

Gov Hub

Um guia prático para
integração e qualificação
de dados públicos

Governo Federal
Ministério do Planejamento e Orçamento
Ministra Simone Nassar Tebet



Fundação pública vinculada ao Ministério do Planejamento e Orçamento, o Ipea fornece suporte técnico e institucional às ações governamentais – possibilitando a formulação de inúmeras políticas públicas e programas de desenvolvimento brasileiros – e disponibiliza, para a sociedade, pesquisas e estudos realizados por seus técnicos.

Presidenta

Luciana Mendes Santos Servo

Diretor de Desenvolvimento Institucional

Fernando Gaiger Silveira

**Diretora de Estudos e Políticas do Estado,
das Instituições e da Democracia**

Luseni Maria Cordeiro de Aquino

**Diretor de Estudos e Políticas
Macroeconômicas**

Cláudio Roberto Amitrano

**Diretor de Estudos e Políticas Regionais,
Urbanas e Ambientais**

Aristides Monteiro Neto

**Diretor de Estudos e Políticas Setoriais,
de Inovação, Regulação e Infraestrutura**

Pedro Carvalho de Miranda

Diretora de Estudos e Políticas Sociais

Letícia Bartholo de Oliveira e Silva

Diretora de Estudos Internacionais

Keiti da Rocha Gomes

Chefe de Gabinete

Alexandre dos Santos Cunha

**Coordenadora-Geral de Imprensa e
Comunicação Social**

Gisele Amaral de Souza

GERIO

Andrea Bossle de Abreu

Direção e fiscalização do Termo de Execução Descentralizada

Fernando Gaiger Silveira

Coordenação operacional

Gustavo Camilo Baptista

Diretoria de Desenvolvimento Institucional**CGCAP - Coordenação-Geral de Serv. Coop. e Apoio à Pesquisa**

Marinésia Lemos Souto

Arno Luiz Guedes Corrêa Junior

Karen da Silva Rodrigues do Prado

Cristiane de Paula Soto Santo

Renata Gonçalves de Oliveira

Manoel de Ribamar Cardoso Barroso

CGPES - Coordenação-Geral de Gestão de Pessoas

Anna Carolina Mendonça Lemos Ribeiro

Everton Gomes Baier

Leone Aquino Ramos

Antonio Ximenes Feijão

Viviane de Faria Gonçalves Diniz

Elaine Calazans de Souza

Katia de Oliveira Fernandes

Coordenação Geral de Planejamento e Orçamento

Wagner Cardoso dos Santos

Darci De Borba Santos Júnior

Luiz Guilherme Ferreira Deud

Manoel Messias Ramos da Silva

Ivanildo Alves da Silva

Coordenação Geral de Desenvolvimento e Tecnologia de Informações

Lucas Ferreira Mation

Fernando José Brustolin

Leonardo de Araujo Silva

Luis Sergio Teixeira Mesquita

Ouvidoria do IPEA

Gisella Garritano de Deus

Saulo Armando de Oliveira Victoriano

Gabriella Tavares Ramos

Ouvidoria: <http://www.ipea.gov.br/ouvidoria>

URL: <http://www.ipea.gov.br>



Universidade de Brasília

Reitora

Rozana Reigota Naves

Decana de Pesquisa e Inovação

Renata Aquino da Silva

Projeto de Pesquisa

Ecosistema de Software Livre para Gestão de Dados
- Solução de software para extração automatizada
de dados e visualização da informação

Autores

Isaque Alves de Lima
Laila Bellix
João Henrique Egewarth
Carla Rocha Silva Aguiar
Davi de Aguiar Vieira
Mateus de Castro Santos
Arthur Alves Melo
Alex Reis

Colaboração

Pedro Gonçalves Rossi Rodrigues
Victor Suzuki

Projeto gráfico e diagramação

Isis Reis

Dashboards

Pedro Rodrigues

Coordenação Geral

Alex Reis
Carla Rocha Silva Aguiar

URL: www.unb.br

URL: <https://www.gov-hub.io/>

Este livro sintetiza os resultados de desenvolvimento científico e tecnológico do Termo de Execução Descentralizada celebrado entre a Universidade de Brasília e o Instituto de Pesquisa Econômica Aplicada.

Gov Hub : um guia prático para integração e qualificação de dados públicos / Isaque Alves de Lima, Laila Bellix, João Henrique Egewarth, Carla Rocha Silva Aguiar, Davi de Aguiar Vieira, Mateus de Castro Santos, Arthur Alves Melo, Alex Reis. – Brasília : UnB : Ipea, 2025.

91 p. : il.

Inclui referências bibliográficas.

ISBN: 978-65-5635-087-5

1. GovHub. 2. Gestão Pública. 3. Dados Públicos. 4. Plataforma Digital. 5. Software Livre. 6. Governança Digital. 7. Brasil. I. Lima, Isaque Alves de. II. Bellix, Laila. III. Egewarth, João Henrique. IV. Aguiar, Carla Rocha Silva. V. Vieira, Davi de Aguiar. VI. Santos, Mateus de Castro. VII. Melo, Arthur Alves. VIII. Reis, Alex. IX. Título. X. Universidade de Brasília. XI. Instituto de Pesquisa Econômica Aplicada.

CDD 354.350285

Ficha catalográfica elaborada por Andréa de Mello Sampaio CRB-1/1650

Como citar:

LIMA, Isaque Alves de et al. **Gov Hub**: um guia prático para integração e qualificação de dados públicos. Brasília: UnB: Ipea, 2025. 91 p.

Gov Hub:

Um guia prático para
integração e qualificação
de dados públicos



SUMÁRIO

08 APRESENTAÇÃO

12 INTRODUÇÃO

18 CAPÍTULO 1 | MAPEAMENTO DOS DESAFIOS NO USO DE DADOS

Entrevistas com coordenadores e gestores 20

Entrevistas por setor 21

Entrevistas focadas 22

Workshop sobre Gestão de Dados 23

Fluxo orçamentário e financeiro 25

Fluxo de gestão de recursos humanos 26

Resultados do mapeamento 29

Como replicar essa etapa? 31

32 CAPÍTULO 2 | ANÁLISE EXPLORATÓRIA DOS DADOS E VIABILIDADE DE INTEGRAÇÃO DE SISTEMAS

Análise de Integração 41

Integração de dados: Compras.gov.br e SIAFI 44

Integração de dados: TransfereGov e SIAFI 46

Resultados da análise exploratória e viabilidade de integração 48

Como replicar essa etapa? 50

52 **CAPÍTULO 3: MODELAGEM, LEVANTAMENTO DE REQUISITOS TÉCNICOS E MAPEAMENTO DE INICIATIVAS**

Resultados da etapa de modelagem, requisitos e iniciativas **56**

Como replicar essa etapa? **59**

62 **CAPÍTULO 4 - GOVHUB: A SOLUÇÃO PARA UMA GESTÃO PÚBLICA ORIENTADA POR DADOS**

Arquitetura Data Mesh e Medallion **66**

Análises Viabilizadas pelo GovHub **71**

Aprendizados **80**

Como replicar essa etapa? **82**

Como saber mais e como contribuir? **85**

86 **CONCLUSÃO**

88 **REFERÊNCIAS BIBLIOGRÁFICAS**

APRESENTAÇÃO



Os países democráticos têm enfrentado uma crescente exigência de transparência como um aspecto crucial para a manutenção da governabilidade. A sociedade, cada vez mais, clama pelo acesso a diferentes peças que constituem o quebra-cabeça do funcionamento estatal, tais como os gastos públicos, o desenho de políticas públicas, os quantitativos de servidores públicos, dentre outros.

As respostas variaram de acordo com cada país. No contexto brasileiro, um importante marco legal é a Lei nº 12.527, de 18 de novembro de 2011, que, dentre outras iniciativas, introduziu no país o conceito de transparência. O significado mais usual desta palavra é associado a objetos translúcidos, isto é, que possuem a propriedade que lhes permitem não impedir, diminuir ou alterar a passagem de luz por eles. Uma segunda acepção, derivada desta, é a da qualidade da franqueza, de transmitir a verdade sem nada a esconder.

Essa lei é um dos diversos degraus de um lento processo de publicização das ações do Estado que, no nosso contexto, atravessa dois séculos. Esse processo teve início no Brasil ainda na época do Reino Unido com Portugal e Algarves, quando as primeiras decisões públicas passaram a ser publicadas em jornais comuns — o que, mais tarde, levou à criação do Diário Oficial, já no reinado de Dom Pedro II.

A referida lei elenca duas formas de transparência: a passiva e a ativa. Enquanto na transparência passiva o cidadão solicita informações de seu interesse por meio de mecanismos institucionais de recepção padronizados (ouvidorias, Sistemas de Informação ao Cidadão, dentre outros), na transparência ativa cada órgão do governo tem o dever de oferecer em locais de fácil acesso dados sobre suas competências e estrutura organizacional, endereços e telefones das respectivas unidades e horários de atendimento ao público, registros de quaisquer repasses ou transferências de recursos financeiros, registros das despesas e informações concernentes a procedimentos licitatórios, inclusive os respectivos editais e resultados, bem como a todos os contratos celebrados, dentre outros.

Tanto os instrumentos e mecanismos de transparência passiva quanto ativa se difundiram no Brasil por força desta legislação. A referida lei levou à ampla divulgação, nos sites oficiais dos órgãos da Administração Pública Federal, de listas e bases de dados com registros de despesas, parcerias, licitações e remunerações de servidores.

Todavia, em que pese listas e bases de dados brutas serem adequadas para uma parcela das demandas da sociedade, novos desafios surgem diante da complexidade dos temas e da dimensão da Administração Pública, assim como da diversidade dos sistemas de controle que geram esses dados. Soma-se a isso o emaranhado terminológico típico da máquina estatal brasileira, no qual expressões tais como empenho, liquidação, nota de crédito, pré-empenho, termo de execução descentralizada, dentre outros, não são de domínio da maioria dos cidadãos.

É neste contexto que se desenvolveu a parceria entre o Instituto de Pesquisa Econômica Aplicada (Ipea) e a Universidade de Brasília, por meio de seu Laboratório Avançado de Produção, Pesquisa e Inovação em Software (Lab Livre) da Faculdade do Gama (FGA). A parceria com este Laboratório, que já possui experiência pregressa no desenvolvimento de projetos importantes do governo federal (como o Portal do Software Público, Participa.br, Aplicativo da Conferência da Juventude, Empurrando Juntas e Dialoga Brasil) visou o desenvolvimento científico e tecnológico de solução em software para subsidiar os processos de monitoramento e de tomada de decisão por meio da extração automatizada de dados e visualização de informações advindas de sistemas estruturantes empregados pela Diretoria de Desenvolvimento Institucional (Dides) do Instituto de Pesquisa Econômica Aplicada. Ela foi formalizada por meio do Termo de Execução Descentralizada 1/2024, que previu, em um primeiro momento, o repasse de R\$ 902.400,00 para sua consecução.

Desenvolvida ao longo de um ano, a solução, denominada GovHub, é um software hospedado no Ipea que viabiliza a coleta e o cruzamento de dados provenientes dos sistemas estruturantes do governo federal. Atualmente, esses sistemas permitem que o Ministério do Planejamento e Orçamento, o Ministério da Fazenda e o Ministério da Gestão e Inovação em Serviços Públicos realizem o controle de: a) orçamento federal, por meio do Sistema Integrado de Administração Financeira do Governo Federal (SIAFI) e do Tesouro Gerencial (TG); b) transferências e parcerias, com o Portal de Transferências e Parcerias (Transferegov); c) contratações, por meio do Portal de Compras do Governo Federal (Compras.gov.br); d) gestão de servidores públicos federais, com o Sistema Integrado de Administração de Pessoal (SIAPE); entre outros.

O GovHub permite o tratamento de bases de dados brutas, eliminando seus problemas de formatação, de erros ou de lacunas, dando condições para que estes sejam elegíveis para cruzamento. Novas bases de dados são produzidas, em um processo de qualificação seriado no qual são obtidas “bases bronze” e “bases prata”, e que resultam no término desta sequência em “bases ouro”, adequadas para cruzamentos e desenvolvimento de painéis gerenciais.

A disponibilização do GovHub na página do Ipea, por conseguinte, é uma inovação que vai além do previsto na Lei de Transparência. Em vez de oferecer ao cidadão uma imensa quantidade de pequenas peças soltas, a solução pode, dependendo da demanda, entregar conjunto de informações já estruturadas, como quebra-cabeças parcialmente ou totalmente montados. Entre as diversas possibilidades de uso, o cidadão poderá acompanhar online tanto a quantidade de recursos que o Ipea gasta com contratos de pessoal, contas de luz e pagamento de bolsas de pesquisa, quanto cruzar essas informações ao longo do tempo — comparando, por

exemplo, os gastos de diferentes anos desde a implantação do sistema — e entre modalidades, como a comparação entre despesas com servidores e custos indiretos, tais como contas de luz e de água das sedes do Ipea.

Ao promover transparência ativa, o governo deve ir além da letra fria da lei, buscando também oferecer eficiência e uma experiência agradável para o cidadão, evitando que ele se sinta em um labirinto de informações. Cabe ao Estado conferir maneiras de qualificar, distinguir e disponibilizar seus dados, de forma que a sociedade civil consiga identificá-los e utilizá-los com facilidade.

Assim, esta publicação apresenta a solução desenvolvida por meio do esforço conjunto das equipes do Ipea e da UnB envolvidas no projeto. Trata-se de uma experiência pioneira, que pode auxiliar na construção mais célere não apenas de um, mas de alguns degraus nesta longa escada da publicização das ações do Estado.

Manifestamos nossa gratidão aos órgãos responsáveis pelos sistemas estruturantes que subsidiam o GovHub, pelo apoio na disponibilização de acessos e informações. Acreditamos que novos passos poderão ser dados em conjunto no futuro.

Boa leitura!

FERNANDO GAIGER SILVEIRA

Diretor de Desenvolvimento Institucional do Ipea

INTRODUÇÃO



Dados, informações e indicadores são insumos estratégicos que qualificam a atuação dos governos e a participação da sociedade. Os usos desses recursos são fundamentais para a formulação, implementação e avaliação de políticas públicas mais efetivas e com impacto direto na vida das pessoas.

No âmbito da gestão pública, há uma série de sistemas de informação que organizam dados públicos. O governo federal brasileiro possui **sistemas estruturantes** que reúnem dados como informações sobre orçamento, estrutura administrativa e pessoal, sendo usados por diferentes órgãos e setores da Administração Pública.



O que são e para que servem os Sistemas Estruturantes (SEs)?

Os sistemas estruturantes são plataformas de software mantidas pelo Governo Federal para centralizar a execução das atividades de gestão dos órgãos públicos. Esses sistemas são essenciais, pois todos os órgãos os utilizam para executar tarefas fundamentais do dia a dia.

Em relação à sua governança, existe um órgão central responsável por coordenar e supervisionar o desenvolvimento das funcionalidades, a manutenção e a evolução desses sistemas, garantindo assim a padronização e a eficiência no uso.

Os outros órgãos, chamados de setoriais, usam os sistemas estruturantes seguindo as orientações, normas e processos estabelecidos pelo órgão central. Dessa forma, mantém-se a coerência e a qualidade dos serviços prestados, além de assegurar que todos trabalhem de forma integrada e alinhada às diretrizes do Governo Federal.

O SIAFI (Sistema de Administração Financeira Federal) é um exemplo de sistema estruturante usado pelo Governo Federal para registrar, acompanhar e controlar a execução orçamentária, financeira e patrimonial dos órgãos públicos. Sua coordenação é do Tesouro Nacional, vinculado ao Ministério da Fazenda.

Outros sistemas estruturantes incluem o SIOP (Sistema Integrado de Planejamento e Orçamento), o SIORG (Sistema de Organização e Inovação Institucional do Governo Federal) e o SIAPE (Sistema Integrado de Administração de Recursos Humanos). A lista com informações sobre esses sistemas está disponível no portal do Ministério da Gestão e da Inovação em Serviços Públicos <https://www.gov.br/governodigital/pt-br/sistemas-internos-de-governo>.

No entanto, há inúmeros desafios na gestão desses sistemas que afetam o dia a dia dos servidores públicos e, consequentemente, a formulação das políticas públicas. Problemas como redundância de informações, inconsistências nos dados entre diferentes sistemas e barreiras à interoperabilidade comprometem a tomada de decisões baseadas em evidências e o processo de abertura de dados.

Além dos impactos nas atividades dos servidores, gestores e no acompanhamento de políticas públicas, a ausência de integração de dados dificulta a transparência e o controle social por parte da população e, por consequência, afetam diretamente a adoção da gestão orientada a dados no setor público.

Embora os governos já possuam grande parte dos dados necessários para melhorar os serviços públicos e otimizar recursos, o pleno aproveitamento desse potencial é impedido pela falta de interoperabilidade entre os sistemas estruturantes, que em sua maioria são legados, de baixa qualidade e com pouca integração de dados.

Com o objetivo de enfrentar esses desafios, o Laboratório de Competência em Software Livre (Lab Livre) da Universidade de Brasília (UnB), em parceria com a Diretoria de Desenvolvimento Institucional (DIDES) do Instituto de Pesquisas Econômicas Aplicadas (Ipea), desenvolveu um projeto voltado à qualificação e integração dos dados provenientes



Gestão Orientada a Dados

A gestão orientada a dados é uma abordagem estratégica que reconhece os dados como ativos indispensáveis, com capacidade de transformar informações baseadas em evidências em decisões e ações administrativas mais eficientes, transparentes e responsáveis (OCDE, 2019). Essa prática é essencial para melhorar resultados organizacionais e fortalecer a confiança pública. Entre os principais benefícios da gestão orientada a dados, destacam-se:

- **Decisões baseadas em evidências:** As escolhas são fundamentadas em análises quantitativas e qualitativas, reduzindo a dependência de suposições ou intuições.
- **Integração e interoperabilidade de dados:** Reúne informações de diferentes sistemas e áreas, promovendo padrões e arquiteturas que proporcionem uma visão ampla e coerente para subsidiar a gestão.
- **Mensuração de resultados:** Enfatiza o uso de métricas e indicadores de desempenho para monitorar o progresso, avaliar resultados e ajustar estratégias de forma precisa.
- **Cultura organizacional orientada a dados:** Estimula a alfabetização em dados, incentivando práticas éticas, uso seguro e compartilhamento transparente de informações em todos os níveis da organização.

- **Capacidade de adaptação e inovação:**

Fortalece a habilidade de organizações para responder a mudanças, explorar tecnologias emergentes e inovar em ambientes complexos e dinâmicos.

- **Construção de confiança e responsabilidade pública:**

O uso consistente e responsável de dados aumenta a credibilidade das decisões e fortalece a transparência e a accountability perante a sociedade.

Para isso, é fundamental garantir a interoperabilidade e governança dos sistemas, permitindo previsão de tendências e ajustes baseados em dados.

Como se aplica no setor público?

A gestão orientada por dados no setor público consiste no uso estratégico de dados e análises para embasar a tomada de decisões. Essa abordagem permite que cidades e órgãos governamentais tomem decisões fundamentadas nas demandas reais da sociedade, implicando em melhorias da qualidade dos serviços prestados aos cidadãos e em otimizações na alocação de recursos. Entre os principais benefícios, destaca-se o aprimoramento do processo decisório, o aumento da transparência e da responsabilização, a alocação mais eficiente de recursos, o fortalecimento do engajamento com os cidadãos e o ganho em eficiência e produtividade.

dos sistemas estruturantes do governo federal, com foco no fortalecimento da gestão pública e da transparência. A iniciativa reuniu o conhecimento científico e potencial de inovação da universidade com as necessidades reais da administração pública, resultando na criação do GovHub — uma plataforma de software livre que aprimora a qualidade das bases de dados e possibilita o cruzamento de informações para gerar dados mais acessíveis.

Por meio de metodologias inovadoras e tecnologias avançadas, o projeto estruturou uma solução robusta que transforma dados fragmentados dos sistemas estruturantes em informações estratégicas, organizadas progressivamente para assegurar qualidade, transparência, governança e reutilização. Essa estrutura qualifica a atuação dos gestores públicos, cria uma base sólida para que dados governamentais sejam convertidos em ativos estratégicos e impulsiona a efetividade administrativa, fortalecendo a accountability perante a sociedade.

Este e-book tem como propósito sistematizar as principais atividades realizadas e as experiências obtidas ao longo do desenvolvimento do **GovHub** junto ao Ipea. O objetivo é apresentar a trilha metodológica adotada, de forma a possibilitar que outros órgãos desenvolvam soluções de qualificação e integração de dados. Para isso, as próximas páginas detalham todas as etapas de desenvolvimento do projeto. Para fins de apresentação e compreensão, as etapas

foram divididas em capítulos, muito embora, na prática, elas tenham ocorrido de forma simultânea.

No capítulo 1, é possível conhecer a etapa inicial de mapeamento dos desafios vivenciados pelo órgão, no contexto da gestão orientada por dados. No capítulo 2, estão apresentados os processos de análise exploratória e a avaliação da viabilidade de integração de sistemas. Já o capítulo 3 aborda a modelagem e levantamento de requisitos técnicos para o desenvolvimento da solução e o mapeamento de iniciativas similares, incluindo um benchmarking, etapa transversal a todo o projeto. Esses 3 primeiros capítulos compõem o diagnóstico realizado pelo projeto, considerando diferentes camadas: gestão, dados e arquitetura. O capítulo 4 detalha a solução desenvolvida, o **GovHub**, com apresentação da metodologia, engenharia e governança de dados. Os aprendizados desse projeto e as informações sobre como contribuir e replicar a plataforma estão disponibilizados neste capítulo. Por fim, este capítulo é seguido por uma conclusão que sintetiza todo o processo.

O material busca oferecer uma visão ampla dos desafios e soluções para a gestão de dados nos sistemas estruturantes do governo federal, fornecendo métodos e caminhos práticos e estratégicos para gestores, acadêmicos e demais interessados na temática.



CAPÍTULO 1

MAPEAMENTO DOS DESAFIOS NO USO DE DADOS



0

100

O processo de mapeamento dos desafios no uso dos dados é o primeiro passo do projeto e busca identificar demandas e problemas que o órgão enfrenta no cotidiano da gestão de dados e informações. Esse processo possibilita ter uma visão compartilhada sobre os gargalos institucionais em relação à gestão de dados com base na vivência dos (as) gestores (as) públicos (as).

Este capítulo detalha as ações realizadas para identificar, organizar e compreender os dados dos sistemas estruturantes e o uso pelos usuários, estabelecendo a base conceitual e técnica para todas as análises e soluções posteriormente desenvolvidas.

Para alcançar tal objetivo, foram combinadas diferentes metodologias qualitativas e quantitativas, como entrevistas com gestores e técnicos, levantamento de sistemas, oficinas colaborativas e o mapeamento dos fluxos de dados e informações. Dessa forma, foi possível caracterizar a utilização dos sistemas estruturantes e conhecer as bases de dados existentes, bem como aproximar-se dos problemas reais enfrentados pelo órgão, compreender suas rotinas e desafios e, a partir disso, propor soluções mais alinhadas às necessidades institucionais.

O resultado desta fase materializou-se em três produtos fundamentais:

- **Diagnóstico das necessidades e dificuldades** na gestão de dados e informações dos sistemas estruturantes
- **Aprofundamento das informações relativas aos sistemas estruturantes, incluindo dados, dicionários e suas relações**
- **Detalhamento dos principais fluxos de dados relacionados a planejamento orçamentário, execução financeira, gestão organizacional e de recursos humanos.**

A etapa de mapeamento teve por objetivo conhecer o funcionamento da Diretoria de Desenvolvimento Institucional do Ipea, com enfoque nas rotinas das equipes, os fluxos de trabalho para a realização de ações para funcionamento do órgão e o nível de familiaridade com o uso de dados na gestão. Isso permitiu caracterizar os sistemas de software utilizados, os dados considerados relevantes para acompanhamento de ações, elaboração de relatórios e tomada de decisões, e os principais obstáculos e dificuldades para gerar informações.

Além disso, foi possível mapear como os dados circulam internamente no órgão: quais são os fluxos e processos envolvidos, de que forma são geridos e com o apoio de quais sistemas. Outro aspecto central foi a avaliação do grau de maturidade do Ipea em relação à gestão orientada por dados. Buscou-se identificar se já existem iniciativas estruturadas nessa área, bem como compreender como as equipes percebem e praticam a gestão de dados em seu cotidiano

Assim, o percurso metodológico adotado envolveu uma série de processos de escuta e investigação. Primeiro foram realizadas entrevistas amplas com coordenadores e gestores para captar a visão institucional e os principais desafios, seguido de entrevistas com servidores de diferentes setores, para detalhar fluxos, sistemas e práticas de gestão de dados. Por fim, foram feitas entrevistas focadas com o intuito de esclarecer aspectos técnicos e validar interpretações. Esse movimento de afunilamento permitiu aprofundar a compreensão do problema e subsidiar a elaboração de um “Workshop sobre Gestão Orientada a Dados”, em que, de forma conjunta, foram discutidos os sistemas e dados, a relação entre informações, e que permitiu caracterizar as dificuldades existentes na extração de dados e posterior manipulação.

1 Entrevistas com coordenadores e gestores

O primeiro ciclo de entrevistas foi conduzido com diretores e coordenadores do Ipea, buscando captar uma visão geral sobre os desafios enfrentados pela instituição na gestão de dados, bem como as expectativas quanto à integração e qualificação desses dados e informações.

Os gargalos identificados concentram-se em dois eixos principais: i) a confiabilidade e qualidade dos dados dos sistemas estruturantes; e ii) a integração e o uso estratégico das informações para apoiar a tomada de decisão e promover a transparência.

No primeiro eixo, referente à confiabilidade, os gestores relataram dificuldade em garantir a qualidade dos dados extraídos dos sistemas. Foram identificados problemas como informações desatualizadas, preenchimento incorreto e baixa clareza dos dados — fatores que comprometem a confiança nas informações utilizadas.

No segundo eixo, referente à integração e uso dos dados, foram identificadas dificuldades relacionadas à integração de dados de diferentes sistemas para a tomada de decisão. Isso significa dizer que os tomadores de decisão do órgão necessitam de informações que, muitas vezes, estão dispersas em bases distintas e que devem ser cruzadas para gerar valor estratégico na gestão das políticas públicas.

Nesse sentido, perguntas estratégicas, detalhadas abaixo, fundamentais para o planejamento e monitoramento das atividades e entregas do órgão eram difíceis de se responder com confiabilidade, agilidade e segurança. Na ausência de ou não confiança nas respostas, o trabalho dos tomadores de decisão e o controle por parte da sociedade são diretamente impactados.

Perguntas estratégicas:

Qual o orçamento total do Ipea? Desse total, quanto foi gasto até o momento? Com o que foi gasto? Quanto ainda tem disponível? Quantos contratos vigentes o Ipea tem? Quais são os principais contratos? Qual a vigência e valores desses contratos? Quantos Termos de Execução Descentralizada (TEDs) o Ipea recebeu e fez? Qual a vigência e valores? Quantos funcionários ativos o Ipea tem? Como se distribuem (bolsistas/contratados, etc)? Desses funcionários ativos, quantos devem se aposentar nos próximos anos?

Esse primeiro levantamento, focado nas demandas e dificuldades dos gestores, foi estratégico para as etapas seguintes. No entanto, era necessário compreender, de modo mais detalhado, como o fluxo de dados e informações era vivenciado pelas equipes técnicas responsáveis pelo uso desses sistemas. Por isso, avançou-se para uma investigação mais aprofundada, por meio de reuniões específicas com representantes das principais áreas envolvidas: tecnologia, gestão, orçamento e pessoal.

2 Entrevistas por setor

O segundo ciclo de entrevistas aprofundou a compreensão sobre a estrutura organizacional do Ipea e o uso rotineiro dos dados em cada setor. Para isso, buscou-se entender como os dados são utilizados nos processos de trabalho, quais são as principais fontes de informação, de que forma e com que frequência os relatórios são gerados, quais ferramentas são utilizadas, quem são os responsáveis pela produção e quem consome esses dados no apoio à gestão.

Foram realizadas entrevistas baseadas na escuta ativa, nas quais as equipes apresentaram seus fluxos de trabalho, os sistemas estruturantes utilizados na execução de atividades administrativas do órgão, as planilhas de acompanhamento e manipulação da informação e os relatórios gerados, além da organização das informações. Essa estratégia foi fundamental para revelar não apenas a estrutura operacional dos processos, mas também a forma como cada setor se relaciona com os dados: como eles são obtidos, tratados, interpretados e utilizados para apoiar as decisões. Além disso, foi possível acessar os relatórios, planilhas e documentos, que ilustram a manipulação e o uso dos dados em cada setor. A sistematização dessas informações constituiu uma etapa central do trabalho, permitindo consolidar uma visão integrada e aprofundada dos fluxos informacionais da instituição.

A partir das entrevistas e do material coletado, foi possível iniciar a análise dos sistemas utilizados como fontes de dados, identificando suas funcionalidades, interdependências e limitações. Além disso, foi elaborada a versão inicial do dicionário de dados, que identificou as variáveis mais relevantes, seus formatos, origens e as principais inter-relações entre elas. Esse processo também permitiu reconhecer as principais dificuldades enfrentadas pelas equipes na produção e qualificação das informações.

Por fim, iniciou-se a formulação de hipóteses e fluxos de dados, que posteriormente foram validados no workshop colaborativo. Entretanto, algumas nuances e pontos específicos ainda exigiam maior compreensão, o que tornou necessário conduzir uma nova rodada de entrevistas, desta vez com foco no detalhamento de aspectos relevantes.

3 Entrevistas focadas

O terceiro ciclo de entrevistas aprofundou aspectos específicos identificados na sistematização inicial. Retornou-se aos grupos diretamente envolvidos na produção e qualificação das informações, com o propósito de detalhar os processos de transformação manual dos dados, compreender com mais precisão como ocorrem essas atividades e esclarecer dúvidas sobre a utilização de fontes distintas que, por vezes, armazenam informações semelhantes. Essa etapa foi essencial para confirmar a qualidade e a consistência dos dados, bem como compreender as escolhas feitas pelas equipes na integração e validação das informações.

Nesse ponto, foram conduzidas entrevistas semi-estruturadas, com foco em esclarecer dúvidas pontuais, solicitar demonstrações práticas de procedimentos de extração e qualificação de dados, e validar o entendimento sobre os fluxos e sistemas previamente mapeados. Esse formato permitiu maior flexibilidade para

abordar casos específicos, aprofundar discussões técnicas e confirmar a aderência entre a prática institucional e o diagnóstico até então elaborado.

Como resultado, foi possível aprimorar o diagnóstico sobre as percepções dos principais problemas relacionados à gestão de dados, bem como o mapeamento dos gargalos existentes. Também foi consolidado o desenho dos processos e fluxos de qualificação dos dados, identificando as informações mais relevantes e impactantes para cada setor, e refinada a caracterização das fontes de dados, o que permitiu ampliar e atualizar o dicionário de dados. Além disso, foram delineados os requisitos iniciais para a integração das bases, estabelecendo parâmetros importantes para a interoperabilidade entre os sistemas.

4 Workshop sobre Gestão de Dados

Após as rodadas de entrevistas e levantamentos de informações, foi estruturado um conjunto robusto de hipóteses e um entendimento mais preciso sobre os fluxos de dados institucionais, o que possibilitou avançar para a próxima etapa: a validação colaborativa dessas hipóteses por meio de uma oficina com as equipes da Diretoria de Desenvolvimento Institucional do Ipea.

A realização da oficina foi uma etapa crucial no diagnóstico da extração e manipulação de dados, uma vez que viabilizou a colaboração na definição dos tipos de dados solicitados por diferentes setores do Ipea, mapeou os fluxos de informação entre os sistemas estruturantes e não estruturantes, e possibilitou identificar outras dificuldades existentes relacionadas à gestão, manipulação e produção das informações nos diferentes setores do Ipea.

As atividades do Workshop foram divididas nos seguintes momentos:

- **Sistemas Estruturantes e Dados:** Esse momento permitiu uma revisão dos sistemas empregados pela DIDES/Ipea e uma identificação mais precisa das informações que cada um produz. Como resultado, caracterizou-se os sistemas estruturantes e não estruturantes, bem como a tipologia de dados produzida por cada solução.
- **Relações entre os Sistemas:** As atividades avançaram no sentido de identificar as inter-relações entre os sistemas e validar o fluxo de informações entre eles. Os participantes foram desafiados a visualizar e compreender como os sistemas se interconectam e trocam dados, proporcionando uma visão mais abrangente da infraestrutura de dados do Ipea.
- **Mapeamento de Dificuldades para Extração e Manipulação de Dados:** O último momento da oficina objetivou a caracterização das dificuldades existentes na extração e posterior manipulação de dados, com o objetivo de

gerar informações úteis para a gestão. A partir da mediação dos moderadores, os participantes foram incentivados a elencar os aspectos que influenciam o processamento da informação, bem como as principais demandas solicitadas.

FIGURA 1:
Registros de realização do Workshop de Gestão de Dados



Após a oficina, as informações sobre os sistemas estruturantes, os dados coletados de cada um e as inter-relações entre eles foram sistematizadas, permitindo mapear os desafios operacionais e de gestão nas suas operações.

No que tange à relação entre os sistemas, na sequência da oficina, foi possível desenhar dois grandes ciclos de dados: o primeiro, referente ao fluxo de planejamento orçamentário e execução financeira; e o segundo relacionado a estrutura organizacional e gestão dos recursos humanos. Esses ciclos são detalhados na sequência.

Contexto do Ipea



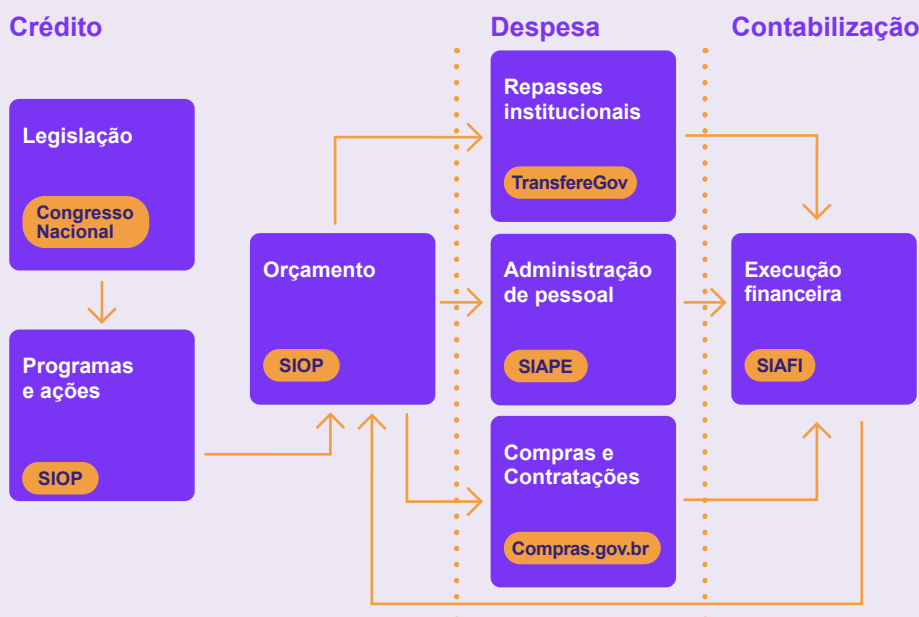
Fluxo orçamentário e financeiro

O fluxo do orçamento e das finanças do Governo Federal envolve diversos sistemas, como o SIOP (planejamento e orçamento), o SIAFI (administração financeira), o Compras.gov.br (compras públicas), o Transferegov (transferências de recursos) e o SIAPE (gestão de pessoal). O SIAFI é o elemento central dessa estrutura, uma vez que toda a execução financeira é consolidada nestes sistema. Por isso, contratos registrados no Compras.gov.br, transferências feitas pelo Transferegov e gastos com pessoal do SIAPE estão todos conectados ao SIAFI.

Como mostra a figura abaixo, o SIOP é o sistema estruturante utilizado pelo Governo Federal para o planejamento do orçamento, sendo utilizado em diversos processos tais como a elaboração do Plano Plurianual (PPA), Projeto de Lei de Diretrizes Orçamentárias (PLDO) e o Projeto de Lei Orçamentária Anual (PLOA). Depois que a LOA é aprovada pelo Congresso Nacional e sancionada pela Presidência da República, os órgãos do governo podem começar a usar os recursos orçamentários em sua operação. As despesas são registradas em diferentes sistemas: o Compras.gov.br, para compras e contratos; o Transferegov, para transferências de recursos; e o SIAPE, que trata dos gastos e dados de pessoal. Todas essas informações são enviadas para o SIAFI, que faz o controle e o registro da execução financeira.

FIGURA 2:

Fluxo do valor do dado - orçamentário/financeiro



Para cada área (contratos, Termos de Execução Descentralizadas (TEDs) e recursos humanos), os indicadores são gerados de formas diferentes. Em muitos casos, é preciso cruzar dados de vários sistemas.

No caso dos contratos, os indicadores refletem a execução em relação ao que foi contratado. A maior parte das informações provém do Compras.gov.br, com dados dos contratos, faturas e notas de empenho. No entanto, como alguns dados estão no SIAFI, é necessário buscá-los nesse sistema também para completar os indicadores. Por isso, os relatórios ainda precisam ser montados manualmente por especialistas.

No caso dos TEDs, os dados principais dizem respeito à execução orçamentária e financeira das parcerias. As informações básicas — como plano de ação e orçamento — provêm do Transferegov. Já os dados financeiros e orçamentários são extraídos do SIAFI, com base nas notas de crédito (NC), notas de empenho (NE) e programações financeiras (PF). Essas informações são consultadas separadamente e precisam ser reunidas manualmente por um técnico, geralmente em planilhas, para ter uma visão completa da execução do TED.

Fluxo de gestão de recursos humanos

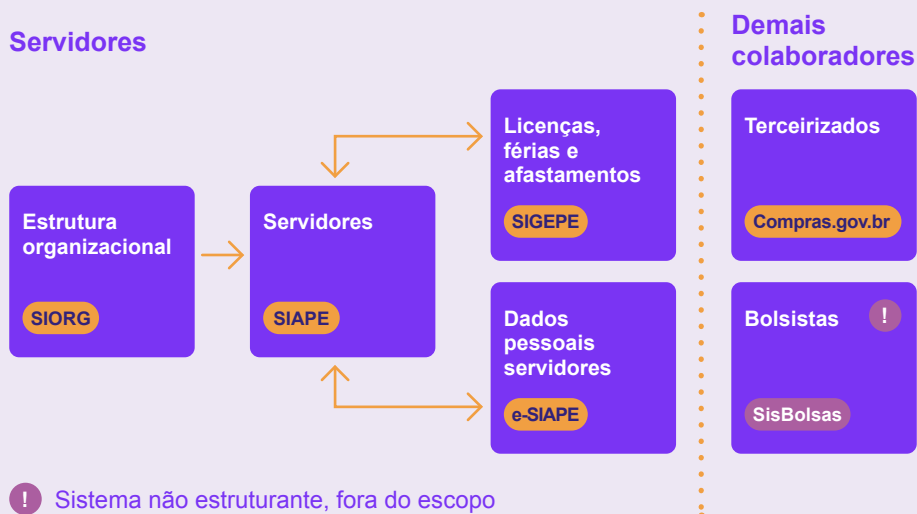
A gestão de recursos humanos também utiliza vários sistemas estruturantes, como o SIORG (estrutura organizacional), o SIAPE (gestão de pessoal), o SIGEPE (gestão de servidores), o Compras.gov.br (contratos de terceirizados) e o SisBolsas (usado internamente pelo Ipea para gerenciar bolsistas).

Essa gestão acontece em dois contextos separados: um para os servidores públicos (estatutários) e outro para terceirizados e bolsistas, que no caso do Ipea, têm regras e sistemas diferentes.

Como mostra a Figura 2, o processo começa com a definição da estrutura de cargos e departamentos no SIORG. Depois, o SIAPE registra a alocação de servidores, as nomeações para chefias e as movimentações, como cessões e requisições. O SIGEPE é responsável pelo controle de licenças, férias e afastamentos, enviando essas informações ao SIAPE. Já o portal e-SIAPE permite acompanhar dados pessoais, ocorrências e tempo de serviço dos servidores.

Os registros dos terceirizados, por sua vez, são feitos no Compras.gov.br, ligados aos contratos, enquanto os dados dos bolsistas são gerenciados pelo SisBolsas, sistema próprio do Ipea.

FIGURA 2:
Fluxo do valor do dado - força de trabalho



A maioria dos indicadores de pessoal é gerada a partir do SIAPE DW, um banco de dados que reúne informações dos sistemas de gestão de pessoas (SIAPE, SIGEPE e e-SIAPE). Ele permite fazer consultas detalhadas sobre os servidores.

No entanto, alguns dados importantes — como o número de cargos vagos e ocupados — também exigem informações do SIORG. Nesse caso, é preciso cruzar manualmente os dados dos dois sistemas, tarefa que normalmente é feita por especialistas.

Para determinar o total da força de trabalho do órgão, é necessário consultar os sistemas que gerenciam terceirizados e bolsistas, pois essas informações não estão disponíveis no SIAPE. Por isso, o processo de geração de indicadores de pessoal depende da extração de dados de vários sistemas e da elaboração manual de relatórios em planilhas, feitos sob demanda.

Esses dois fluxos evidenciam a complexidade na gestão de dados governamentais, abrangendo a sua obtenção, qualificação e integração. Esses desafios impactam diretamente na tomada de decisão por parte dos gestores e afetam a dinâmica de trabalho do órgão.

Durante esse processo de diagnóstico, foram identificados desafios de diferentes naturezas: alguns relacionados à infraestrutura e limitações dos próprios sistemas estruturantes; outros, correlatos aos procedimentos internos de gestão da informação; e, por fim, aqueles associados ao uso efetivo dos dados pelos gestores e servidores. Essas questões foram organizadas e categorizadas em diferentes frentes de ação, compondo um *roadmap* inicial a ser ajustado e priorizado pelo órgão, de acordo com suas capacidades institucionais e estratégias de atuação.

Desse processo, vale destacar que o mapeamento evidenciou um conjunto de problemas estruturais que comprometem a confiabilidade, a transparência e a usabilidade dos dados nos sistemas estruturantes do governo federal. A fragmentação entre sistemas, a ausência de padronização e a falta de integração automatizada exigem que a geração de indicadores dependa de cruzamentos manuais e interpretações especializadas, o que amplia o risco de erros, retrabalho e limita a agilidade na tomada de decisão. A baixa qualidade das informações — marcada por registros inconsistentes, campos não estruturados e ausência de identificadores únicos — afeta diretamente a construção de análises confiáveis, tanto para o público interno quanto para o externo.

Essas dificuldades se manifestam de forma concreta no cotidiano da Diretoria de Desenvolvimento Institucional do Ipea, em que, apesar da quantidade de dados gerados por diferentes áreas, muitas perguntas estratégicas demandam esforços intersetoriais para serem respondidas. A circulação da informação se torna morosa e os relatórios são, muitas vezes, pouco acessíveis ou compreensíveis para gestores e técnicos. A realidade observada é comum a diversos órgãos que operam com os mesmos sistemas estruturantes e enfrentam limitações semelhantes em termos de interoperabilidade e governança.

Com base nessas informações, o mapeamento dos problemas cumpriu um papel decisivo ao consolidar uma visão institucional compartilhada sobre os principais obstáculos enfrentados na produção e no uso de dados para a gestão pública. A sistematização de evidências, combinada com a participação ativa das equipes técnicas, permitiu compreender os desafios relacionados à gestão de dados e informações nos sistemas estruturantes.

Resultados do mapeamento

Essa fase foi essencial para construir uma compreensão compartilhada entre as equipes técnicas e de gestão, identificar os principais gargalos e alinhar as expectativas para as próximas fases do projeto.

O resultado deste processo materializou-se em três produtos fundamentais:

1 Diagnóstico das principais necessidades e dificuldades para a gestão de dados e informações:

Esse diagnóstico constituiu a base para o desenvolvimento da solução e o aprimoramento de processos de gestão orientada por dados.

2 Aprofundamento das informações sobre os sistemas estruturantes, dados e dicionários:

Foi realizado um mapeamento detalhado dos sistemas estruturantes utilizados pela DIDES/Ipea, com a identificação de escopo e dicionários de dados, além de mais de 50 relatórios institucionais utilizados pelas equipes e as relações entre os distintos sistemas. Esse levantamento permitiu identificar lacunas e inconsistências na utilização dos sistemas. O aprofundamento do mapeamento possibilitou a identificação das relações entre as bases de dados e a necessidade de maior alinhamento técnico entre as áreas responsáveis por cada sistema.

Além da representação técnica dos fluxos, foi criado um diagrama específico para ilustrar o fluxo de valor dos dados. Esse material apresenta como os dados, gerados em diferentes etapas do processo, produzem valor institucional ao se transformarem em informações qualificadas para a tomada de decisão, contribuindo para uma gestão mais eficiente e transparente.

Por fim, foi organizada uma lista das bases de dados e sistemas envolvidos nos processos mapeados, com indicação dos principais tipos de dados fornecidos por cada um.

3 Detalhamento dos principais fluxos de dados sobre planejamento orçamentário, execução financeira e gestão institucional:

Considerando que os processos relacionados ao orçamento, à execução financeira e à gestão de pessoas constituem áreas críticas para a administração do órgão, foi realizado um detalhamento específico dos fluxos de dados relacionados a esses temas. Esse exercício evidenciou pontos de atenção, como redundâncias na coleta de dados, inconsistências na nomenclatura e na periodicidade das atualizações, além da necessidade de automatizar os processos hoje realizados manualmente.

Foram elaborados diagramas detalhados representando os caminhos percorridos pelos dados, desde sua coleta, passando pelos processos de integração entre sistemas, até a geração de relatórios e indicadores. Esse artefato tornou visíveis as interdependências entre áreas, os pontos críticos de entrada e saída de dados e os processos manuais ainda existentes.

A conclusão da Etapa 1 oferece uma visão estratégica dos desafios enfrentados pelo órgão na gestão de dados e informações, ao mesmo tempo em que fornece insumos concretos para superá-los. O diagnóstico, os mapas e os fluxos produzidos representam uma base sólida para a construção das soluções, orientadas a garantir maior integração, qualidade e uso inteligente dos dados no apoio à gestão institucional. Esse entendimento permite avançar para a análise exploratória dos dados, a qual será descrita no próximo capítulo.



Importância da etapa de mapeamento do problema

Toda solução, seja ela tecnológica ou não, precisa atender a uma necessidade. Assim, a compreensão aprofundada do problema é uma etapa decisiva para o desenvolvimento de soluções adequadas. Nesse processo, a escuta ativa dos diferentes usuários, de forma a identificar suas demandas centrais e, sobretudo, entender as reais dificuldades enfrentadas na realização de diferentes tarefas, são essenciais para definir o escopo e o desenho da solução. Por essa razão, a etapa de mapeamento é estratégica para garantir o impacto e o resultado do que se pretende criar. No desenvolvimento do GovHub, embora a escuta tenha sido constante, esse momento foi particularmente pedagógico para as equipes da DIDES/Ipea e Lab Livre / UnB, pois proporcionou uma troca valiosa e uma compreensão dos desafios que são comuns a diversas instituições.



Como replicar essa etapa?

Para aplicar a metodologia do mapeamento dos desafios descrita neste capítulo, siga os passos abaixo:

Mapeamento dos Dados e Indicadores

- Realize entrevistas semi estruturadas com equipes estratégicas que utilizam os relatórios para tomada de decisão, com o objetivo de identificar as perguntas estratégicas que buscam ser respondidas pelos relatórios.

- Realize reuniões/entrevistas com as equipes que produzem os relatórios, **para entender:**

- Identifique todos os relatórios institucionais relevantes.
- Liste os indicadores e variáveis utilizados em cada relatório.
- Identifique a fonte dos dados utilizados por cada relatório.
- Sistematize como os dados dos sistemas se relacionam para a produção dos relatórios.
- Sistematize os fluxos de coleta de dados.
- Sistematize os dados necessários de cada sistema, bem como a metodologia de cálculo dos indicadores.
- Valide a sistematização com a equipe operacional/estratégica.

- A frequência com que os relatórios são produzidos;
- O processo de obtenção de dados e produção dos relatórios;
- As principais dificuldades encontradas na produção dos relatórios (dados não qualificados, falta de dados, etc);
- Como inserem dados nos sistemas;
- Como consomem os dados dos sistemas.

Artefatos produzidos

- Sistematização de questões estratégicas respondidas por meio dos sistemas estruturantes;
- Diagrama com os fluxos dos dados, desde a coleta à integração de bases e geração de relatório;
- Diagrama com o fluxo de valor do dado;
- Lista das bases de dados/sistemas necessários para gestão orientada a dados, com os principais dados fornecidos por cada sistema;

CAPÍTULO 2

ANÁLISE EXPLORATÓRIA DOS DADOS E VIABILIDADE DE INTEGRAÇÃO DE SISTEMAS

A etapa de exploração e estudos para cruzamento e qualificação de dados representa o momento mais extenso de todo o ciclo de desenvolvimento. É nela que se torna possível compreender a estrutura, a qualidade e a lógica dos dados e sistemas.

O trabalho iniciou-se com o acesso às diversas bases de dados, seguido pela construção de um dicionário técnico que documentou campos, formatos e metadados essenciais. Simultaneamente, foi mapeado o fluxo de valor organizacional, identificando onde cada informação gera impacto real no negócio.

Combinando rigor científico e conhecimento de domínio, as análises exploratórias revelaram o panorama real dos dados: lacunas de completude, inconsistências estruturais e outliers significativos. Essa investigação orientou os processamentos iniciais de limpeza e padronização, estabelecendo modelos de qualidade sustentáveis.

Paralelamente, foram realizados estudos com enfoque na integração entre fontes heterogêneas, verificando a compatibilidade de chaves relacionais, formatos de dados e políticas institucionais de compartilhamento. Essa análise previne gargalos futuros e garante escalabilidade.

O resultado desta fase materializou-se em três produtos fundamentais:

- **Diagnóstico técnico detalhado** do ecossistema de dados;
- **Requisitos refinados** de integração das bases de dados, baseados em evidências empíricas;
- **Parecer de viabilidade** de integração das bases de dados, que orienta as decisões arquiteturais.

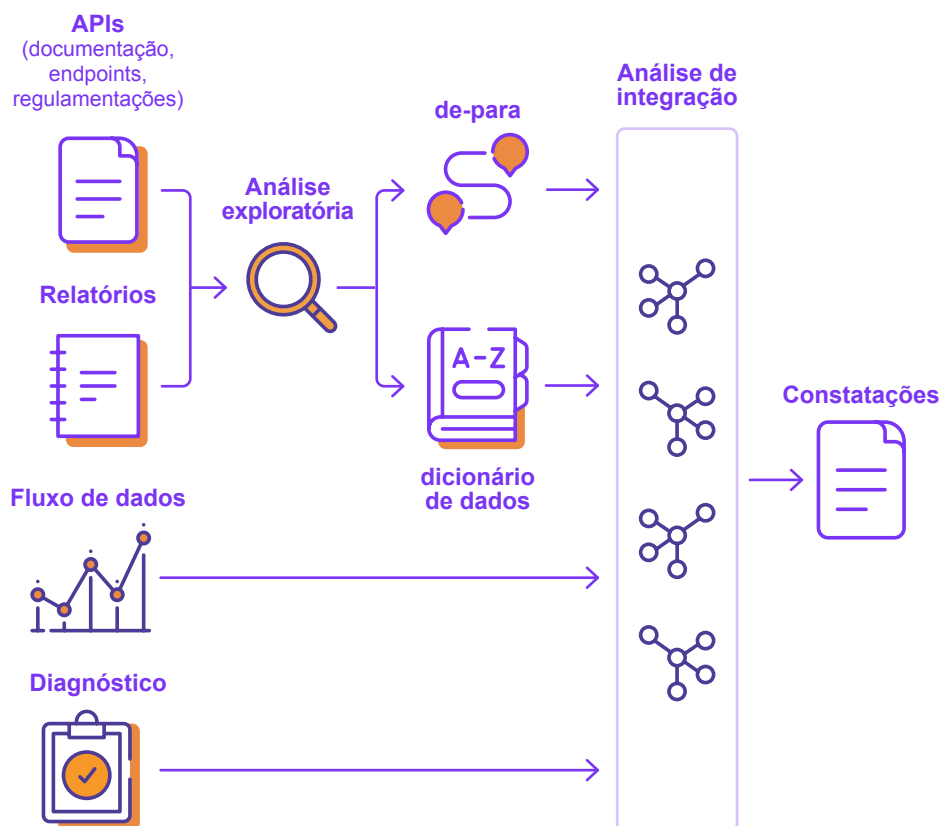
Essa etapa avaliou a qualidade das informações, o tipo de dado existente em cada sistema e localizou suas ocorrências nos fluxos de trabalho. Isso possibilitou rastrear o ciclo de vida das informações e, assim, revelar as conexões entre as diferentes bases. Ela também foi fundamental para compreender se os problemas encontrados eram sistêmicos ou operacionais.

Como descrito na Figura 3, este capítulo descreve a metodologia de análise exploratória e de integração dos dados. A análise exploratória visa estudar toda a documentação técnica e as informações sobre os sistemas estruturantes, de forma a entender o seu funcionamento, como acessá-los e a qualidade, consistência e confiabilidade desses dados. Na segunda fase, com enfoque em estudos de viabilidade de integração, a equipe buscou formas de identificar e acompanhar a evolução dos dados nos sistemas por meio de chaves primárias e identificadores, além de explorar os caminhos para viabilizar a interoperabilidade entre as bases. Essa investigação possibilitou identificar em cada API (*Application Programming Interface*, que significa Interface de Programação de Aplicações, que permite a comunicação entre sistemas e aplicativos) os *endpoints* (ponto de acesso dentro da API que permite consultar ou enviar dados) relacionados com as informações necessárias para a emissão dos relatórios e, por consequência, avaliar se os dados eram consistentes e atualizados.



FIGURA 3:

Fluxo metodológico para a análise exploratória e estudo de integração

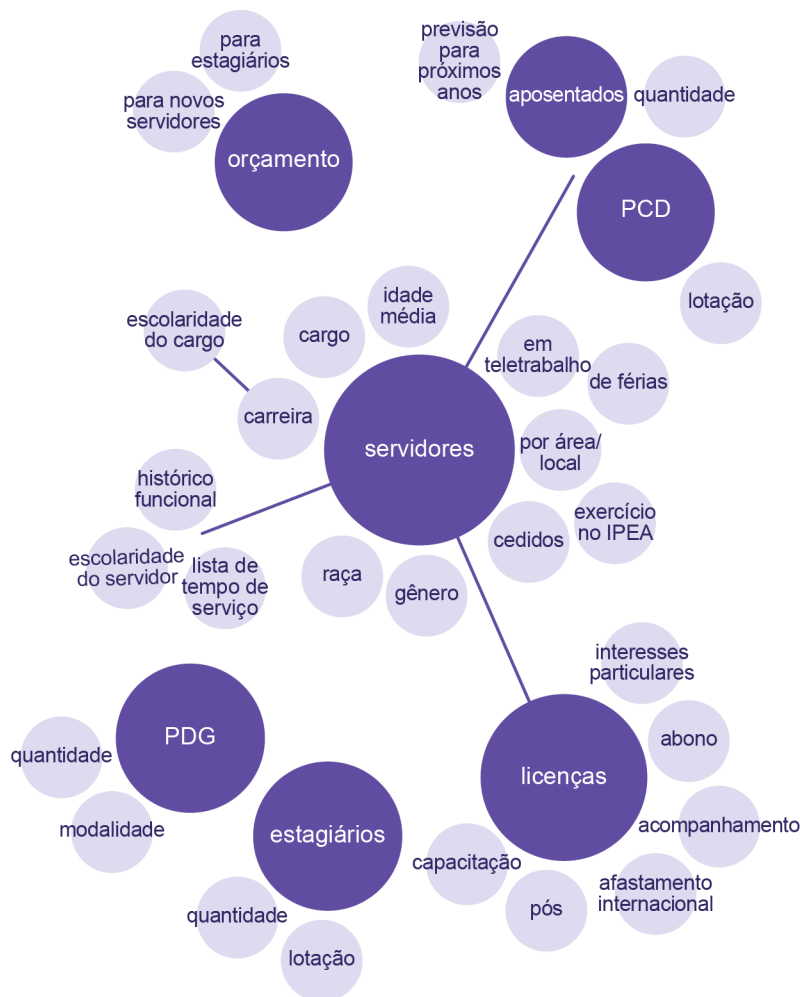


1 Análise exploratória

Após o mapeamento dos fluxos de dados e a identificação das principais dificuldades, no contexto da Diretoria de Desenvolvimento Institucional do Ipea, realizados na etapa anterior, iniciou-se a exploração dos dados e a análise das questões técnicas relacionadas ao acesso e à manipulação das informações. O primeiro passo foi mapear os indicadores e dados presentes em cada um dos mais de 50 relatórios disponibilizados durante as entrevistas. Em seguida, as informações utilizadas para a elaboração de tais relatórios foram identificadas nos respectivos sistemas estruturantes. As Figuras 4, 5 e 6 sintetizam os dados/indicadores utilizados pela Coordenação-Geral de Gestão de Pessoas (CGPES), a Coordenação-Geral de Gestão Estratégica e Orçamento (CGPGO) e a Coordenação-Geral de Serviços Corporativos e Apoio à Pesquisa (CGCAP).

FIGURA 4:

Mapeamento dos dados/indicadores necessário para gestão de pessoas

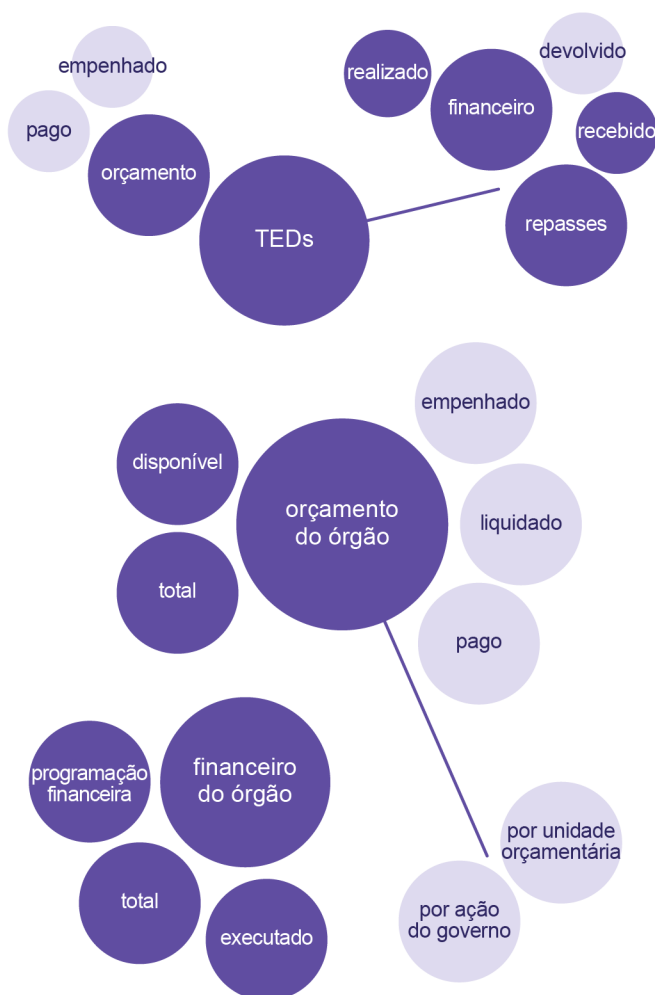


A análise exploratória avançou no mapeamento de recursos tecnológicos para caracterizar a tipologia do dado que pode ser extraída de um sistema estruturante e o mapeamento de comandos a serem utilizados por soluções de software externas. Para tanto, foi necessário estudar as APIs, as documentações técnicas e os procedimentos exigidos para concessão de acesso, com intuito de realizar a extração de informações de forma automatizada. Todavia, cabe destacar que o acesso aos sistemas e, consequentemente, aos dados não ocorreu de maneira uniforme e simples: os sistemas

estruturantes são mantidos por diferentes órgãos, cada um com políticas específicas de acesso e solicitação. Por exemplo, os dados do SIORG estavam disponíveis por meio de uma API pública e bem documentada, o que permitiu iniciar a exploração de forma imediata. Já no caso do SIAFI, foi necessária uma solicitação formal ao órgão responsável, com justificativas técnicas e aprovação prévia. Essas diferenças impactam diretamente o acesso, o cronograma e a estratégia de análise para cada sistema.

FIGURA 5

Mapeamento dos dados/indicadores para gestão do planejamento a execução de TEDs

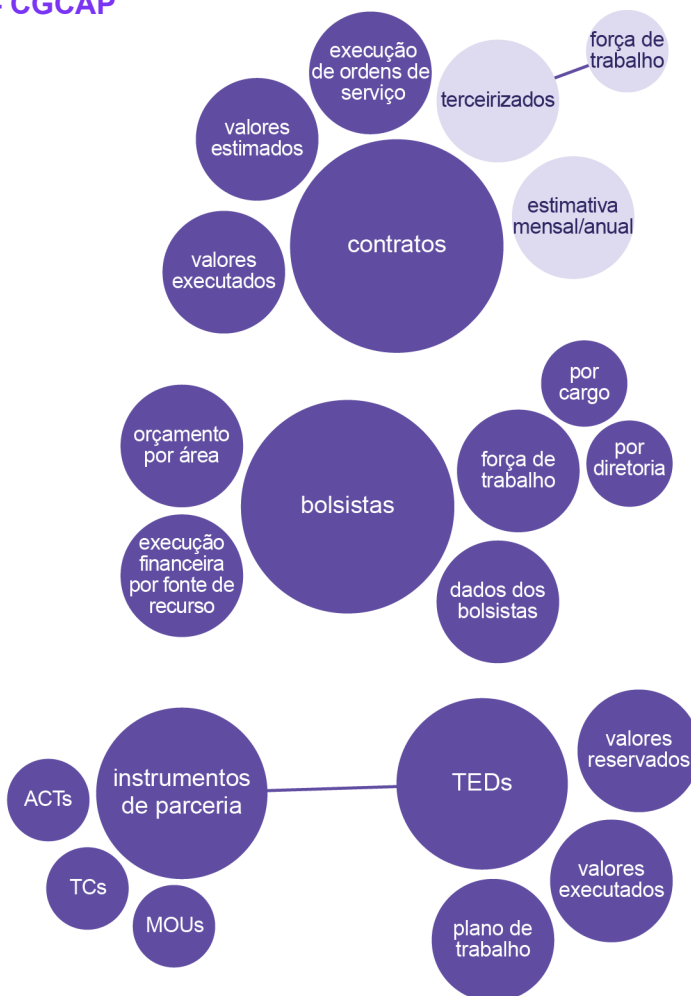


Durante essa análise, parte da documentação das APIs estava desatualizada ou incompleta. Relato similar foi dado pela equipe que opera os sistemas estruturantes, que apontou que os tutoriais de uso são defasados e atualizações nos sistemas não são devidamente comunicadas. Para contornar esse desafio, foi necessário explorar os dados de forma experimental, observando diretamente as informações retornadas. Esse esforço resultou em um relatório técnico (disponível em: <https://www.gov-hub.io/>) que detalha como acessar as APIs, além de apresentar os processos, regras e especificidades de cada sistema.

FIGURA 6:

Mapeamento dos dados/indicadores para gestão do planejamento a execução financeira

Dados - CGCAP





Acessando as APIs do Governo Federal

O acesso às APIs dos sistemas estruturantes é controlado pelos mantenedores e proprietários de cada sistema. O acesso pode ser solicitado por órgãos e entidades governamentais que necessitam integrar ou consumir informações disponíveis nos sistemas estruturantes. Dependendo da API, o pedido pode ser feito diretamente pela equipe técnica ou administrativa do órgão interessado.

Como solicitar acesso?

- Identifique o tipo de API
 - APIs Abertas: Não exigem processos complexos de autorização e podem ser acessadas diretamente por meio do catálogo de serviços;
 - APIs Fechadas: Exigem uma solicitação formal, que pode incluir aprovação interna, assinatura de termos de adesão ou justificativas técnicas.
- Consulte o Catálogo de APIs Governamentais, disponível no portal Conecta Gov.br, principal referência para informações detalhadas sobre as APIs.
- Formalize a solicitação.
- Entre em contato com o responsável pela API utilizando as informações fornecidas no catálogo.
- Prepare os documentos necessários, como ofícios formais ou termos de adesão, dependendo das exigências da API.
- Inclua informações detalhadas, como o objetivo do uso e o código do órgão solicitante.
- Acompanhe e configure o acesso após a aprovação do pedido.
- Configure as chaves de acesso e os tokens de autenticação fornecidos.
- Realize testes iniciais para garantir que os dados estão sendo coletados corretamente.
- Ajuste a infraestrutura de consumo de dados para atender às demandas do órgão.

Recomendações

- Planeje o uso: Antes de solicitar acesso, defina claramente como os dados serão utilizados e documente as etapas do processo técnico.
- Utilize ferramentas de código aberto: Ferramentas reconhecidas no mercado, como Apache Superset e DBT, podem facilitar a integração e a análise dos dados.
- Garanta a segurança: Proteja as chaves de acesso e siga as políticas de conformidade estabelecidas pelos responsáveis pelas APIs.

Após obter acesso, foi possível mapear os *endpoints* que continham as informações necessárias para gerar relatórios, os identificadores únicos e as regras de relacionamento entre diferentes sistemas. Com esse mapeamento, elaborou-se um dicionário de dados, contendo definições das variáveis, suas nomenclaturas, estruturas, formatos, periodicidades e regras de cálculo e interpretação de cada indicador. Além disso, também foi gerada uma tabela de correspondência entre os campos das APIs e os elementos dos relatórios institucionais. A Figura 7 apresenta exemplos do mapeamento realizado, relacionando as colunas Variável (xml) e XPath (indicando a estrutura dos dados nas APIs) com os campos utilizados nos relatórios emitidos pela DIDES/Ipea.

FIGURA 7
Correlação entre os dados usados nos relatórios com as variáveis identificadas nas APIs

Variável (XML)	de-para	Relatório	XPath
grMatricula	Matrícula	Relatório 10 - Afastamentos e Licença-.pdf	/soap:Envelope/soap:Body/ns1:consultaDadosAfastame...
ferias	Férias	Relatório 10 - Afastamentos e Licença-.pdf	/soap:Envelope/soap:Body/ns1:consultaDadosAfastame...
lpa	LPA	Relatório 10 - Afastamentos e Licença-.pdf	/soap:Envelope/soap:Body/ns1:consultaDadosAfastame...
ocorrencias / DadosOcorrencias	Ocorrências de Afastamento	Relatório 10 - Afastamentos e Licença-.pdf	/soap:Envelope/soap:Body/ns1:consultaDadosAfastame...

De modo a consolidar essa etapa, foi realizada uma avaliação da qualidade dos dados, considerando sua completude, consistência, precisão e atualização. A análise revelou campos frequentemente nulos ou inconsistentes, além de variações nos padrões de preenchimento entre diferentes órgãos. Também foi observada a ausência de identificadores únicos e o uso de campos não estruturados, especialmente em bases que exigiam cruzamentos entre sistemas distintos, como no caso dos contratos, que envolvem a articulação entre o Compras.gov.br e o SIAFI.

2 Análise de Integração

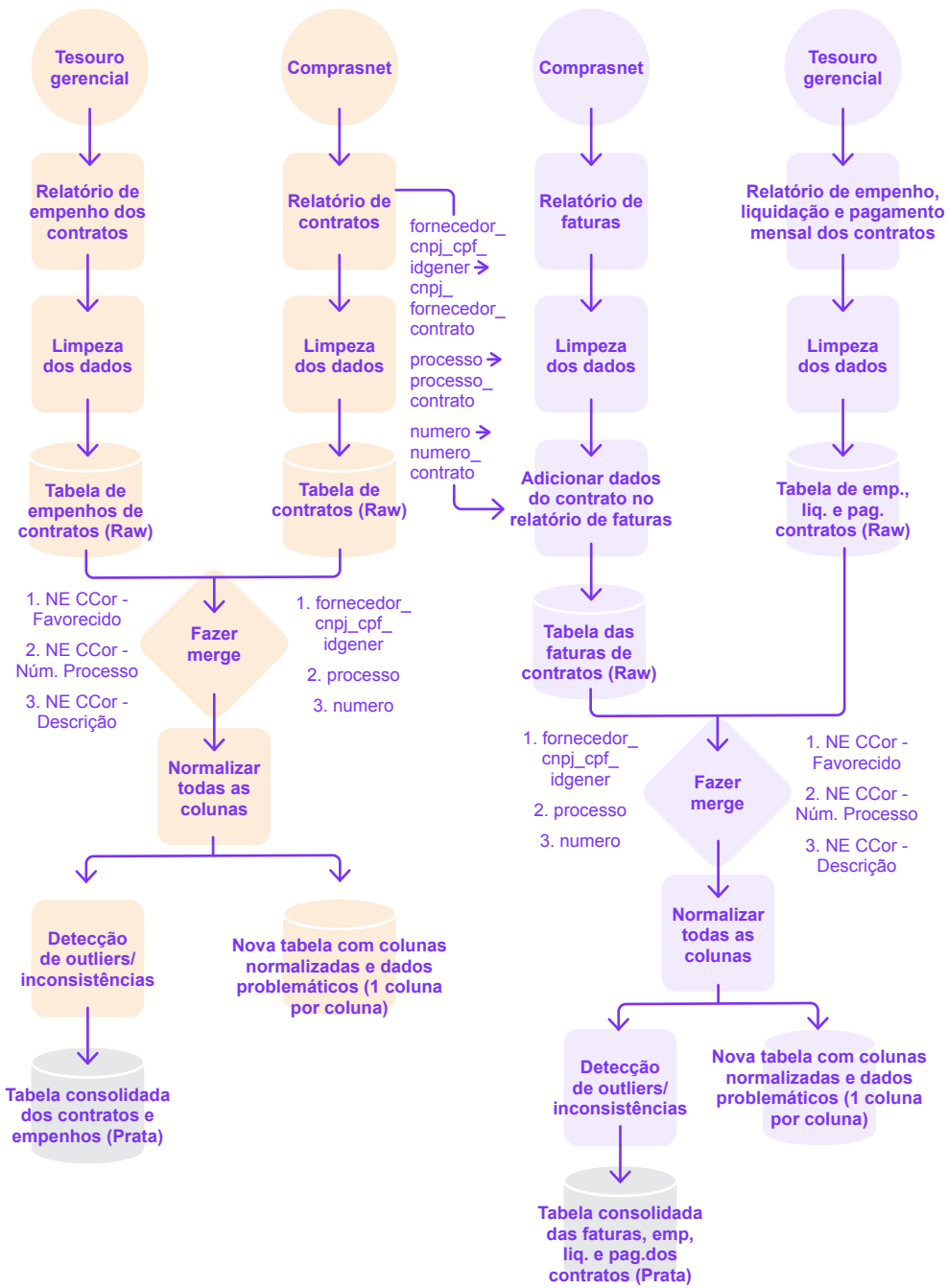
A análise de integração busca caracterizar a viabilidade das conexões entre sistemas estruturantes distintos, de modo que as informações possam ser integradas de maneira confiável e sustentável. Assim, essa etapa do trabalho focou em sistematizar todas as informações coletadas durante as fases de trabalho precedentes. Por essa razão, é fundamental que as etapas anteriores tenham sido realizadas com precisão a fim de garantir informações qualificadas e validadas pelo órgão.

A primeira tarefa foi localizar os identificadores-chave que conectam tabelas dentro de um mesmo sistema, a partir da análise das estruturas dos dados retornados pelas APIs e da documentação disponível. Com esses identificadores, foi possível mapear as relações internas das bases e compreender as conexões entre diferentes conjuntos de dados.

Em seguida, a análise avançou para identificar relações entre sistemas distintos, buscando identificadores comuns que possibilitam a integração entre bases diferentes, garantindo a combinação consistente das informações. Esse mapeamento exigiu uma avaliação detalhada dos padrões de nomenclatura e da estrutura dos identificadores em cada sistema, já que nem sempre havia correspondência direta entre eles.

Para facilitar a visualização dessas conexões, fluxogramas de integração dos dados, como apresentado na Figura 8, foram desenvolvidos. Esses diagramas ajudaram a estruturar o fluxo de coleta de dados e a compreender melhor as relações internas e externas entre sistemas. Um dos modelos desenvolvidos foi o de ETL (Extração, Transformação e Carga), utilizado como base para organizar o fluxo de integração. Outros fluxos de integração de dados podem ser acessados no seguinte link: <https://www.gov-hub.io>

FIGURA 8
Fluxo de integração dos dados do Compras.gov e Tesouro Gerencial.



Essa fase também incluiu a verificação da viabilidade de integração entre as bases de dados, considerando políticas institucionais de acesso, formatos de dados, chaves relacionais e possíveis limitações técnicas.

A partir dessas análises, estruturou-se um processo de integração entre diferentes bases de dados. No âmbito da parceria entre UnB e Ipea, os sistemas analisados abrangem SIORG, SIAFI, SIAPE, TransfereGov e Compras.gov.br, cada um com características específicas que demandaram abordagens diferenciadas para integração.

Ficou evidente que a análise exploratória e a análise de integração são processos complementares. A exploração revelou quais informações estavam disponíveis e forneceu *insights* sobre as relações entre elas, sistematizando todas as informações coletadas. Essa abordagem permitiu criar um modelo de integração sólido e sustentável, garantindo o uso eficiente dos dados.



Exemplo:

A expectativa de interoperabilidade baseada em identificadores comuns não se confirmou na análise exploratória. Um caso emblemático foi o do SIORG: inicialmente, a Unidade Gestora (UG) foi considerada como possível chave de integração entre sistemas, mas a equipe identificou que não há correspondência direta entre os códigos utilizados. A mesma UG aparece com identificadores distintos em diferentes sistemas, inviabilizando o cruzamento automático. Por outro lado, essa análise revelou que, embora eventualmente haja trocas de informações, como o Compras.gov.br, que aciona operações no SIAFI, essas comunicações são específicas, fragmentadas e não sustentam uma interoperabilidade ampla ou automática. Na prática, os dados são replicados em bases distintas e não existe integração nativa entre os sistemas.

Contexto do Ipea

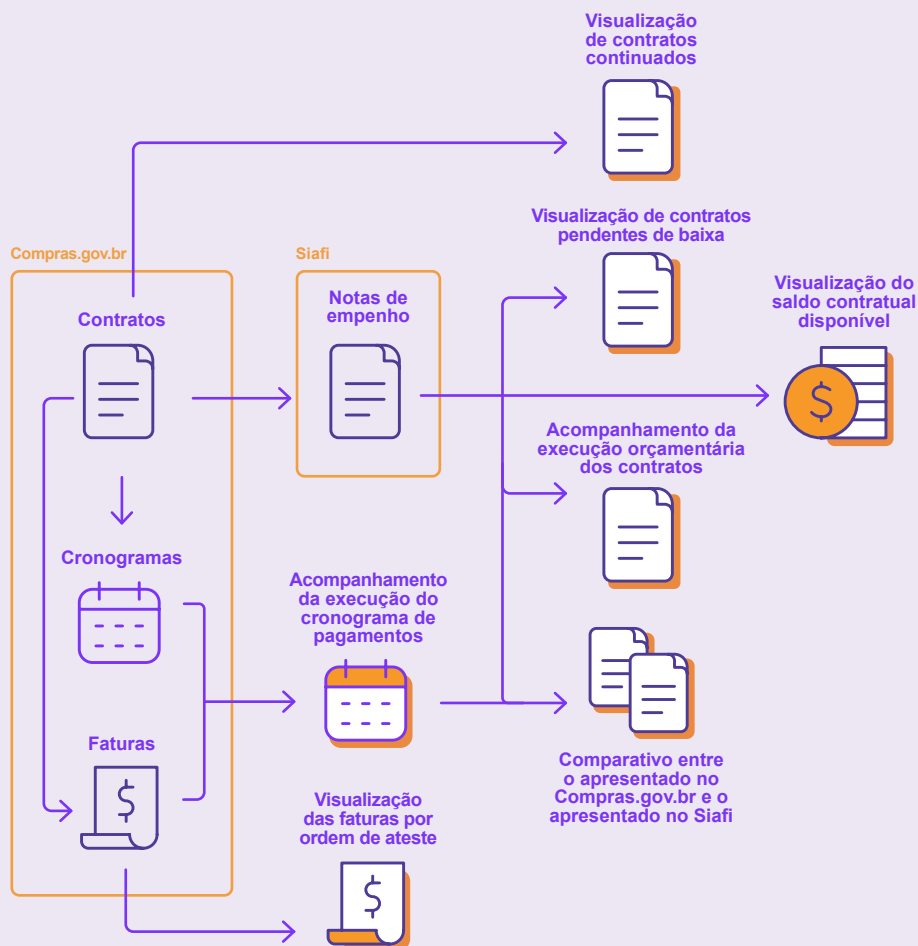


Integração de dados: Compras.gov.br e SIAFI

Ao fazer o cruzamento de dados e informações relativas aos contratos, foi necessário estabelecer um fluxo de integração entre o Compras.gov.br e o SIAFI, conforme demonstra a Figura 9.

FIGURA 9

Fluxo de integração entre Compras.gov.br e SIAFI, destacando a relação entre contratos, cronogramas, faturas e notas de empenho



Nesse processo, constatou-se cinco situações que dificultam o processo de integração e seus impactos:

1 Contratos permanecem abertos mesmo após a execução financeira estar concluída

Se o contrato já teve todas as faturas atestadas e os pagamentos registrados no SIAFI, então seu status deveria ser atualizado automaticamente como “encerrado” no Compras.gov.br. Caso contrário, há uma falha operacional na baixa contratual, o que compromete a análise da execução e gera distorções nos indicadores.

2 O cronograma de pagamento não reflete a execução real das faturas

Se os cronogramas cadastrados não são atualizados ou seguidos durante a execução, então as faturas podem ser emitidas fora do previsto, tornando ineficaz o acompanhamento da execução planejada. Isso compromete a capacidade de prever desembolsos e pode gerar divergências entre planejamento e execução.

3 Há divergência entre os valores registrados no Compras.gov.br e os valores executados no SIAFI

Se os dados financeiros dos contratos não são sincronizados entre os dois sistemas, então os saldos contratuais apresentados podem estar incorretos. Isso prejudica a confiabilidade da informação gerada tanto para o gestor quanto para o controle social.

4 Nem todas as faturas estão vinculadas diretamente a seus cronogramas e empenhos

Se há faturas emitidas sem vinculação explícita ao cronograma ou à nota de empenho correspondente, então há falhas na rastreabilidade documental do contrato. Isso torna o acompanhamento e a auditoria do contrato mais complexos e sujeitos a erro.

5 A ausência de identificadores únicos compromete o cruzamento automático entre os sistemas

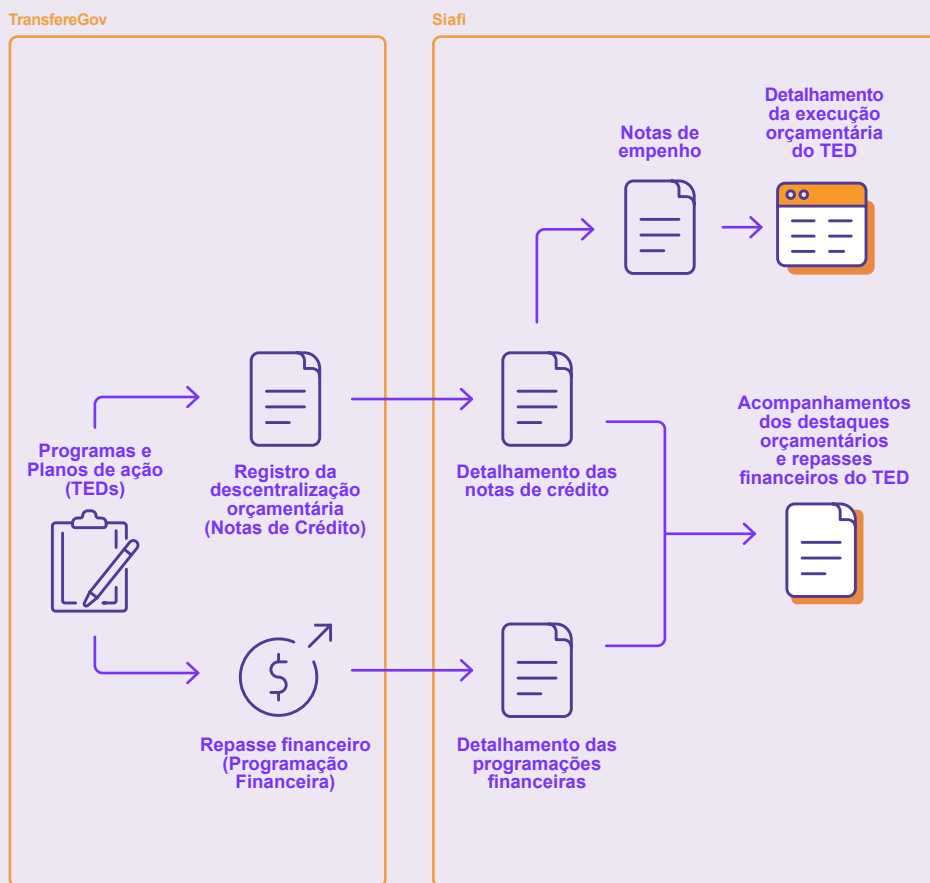
Se contratos, empenhos e faturas não compartilham identificadores consistentes entre os sistemas Compras.gov.br e SIAFI, então é necessário realizar cruzamentos manuais, o que eleva o risco de erro e retrabalho. Isso reduz a viabilidade da automação da geração de indicadores.

Integração de dados: TransfereGov e SIAFI

No que tange às informações de Termos de Execução Descentralizada (TED) realizados e recebidos pelo órgão, foi necessário observar a integração entre TransfereGov e SIAFI, conforme mostra a Figura 10.

FIGURA 10

Fluxo de integração entre TransfereGov (módulo de TED) e SIAFI



Nesse processo, identificou-se as seguintes constatações que dificultam a integração:

1 A execução orçamentária dos TEDs não é sistematizada no TransfereGov

Se os TEDs são firmados e registrados no TransfereGov, mas a execução orçamentária (empenho, liquidação, pagamento) só é detalhada no SIAFI, então não há um mecanismo unificado de acompanhamento da execução dentro do próprio TransfereGov. Isso obriga os gestores a fazerem conciliações manuais e dificulta a geração de indicadores automáticos.

2 O acompanhamento dos destaques orçamentários e repasses financeiros ocorre de forma fragmentada

Se o detalhamento das notas de crédito e das programações financeiras é feito apenas no SIAFI, então os usuários do TransfereGov não conseguem acompanhar em tempo real o fluxo completo do recurso descentralizado. Isso compromete o monitoramento e o planejamento da execução dos programas e planos de ação.

3 Não há visualização consolidada dos repasses financeiros

Se os dados sobre as programações financeiras existem no SIAFI, mas não são consolidados em relatórios ou painéis acessíveis no TransfereGov, então os gestores não conseguem antecipar os repasses pendentes. Essa ausência prejudica o controle de fluxo de caixa e a capacidade de reprogramação dos órgãos recebedores.

4 Ausência de integração entre a programação orçamentária e financeira afeta a rastreabilidade do TED.

Uma vez que os registros de descentralização (nota de crédito) e o repasse financeiro (programação financeira) não são integrados no TransfereGov ou no SIAFI, o rastreamento do ciclo completo de vida do TED, do compromisso à execução, apresenta limitações. Isso fragiliza a transparência e a prestação de contas dos instrumentos de descentralização.

5 Falta de identificadores consistentes entre os sistemas dificulta o cruzamento entre o plano de ação e os registros de execução

Se o identificador do TED e plano de ação no TransfereGov não é o mesmo utilizado para registrar os empenhos e programações no SIAFI, então é necessário criar mecanismos manuais de correspondência. Isso aumenta o risco de erro e inviabiliza o acompanhamento automatizado da execução por instrumento.

A análise exploratória dos dados e a caracterização de sua integração em diferentes sistemas se consolidou como uma etapa fundamental para estruturar o processo de qualificação permanente dos dados. Para além de compreender e detalhar as informações que podem ser extraídas a partir de API, o padrão de representação de informação em cada um dos sistemas estruturantes e possibilidades de integração, foi possível contribuir com documentações, atualizações e sistematizações que podem ser úteis para distintos órgãos públicos do governo federal.

O mapeamento dos fluxos de trabalho permitiu identificar os mecanismos para a geração de informações, como eles circulam entre setores e quais conexões podem ser estabelecidas entre sistemas estruturantes. Mesmo sem integrações nativas, a análise revelou que é viável cruzar bases de diferentes fontes para produzir novos indicadores e responder a demandas específicas da gestão. Essa compreensão permitiu não apenas identificar gargalos, mas também formular caminhos para a reorganização da informação de forma estratégica.

Essa segunda etapa foi essencial para distinguir problemas de processo e problemas/limitações de sistemas estruturantes, fornecendo uma base sólida para propor soluções alinhadas com as necessidades reais do órgão federais. Com base nos resultados obtidos, foi possível identificar constatações e lacunas que orientaram o refinamento dos requisitos para a integração dos dados, além de identificar os limites técnicos e institucionais que impactam a viabilidade das integrações propostas.

Essa etapa consolidou-se como a mais longa e complexa de todo o ciclo do projeto, marcando o amadurecimento das estratégias e o aprofundamento técnico necessário para a construção de soluções sustentáveis de integração de dados.

Resultados da análise exploratória e viabilidade de integração

Inicialmente, este projeto partiu da hipótese de que seria possível integrar automaticamente os sistemas estruturantes diretamente ao nível dos dados, aproveitando as bases existentes. Essa suposição fundamentava-se no fato de que os sistemas tratam os mesmos objetos centrais (contratos, unidades gestoras e empenhos), o que, em princípio, sugeria uma possibilidade natural de conexão direta entre eles.

Contudo, as análises exploratórias realizadas revelaram uma realidade mais complexa. Por serem geridos e desenvolvidos de forma descentralizada, os sistemas estruturantes carecem de uma chave única que permitisse a integração completa de suas bases de dados. Essa ausência de padronização impõe desafios significativos ao processo de integração, exigindo o desenvolvimento de heurísticas com múltiplas regras distintas para conectar efetivamente as diferentes bases de dados.

Essa complexidade, no entanto, reflete a realidade da maioria dos sistemas distribuídos e dos serviços governamentais. A natureza descentralizada da gestão pública brasileira, embora traga benefícios em termos de autonomia e especialização, inevitavelmente resulta em heterogeneidade de padrões e estruturas de dados.

Como resultado desse processo, foi produzido:

1 Refinamento do diagnóstico técnico detalhado do ecossistema de dados do Ipea, contendo informações sobre as características das bases de dados, suas interdependências e as limitações atuais para integração.

Esse diagnóstico gerou uma visão abrangente das condições técnicas e operacionais que influenciam a gestão e a interoperabilidade das informações.

2 Requisitos refinados para a integração entre as bases de dados, considerando tanto aspectos técnicos quanto institucionais.

Esses requisitos indicam adaptações ou desenvolvimentos necessários para garantir a consistência, segurança e atualização contínua dos dados, reduzindo a dependência de processos manuais e aumentando a confiabilidade das informações utilizadas na tomada de decisão.

3 Documentação e código de software sobre a viabilidade técnica para a integração das bases de dados:

documento que orienta as escolhas futuras em relação à arquitetura de sistemas, recursos tecnológicos e processos organizacionais. Esse parecer indica as opções mais viáveis do ponto de vista técnico, aponta limitações que precisam ser superadas e oferece subsídios concretos para que as próximas etapas avancem de maneira segura, estruturada e eficiente.

Essa etapa foi decisiva para consolidar o entendimento sobre o funcionamento atual do ecossistema de dados e estabelecer uma base técnica sólida para o desenvolvimento de soluções orientadas à integração, qualificação e uso inteligente das informações institucionais.



Como replicar essa etapa?

Para aplicar a metodologia de análise exploratória e estudos de viabilidade de integração de sistemas no seu órgão, recomenda-se seguir as seguintes etapas práticas:

1 Preparação para a Análise Exploratória

- Tenha em mãos os relatórios institucionais relevantes e identifique os principais indicadores e variáveis utilizados pela sua equipe de gestão.
- Verifique a documentação técnica existente sobre os sistemas estruturantes que pretende analisar (SIAFI, SIORG, TransfereGov, Compras.gov.br, SIAPE, etc.).
- Solicite ou garanta o acesso às APIs oficiais de cada sistema estruturante, respeitando as políticas de acesso estabelecidas (abertas ou fechadas).

2 Análise Exploratória

- Acesse as APIs, estude a documentação disponível e teste os endpoints relevantes, validando as informações retornadas na prática.
- Avalie a qualidade dos dados: verifique completude, consistência, precisão e atualização.
- Construa um dicionário de dados, documentando cada variável identificada (nome, formato, periodicidade, descrição e vínculo com os relatórios institucionais).

3 Análise de Integração

- Encontre os identificadores únicos e as chaves relacionais de cada sistema, garantindo a rastreabilidade entre tabelas e relatórios.
- Modele fluxos de integração usando diagramas ou fluxogramas para representar as relações entre as bases de dados.

4 Estudo da Viabilidade de Integração

- Avalie as possíveis integrações entre os sistemas, verificando a compatibilidade de identificadores (chaves) e padrões de nomenclatura.
- Analise as políticas institucionais de acesso e de compartilhamento de dados para cada sistema.

Artefatos produzidos:

- Dicionário de dados de cada base utilizada
- Documentação/código fonte das inferências mapeadas durante a exploração dos dados
- Documentação das chaves para integração das bases de dados, dois a dois
- Evolução dos diagramas com o fluxo de valor do dado
- Documentação como ter acesso às API/Bancos de dados dos sistemas utilizados

CAPÍTULO 3

MODELAGEM, LEVANTAMENTO DE REQUISITOS TÉCNICOS E MAPEAMENTO DE INICIATIVAS



Essa etapa compreende a modelagem e definição de requisitos, que organizam tecnicamente os fluxos e variáveis mapeadas, realizados com base na análise dos processos e das necessidades institucionais. Para isso, foi feito um levantamento das ferramentas existentes na infraestrutura de serviços de TI, da arquitetura técnica e dos dados estratégicos necessários para a solução de qualificação e integração de dados dos sistemas estruturantes. Também foram identificadas as restrições relacionadas à regulamentação, aos níveis de acesso e à governança dos dados.

Além do levantamento das informações técnicas, foi realizado um mapeamento de práticas e soluções para o uso, a qualificação e a integração dos dados dos sistemas estruturantes. Esse processo foi importante para identificar iniciativas que tenham como objetivo integrar e qualificar os dados públicos. Com isso, foi possível definir requisitos técnicos e modelar uma solução mais aderente à realidade do Ipea e, ao mesmo tempo, inovadora, ao propor alternativas ainda não testadas ou adotadas por outros órgãos.

O resultado desta fase materializou-se em três produtos fundamentais:

- **Dicionário de variáveis**
- **Requisitos técnicos para a solução**
- **Mapeamento de iniciativas similares**



A etapa de modelagem, definição de requisitos e levantamento de iniciativas foi estruturada com base em dois mapeamentos: o das ferramentas e práticas internas do Ipea e o de iniciativas similares em outros órgãos públicos. Em relação ao primeiro mapeamento, o objetivo foi identificar as ferramentas, linguagens de programação, práticas adotadas e a equipe responsável pela manutenção dos sistemas e serviços do Ipea. O propósito central dessa etapa foi conhecer a capacidade técnica da instituição no uso, desenvolvimento e absorção de soluções tecnológicas, a fim de garantir que a adoção de novas ferramentas e dependências tecnológicas estivesse alinhada com as práticas, ferramentas, e conhecimento já adotadas pela equipe de desenvolvimento.

Esse percurso foi importante para estabelecer restrições sobre as escolhas tecnológicas e compreender o funcionamento dos serviços de TI do Ipea. Assim, o levantamento abrangeu desde a organização departamental até os processos de manutenção e evolução das soluções hospedadas na instituição, incluindo ferramentas utilizadas e práticas adotadas.

Como ocorre na maioria dos órgãos federais, a equipe de TI concentra seus esforços prioritariamente na sustentação dos serviços utilizados pelo órgão, dedicando menos recursos à manutenção e evolução desses sistemas, cuja responsabilidade é de órgãos centrais. Em termos de ferramentas de engenharia de dados e infraestrutura, o Ipea adota as principais soluções open source utilizadas no mercado: Kubernetes para gestão da infraestrutura e implantação de serviços, Airflow para automações e processamento de dados, e Superset para criação de dashboards de BI.

Com base nesse cenário, foi proposta uma solução de software cuja arquitetura de dados utilizasse essas ferramentas já consolidadas na infraestrutura da instituição, priorizando o desenvolvimento de automações para viabilizar a implantação e sustentação da solução, sem gerar sobrecarga adicional para a equipe.

O segundo mapeamento buscou identificar as iniciativas e soluções, em desenvolvimento ou já implantadas por outros órgãos públicos e empresas estatais, voltadas ao apoio à interoperabilidade, ao uso de dados e à gestão estratégica das informações nos sistemas estruturantes. O interesse central desse processo de *benchmarking* foi compreender como essas soluções similares poderia ser aproveitadas para fortalecer o reuso e a colaboração, acelerando o amadurecimento da proposta desenvolvida no âmbito da parceria UnB e Ipea, evitando esforços duplicados.

De forma complementar, esse levantamento também permitiu identificar ferramentas e plataformas, processos e linguagens de programação adotados por outros órgãos, contribuindo para a formação de um panorama mais amplo sobre as ações e experiências da gestão orientada por dados no setor público.

Para mapear as soluções de integração e qualificação de dados dos sistemas estruturantes existentes ou já realizadas por outros órgãos, foram feitos contatos diretos com a Serpro e a Dataprev. Além de levantar o que estava sendo desenvolvido, buscou-se compreender e discutir os sistemas sob responsabilidade dessas empresas. A partir do diálogo com o corpo técnico, constatou-se a inexistência de uma previsão ou arquitetura desenhada para viabilizar a interoperabilidade entre os sistemas estruturantes.

Cabe ressaltar que as conversas com os profissionais envolvidos na construção dos sistemas estruturantes foram valiosas para a compreensão, em alto nível, da arquitetura dessas soluções. Um aspecto técnico relevante identificado foi o uso do protocolo SOAP (Simple Object Access Protocol) como padrão para a troca de mensagens com as APIs dos sistemas estruturantes.

Além disso, outros órgãos públicos que pudessem ter atuação mais avançada em engenharia de dados foram procurados. Destaca-se, nesse processo, o diálogo contínuo com o Ministério da Gestão e Inovação em Serviços Públicos (MGI), que apresentou os avanços nas análises de dados e integração do SIORG, bem como os desafios enfrentados para a interoperabilidade entre os sistemas estruturantes.

Ao final desse processo exploratório, concluiu-se que ainda não existem iniciativas em larga escala sendo promovidas de forma consolidada e ativa para resolver as lacunas de interoperabilidade entre os sistemas estruturantes. Essa constatação reforçou o caráter inovador da solução desenvolvida, o GovHub, que pode ser aplicada de forma adaptável aos diferentes contextos e necessidades dos distintos órgãos governamentais.

Em ambos os mapeamentos, foram observados os seguintes pontos:

- **Ferramentas adotadas** - Tecnologias e plataformas utilizadas para o processamento, armazenamento e visualização de dados, incluindo bancos de dados, linguagens de programação, frameworks e software de análise.
- **Dependências** - Infraestrutura, bibliotecas e serviços externos necessários para o funcionamento adequado da solução, como APIs de terceiros, conectores de bancos de dados e requisitos de hardware.
- **Exemplo de relatórios, e visualizações gerados manualmente** - Modelos de relatórios e dashboards utilizados atualmente, que podem servir de referência para a automação e padronização na nova solução.
- **Documentação técnica** - Registro detalhado da solução desenvolvida, incluindo a arquitetura, a estrutura de dados, os procedimentos de integração, o uso das ferramentas e as instruções para manutenção e evolução do sistema.

Desafios técnicos de acesso aos dados

Os desafios de acesso aos dados não se restringem ao diálogo institucional, manifestando-se também nos próprios fluxos técnicos de acesso aos dados. Por exemplo, conforme detalhado no relatório de acesso às APIs (disponível em: www.gob-hub.io), mesmo quando as APIs estão formalmente disponibilizadas, o processo de contratação costuma ser burocrático e envolve diversas etapas: aquisição de certificados digitais, validação de credenciais, preenchimento de formulários e prazos extensos para liberação do acesso.

Para ilustrar essa complexidade, o acesso aos dados do SIAFI via API exigiu a contratação do serviço junto ao Serpro, a emissão de certificado digital e a designação de responsáveis técnicos vinculados ao órgão solicitante. Além disso, enfrentam-se obstáculos como documentação desatualizada e suporte técnico limitado, resultando em um processo longo até a liberação efetiva do acesso.

Adicionalmente, mapeou-se que os dados disponibilizados pelas APIs variam de acordo com o módulo da API, exigindo conhecimento técnico específico e infraestrutura adequada para consumi-los de forma confiável. Essa fragmentação de serviços e a falta de documentação adequada impactam no cronograma de desenvolvimento das soluções de software, que depende dessas integrações, dificultando a criação de soluções interoperáveis de forma ágil e segura. Diante desse cenário, a equipe do projeto avançou na documentação de todo o processo, com o objetivo de facilitar futuras integrações e garantir maior segurança técnica.

Apesar de todas essas barreiras, a base técnica das soluções, considerando as ferramentas e a arquitetura utilizada pelo MGI e pelo Ipea, é semelhante e já consolidada, o que facilita a adoção de soluções interoperáveis do ponto de vista tecnológico. Essa convergência técnica foi considerada, desde o início, como uma prioridade para orientar as conversas com os principais atores e viabilizar sua adoção futura.

Resultados da etapa de modelagem, requisitos e iniciativas

Esta etapa, que combinou diferentes metodologias e processo, foi determinante para o desenvolvimento da solução GovHub. A modelagem e a definição dos requisitos partiram da escuta e observação: reconhecer o que já existe, como é utilizado e o que ainda permanece sem resposta clara.

Esse conhecimento permitiu construir requisitos que não só resolveram problemas específicos do Ipea, mas também se mostraram aplicáveis a outros órgãos federais.

Como resultado desse processo, foram produzidos os seguintes produtos:

1 Ferramentas propostas:

Com base nos levantamentos obtidos nesta etapa, o modelo arquitetural proposto adota ferramentas tecnológicas baseadas em soluções código aberto, incluindo:

- **Apache Airflow** para a orquestração de pipelines de dados;
- **DBT** para transformação e modelagem de informações;
- **Apache Superset** para visualização e exploração de dados;
- **PostgreSQL** como banco de dados relacional;
- **Docker** para containerização e implantação.

Essas ferramentas proporcionam maior flexibilidade, escalabilidade e integração com diferentes sistemas governamentais. Além disso, já são amplamente conhecidas pela equipe técnica do Ipea e contam com comunidades de desenvolvimento ativas.

2 Requisitos do sistema:

A partir das reflexões e observações realizadas, foram formulados quatro principais requisitos que nortearam a construção da solução:

- **Requisito 1: Capacidade de integração com sistemas estruturantes:** Conectar aos sistemas estruturantes (como SIAFI, SIORG, SIAPE, Compras.gov.br e TransfereGov) sem exigir a substituição ou modificação desses sistemas, garantindo a continuidade operacional e respeitando as particularidades de cada órgão.
- **Requisito 2: Flexibilidade para configurar indicadores e visualizações:** Permitir o uso de diferentes ferramentas de análises e dashboards, utilizando os dados já existentes nos sistemas estruturantes, de modo que cada setor possa responder às suas próprias demandas de gestão.
- **Requisito 3: Governança e rastreabilidade dos dados:** Garantir a rastreabilidade e política de qualidade dos dados, com histórico de transformações e políticas de modificações. Isso garante a confiabilidade das informações para tomada de decisões, além de facilitar auditorias e revisões técnicas.
- **Requisito 4: Sustentabilidade e criação colaborativa:** Adotar tecnologias abertas (*software* livre) e práticas alinhadas à realidade técnica da equipe do órgão. Essa estratégia facilita a adoção e adaptação e incentiva a colaboração

entre órgãos na evolução da plataforma. Isso amplia o potencial de reutilização e economiza recursos públicos ao evitar retrabalho.

Este capítulo consolidou a base conceitual do método adotado: partir da realidade institucional, identificar os pontos de fricção na circulação da informação e estruturar soluções conectadas aos processos reais. Essa abordagem sustenta a proposta do GovHub, detalhada na sequência, como uma solução que não substitui os sistemas existentes, mas articula o que já está em uso, promovendo a reutilização inteligente e o uso qualificado dos dados, sempre com foco na visão integrada.

Compreender os fluxos, as perguntas da gestão e os dados disponíveis é o que dá sentido à solução técnica proposta. Além disso, também é o que garante que qualquer automação ou visualização futura responda efetivamente às necessidades das pessoas envolvidas na gestão pública. Ao final desta etapa, destaca-se que o impacto da proposta não reside apenas na ferramenta desenvolvida, mas também no próprio processo de construção — pautado pela convicção de que dados são ativos vivos, atravessam setores e refletem escolhas institucionais.





Como replicar essa etapa?

Fase 1: Diagnóstico e Mapeamento Interno

1 Sistematização do diagnóstico técnico

- Compile os dados dos capítulos anteriores em um documento único;
- Detalhe a infraestrutura atual (servidores, redes, capacidade);
- Mapeie as ferramentas tecnológicas em uso;
- Identifique dependências institucionais e parcerias existentes;
- Documente limitações e gargalos conhecidos.

2 Mapeamento de ferramentas existentes

- Liste todos os bancos de dados utilizados (tipo, versão, finalidade);
- Identifique as linguagens de programação dominantes na equipe;
- Catalogue frameworks e bibliotecas em uso no órgão;
- Mapeie os software de análise e visualização disponíveis;
- Avalie o nível de conhecimento da equipe em cada tecnologia.

3 Análise de relatórios e dashboards existentes

- Colete exemplos dos relatórios mais utilizados;
- Identifique dashboards críticos para a gestão;
- Analise a frequência de uso e a audiência de cada visualização;
- Documente os pontos fortes e as limitações das soluções atuais;
- Entreviste usuários finais sobre necessidades não atendidas.

Fase 2: Definição de Requisitos e Arquitetura

1 Consolidação de Requisitos Técnicos

Prioridades definidas:

- **Integração:** Compatibilidade com sistemas existentes;
- **Flexibilidade:** Configuração ágil de indicadores e visualizações;
- **Governança:** Rastreabilidade e qualidade dos dados;
- **Sustentabilidade:** Uso de tecnologias abertas e comunidade ativa.

Ações práticas:

- Defina requisitos funcionais específicos;
- Estabeleça requisitos não funcionais (performance, segurança);
- Priorize os requisitos por importância e viabilidade;
- Valide os requisitos com stakeholders técnicos e de negócio.

2 Modelagem da Arquitetura Inicial

Stack tecnológico baseado em soluções livres:

- **Apache Airflow:** Orquestração de pipelines;
- **DBT:** Transformação e modelagem de dados;
- **PostgreSQL:** Banco de dados relacional;
- **Docker:** Containerização e deploy.
- **Apache Superset:** Visualização e exploração;

Ações práticas:

- Desenhe o diagrama da arquitetura proposta;
- Valide a compatibilidade com a infraestrutura existente;
- Estime os recursos necessários para a implementação;
- Planeje a estratégia de migração gradual.

Fase 3: Análise de iniciativas e Ecossistema Externo

1 Mapeamento de soluções disponíveis

- Pesquise APIs governamentais disponíveis (ex.: dados.gov.br);
- Mapeie plataformas de integração existentes;
- Investigate serviços de empresas estatais (Serpro, Dataprev);
- Documente limitações e restrições de cada serviço;
- Avalie custos e modelos de contratação.

Colete exemplos de relatórios e dashboards existentes:

Analise como os relatórios e visualizações são atualmente gerados e utilizados pelas equipes técnicas e de gestão, destacando pontos fortes e limitações.

Defina e consolide requisitos técnicos:

Sistematize os requisitos técnicos e funcionais da solução, priorizando:

- Capacidade de integração com os sistemas existentes;
- Governança e rastreabilidade dos dados;
- Flexibilidade para configurar indicadores e visualizações;
- Sustentabilidade e uso de tecnologias abertas.

Inicie a modelagem da arquitetura:

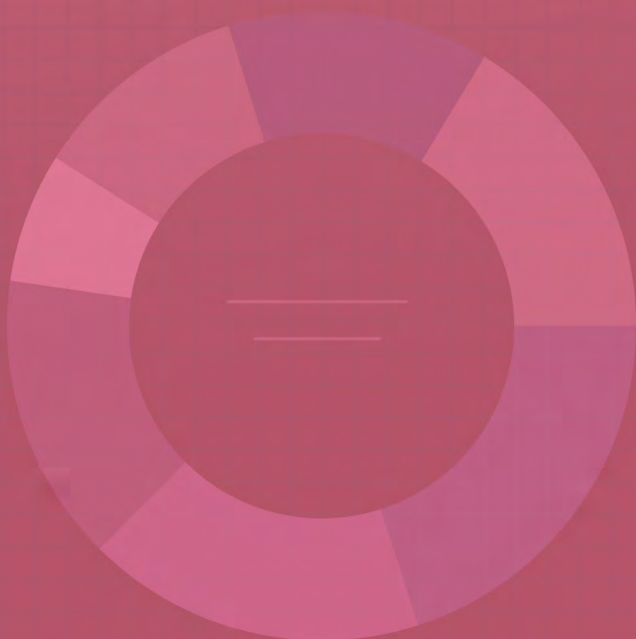
Adote ferramentas de código aberto que permitam orquestração, transformação, visualização e armazenamento de dados de forma escalável e segura (por exemplo: Apache Airflow, DBT, Apache Superset, PostgreSQL e Docker).

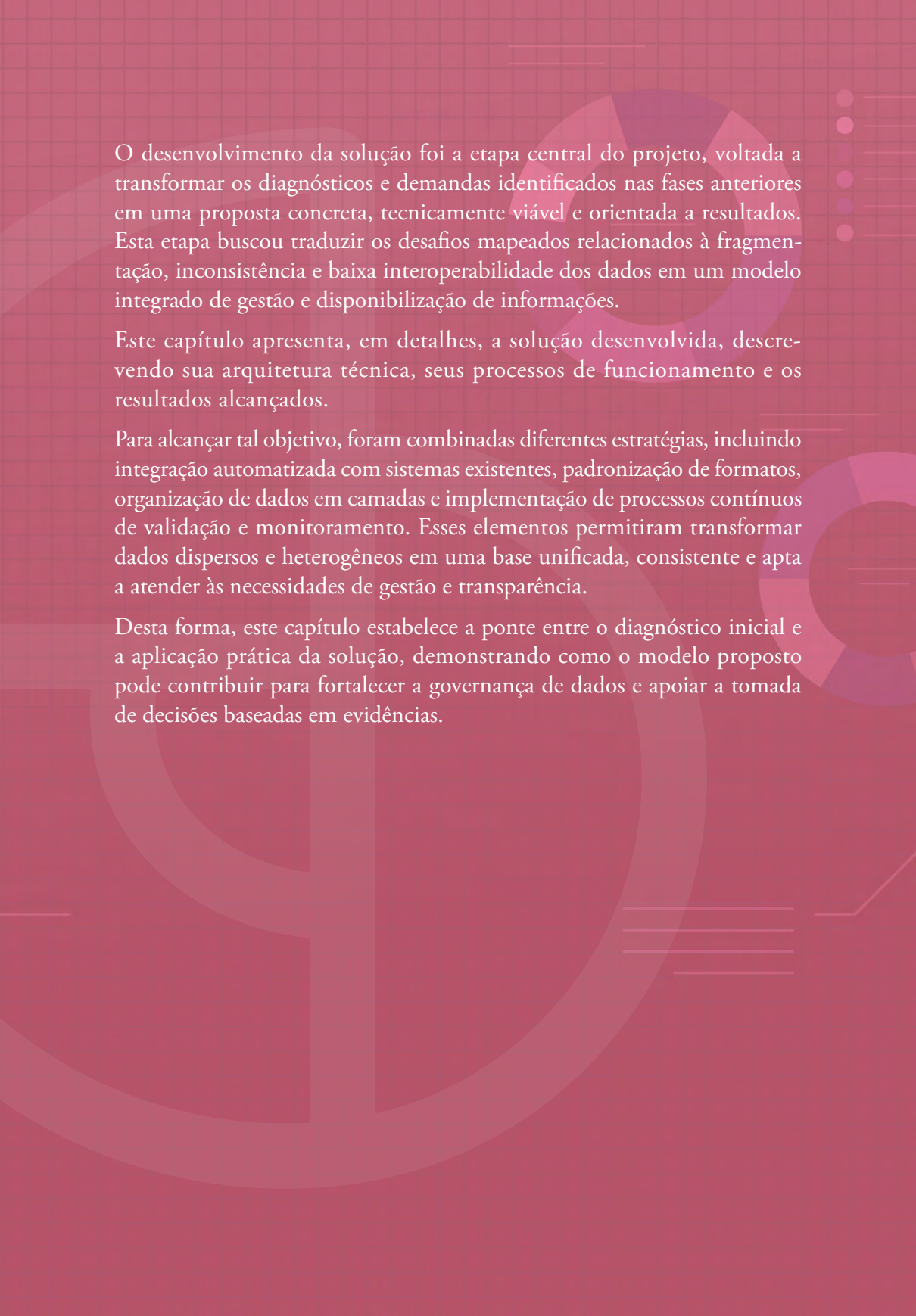
Artefatos produzidos:

- Documento de requisitos de qualificação dos dados;
- Documentação da versão inicial da arquitetura;
- Documento de requisitos para integração das bases de dados;
- Proposta de ferramentas e tecnologias adotadas.

CAPÍTULO 4

GOVHUB: A SOLUÇÃO PARA UMA GESTÃO PÚBLICA ORIENTADA POR DADOS





O desenvolvimento da solução foi a etapa central do projeto, voltada a transformar os diagnósticos e demandas identificados nas fases anteriores em uma proposta concreta, tecnicamente viável e orientada a resultados. Esta etapa buscou traduzir os desafios mapeados relacionados à fragmentação, inconsistência e baixa interoperabilidade dos dados em um modelo integrado de gestão e disponibilização de informações.

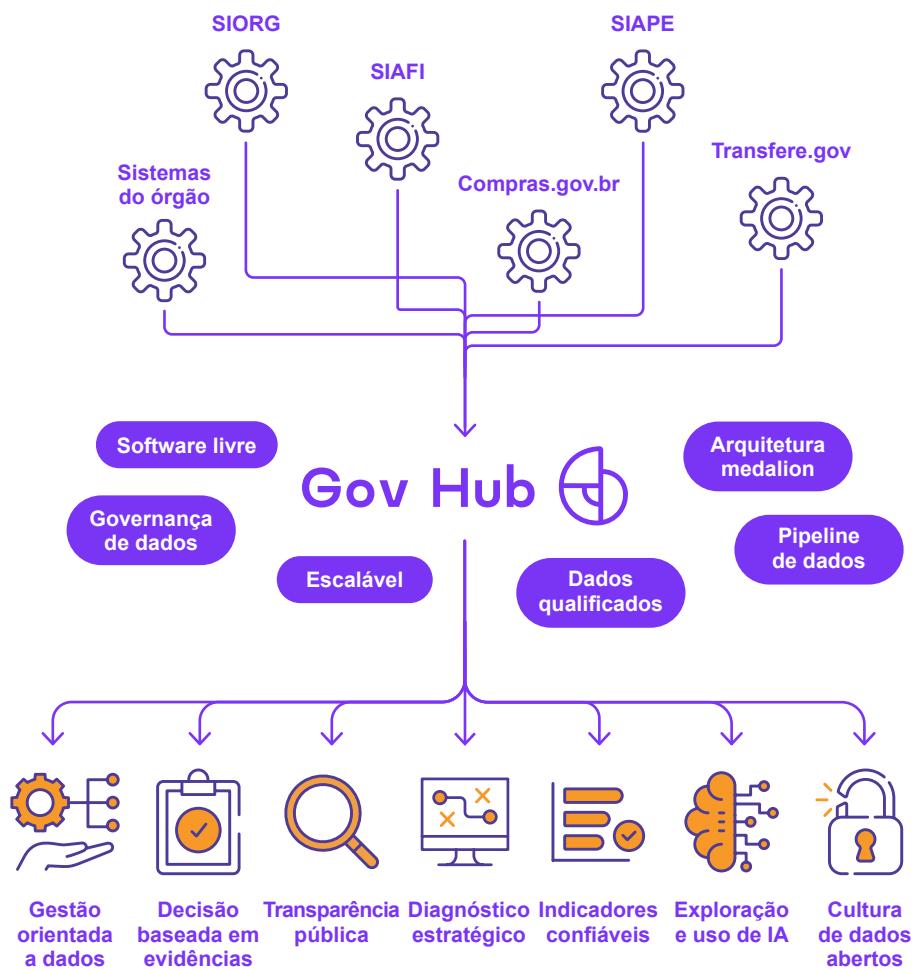
Este capítulo apresenta, em detalhes, a solução desenvolvida, descrevendo sua arquitetura técnica, seus processos de funcionamento e os resultados alcançados.

Para alcançar tal objetivo, foram combinadas diferentes estratégias, incluindo integração automatizada com sistemas existentes, padronização de formatos, organização de dados em camadas e implementação de processos contínuos de validação e monitoramento. Esses elementos permitiram transformar dados dispersos e heterogêneos em uma base unificada, consistente e apta a atender às necessidades de gestão e transparência.

Desta forma, este capítulo estabelece a ponte entre o diagnóstico inicial e a aplicação prática da solução, demonstrando como o modelo proposto pode contribuir para fortalecer a governança de dados e apoiar a tomada de decisões baseadas em evidências.

FIGURA 11

Visão geral da solução Gov Hub



O GovHub surge como uma resposta prática ao desafio da fragmentação e inconsistência de dados. Mais do que uma ferramenta, trata-se de uma infraestrutura de dados. Sua função é atuar como um *hub* inteligente: um ponto central que conecta, organiza e compartilha dados provenientes de diferentes sistemas. Com base em coletas automatizadas, tratamento técnico e organização por camadas, a plataforma transforma dados brutos e fragmentados em informações úteis e prontas para uso.



A Metáfora da Netflix

Imagine a Netflix dos dados: assim como você acessa filmes de diferentes estúdios em uma única plataforma, o GovHub permite que você acesse dados de diferentes órgãos e sistemas estruturantes a partir de uma interface unificada. Não é necessário saber se a informação vem do sistema de gestão de pessoas, execução financeira ou compras - tudo está organizado e disponível no mesmo lugar.

Catálogo Descentralizado: Assim como a Netflix não produz todo o conteúdo que oferece (há filmes da Disney, séries da BBC, documentários independentes, etc.), o GovHub não produz os dados brutos, mas centraliza o acesso em um só lugar. Cada “domínio” (assim como cada estúdio) mantém a propriedade de seus dados, mas os disponibiliza por meio da plataforma.

Experiência Unificada: Você não precisa saber se um filme é da Warner, Netflix Originals ou Paramount — a interface é a mesma. No GovHub, é possível acessar dados de pessoal, financeiros ou organizacionais pela mesma interface, sem se preocupar com onde estão armazenados.

Qualidade Garantida: Assim como a Netflix tem padrões de qualidade para o que entra no catálogo (resolução, áudio, legendas), o GovHub garante que os dados sigam padrões de qualidade, governança e formato.

Acesso Personalizado: Diferentes usuários veem conteúdos distintos na Netflix (ficção, drama, por região). No GovHub, diferentes usuários acessam diferentes conjuntos de dados, conforme suas permissões e necessidades.

Busca e Descoberta: Como encontrar aquele filme que você não lembra o nome? A Netflix facilita isso. O GovHub facilita encontrar “aquele dado de licitação do trimestre passado”, mesmo sem saber exatamente onde está.

Essa metáfora funciona porque ambos transformam complexidade em simplicidade: a Netflix esconde a complexidade de negociar com centenas de estúdios, enquanto o GovHub esconde a complexidade de integrar dezenas de sistemas diferentes.

Essa solução foi pensada para apoiar gestores públicos em aproveitar melhor os dados que já possuem. Além disso, torna-se mais fácil garantir a rastreabilidade entre bases, aplicar regras de negócio com clareza, compartilhar indicadores de forma segura e até publicar dados abertos com qualidade, sem tirar a autonomia do órgão.

Arquitetura Data Mesh e Medallion

A arquitetura do GovHub estabelece regras claras para o ciclo de vida da engenharia de dados: partindo da extração automatizada das bases originais, prosseguindo com as transformações e cruzamentos para qualificação da informação e, por conseguinte, disponibilizando a informação final para uso. Baseada na **Arquitetura Medallion**, esse ciclo garante qualidade e confiabilidade em cada etapa, agregando serviços subjacentes de segurança da informação, governança e integridade de dados, caracterização de inconsistências, dentre outros.

A arquitetura adota o paradigma Lakehouse, estruturado em quatro camadas de processamento de dados. Essa abordagem fornece um alicerce técnico e organizacional para transformar dados brutos em inteligência acionável, com controle, escalabilidade e governança descentralizada.

A Arquitetura Medallion estrutura os dados em camadas progressivas de maturidade: *raw*, *bronze*, *silver* e *gold*.



Camada Raw - Extração

- Coleta automatizada dos dados nas bases originais;
- Preservação da estrutura e formato nativos;



Camada Bronze - Padronização

- Limpeza e padronização dos dados coletados;
- Criação do histórico completo de mudanças.



Camada Silver - Integração

- Aplicação de regras de qualidade e validação;
- Integração entre diferentes bases de dados.



Camada Gold - Qualificação e disponibilização

- Criação de visões consolidadas e prontas para análise;
- Disponibilização dos dados por meio de APIs e interfaces amigáveis;
- Fonte de dados para análises (dashboards de BI - Business Intelligence), algoritmos de machine learning, dados abertos e transparência ativa.

Essa organização em camadas viabiliza a criação de uma base sólida e reutilizável de dados, para diferentes contextos institucionais. Trata-se de um repositório versionado e governável, com informações qualificadas e integradas, com potencial para ações diversas: análises manuais, automações, uso de bases compartilhadas e

em modelos preditivos, transparência ativa, publicação em repositórios de dados abertos e relatórios oficiais.

O versionamento entre camadas garante que cada transformação feita sobre os dados possa ser auditada, revertida ou evoluída, sem comprometer o histórico. Nesse sentido, mesmo que os dados de uma camada *Gold* parecem semelhantes às camadas anteriores, sua função é representar a versão curada, confiável e validada por regras de negócio, pronta para consumo interno ou disponibilização para o público externo. Essa organização evita divergências de informações entre relatórios, divergência de lógicas de tratamento da informação e interpretações inconsistentes.

Ao adotar a arquitetura Medallion, o GovHub estabelece um caminho escalável para a governança descentralizada dos dados, o reuso estruturado e a multiplicidade de aplicações — desde painéis de BI até soluções em inteligência artificial. A plataforma ganha autonomia técnica e confiança institucional para ser usada, expandida e compartilhada em diferentes contextos da gestão pública.

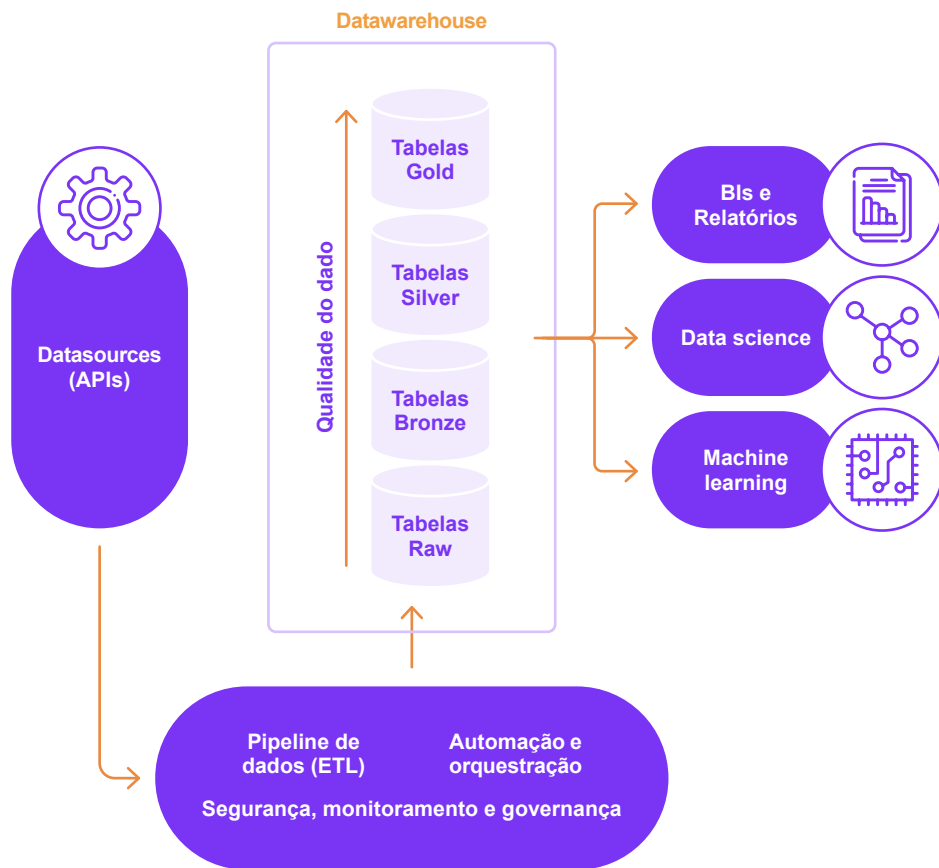
O GovHub oferece uma infraestrutura de dados pública que é:

- **Adaptável**, com pipelines automatizados e escaláveis;
- **Confiável**, com versionamento e rastreabilidade entre camadas;
- **Reutilizável**, para múltiplos fins institucionais e públicos;
- **Aberta**, com possibilidade de expansão por qualquer órgão.

A arquitetura do projeto articula os conceitos apresentados em uma solução concreta e escalável, refletida na forma como os dados são coletados, organizados e utilizados. A Figura 12, apresenta uma visão estruturada da arquitetura adotada. Esta imagem traduz com fidelidade a organização do GovHub: desde a ingestão de dados brutos oriundos de APIs oficiais dos sistemas estruturantes até sua transformação em informações qualificadas, disponíveis para análise, automação, aprendizado de máquina e dados abertos.

FIGURA 12

Visão geral da arquitetura do Gov Hub



Na sequência, são apresentados os detalhes de cada uma das etapas do ciclo de vida dos dados na plataforma: a coleta automatizada, o processo de qualificação em camadas e as interfaces de visualização e reuso estratégico.

1 Coleta automatizada de dados

A integração do GovHub com os sistemas estruturantes é realizada por meio das APIs oficiais dos sistemas estruturantes, respeitando os padrões e protocolos definidos por cada plataforma. A coleta ocorre de forma automatizada, o que reduz a dependência de processos manuais e garante a atualização periódica das informações. Esse processo é operado por pipelines de dados desenvolvidos com ferramentas como Apache Airflow e scripts em Python, garantindo não apenas transparência e rastreabilidade, mas também auditabilidade e robustez técnica na manutenção do fluxo contínuo de ingestão.

2 Qualificação e estruturação

A partir da ingestão e armazenamento, os dados brutos passam por processos de limpeza, padronização e organização em camadas progressivas (raw, bronze, silver e gold), conforme o modelo da Arquitetura Medallion. Essa estrutura permite rastrear a origem dos dados, aplicar regras de negócio de forma transparente e gerar tabelas otimizadas para análise e uso institucional.

O grande diferencial está na capacidade de promover “conversas” entre sistemas que antes não se comunicavam. Por exemplo: é possível relacionar um contrato registrado no sistema Compras.gov.br com os pagamentos correspondentes registrados no SIAFI. Isso permite obter uma visão integrada e completa de todo o ciclo do processo.

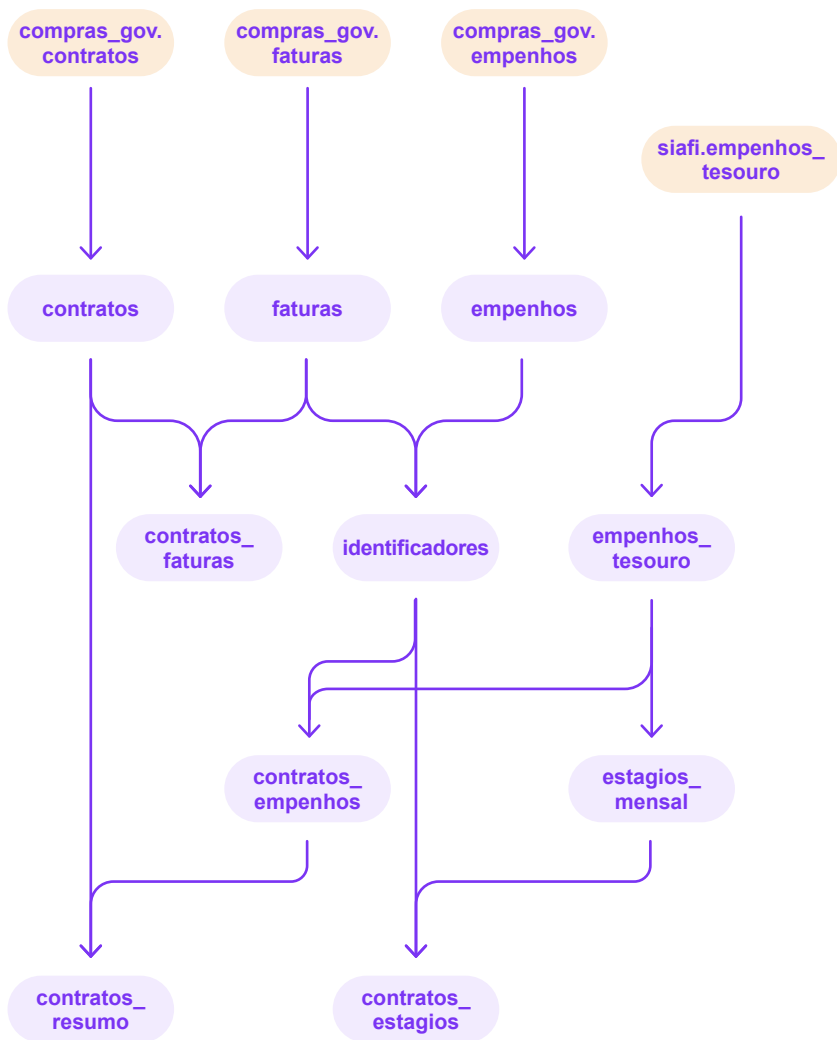
Para fazer essas transformações, foi utilizado o processo denominado ELT, que significa:

- **Extrair:** pegar os dados dos sistemas originais;
- **Load (Carregar):** colocar esses dados no nosso sistema;
- **Transformar:** limpar, organizar e conectar as informações.

É como uma esteira de fábrica automatizada que coleta dados brutos, os carrega no sistema e, em seguida, os transforma em informações úteis. Para isso, foi utilizada a ferramenta DBT (Data Build Tool), que documenta automaticamente cada mudança feita nos dados. Assim, tem-se um histórico completo de todas as alterações realizadas, o que garante total transparência no processo. A Figura 13 exemplifica o processo de transformações implementadas no DBT.

FIGURA 13

Representação de ETL dos dados de contratos no DBT



3 Visualização e uso estratégico

A camada final da arquitetura é dedicada à análise e ao uso estratégico dos dados. Para isso, o GovHub disponibiliza dashboards interativos desenvolvidos com Apache Superset, permitindo que diferentes perfis de usuários acessem e explorem as informações de forma intuitiva, sem a necessidade de conhecimento técnico avançado. As visualizações são personalizáveis e estão organizadas por áreas temáticas, abrangendo a execução de TEDs, monitoramento de contratos, gestão de recursos humanos e planejamento orçamentário.

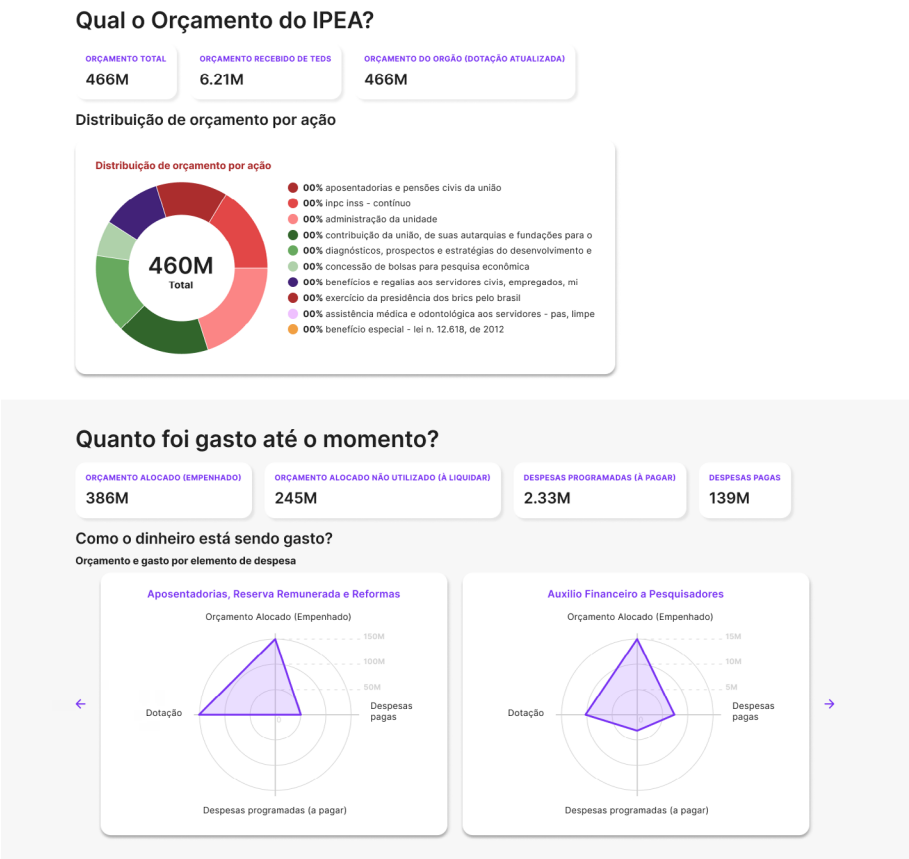
Além da interface visual, todos os dados qualificados, existentes na podem ser exportados ou acessados diretamente por meio de uma API, o que amplia suas possibilidades de uso por outras ferramentas institucionais e iniciativas externas. Neste particular, cabe destacar as possibilidades de integrar serviços voltados à transparência ativa do órgão, publicação de dados abertos, integrações automatizadas ou soluções baseadas em inteligência artificial para apoio à gestão ou predição de comportamentos futuros.

Análises Viabilizadas pelo GovHub

O GovHub é uma solução pronta para uso que oferece **análises essenciais para a gestão pública**. Desenvolvido inicialmente para atender as necessidades da Diretoria de Desenvolvimento Institucional do Ipea, o sistema pode ser facilmente adaptado para qualquer órgão público.

Na camada estratégica, o GovHub apresenta os **indicadores-chave do Ipea**, oferecendo os números mais importantes para subsidiar o acompanhamento e tomada de decisão. Isso inclui, por exemplo, a **situação orçamentária** completa, com informações sobre as despesas liquidadas, o que está planejado e o andamento do empenho dos recursos. Além disso, fornece um **panorama da força de trabalho**, com informações consolidadas sobre o quadro de servidores ativos, aposentados e terceirizados, situações envolvendo afastamento e férias, dentre outros.

FIGURA 14
Exemplo de dashboard com dados estratégicos



No nível operacional, o sistema oferece acompanhamento detalhado de **contratos administrativos**, garantindo controle eficiente de contratos e fornecedores. Também gerencia **TEDs (Termos de Execução Descentralizada)**, facilitando o controle de convênios e parcerias. Por fim, disponibiliza **análises aprofundadas de gestão de pessoas**, com informações detalhadas sobre recursos humanos.



Problema 1:

Durante a análise dos sistemas Compras.gov.br e SIAFI, foi identificado que muitos contratos permanecem abertos mesmo após o encerramento da execução financeira. Os cronogramas de pagamento não são atualizados conforme a execução real, e há divergências frequentes entre os valores registrados nos dois sistemas.



Solução:

Esses problemas, que antes exigiam conciliações manuais e interpretativas, agora são resolvidos automaticamente pelo GovHub, por meio do cruzamento estruturado entre os registros do Compras.gov.br e os dados de execução financeira disponíveis no SIAFI. Com isso, a plataforma consegue identificar, por exemplo, contratos que seguem ativos mesmo após o pagamento integral de suas faturas, apontando potenciais falhas operacionais na baixa contratual. Também é possível analisar a aderência entre o cronograma planejado e a execução efetiva das faturas, revelando distorções que afetam o planejamento de desembolsos.



Problema 2:

Outras constatações, como a falta de vínculos consistentes entre faturas, cronogramas e empenhos, ou a ausência de identificadores únicos entre os sistemas, comprometem a rastreabilidade documental e inviabilizam o cruzamento automatizado dos dados. O GovHub resolve essas lacunas por meio da criação de regras de correspondência e tabelas intermediárias, que permitem reconstruir essas ligações de forma lógica e auditável.



Solução:

Os dashboards implementados, conforme exemplificado nas Figuras 14, permitem identificar inconsistências contratuais, acompanhar a execução orçamentária em tempo real, e compreender padrões de contratação por unidade, objeto ou frequência. Com isso, gestores ganham uma base sólida para decisões estratégicas de planejamento, contingenciamento e aprimoramento dos processos internos.

FIGURA 15

Exemplo de visualização dos dados de acompanhamento orçamentário do contrato.

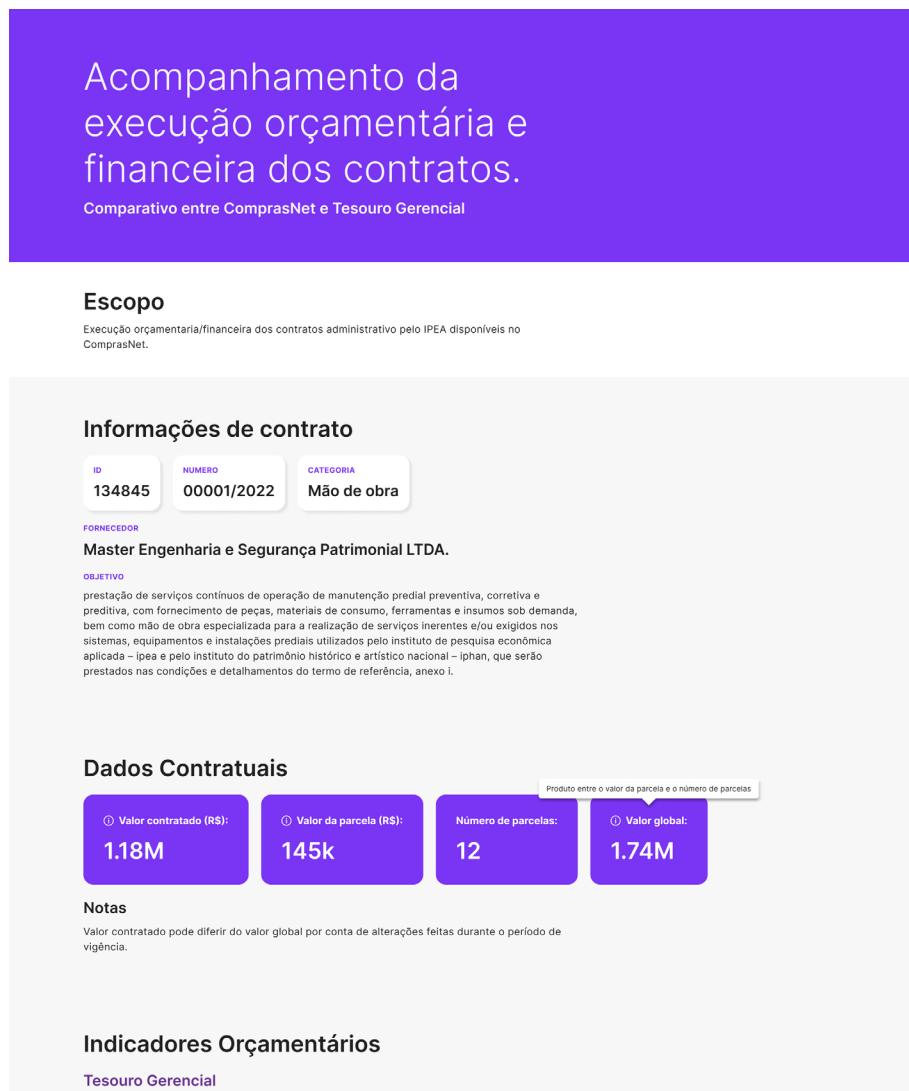
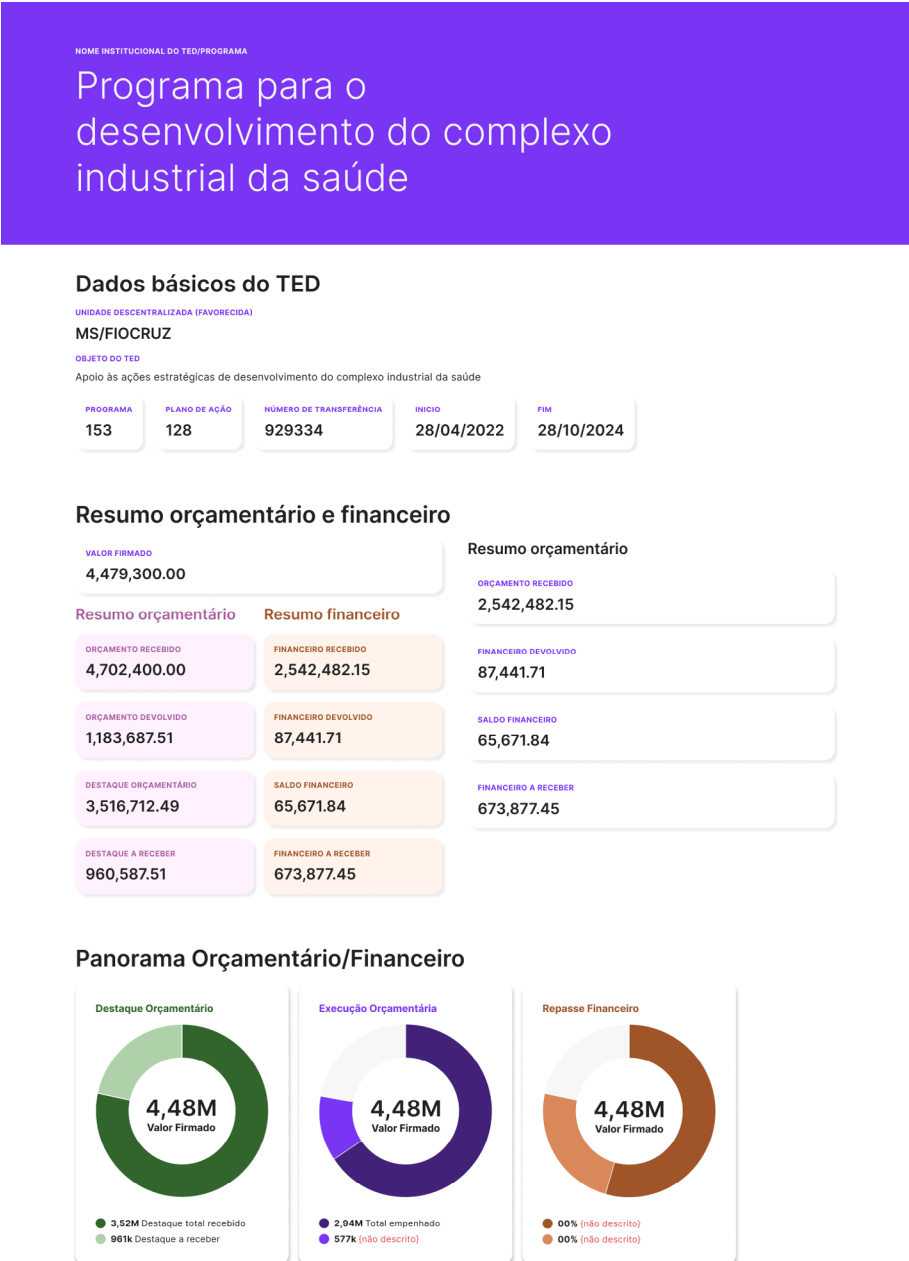


FIGURA 16
Exemplo de visualização dos dados de acompanhamento da execução orçamentária e financeira dos TEDs





Problema 3:

A análise dos dados relacionados aos TEDs revelou graves limitações na capacidade de rastrear a execução dos recursos descentralizados. O TransfereGov, sistema responsável por formalizar os TEDs, não disponibiliza informações sistematizadas sobre a execução orçamentária e financeira desses instrumentos. Já o SIAFI, onde os registros de empenho, liquidação, pagamento e devolução são feitos, não possui identificadores consistentes que permitam relacionar automaticamente essas operações ao TED correspondente.

Como resultado, os gestores públicos têm dificuldade para acompanhar a execução dos instrumentos de descentralização e precisam recorrer a cruzamentos manuais entre sistemas, o que compromete a agilidade, aumenta o risco de erro e prejudica a transparência e a prestação de contas.



Solução:

O GovHub resolve esse problema ao integrar os dados do TransfereGov e do SIAFI por meio de regras de correspondência, permitindo rastrear o ciclo completo do recurso descentralizado — desde a assinatura do TED até a sua aplicação final. Conforme evidenciado na Figura 15, a plataforma relaciona os valores firmados com os registros efetivos de disponibilização orçamentária, empenho, liquidação e pagamento, consolidando essas informações em dashboards interativos que oferecem uma visão clara e confiável da execução.

Além disso, o GovHub permite identificar lacunas específicas que dificultam o monitoramento dos TEDs, tais como a ausência de visualizações consolidadas dos repasses “ainda a receber”, a fragmentação dos dados sobre notas de crédito e programações financeiras e a falta de integração lógica entre a programação orçamentária e execução financeira. Ao consolidar esses registros e permitir seu acompanhamento em tempo real, a plataforma fortalece a capacidade dos órgãos públicos de planejar, reprogramar e prestar contas de forma segura e eficaz.



Problema 4:

Embora existam bases consolidadas sobre vínculos e remunerações, as informações estão muitas vezes desorganizadas, pouco integradas e limitadas em termos analíticos, dificultando uma visão ampla e estratégica sobre o quadro de servidores. Faltam, por exemplo, mecanismos para mapear ocupações por função ou unidade, identificar padrões de afastamentos, visualizar projeções de aposentadorias ou monitorar cargos vagos e acumulados de forma confiável.



Solução:

O GovHub responde a essas lacunas ao consolidar e qualificar dados de pessoal de maneira estruturada, criando uma base confiável para compreender a composição, movimentação e evolução do quadro funcional. A plataforma permite analisar a distribuição geográfica, funcional e sociodemográfica dos servidores, combinando variáveis como idade, gênero, localidade, função exercida e tipo de vínculo. Isso possibilita leituras aprofundadas sobre quem ocupa cargos de chefia, quais perfis apresentam maior incidência de licenças e como se comportam as movimentações internas ao longo do tempo.

FIGURA 17
Exemplo de visualização dos dados gerais de pessoal, com o perfil dos funcionários do Ipea

Contratos Comprasnet
x Tesouro Gerencial
2025_06_10T12

Visão geral

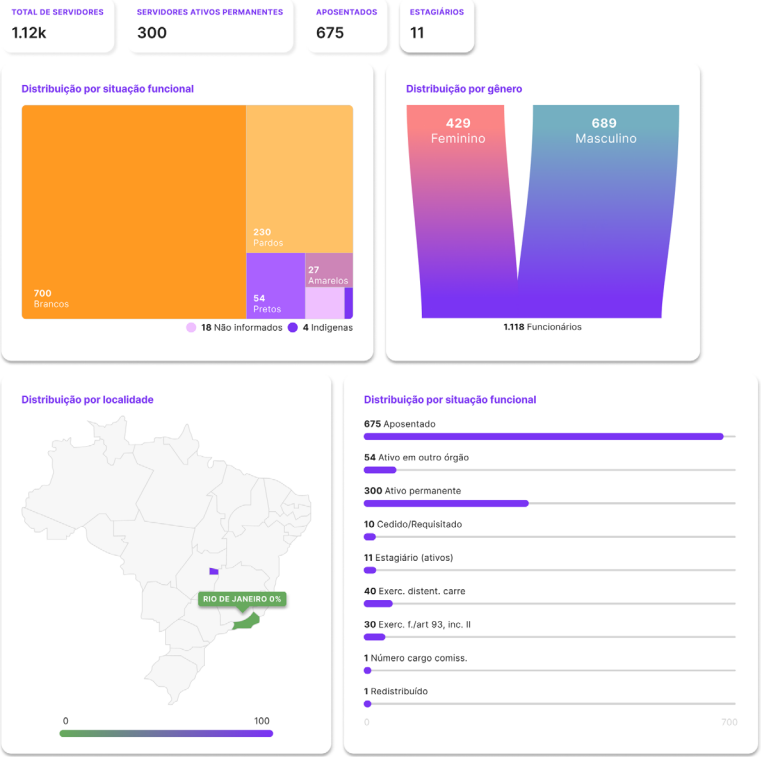


Tabela resumo de servidores
por cargo, gênero, situação funcional e localidade

nome ▾	valor em real ▾	%	cidade	estado	total de servidores
Técnico de planejamento e pesquisa	Masculino	Ativo Permanente	Brasília	DF	185
Técnico de planejamento e pesquisa	Masculino	Aposentado	Brasília	DF	172
Técnico de planejamento e pesquisa	Feminino	Ativo Permanente	Brasília	DF	74

Além disso, os dashboards apoiam análises operacionais relevantes, como o acompanhamento de cargos vagos e ocupados, a detecção de acúmulo de funções e a projeção de desligamentos por aposentadoria. Essas análises são fundamentais para planejar concursos, reorganizar equipes, priorizar nomeações e alinhar a gestão de pessoal às metas institucionais.

Todos os dados são qualificados automaticamente em pipelines que garantem padronização, rastreabilidade e atualização periódica, permitindo que os painéis estejam sempre alinhados à realidade do órgão. As visualizações podem ser filtradas, exportadas e personalizadas conforme o interesse de cada gestor e a arquitetura aberta da plataforma permite a criação de novos painéis a partir das mesmas bases, expandindo a capacidade de resposta conforme os desafios de cada instituição.

Aprendizados

O primeiro desafio, e principal motivador deste projeto, é que **os sistemas estruturantes não foram projetados para interoperabilidade entre si**. Isso se reflete na ausência de uma chave única de integração entre as bases de dados, o que torna a conexão entre sistemas um desafio estrutural. Para contornar essa limitação, foram criadas regras empíricas e constatações de integração, baseadas na observação de como os dados são registrados e utilizados no dia a dia.

Essas constatações partem da observação do padrão de uso dos sistemas, sendo continuamente refinadas conforme novas exceções são descobertas. Esse processo retroalimenta o aprimoramento tanto dos sistemas quanto dos processos internos. Quando uma constatação se torna robusta o suficiente, ela pode embasar uma mudança no sistema, como a criação de um campo estruturado, ou orientar ajustes no processo que corrijam o uso do sistema, impactando positivamente na padronização dos dados.



Adotar a arquitetura Medallion garantiu uma estrutura robusta para a ingestão e qualificação dos dados. Diferentemente de abordagens em que os dados são transformados diretamente em *dashboards* ou por meio de consultas ad hoc, o modelo Medallion organiza os dados em camadas progressivas de qualidade, promovendo maior confiabilidade, rastreabilidade e padronização. Ainda que os dados estejam tecnicamente centralizados, a arquitetura favorece uma governança descentralizada e viabiliza a evolução contínua do ecossistema de dados, garantindo controle, autonomia e escalabilidade. Equipes e usuários podem criar indicadores e conjuntos derivados sem comprometer a integridade do repositório central. Com boas práticas de governança e documentação, o modelo incentiva uma cultura de dados abertos e qualificados, ampliando sua aplicação em diferentes níveis de decisão.



Um dos impactos mais relevantes da solução foi viabilizar a abertura e a qualificação dos dados do planejamento orçamentário e execução financeira. Ao centralizar, padronizar e integrar dados de sistemas estruturantes como SIAFI, Compras.gov.br e TransfereGov, a plataforma oferece uma visão consolidada e transparente de todo o fluxo orçamentário, do planejamento à execução. Ao tornar os dados acessíveis, auditáveis e reutilizáveis, fortalece o controle institucional, a fiscalização e a atuação de pesquisadores, jornalistas e qualquer pessoa interessada. A iniciativa reforça os princípios do governo digital ao promover o uso efetivo da informação pública para análise, decisão e participação social.

O projeto também gerou **documentações atualizadas das APIs governamentais.** Durante o processo de qualificação e integração dos dados, foi realizada uma análise aprofundada das APIs disponíveis nos SEs, incluindo a verificação de *endpoints* e parâmetros, bem como a testagem dos dados retornados e sua compatibilidade com outras fontes. Em muitos casos, as documentações oficiais estavam desatualizadas, incompletas ou ausentes, especialmente em sistemas legados.

Diante dessas lacunas, foi necessário reconstruir parte da lógica das APIs por meio de engenharia reversa, mapeando o tráfego de dados, os campos utilizados e as transformações aplicadas no backend. Esse processo resultou em documentações revisadas com propostas de melhoria e indicou caminhos mais acessíveis para futuras iniciativas de interoperabilidade. Além disso, a equipe enfrentou barreiras operacionais e institucionais, como exigência de certificados digitais, credenciais e ausência de orientações técnicas e contratuais. Para mitigar esses entraves e facilitar futuras integrações, **elaborou-se um guia prático de acesso a dados públicos via APIs** com instruções passo a passo, pontos de atenção e cobrindo aspectos técnicos, administrativos e jurídicos.



Como replicar essa etapa?

O GovHub foi desenvolvido inicialmente em parceria com a Diretoria de Desenvolvimento Institucional do Ipea, sendo concebida como uma plataforma de software livre, aberta, replicável e sustentável, pensada para ser adotada por qualquer órgão público que utilize sistemas estruturantes ou deseje integrar seus próprios dados internos de forma independente. A solução é baseada em tecnologias de código aberto, permitindo às instituições adaptar e expandir a solução conforme suas necessidades, evitando dependências de fornecedores e restrições institucionais.

Esse é um projeto em constante evolução, mantido como software livre e aberto à colaboração de órgãos públicos e parceiros institucionais. Isso significa que qualquer instituição pode:

- Instalar e adaptar a solução conforme suas próprias necessidades;
- Conectar seus sistemas internos e cruzar suas bases com os dados estruturantes federais;
- Sugerir melhorias, reportar problemas ou contribuir com novos módulos e dashboards por meio de pull requests.

Atualizações periódicas trazem novas conexões com sistemas públicos, melhorias de usabilidade, dashboards adicionais e boas práticas de gestão orientada a dados. Cada contribuição, advinda de pesquisadores do Lab Livre da UnB ou da comunidade de usuários, reforça a capacidade do GovHub de atender a diferentes realidades institucionais.

Os dashboards atualmente existentes no GovHub foram desenhados para serem práticos e personalizáveis, com potencial de customizações para cada instituição. Também é disponibilizada documentação técnica completa, com tutoriais, exemplos e instruções passo a passo para instalação, integração e personalização.

Ao adotar o GovHub, o órgão passa a ter maior potencial de utilização de seus dados, reduzindo a dependência de processos manuais e fortalecendo sua capacidade de gestão. A plataforma já foi testada em um ambiente real e demonstrou ser capaz de transformar dados dispersos em diagnósticos rápidos, confiáveis e acionáveis.

O código-fonte da plataforma está disponível publicamente em: <https://github.com/GovHub-br>

Se uma instituição deseja adotar, adaptar ou contribuir com o GovHub, acesse a documentação disponível e entre em contato com a equipe do Lab Livre da UnB, nos contatos indicados no repositório aberto do projeto. Juntos, é possível fortalecer a gestão pública orientada a dados, impulsionando a eficiência, a transparência e a inovação no setor público.

Artefatos produzidos:

- | | |
|---|---|
| • Repositório público com licença livre com todo o código desenvolvido; | ambiente de desenvolvimento, homologação e produção; |
| • Documentação automatizada e disponível dos processamentos realizados nos dados; | • Tutoriais de uso e criação de novos dashboards; |
| • Documentação técnica para executar a solução em | • Adoção de boas práticas e padrões de comunidades de software livre. |



O que são os software livres?

Sistemas de software livre são regidos por licenças que oferecem quatro liberdades fundamentais: executar, adaptar, redistribuir e modificar o código sem restrições. Com isso, órgãos ou instituições que enfrentam desafios semelhantes podem utilizar o mesmo software e desenvolver, de forma colaborativa, novas funcionalidades. Esse modelo fomenta uma comunidade ativa que compartilha conhecimento e reduz custos.

Para a administração pública, o uso de software livre representa a aplicação de recursos públicos em um código que também é público, garantindo maior transparência, economia e sustentabilidade. Ao reutilizar e aprimorar soluções existentes, reduz-se o custo de desenvolvimento e o risco de depender de um único fornecedor, evitando o chamado “aprisionamento tecnológico” (Vendor Lock-in). Além disso, a maior concorrência entre prestadores de serviços eleva a qualidade dos sistemas e estimula o fortalecimento da indústria tecnológica do país.

O governo brasileiro tem papel estratégico nesse cenário, incentivando o uso de software livre de forma planejada e de longo prazo. Muitos gestores públicos já priorizam padrões abertos e licenças livres, e há diretrizes de licitação que estimulam a adoção de tais soluções, reduzindo dependências e custos. Iniciativas federais reforçam esse compromisso, como a adoção do Plone (como CRM), do Mapas Culturais e do SIMEC, entre outros exemplos bem-sucedidos.

Como saber mais e como contribuir?

O GovHub foi projetado para ser facilmente adotado por qualquer órgão público, sem a necessidade de licenças comerciais ou estruturas complexas. A plataforma é aberta, flexível e pode ser instalada e executada por equipes técnicas internas que possuam conhecimentos básicos em ambientes de dados.

Atualmente, o GovHub utiliza tecnologias amplamente conhecidas e de uso consolidado no setor público, como Docker, Kubernetes, PostgreSQL, Python, Apache Airflow e Apache Superset. Essas ferramentas foram escolhidas por sua robustez, escalabilidade e alinhamento com os princípios de software livre. Vale ressaltar que, por estar em constante evolução, a plataforma poderá incorporar novas ferramentas ou versões atualizadas, sempre priorizando a transparência, rastreabilidade e sustentabilidade tecnológica.

Além disso, a plataforma já disponibiliza pipelines prontos para coleta e qualificação de dados, assim como dashboards iniciais que podem ser personalizados de acordo com as necessidades e prioridades de cada órgão. Toda a documentação técnica, incluindo manuais de instalação, guias de configuração e exemplos práticos, está disponível no site oficial: www.gov-hub.io.



CONCLUSÃO

—
—



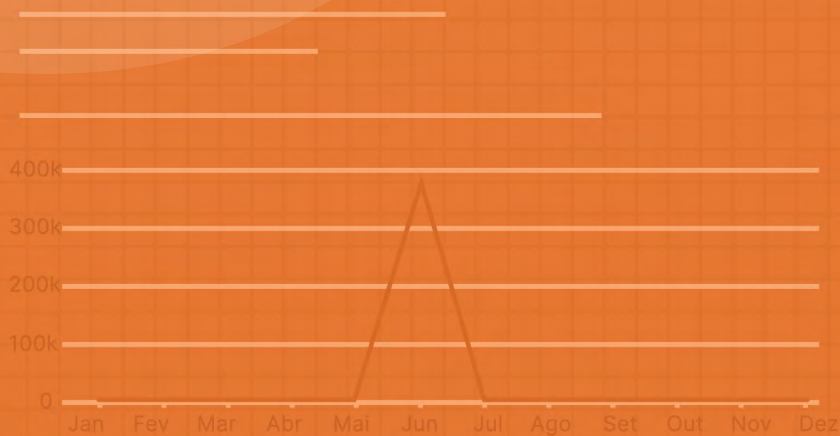
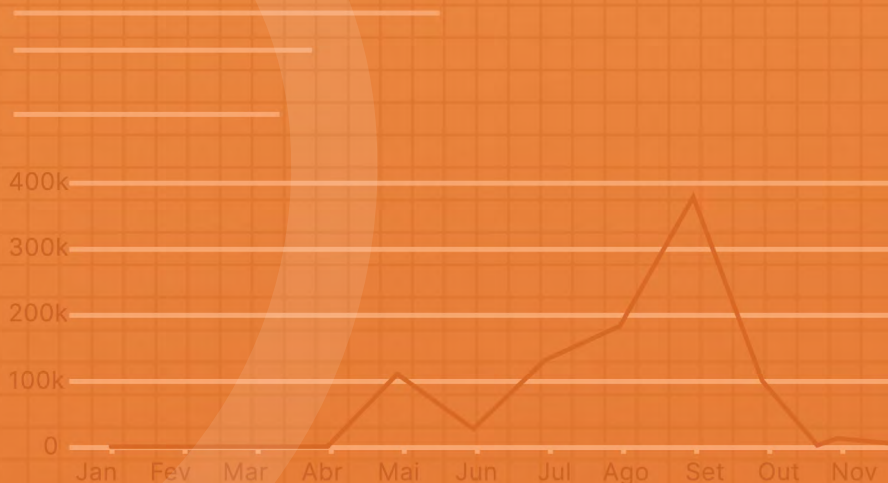
A experiência relatada neste e-book demonstra como o uso estratégico de dados pode transformar a gestão pública. A construção do GovHub — plataforma desenvolvida em parceria com a Diretoria de Desenvolvimento Institucional do Ipea — representa um avanço concreto no caminho por uma administração pública mais eficiente, autônoma e transparente. O projeto enfrentou desafios técnicos e institucionais significativos, mas mostrou que é possível superá-los com soluções baseadas em software livre, arquitetura robusta e compromisso com a abertura e a reutilização de dados públicos.

Ao integrar e qualificar informações oriundas de sistemas estruturantes, o GovHub fortalece a capacidade dos órgãos públicos de compreender e gerir seus próprios dados, reduzindo a dependência de fornecedores e promovendo maior autonomia tecnológica. A plataforma também contribui para a padronização e rastreabilidade dos dados, ampliando seu potencial de uso tanto na formulação de políticas públicas quanto no controle social.

A parceria com a DIDES/Ipea foi fundamental para unir capacidades técnicas, institucionais e analíticas em torno de um propósito comum: tornar os dados públicos mais acessíveis, utilizáveis e impactantes. Juntas, as equipes desenvolveram uma solução replicável, com documentação aberta e aderente às necessidades reais de gestão, superando limitações históricas de interoperabilidade e fomentando uma cultura de inovação baseada na colaboração entre governo e sociedade.

Os aprendizados deste projeto apontam para caminhos promissores na construção de políticas públicas orientadas por evidências, com maior transparência e participação. O GovHub se consolida, assim, como uma infraestrutura pública de dados que pode ser expandida, compartilhada e adaptada por outros órgãos, contribuindo para o fortalecimento da democracia e da governança digital no Brasil.

REFERÊNCIAS BIBLIOGRÁFICAS



GOVHUB. GovHub: repositório com código e documentação da plataforma livre de qualificação e integração de dados e informações governamentais [online]. Disponível em: <https://www.gov-hub.io/>.

ALVES, Isaque et al. Practices for managing machine learning products: A multivocal literature review. **IEEE Transactions on Engineering Management**, v. 71, p. 7425-7455, 2023.

ALVES, I.; EGEWARTH, J. H.; BELLIX, L.; REIS, A.; AGUIAR, C. S. R. Plataforma de Integração de Sistemas Estruturantes da Execução Financeira Federal. In: **Latin American Symposium on Digital Government**, 2025.

COSTA, L.; DUTRA, M. T.; OLIVEIRA, G.; SILVA, M.; SOARES, D. C.; DE CÁSSIA SILVA FARIA, L.; JR., W. M.; PAPP, G. Ciência de dados e transparência: experiências com dados públicos do sicom. In: **Anais Estendidos do XXXIX Simpósio Brasileiro de Bancos de Dados**, Porto Alegre, RS, p. 246–252, 2024.

SILVA CRISTÓVAM, J. S. da; SAIKALI, L. B.; SOUSA, T. P. de. Governo digital na implementação de serviços públicos para a concretização de direitos sociais no Brasil. **Sequência (Florianópolis)**, v. 43, n. 89, p. 209–242, 2020.

DATA, M. C.; KOMOROWSKI, M.; MARSHALL, D. C.; SALCICCIOLI, J. D.; CRUTAIN, Y. Exploratory data analysis. In: **SECONDARY ANALYSIS OF ELECTRONIC HEALTH RECORDS**, p. 185–203, 2016.

DUNLEAVY, P.; MARGETTS, H.; BASTOW, S.; TINKLER, J. New public management is dead—long live digital-era governance. **Journal of Public Administration Research and Theory**, v. 16, n. 3, p. 467–494, 2006.

GIL-GARCIA, J. R.; DAWES, S. S.; PARDO, T. A. Digital government and public management research: finding the crossroads. **Public Management Review**, v. 20, n. 5, p. 633–646, 2018.

MACHADO, I. A.; COSTA, C.; SANTOS, M. Y. Data mesh: concepts and principles of a paradigm shift in data architectures. **Procedia Computer Science**, v. 196, p. 263–271, 2022.

MINISTÉRIO DO PLANEJAMENTO. Introdução ao siop – sistema integrado de planejamento e orçamento. Brasília, 2010.

NADAL, S.; ROMERO, O.; ABELLÓ, A.; VASSILIADIS, P.; VANSUMMEREN, S. An integration-oriented ontology to govern evolution in big data ecosystems. **Information Systems**, v. 79, p. 3–19, 2019.

RUSSOM, P. et al. Data lakes: purposes, practices, patterns, and platforms. **TDWI White Paper**, 2017.

VIANA, A. C. A. Transformação digital na administração pública: do governo eletrônico ao governo digital. **Revista Eurolatinoamericana de Derecho Administrativo**, v. 8, n. 1, p. 115–136, 2021.

FREE SOFTWARE FOUNDATION EUROPE. *Modernising with Free Software: Guia para modernização com software livre* [online]. Disponível em: https://download.fsfe.org/campaigns/pmpc/PMPC-Modernising-with-Free-Software.pt_br.pdf.

GOVOCAL. *Buying a Platform* [e-book]. Disponível em: https://res.cloudinary.com/govocal/image/upload/v1729160209/Guides/Buying%20a%20platform/Buying_a_platform_ENUS.pdf.

OPEN KNOWLEDGE BRASIL. *Publicadores de Dados 2* [online]. São Paulo, maio 2021. Disponível em: <https://www.ok.org.br/wp-content/uploads/2021/05/PublicadoresDeDados2.pdf>.

SERPRO. *Serpro Multicloud* [e-book]. Brasília, DF: SERPRO, n.d. Disponível em: https://www.serpro.gov.br/serpromulticloud_ebook.pdf.

MACHADO, N.; HOLANDA, V. B. de. Diretrizes e modelo conceitual de custos para o setor público a partir da experiência no governo federal do Brasil. *Revista de Administração Pública*, Rio de Janeiro, v. 44, n. 4, p. 687–708, ago. 2010. Disponível em: <https://doi.org/10.1590/S0034-76122010000400003>.

SUZART JA DA S. End-user satisfaction with the Integrated System of the Federal Government Financial Administration (SIAFI): a case study. *JISTEM JInfSyst Technol Manag* [Internet]. 2013Jan;10(1):145–60. Disponível em: <https://www.scielo.br/j/jistm/a/tzsBrMWjsnZW63tjn9Y9Lwv/?lang=en>.

BRASIL. Ministério da Economia. Secretaria do Tesouro Nacional. *SIAFI: Sistema Integrado de Administração Financeira* [online]. Disponível em: <https://www.gov.br/tesouronacional/pt-br/siafi>.

ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICOS (OCDE). *The path to becoming a data-driven public sector*. Paris: OECD Publishing, 2019. Disponível em: https://www.oecd.org/en/publications/2019/11/the-path-to-becoming-a-data-driven-public-sector_9ed7e867.html.

Acesse nossas publicações



Acompanhe nossas redes sociais





UnB



Lab Livre

ipea

Instituto de Pesquisa
Econômica Aplicada

Gov Hub

