwan). Essa variação impõe ao consumidor o ônus de adaptar sua estratégia de busca por portal.

4. **Multilinguismo não padronizado:** Três abordagens estruturalmente incompatíveis foram encontradas para campos multilíngues: dicionário indexado por idioma (Suíça), lista de objetos com atributo de idioma (Espanha) e dicionário de listas por idioma para tags (Finlândia). Cada abordagem requer lógica de extração diferente e tratamento de *fallback* de idioma.
5. **Autenticação obrigatória para leitura:** O portal brasileiro exige o envio de uma chave de API em todas as requisições, inclusive de leitura. A ausência do header não gera mensagem de erro descritiva (o servidor retorna uma cadeia de redirecionamentos HTTP) o que dificulta o diagnóstico e contradiz os princípios de DGA.

Em conjunto, essas evidências demonstram que as barreiras de interoperabilidade se manifestam de forma concreta e recorrente, reforçando a necessidade de uma camada de abstração unificada, como a implementada na Hlib.

Uso em Sistemas Baseados em IA

A Hlib foi projetada de modo a servir como ferramenta nativa em arquiteturas de agentes de IA e sistemas RAG (*Retrieval-Augmented Generation*). Ao encapsular toda a complexidade de integração com portais heterogêneos, a biblioteca permite que LLMs consultem dados governamentais de múltiplas fontes sem nenhum conhecimento das APIs subjacentes. Barcellos et al. [Barcellos et al. 2024] demonstram que LLMs já são capazes de aprimorar a interpretabilidade de dados abertos governamentais quando dotados de acesso programático adequado; a Hlib fornece exatamente a camada de acesso de forma portátil e extensível. *Frameworks* de agentes de IA como *LangChain*, *LlamaIndex* e o *Claude Agent SDK* permitem registrar funções Python como ferramentas disponíveis ao modelo. Diante desse cenário, as funções `search_data()` e `fetch_dataset_data()` da Hlib podem ser expostas diretamente como ferramentas de um agente, viabilizando fluxos como: (i) o usuário descreve em linguagem natural uma necessidade de dados; (ii) o LLM formula uma query e invoca `search_data()` para explorar os portais; (iii) o modelo seleciona o dataset mais relevante e invoca `fetch_dataset_data()` para obter os dados em formato tabular; (iv) o LLM analisa e responde com base nos dados reais recuperados. Esse padrão materializa o cenário de governo orientado a dados descrito por Akatkin e Yasinovskaya [Akatkin and Yasinovskaya 2020] e discutido por Ramlan et al. [Ramlan et al. 2025], onde ferramentas de IA operam sobre DGA como insumo para geração de valor.

5. Conclusão

Este trabalho investigou as barreiras de interoperabilidade presentes em portais de dados abertos governamentais e propôs a biblioteca Python Hlib como uma camada de interoperabilidade programática para sua mitigação. A revisão rápida da literatura evidenciou que essas barreiras são multidimensionais e persistentes: no plano técnico, manifestam-se na heterogeneidade de APIs, formatos e esquemas de metadados; no plano semântico, na ausência de vocabulários compartilhados e na incompletude de metadados; e no plano

organizacional, na descontinuidade de plataformas, na fragmentação de governança e em políticas de autenticação incompatíveis com os princípios do dado aberto.

A avaliação empírica da Hlib, conduzida sobre 14 portais nacionais, corroborou esses achados ao demonstrar, de forma concreta, a recorrência dessas barreiras em cenários reais. Dos portais investigados, apenas 10 puderam ser integrados, sendo os demais exemplos diretos de limitações técnicas e organizacionais. Mesmo entre os portais integrados, observou-se a ausência de padronização de metadados e inconsistências significativas em funcionalidades essenciais, como busca textual e acesso programático. Nesse contexto, a Hlib demonstra que a adoção de uma camada de abstração em camadas, é capaz de encapsular tais heterogeneidades e viabilizar o consumo unificado de DGA, inclusive em cenários que envolvem sistemas baseados em inteligência artificial.

Apesar das contribuições apresentadas, este trabalho possui limitações que devem ser consideradas. A Revisão Rápida da Literatura pode ter excluído trabalhos relevantes devido às restrições de bases e critérios de seleção adotados. A avaliação da Hlib concentrou-se exclusivamente no módulo de recuperação de dados, não contemplando os módulos de enriquecimento semântico e visualização do *framework* original. A cobertura geográfica, embora diversificada, permanece limitada a um conjunto específico de países em um recorte temporal único. Uma limitação estrutural relevante também diz respeito à escalabilidade da abordagem: cada novo portal demanda uma implementação manual específica (interface `Portal` + `adapter` adequado), e qualquer mudança na infraestrutura de um portal existente, como a migração de plataforma observada em Áustria e Chipre, invalida a integração correspondente sem aviso. Esse padrão de manutenção manual reflete a heterogeneidade intrínseca e volátil do ecossistema de DGA, e não constitui uma limitação exclusiva da Hlib, mas um desafio inerente a qualquer camada de abstração sobre portais governamentais não padronizados.

Como direções futuras, destacam-se três frentes principais. A primeira consiste na validação empírica da integração da Hlib com grandes modelos de linguagem, por meio de experimentos que avaliem quantitativamente seu impacto na recuperação e uso de dados governamentais. A segunda refere-se à extensão da biblioteca com os demais módulos do *framework* originalmente proposto, incorporando capacidades de enriquecimento semântico e visualização. A terceira, motivada pela limitação de escalabilidade identificada, é a investigação de mecanismos de detecção automática de esquema de API, que permitam que novos portais sejam integrados com configuração mínima, reduzindo a dependência de implementações manuais e tornando a biblioteca mais resiliente.

Em síntese, este trabalho demonstra que as barreiras de interoperabilidade em DGA não são apenas conceituais, mas se manifestam de forma concreta e recorrente na prática. Nesse sentido, a proposta de uma camada de interoperabilidade programática representa um passo fundamental para reduzir a fragmentação atual e viabilizar o uso efetivo desses dados em aplicações avançadas. Sem abordar a interoperabilidade no nível programático, o potencial dos DGA para apoiar serviços públicos inteligentes permanece fundamentalmente limitado.

References

Akatkin, Y. and Yasinovskaya, E. (2020). *Data-Driven Government in Russia: Linked Open Data Challenges, Opportunities, Solutions*, pages 245–257. Springer.

- Barcellos, R., Bernardini, F., Viterbo, J., and Zuiderwijk, A. (2023). Hippolyta: a framework to enhance open data interpretability and empower citizens. In *Proceedings of the 24th Annual International Conference on Digital Government Research*, DGO '23. ACM.
- Barcellos, R., Bernardini, F., Zuiderwijk, A., and Viterbo, J. (2024). Exploring interpretability in open government data with ChatGPT. In *Proceedings of the 25th Annual International Conference on Digital Government Research*, DGO '24. ACM.
- Beluzo, J., Dias, B., Craveiro, G., and Araujo, R. (2021). Re-public: workflow to publish and reuse linked open government data. In *Anais do XVII Simpósio Brasileiro de Sistemas de Informação*.
- Beluzo, J. R. (2015). IDEO – integrador de dados da execução orçamentária brasileira: Um estudo de caso da integração de dados das receitas e despesas nas esferas federal, estadual – governo de são paulo, e municipal – municípios do estado de são paulo. Dissertação de mestrado em ciências – programa de pós-graduação em sistemas de informação, Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, Brazil. Orientadora: Profa. Dra. Gisele da Silva Craveiro.
- Ciancarini, P., Giancarlo, R., and Grimaudo, G. (2024). Digital transformation in the public administrations: A guided tour for computer scientists. *IEEE Access*, 12:22841–22865.
- Diniz, P. H., Reis, L. P., and Cardoso, H. L. (2025). E-government systems to improve citizen-to-public administration communication: A literature review. *Digital Government: Research and Practice*, 6(4).
- Dobbins, M. (2017). Rapid review guidebook. *Natl Collab Cent Method Tools*, 13:25.
- Folmer, E., Ronzhin, S., Van Hillegersberg, J., Beek, W., and Lemmens, R. (2020). Business rationale for linked data at governments: A case study at the netherlands' kadaster data platform. *IEEE Access*, 8:70822–70835.
- Heise, A. and Naumann, F. (2012). Integrating open government data with stratosphere for more transparency. *Journal of Web Semantics*, 14:45–56.
- Jetzek, T. (2016). Managing complexity across multiple dimensions of liquid open data: The case of the danish basic data program. *Government Information Quarterly*, 33(1):89–104.
- Kato, F., Takeda, H., Tashiro, S., Hiramoto, K., and Matsuzawa, Y. (2018). *IMI: A Common Vocabulary Framework for Open Government Data*, pages 336–351. Springer International Publishing.
- Lehmberg, O., Bizer, C., and Brinkmann, A. (2017). WInte.r – a web data integration framework. In *Proceedings of the ISWC 2017 Posters & Demonstrations and Industry Tracks*, volume 1963 of *CEUR Workshop Proceedings*. CEUR-WS.org.
- Lodato, T., French, E., and Clark, J. (2018). Open government data in the smart city: Interoperability, urban knowledge, and linking legacy systems. *Journal of Urban Affairs*, 43(4):586–600.

- Máchová, R. and Lněnička, M. (2017). Evaluating the quality of open data portals on the national level. *Journal of Theoretical and Applied Electronic Commerce Research*, 12:21–41.
- Margariti, V., Stamati, T., Anagnostopoulos, D., Nikolaidou, M., and Raptis, A. (2025). Evaluation of reliability and consistency of a new model for organizational interoperability assessment. *Digital Government: Research and Practice*, 6(3):1–16.
- Masoumi, H., Farahani, B., and Aliee, F. S. (2020). An ontology-based open data interoperability approach for cross-domain government data services. In *2020 25th International Computer Conference, Computer Society of Iran (CSICC)*, pages 1–8. IEEE.
- Meyerhoff Nielsen, M. and Jordanoski, Z. (2023). Digital transformation, governance, and coordination in times of crisis: An analysis of australia, denmark, and the republic of korea. *Digital Government: Research and Practice*, 4(4):1–20.
- Musyaffa, F. A., Engels, C., Vidal, M.-E., Orlandi, F., and Auer, S. (2017). Experience: Open fiscal datasets, common issues, and recommendations. *Journal of Data and Information Quality*, 9(4):1–10.
- Neumaier, S., Umbrich, J., and Polleres, A. (2016). Automated quality assessment of metadata across open data portals. *Journal of Data and Information Quality*, 8(1):1–29.
- Nikiforova, A. and McBride, K. (2021). Open government data portal usability: A user-centred usability analysis of 41 open government data portals. *Telematics and Informatics*, 58:101539.
- Penteado, B., Isotani, S., Bittencourt, I. I., Mello, R. F., and Bittencourt, I. (2022). An architecture for monitoring public educational policies based on big open linked data. In *Anais do XXXIII Simpósio Brasileiro de Informática na Educação*.
- Ramlan, N., Fauzia, F., Hesti, L., Ramlan, N., and Sensuse, D. (2025). Developing a big data strategy: A public sector perspective from the ministry of tourism and creative economy. *Digital Government: Research and Practice*, 6(4).
- Shah, S. I. H., Peristeras, V., and Magnisalis, I. (2021). Government big data ecosystem: Definitions, types of data, actors, and roles and the impact in public administrations. *Journal of Data and Information Quality*, 13(2):1–25.
- Sheoran, S., Mohanasundaram, S., Kasilingam, R., and Vij, S. (2023). Usability and accessibility of open government data portals of countries worldwide: An application of TOPSIS and entropy weight method. *International Journal of Electronic Government Research*, 19(1).
- Sánchez, L., Lanza, J., Santana, J. R., Sotres, P., González, V., Martín, L., Solmaz, G., Kovacs, E., Dietzel, M., Summa, A., Jafari, A. R., Minerva, R., and Crespi, N. (2023). Data enrichment toolchain: A data linking and enrichment platform for heterogeneous data. *IEEE Access*, 11:103079–103091.
- Toro, J. F., Carrion, D., Albertella, A., and Brovelli, M. A. (2019). Cross-border open data sharing: Gioconda project. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-4/W14:233–238.