

Gov Hub: Plataforma livre de integração e qualificação de dados governamentais

Davi Vieira

Faculdade de Ciências e Tecnologias
em Engenharia, Universidade de
Brasília (FCTE/UnB)
Brasília, Brasil
davideaguiarvieira@gmail.com

Isaque Alves

Instituto de Matemática e Estatística,
Universidade de São Paulo (IME/USP)
São Paulo, Brasil
isaque.alves@ime.usp.br

Mateus Santos

Faculdade de Ciências e Tecnologias
em Engenharia, Universidade de
Brasília (FCTE/UnB)
Brasília, Brasil
mateus.castro3@gmail.com

Arthur Alves Melo

Faculdade de Ciências e Tecnologias
em Engenharia, Universidade de
Brasília (FCTE/UnB)
Brasília, Brasil
arthuralves1538@gmail.com

Milena Rocha

Faculdade de Ciências e Tecnologias
em Engenharia, Universidade de
Brasília (FCTE/UnB)
Brasília, Brasil
milenafernandes594@gmail.com

Wallyson Souza

Faculdade de Ciências e Tecnologias
em Engenharia, Universidade de
Brasília (FCTE/UnB)
Brasília, Brasil
souzawpcs@gmail.com

Carla Silva Rocha Aguiar

Faculdade de Ciências e Tecnologias
em Engenharia, Universidade de
Brasília (FCTE/UnB)
Brasília, Brasil
caguiar@unb.br

Resumo

A atual falta de interoperabilidade entre os sistemas estruturantes do Governo Federal tem dificultado a consolidação de um governo digital mais transparente, eficiente e orientado a dados. Essa fragmentação resulta em alto custo operacional para a obtenção, qualificação e análise de informações necessárias à tomada de decisões estratégicas. Este artigo apresenta o Gov Hub, uma plataforma de software livre desenvolvida em parceria com o Instituto de Pesquisa Econômica Aplicada (Ipea) para mitigar essas barreiras de integração. A partir do estudo e mapeamento dos fluxos da política pública e da jornada dos dados, a plataforma combina práticas de governança com ferramentas de orquestração, extração, carga, transformação e visualização de dados. O artigo detalha a arquitetura do sistema e apresenta um cenário de uso real. Como resultado, o Gov Hub possibilitou uma visão integrada da jornada dos dados, ampliando a qualidade das informações e qualificando metadados para futuras aplicações de inteligência artificial. Os feedbacks coletados indicam a automação de processos anteriormente manuais e a atualização frequente de painéis de gestão, sugerindo que a ferramenta reduz o esforço operacional e estabelece uma nova perspectiva estrutural para iniciativas de transparência e governo digital.

Vídeos de demonstração: <https://zenodo.org/records/20359710> e <https://www.youtube.com/watch?v=Lo9o900UG1g>.

Palavras-chave

Governo digital, Interoperabilidade, Governança de dados, Engenharia de dados, Transparência pública; Gov Hub.

1 Introdução

A modernização da administração pública ultrapassou a barreira da mera digitalização de rotinas burocráticas e ingressou na era do Governo Digital [9, 23]. Esse novo paradigma não se limita a oferecer formulários e serviços online, mas exige uma mudança na lógica de atuação do Estado, que passa a ter o uso estratégico da informação como pilar central. Sob essa ótica, a transparência deixa de ser apenas o cumprimento de uma obrigação legal de publicar dados brutos e passa a representar a necessidade de informações claras e acionáveis para a sociedade e para os próprios gestores [3, 15].

No entanto, a transição para uma gestão baseada em evidências esbarra em um gargalo tecnológico profundo: a fragmentação das plataformas governamentais. No âmbito federal, os Sistemas Estruturantes (SE) responsáveis pelos ciclos orçamentário, financeiro e de gestão de pessoas, concebidos em diferentes épocas e com arquiteturas distintas, operam predominantemente em silos. Essa limitação compromete a interoperabilidade entre plataformas governamentais de troca de dados, dimensão apontada como central para a consolidação de iniciativas de governo digital [16]. Na prática, a ausência de mecanismos nativos de integração cria um cenário em que o dado nasce, transita e morre isolado, dificultando o compartilhamento, o reuso e a articulação institucional necessários à colaboração de dados no setor público [22]. Como consequência, os órgãos públicos mantêm baixa capacidade de consolidar uma visão unificada sobre a execução de suas políticas públicas.

Para contornar esse obstáculo sem a necessidade de substituir abruptamente sistemas legados complexos, as abordagens orientadas a dados têm se consolidado como a alternativa de maior impacto. Ao focar na extração, no tratamento e no cruzamento direto

das informações geradas por esses sistemas, o Estado consegue estabelecer uma sólida governança de dados, associando a qualidade e a gestão da informação a dimensões centrais do processo [4]. Essa estratégia cria uma camada de integração analítica que diagnostica falhas nos processos operacionais e devolve ao gestor o domínio sobre a informação.

Nesse contexto, este artigo apresenta o relato de desenvolvimento e a estrutura do Gov Hub, uma plataforma de software livre projetada para integrar e qualificar dados da execução orçamentária e financeira do Governo Federal (incluindo sistemas como SIOP, SIAFI, TransfereGov, ComprasGov e SIAPE). Utilizando arquiteturas modernas de engenharia de dados, a ferramenta centraliza as informações, viabiliza o consumo por meio de dashboards interativos e cria um ambiente propício à gestão pública orientada a dados.

Além desta introdução, o restante deste artigo está organizado da seguinte forma, a Seção 2 discute os desafios relacionados à ausência de governança e de interoperabilidade entre os sistemas estruturantes do Governo Federal, a Seção 3 apresenta a plataforma Gov Hub, abrangendo o mapeamento do fluxo da política pública e da jornada dos dados, bem como sua arquitetura de engenharia e governança de dados, a Seção 4 descreve um cenário de uso da plataforma no contexto do Ipea e apresenta os resultados observados a partir de sua adoção e a Seção 5 apresenta as conclusões e as perspectivas decorrentes do trabalho. Por fim, uma seção dedicada à disponibilidade do artefato, na qual são apresentadas as condições para sua utilização e evolução.

2 A Falta de Governança de Dados

A eficiência operacional da administração pública depende de uma comunicação fluida entre os Sistemas Estruturantes (SE), como o SIOP, o SIAFI, o ComprasGov e o TransfereGov. No entanto, por terem sido projetadas em contextos tecnológicos distintos e legados, essas plataformas operam nativamente como silos independentes. A principal consequência dessa arquitetura fragmentada é a ausência de uma governança de dados estruturada entre os órgãos [17]. Na prática, as bases de dados encontram-se pulverizadas e carecem de padronização, não compartilham metadados nem chaves primárias.

Essa desconexão estrutural compromete a rastreabilidade ao longo da jornada do dado governamental. Por exemplo, o fluxo de um recurso se inicia com o planejamento no SIOP, avança para a geração da despesa em plataformas específicas (como o ComprasGov ou TransfereGov) e, finalmente, deve ter seus reflexos contábeis consolidados no SIAFI. Como não existem amarrações lógicas diretas entre o sistema que gera a despesa e o sistema que a contabiliza, o acompanhamento integral dessa execução exige processos manuais por parte dos gestores públicos, frequentemente apoiados em planilhas e relatórios.

Sem um dicionário de dados centralizado e fluxos automatizados, torna-se inviável realizar consultas integradas ou análises sistêmicas complexas [18]. A dificuldade em estruturar a informação bruta inviabiliza a geração de indicadores gerenciais precisos, tornando o planejamento financeiro e a fiscalização processos morosos. Para superar esse bloqueio, a criação de uma camada de engenharia capaz de unificar e qualificar essas fontes fragmentadas torna-se um pré-requisito para qualquer avanço significativo na gestão pública.

Soluções como CKAN [7], OpenSpending [20], OpenFisca [19] e o Open Contracting Data Standard (OCDS) [1] contribuem para diferentes aspectos do ecossistema de dados governamentais, como publicação de dados abertos, transparência fiscal, modelagem de políticas públicas e padronização de informações sobre contratações. Entretanto, essas iniciativas atuam sobre dados já estruturados ou em domínios específicos, não abordando a interoperabilidade entre os diversos Sistemas Estruturantes da administração pública. Essa lacuna motiva o Gov Hub, que provê uma arquitetura aberta para integrar, governar e disponibilizar dados qualificados provenientes de múltiplos sistemas governamentais.

3 A Plataforma Gov Hub

3.1 Introdução da Ferramenta

O Gov Hub é uma plataforma de software livre desenvolvida por meio de uma parceria entre o Instituto de Pesquisa Econômica Aplicada (Ipea) e o Laboratório de Competência em Software Livre (Lab Livre) da Universidade de Brasília (UnB). A ferramenta foi concebida com um propósito fundamental duplo: atuar como um instrumento prático de apoio à gestão pública e elevar o nível de transparência dos gastos governamentais para a sociedade civil. Ao romper os silos tecnológicos e consolidar informações dispersas, a integração de dados promovida pelo Gov Hub serve a diversos propósitos institucionais, permitindo desde a geração de diagnósticos estratégicos baseados em evidências até o fortalecimento de uma cultura de dados abertos no âmbito federal.

Além da parceria com o Ipea, a arquitetura de integração e qualificação de dados do Gov Hub tem sido adotada por outros órgãos da esfera federal, que passaram a utilizar a infraestrutura do projeto como alicerce para iniciativas próprias de governança e de análise de informações. Entre as expansões e adoções recentes, destacam-se: o Ministério das Cidades (MCID), usando a plataforma para apoiar o acompanhamento da conjuntura imobiliária, no monitoramento das linhas de financiamento e subsídios dos programas habitacionais, além do acompanhamento do impacto social dos projetos; o Ministério da Gestão e da Inovação em Serviços Públicos (MGI) para estruturar e otimizar as análises do Observatório de Compras Públicas; e o Ministério da Igualdade Racial (MIR) para embasar o acompanhamento, o monitoramento e a avaliação de políticas públicas específicas da pasta.

3.2 Mapeamento do Fluxo da Política e dos Dados

Uma metodologia fundamental da solução consiste na modelagem e estruturação do domínio governamental, etapa analítica que precede e orienta a própria engenharia de dados. Na prática, identificou-se que os próprios gestores, os detentores institucionais e donos dos dados, frequentemente não possuíam uma visão consolidada e sistêmica do trânsito da informação. Dessa forma, o mapeamento do fluxo da política pública e da jornada dos dados consolidou-se como condição necessária à governança, servindo de base para o trabalho da equipe de engenharia. Esse processo exigiu a aproximação entre conhecimento de domínio, engenharia de software e ciência de dados [5], além da compreensão do problema, das responsabilidades institucionais, das regras de negócio e das expectativas

dos atores envolvidos, o que se aproxima de práticas de engenharia de requisitos para sistemas intensivos em dados [2].

A Figura 1 apresenta o modelo conceitual resultante desse mapeamento, organizado em três camadas que refletem a lógica institucional do ciclo de gestão pública:

- **Planejamento e Alocação:** corresponde à etapa inicial, na qual são definidos os marcos legais, a dotação orçamentária e a programação financeira. Nessa camada, os recursos são planejados e alocados formalmente por meio de instrumentos como o Plano Plurianual (PPA), a Lei de Diretrizes Orçamentárias (LDO) e a Lei Orçamentária Anual (LOA), cujos dados são armazenados em sistemas como o SIOP.
- **Execução Operacional:** agrupa os módulos nos quais os recursos planejados são efetivamente consumidos. Essa camada abrange os processos de contratação e de suprimentos (ComprasGov), a gestão de transferências voluntárias e obrigatórias (TransfereGov) e a administração de pessoal (SIAPE). É nesse estágio que as despesas são geradas e os compromissos financeiros formalizados.
- **Consolidação Financeira e Controle:** representa o ponto de convergência contábil, no qual os reflexos financeiros de todas as operações anteriores são registrados e consolidados. O SIAFI atua como sistema integrador dessa camada, centralizando empenhos, liquidações e pagamentos.

O diagrama dispõe as etapas como um ciclo: a consolidação contábil de um exercício realimenta o planejamento do seguinte, mostrando que a gestão do recurso é contínua. Reunir esse percurso em uma única representação, antes disperso entre sistemas, dá à equipe a visão sistêmica do fluxo.



Figure 1: Diagrama do fluxo da política pública.

Compreender essa dinâmica processual foi o pré-requisito para arquitetar a coleta de dados, ilustrada na Figura 2. O diagrama demonstra o sequenciamento técnico exigido para o consumo das informações governamentais, evidenciando um padrão de **extração com dependência hierárquica** (em cascata). Nesse modelo, a ingestão de dados não ocorre de forma paralela e independente: primeiro, é necessário extrair os dados de uma **fonte-base** (entidade principal) e ingeri-los no banco de dados da plataforma. Somente após essa ingestão inicial, de posse dos identificadores obtidos, é possível buscar as **fontes dependentes** (entidades suplementares) vinculadas àquela fonte-base. A título de exemplo concreto, no domínio de contratações públicas do ComprasGov, o pipeline extrai primeiro o cadastro de **contratos** do Ipea, obtendo os respectivos identificadores (`id_contrato`). Somente com esses identificadores já persistidos, a plataforma dispara as consultas subsequentes para extrair as **faturas** e os **empenhos** atrelados a cada contrato, bem como os registros de **terceirizados** vinculados. Essa lógica em cascata garante a integridade referencial desde a origem e é replicada

em cada domínio governamental integrado à plataforma. Apenas após respeitar essas relações de dependência, o conjunto de informações coletadas é encaminhado, de forma íntegra, para a camada bruta do Data Warehouse.

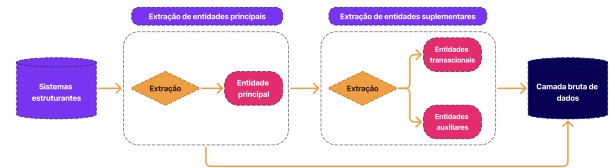


Figure 2: Sequenciamento técnico da coleta de dados.

3.3 Arquitetura de Dados do Gov Hub

Para viabilizar a ingestão, o cruzamento e a disponibilização do grande volume de informações governamentais, o Gov Hub foi construído sobre uma esteira de Extração, Carga e Transformação (ELT), incorporando práticas de engenharia de dados voltadas à organização de fluxos, à preparação de bases de dados e ao suporte a aplicações analíticas [13], conforme ilustrado na Figura 3. O processo inicia-se com a extração de dados diretamente das fontes dos sistemas estruturantes. Após essa coleta, realiza-se a carga (load) do dado em seu estado bruto, preservando o formato original da fonte, na camada raw do Data Warehouse da plataforma [21].

A partir da camada raw, inicia-se a etapa de transformação e estruturação do Data Warehouse, fundamentada no paradigma da **Arquitetura Medallion**. A transição da camada raw para a camada bronze engloba tratamentos iniciais, com foco na garantia de qualidade e na tipagem correta dos dados. Em seguida, a passagem da camada bronze para a silver constitui o núcleo da interoperabilidade da ferramenta: é neste momento que ocorre o cruzamento e a integração das bases provenientes de diferentes fontes. Por fim, os dados evoluem da camada silver para a gold, sendo refinados e qualificados estritamente para responder a perguntas gerenciais e aplicar as regras de negócio da gestão pública.

A orquestração diária desse processo é realizada pelo **Apache Airflow**, enquanto a lógica de modelagem e transformação de dados é executada pelo **dbt (data build tool)**. A sinergia entre essas tecnologias é garantida pelo uso do pacote **Astronomer Cosmos**. Essa integração permite que a linhagem de transformações estabelecida pelo dbt seja mapeada dinamicamente: cada modelo (model) do dbt é convertido em uma tarefa (task) individual em uma DAG do Airflow. Dessa forma, o Cosmos coordena automaticamente a fila e as dependências de execução, reforçando a confiabilidade operacional necessária para ambientes de dados em produção [8]. A Figura 4 apresenta um trecho da interface do usuário do Airflow, ilustrando as tarefas de uma DAG do projeto geradas pelo dbt.

Com o Data Warehouse modelado e atualizado, a arquitetura torna-se agnóstica quanto à aplicação de consumo. Para suprir a necessidade específica do Ipea de relatórios e planilhas gerenciais, integrou-se o **Apache Superset** como ferramenta principal de Business Intelligence (BI) para exploração visual e analítica. Ademais, uma premissa fundamental da ferramenta é a governança: todos os

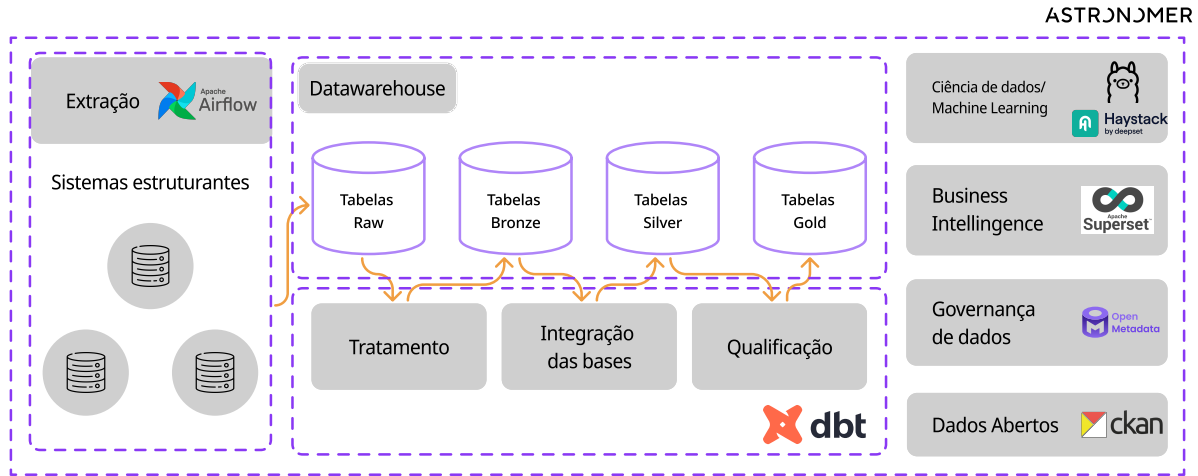


Figure 3: Arquitetura ELT do Gov Hub.

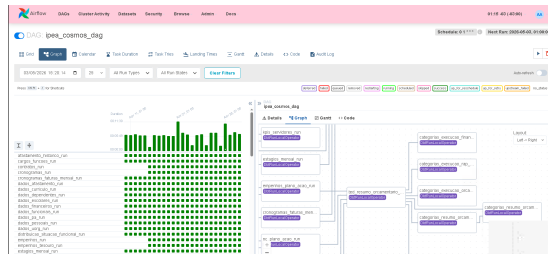


Figure 4: Interface do Apache Airflow com DAG gerada pelo dbt.

modelos são documentados com o **dbt docs**. Essa documentação estruturada é exportada para o **OpenMetadata**, que atua como o motor central de governança de dados da plataforma, centralizando o catálogo de dados, o dicionário de variáveis, a documentação histórica e a linhagem visual de ponta a ponta. A representação explícita de grafos de linhagem tem sido considerada um insumo relevante para pesquisas e práticas de governança de dados [6]. A Figura 5 apresenta um exemplo prático dessa rastreabilidade na interface do OpenMetadata, ilustrando a linhagem (data lineage) das informações referentes a contratos governamentais, desde a extração via API até a composição dos painéis e relatórios finais.

Como parte do pipeline, o Gov Hub realiza a vetorização dos metadados catalogados no OpenMetadata, gerando *embeddings* utilizados na busca semântica e na geração de consultas SQL em linguagem natural. Em vez de indexar os registros transacionais, a abordagem utiliza apenas os metadados do catálogo, reduzindo ruído na recuperação semântica. Os *embeddings* são armazenados

no banco vetorial **Qdrant** e integrados a uma arquitetura Retrieval-Augmented Generation (RAG), permitindo funcionalidades de busca semântica e Text-to-SQL na plataforma.

4 Cenário de Uso

Esta seção apresenta um cenário de uso concreto do Gov Hub no contexto do Ipea, demonstrando como a plataforma resolve um problema operacional real enfrentado pela Diretoria de Desenvolvimento Institucional (DIDES).

4.1 Problema: Consolidação Manual de Relatórios de Contratos

Durante a fase de diagnóstico do projeto, realizada por meio de reuniões setoriais e de workshops com as três coordenações da DIDES (CGPES, CGPGO e CGCAP), identificou-se um gargalo operacional recorrente: a elaboração de relatórios gerenciais sobre a execução de contratos administrativos. Para gerar um único relatório consolidado, os servidores da Coordenação de Compras e Contratos (COCCT/CGCAP) precisavam acessar manualmente o ComprasGov para extrair dados de contratos vigentes, consultar o SIAFI via Tesouro Gerencial para obter os valores empenhados, liquidados e pagos, e cruzar essas informações com as notas de crédito no TransfereGov, caso o contrato envolvesse recursos de TEDs. Todo esse processo era realizado em planilhas, sem padronização, e dependia do conhecimento tácito de poucos servidores.

O diagnóstico revelou que essa situação não era isolada. Dentre as dificuldades mapeadas no workshop com as equipes da DIDES, destacam-se: dados inconsistentes ou desatualizados entre sistemas diferentes, necessidade de dupla checagem manual, relatórios disponibilizados apenas por e-mail à diretoria e sobrecarga dos colaboradores responsáveis pela consolidação [11]. Em síntese, a ausência

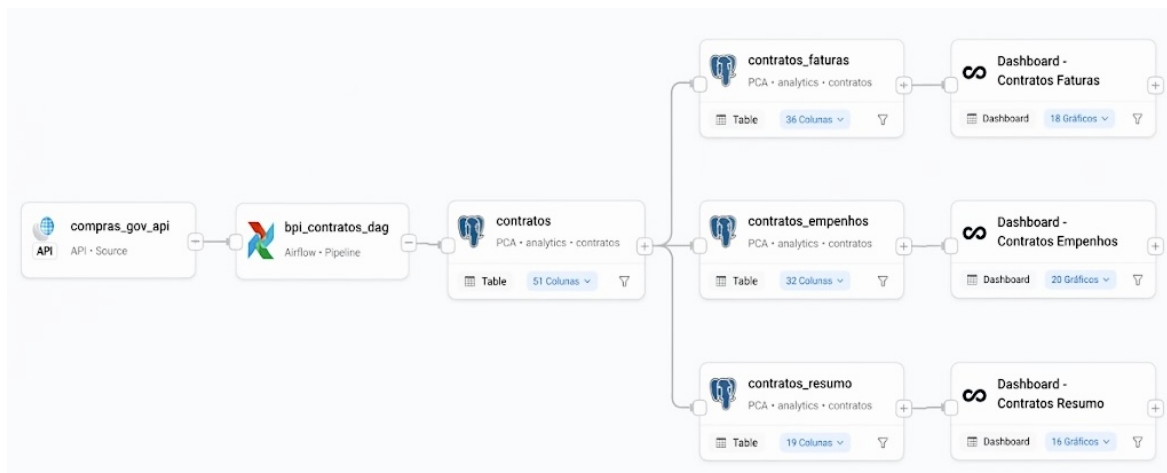


Figure 5: Linhagem de dados no OpenMetadata.

de integração entre os sistemas impunha um custo operacional elevado e comprometia a confiabilidade da informação produzida.

4.2 Ingestão e Tratamento Automatizado

Com o fluxo da política e as dependências entre entidades já mapeados (conforme descrito na Seção 3.2), o Gov Hub automatiza integralmente o ciclo de coleta. No domínio de contratos, a DAG orquestrada pelo Apache Airflow executa o seguinte fluxo: primeiro, conectores REST extraem o cadastro completo de contratos do Ipea a partir da API do ComprasGov, utilizando o código da unidade de origem como parâmetro. Com os identificadores de contrato já persistidos na camada *raw*, o pipeline dispara automaticamente as consultas dependentes para extrair faturas, empenhos e terceirizados vinculados a cada contrato. Paralelamente, outras DAGs executam a ingestão de dados do SIAFI, do SIOP e do TransfereGov, seguindo a mesma lógica de dependência hierárquica.

Na sequência, os modelos dbt realizam a padronização estrutural e tipagem (camada *bronze*), promovem o cruzamento entre as bases originalmente desconectadas (camada *silver*) e aplicam as regras de negócio da gestão pública para produzir indicadores qualificados (camada *gold*). Todo esse processamento, que antes consumia dias de trabalho manual, é executado diariamente de forma autônoma.

4.3 Governança, Linhagem e Auditabilidade

Um dos problemas identificados no diagnóstico foi a ausência de documentação das regras de negócio e a dependência de conhecimento tácito individual na produção de relatórios. O Gov Hub endereça essa fragilidade por meio de governança integrada: todos os modelos de transformação são documentados no **dbt docs** e exportados para o **OpenMetadata**, que centraliza o catálogo de dados, o dicionário de variáveis e a linhagem visual ponta a ponta. Essa ênfase em rastrear a proveniência dos dados em escala dialoga com propostas recentes de sistemas unificados de linhagem [14].

Na prática, qualquer servidor pode inspecionar a linhagem de um contrato específico, identificando desde a API de origem, os pipelines de ingestão e os modelos de transformação aplicados até os dashboards que consomem o dado final. Esse mecanismo substitui

a conferência manual, reduz a dependência de poucos especialistas e torna o processo auditável, ao mesmo tempo em que evita lacunas de observabilidade em fluxos de dados complexos [12].

4.4 Consumo Analítico e Resultados Observados

Com os dados consolidados na camada *gold*, os gestores da DIDES passaram a acessar painéis com visões integradas que antes não existiam. O portal institucional do Gov Hub (Figura 6) oferece uma visão executiva do orçamento do Ipea, consolidando o orçamento total, os recursos recebidos via TEDs e a dotação atualizada do órgão, além de apresentar a distribuição orçamentária por ação governamental. Já na camada analítica, dashboards construídos no Apache Superset (Figura 7) permitem uma exploração detalhada dos dados operacionais, como o quantitativo de faturas segmentado por situação de pagamento, tipo de contratação e evolução temporal, com filtros interativos por contrato, ano de emissão e status.

O painel de execução contratual, por exemplo, correlaciona automaticamente contratos vigentes do ComprasGov com empenhos e pagamentos registrados no SIAFI, permitindo identificar concentrações orçamentárias e faturas pendentes, tudo em uma única interface atualizada diariamente.

Em 20 de julho de 2026, a implantação mantinha 58 DAGs ativas e agendadas. No dbt, 12 fontes lógicas representavam sistemas e domínios governamentais de origem e agrupavam 93 tabelas de entrada. Essas tabelas alimentavam as transformações; a partir delas e de resultados intermediários, o dbt construía 127 modelos para padronizar e organizar os dados. Dos 127, 30 dependiam, direta ou indiretamente, de pelo menos duas fontes lógicas e na camada *gold*, 17 dos 23 modelos (73,9%) apresentavam essa linhagem de múltiplas origens, evidenciando a automação de grande quantidade de integrações que antes seriam realizadas manualmente. Nos 90 dias anteriores, a DAG de transformação do dbt concluiu 84 de 90 execuções com sucesso (93,3%), com duração mediana de 7min49s.

5 Conclusão

Este artigo apresentou o Gov Hub como uma solução para um problema estrutural da administração pública: a fragmentação dos



Figure 6: Portal institucional do Gov Hub com a visão orçamentária do Ipea.

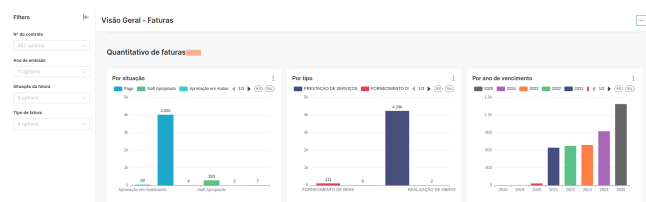


Figure 7: Dashboard analítico de faturas no Apache Superset.

dados governamentais em sistemas legados que, historicamente, não foram projetados para se comunicarem de forma integrada. Ao conectar bases oriundas de diferentes Sistemas Estruturantes, como SIOP, SIAFI, ComprasGov, TransfereGov e SIAPE, a plataforma demonstra que é possível superar barreiras à interoperabilidade sem impor a substituição dos sistemas existentes. Em vez disso, o Gov Hub propõe a criação de uma camada analítica de integração, governança e qualificação dos dados, capaz de preservar os sistemas de origem e, ao mesmo tempo, transformar informações dispersas em ativos estratégicos para a gestão pública.

A experiência desenvolvida no contexto do Ipea foi fundamental para validar a proposta da plataforma em um cenário real, no qual a rastreabilidade, a confiabilidade e a consolidação das informações são condições essenciais para o acompanhamento da execução orçamentária e financeira. Embora o cenário de uso esteja associado a uma instituição federal, os problemas enfrentados refletem desafios recorrentes em diferentes níveis do Estado, incluindo os governos federal, estadual e municipal. Nesse sentido, o Gov Hub contribui não apenas para modernizar processos internos, mas também para fortalecer a transparência ativa à sociedade, ampliar a auditabilidade dos dados públicos e abrir novas possibilidades de uso estratégico da informação, como a melhoria da gestão dos recursos governamentais, a identificação de áreas deficitárias e a priorização de investimentos com base em evidências [10].

Além disso, por ser uma solução de software livre, o Gov Hub foi concebido para permitir sua difusão, adaptação e evolução incremental por diferentes órgãos públicos. Sua arquitetura modular favorece a incorporação progressiva de novas bases, domínios e aplicações analíticas, criando condições para que o sistema cresça à medida que mais instituições o adotem e contribuam para seu aprimoramento. Outro aspecto é que a governança proporcionada pela plataforma é um elemento central nesse processo, pois as decisões públicas dependem diretamente da confiabilidade dos dados utilizados; dados incorretos, incompletos ou mal interpretados podem resultar em escolhas igualmente equivocadas.

A efetiva transformação digital da administração pública exige infraestrutura, governança, interoperabilidade semântica e mecanismos de qualificação contínua da informação. Nesse sentido, a plataforma apresenta-se como uma alternativa viável para apoiar órgãos públicos na construção de ambientes analíticos mais integrados, transparentes e reutilizáveis.

Disponibilidade de Artefatos

O Gov Hub é disponibilizado como software livre sob a MIT License, garantindo ampla liberdade de uso, modificação, redistribuição e incorporação em projetos derivados. Essa escolha de licenciamento reflete o compromisso do projeto com os princípios do software livre, assegurando que qualquer órgão público, universidade ou organização da sociedade civil possa adotar, adaptar e evoluir a plataforma sem restrições proprietárias.

Na prática, isso significa que qualquer instituição que deseje, por exemplo, criar um fórum próprio de governança de dados ou integrar o pipeline da plataforma em seus sistemas internos pode fazê-lo livremente, utilizando o código-fonte como ponto de partida e adaptando-o às suas necessidades específicas. A arquitetura modular do Gov Hub favorece esse tipo de incorporação progressiva, permitindo que novos domínios, fontes de dados e aplicações analíticas sejam adicionados de forma incremental.

O repositório oficial da organização está disponível publicamente no GitHub em <https://github.com/GovHub-br>, onde se encontram o código-fonte, a documentação técnica e as instruções de implantação. A manutenção do projeto em uma plataforma aberta visa fomentar a colaboração entre desenvolvedores, pesquisadores e gestores públicos, criando condições para que uma comunidade ativa contribua para o aprimoramento contínuo da ferramenta.

Agradecimentos

Este estudo foi financiado, em parte, pelo Instituto de Pesquisa Econômica Aplicada, no âmbito do projeto IPEA/UnB/FGA – Ecossistema de Software Livre para Gestão de Dados: Solução de Software para Extração Automatizada de Dados e Visualização da Informação; pelo Ministério da Gestão e da Inovação em Serviços Públicos, no âmbito do projeto Ecossistema de Software Livre para Gestão Orientada a Dados e Integração de Bases do Observatório de Contratações Públicas; e pelo Ministério das Cidades, no âmbito do projeto Desburocratização, Industrialização e Inovação no Setor da Construção Habitacional.

References

- [1] [n. d.]. Open Contracting Data Standard — Open Contracting Data Standard 1.1.5 documentation. <https://standard.open-contracting.org/latest/en/> Accessed: 20

- Jul. 2026.
- [2] Antonio Pedro Santos Alves, Marcos Kalinowski, Daniel Mendez, Hugo Villamizar, Kelly Azevedo, Tatiana Escovedo, and Helio Lopes. 2024. Industrial Practices of Requirements Engineering for ML-Enabled Systems in Brazil. In *Anais do XXXVIII Simpósio Brasileiro de Engenharia de Software*. Sociedade Brasileira de Computação, Porto Alegre, RS, Brasil, 224–233. doi:10.5753/sbes.2024.3371
 - [3] Judie Attard, Fabrizio Orlandi, Simon Scerri, and Sören Auer. 2015. A systematic review of open government data initiatives. *Government Information Quarterly* 32, 4 (2015), 399–418.
 - [4] Bruno Miguel Vital Bernardo, Henrique S. Mamede, João Barroso, and Vitor Santos. 2024. Data governance & quality management: Innovation and breakthroughs across different fields. *Journal of Innovation & Knowledge* 9, 4 (2024), 100598. doi:10.1016/j.jik.2024.100598
 - [5] Gabriel Busquim, Allysson Alex Araújo, Maria Julia Lima, and Marcos Kalinowski. 2024. Towards Effective Collaboration between Software Engineers and Data Scientists developing Machine Learning-Enabled Systems. In *Anais do XXXVIII Simpósio Brasileiro de Engenharia de Software*. Sociedade Brasileira de Computação, Porto Alegre, RS, Brasil, 24–34. doi:10.5753/sbes.2024.3027
 - [6] Yunpeng Chen, Ying Zhao, Xuanjing Li, Jun Zhang, Jichao Long, and Fangfang Zhou. 2024. An open dataset of data lineage graphs for data governance research. *Visual Informatics* 8, 1 (2024), 1–5. doi:10.1016/j.visinf.2024.01.001
 - [7] CKAN. 2026. CKAN Documentation. <https://docs.ckan.org/en/2.11/>. Accessed: 20 Jul. 2026.
 - [8] Luiz Alexandre Costa, Awdren L. Fontão, Eleni Constantinou, Rodrigo Pereira dos Santos, and Alexander Serebrenik. 2025. Building the Landing Zone for Site Reliability Engineering: A Multivocal Literature Review. In *Anais do XXXIX Simpósio Brasileiro de Engenharia de Software*. Sociedade Brasileira de Computação, Porto Alegre, RS, Brasil, 93–103. doi:10.5753/sbes.2025.9772
 - [9] José Sérgio da Silva Cristóvam, Lucas Bossoni Saikali, and Thanderson Pereira de Sousa. 2020. Governo Digital na Implementação de Serviços Públicos para a Concretização de Direitos Sociais no Brasil. *Seqüência (Florianópolis)* 43, 89 (2020), 209–242. doi:10.5007/2177-7055.2020v43n89p209
 - [10] Daniel Pitanguiera de Avelino, João Cláudio Pompeu, and Igor Ferraz da Fonseca. 2021. *Democracia Digital: Mapeamento de Experiências em Dados Abertos, Governo Digital e Ouvidorias Públicas*. Technical Report 2624. Instituto de Pesquisa Econômica Aplicada (Ipea). doi:10.38116/td2624
 - [11] Denise Do Carmo Direito and Valéria Luzia Mota Barros. 2025. *Texto para Discussão 3099*. Instituto de Pesquisa Econômica Aplicada (Ipea). doi:10.38116/td3099-port
 - [12] Francisco A. A. Gomes, Paulo A. L. Rego, and Fernando A. M. Trinta. 2025. Rumo a um Catálogo de Anti-padrões de Observabilidade: Uma Revisão da Literatura. In *Anais do XXXIX Simpósio Brasileiro de Engenharia de Software*. Sociedade Brasileira de Computação, Porto Alegre, RS, Brasil, 104–114. doi:10.5753/sbes.2025.9779
 - [13] Petra Heck. 2024. What About the Data? A Mapping Study on Data Engineering for AI Systems. In *Proceedings of the IEEE/ACM 3rd International Conference on AI Engineering – Software Engineering for AI (CAIN '24)*. Association for Computing Machinery, New York, NY, USA, 43–52. doi:10.1145/3644815.3644954
 - [14] Gabriela Jacques-Silva, Evangelia Kalyvianaki, Katriel Cohn-Gordon, et al. 2025. Unified Lineage System: Tracking Data Provenance at Scale. In *Companion of the 2025 International Conference on Management of Data (SIGMOD/PODS '25 Companion)*. Association for Computing Machinery. doi:10.1145/3722212.3724458
 - [15] Marijn Janssen, Yannis Charalabidis, and Anneke Zuiderwijk. 2012. Benefits, adoption barriers and myths of open data and open government. *Information systems management* 29, 4 (2012), 258–268.
 - [16] Keegan McBride, Sujani Kamalanathan, Sandhra-Mirella Valdma, Taavi Toomere, and Margus Freudenthal. 2022. Digital Government Interoperability and Data Exchange Platforms: Insights from a Twenty Country Comparative Study. In *Proceedings of the 15th International Conference on Theory and Practice of Electronic Governance (ICEGOV '22)*. Association for Computing Machinery, New York, NY, USA, 90–97. doi:10.1145/3560107.3560123
 - [17] Isela Mendoza, Renan Corrêa, and Flavia Bernardini. 2023. Como a governança de dados pode auxiliar na mitigação de barreiras de uso de portais de dados governamentais abertos? uma análise da literatura. In *Workshop de Computação Aplicada em Governo Eletrônico (WCGE)*. SBC, 212–223.
 - [18] Sergi Nadal, Oscar Romero, Alberto Abelló, Panos Vassiliadis, and Stijn Vansumeren. 2019. An integration-oriented ontology to govern evolution in big data ecosystems. *Information systems* 79 (2019), 3–19.
 - [19] OpenFisca. 2026. OpenFisca Documentation. <https://openfisca.readthedocs.io/en/latest/>. Accessed: 20 Jul. 2026.
 - [20] OpenSpending. 2026. OpenSpending Documentation. <https://docs.openspending.org/en/latest/developers/>. Accessed: 20 Jul. 2026.
 - [21] Philip Russom et al. 2017. Data lakes: Purposes, practices, patterns, and platforms. *TDWI white paper* (2017).
 - [22] Juliane Schmeling, Sami al Dakruni, and Ines Mergel. 2025. Data collaboration in digital government research: A literature review and research agenda. *Government Information Quarterly* 42, 3 (2025), 102063. doi:10.1016/j.giq.2025.102063
 - [23] Ana Cristina Aguilar Viana. 2021. Transformação digital na administração pública: do governo eletrônico ao governo digital. *Revista Eurolatinoamericana de Derecho Administrativo* 8, 1 (2021), 115–136. doi:10.14409/redoeda.v8i1.10330