Relatório final de atividades

Projeto: ServiçosPI: inovação e eficiência na recuperação de informações de propriedade intelectual do INPI.

Coordenador do Projeto:	**Washington Luís Ribeiro de Carvalho Segundo**		
Coordenação:	**Coordenação-Geral de Informação Científica e Técnica - CGIC**		
Nome do(a) pesquisador(a):	**Francisco de Assis Boldt**		
Vigência da Bolsa de Pesquisa:	**06/07/2026**	**Até:**	**05/10/2026**
Encerramento em:	**05/10/2026**		
Ref. Fundação:	**31849**	**Ref. Ibict:**	**0677/2024-91**

Descrição das Atividades da Bolsa de Pesquisa:

1 - Estudos para definição e validação da arquitetura tecnológica. Meta 5 Análise para seleção de componentes tecnológicos adequados ao desenvolvimento dos módulos RPI on-line e área de usuário; 2 - Elaborar relatório técnico final de bolsa, consolidando resultados, atividades desenvolvidas e contribuições metodológicas da pesquisa.

1 INTRODUÇÃO

O projeto ServiçosPI: inovação e eficiência na recuperação de informações de propriedade intelectual do Instituto Nacional da Propriedade Industrial (INPI) tem por finalidade desenvolver uma infraestrutura moderna, escalável e eficiente para o gerenciamento e a recuperação de ativos de propriedade intelectual no Brasil.

A arquitetura do projeto foi estruturada em duas grandes frentes de desenvolvimento: a interface (*front-end*) e o *datalake* (*back-end*), este responsável por centralizar, tratar e prover os dados dos cinco principais ativos de propriedade intelectual geridos pelo INPI: patentes, marcas, contratos, programas de computador e desenhos industriais.

No âmbito do *datalake*, adotou-se a arquitetura *Medallion* (Medalhão), composta por três camadas de maturidade de dados:

Camada Bronze: repositório de dados brutos e originais extraídos dos sistemas legados e fontes primárias;

Camada *Silver*: etapa de limpeza, padronização, validação e estruturação dos dados;

Camada *Gold*: camada na qual os dados limpos são transformados mediante a aplicação das regras de negócio específicas para disponibilização às consultas do sistema e consumidores finais.

A atuação do pesquisador durante o período de vigência da bolsa (06/07/2026 a 05/10/2026) deu-se na equipe do *back-end*, com o objetivo focal de implementar as transformações e regras de negócio da camada *Gold* para o ativo Programas de Computador.

Por se tratar de um ingresso com o projeto já em andamento, o plano de trabalho individual sofreu adaptações necessárias ao contexto temporal e operacional das entregas de infraestrutura. As seções a seguir detalham a metodologia empregada, as atividades desenvolvidas, os resultados obtidos e as considerações finais sobre a contribuição técnica do período.

2 MÉTODO

Para o alinhamento e o desenvolvimento das atividades de integração da camada *Gold* no *datalake*, foram utilizados os seguintes métodos, recursos e arcabouço tecnológico:

1. Ambientação Tecnológica e Estudo do Código-Fonte Existente: Análise comparativa do código-fonte e dos pipelines desenvolvidos previamente por bolsistas sêniores para o ativo de Patentes, relacionando-os com os requisitos e especificações da arquitetura do projeto.
2. Stack Tecnológica da Infraestrutura de Dados:
    - MinIO: Utilizado como armazenamento de objetos (*Object Storage*) para hospedagem dos dados do *datalake*
    - Apache Spark: Utilizado como motor de processamento distribuído para leitura, manipulação e preparação dos *pipelines* de transformação
    - Apache Iceberg: Formato de tabela aberto utilizado para gerenciar grandes volumes de dados no *datalake*, garantindo performance e controle de versão das tabelas nas camadas *Silver* e *Gold*.
3. Integração e Rituais de Gestão Ágil:
    - Participação ativa nas reuniões semanais exclusivas da equipe de back-end para acompanhamento de decisões de arquitetura e infraestrutura.
    - Participação em reuniões diárias (*dailies*) e nas reuniões semanais conjuntas de alinhamento entre as equipes de *back-end* e *front-end*.
    - Análise dos casos de uso de interface para dedução prévia e antecipação das necessidades de modelagem da camada *Gold*.

3 RESULTADOS ALCANÇADOS

Durante o período da bolsa, as atividades estiveram direcionadas ao entendimento da infraestrutura global, à preparação do ambiente de desenvolvimento e à análise da viabilidade de implementação do ativo de Programas de Computador.

Devido às interdependências da esteira de dados, a disponibilização dos dados da camada *Silver* para o ativo de Programas de Computador ocorreu apenas no início do último mês de vigência da bolsa. Adicionalmente, as regras de negócio formais para a consolidação dos dados desse ativo na camada *Gold* ainda não haviam sido finalizadas pelas áreas competentes do INPI, existindo até o momento apenas o mapeamento dos casos de uso focados na interface com o usuário.

Frente a este cenário, os resultados alcançados concentraram-se em:

1. Desenvolvimento e Preparação da Estrutura de Código (*Boilerplate*): Construção prévia dos scripts e pipelines em Apache Spark para conexão ao MinIO e consumo das tabelas em formato Apache Iceberg da camada Silver. O código atua como infraestrutura pronta para receber a aplicação das transformações assim que as regras formais forem definidas.
2. Análise de Requisitos e Mitigação de Riscos Técnicos: Durante a participação nas reuniões de alinhamento técnico, a contribuição do pesquisador deu-se por meio do levantamento de questionamentos estruturais referentes a alterações no acesso aos dados e ao formato de disponibilização da camada *Gold*. Esse processo permitiu à equipe antecipar gargalos, identificar potenciais divergências de integração entre as frentes e propor ajustes preventivos na arquitetura.
3. Mapeamento de Compatibilidade entre Ativos: Estudo detalhado dos pipelines existentes do ativo de Patentes e comparação direta com as demandas levantadas para Programas de Computador, identificando padrões reutilizáveis de arquitetura.

4 CONSIDERAÇÕES FINAIS

O período de atuação no projeto ServiçosPI permitiu a integração do pesquisador ao fluxo de trabalho e à infraestrutura do *datalake* do INPI. Embora limitações de dependência cronológica, como o momento de disponibilização da camada *Silver* e a ausência da definição final das regras de negócio do ativo de Programas de Computador, tenham impedido a entrega de artefatos de dados finalizados para produção, o trabalho desempenhado cumpriu papel fundamental na preparação técnica do ativo.

O arcabouço de código desenvolvido para o consumo das tabelas Iceberg via Spark/MinIO e o levantamento de pontos críticos durante os rituais de alinhamento asseguram que as etapas posteriores do projeto contem com uma base sólida e testada para a rápida implementação das transformações de negócio assim que forem formalizadas.

REFERÊNCIAS

gov.br
Documento assinado digitalmente
FRANCISCO DE ASSIS BOLDT
Data: 21/10/2026 18:41:16-0300
Verifique em https://validar.iti.gov.br

Francisco de Assis Boldt

Pesquisador

De acordo:

gov.br
Documento assinado digitalmente
WASHINGTON LUIS RIBEIRO DE CARVALHO SEG
Data: 23/10/2026 10:05:08-0300
Verifique em https://validar.iti.gov.br

Washington Luís Ribeiro de Carvalho Segundo
Coordenador do projeto