

INTEGRAÇÃO DE BASES DE DADOS PARA A GEOESPACIALIZAÇÃO DE CADASTROS IMOBILIÁRIOS FEDERAIS EM VITÓRIA/ES

Integration of Databases for the Geospatial Mapping of Federal Properties in Vitória/ES

Jessica Carvalho Vianna Có
Secretaria do Patrimônio da União
Coordenação-Geral de Demarcação
jessica.co@gestao.gov.br

Resumo:

Este artigo descreve o problema da desatualização dos dados cadastrais dos imóveis federais, armazenados em sistemas antigos e literais, e apresenta uma proposta de fluxo ETL para a compatibilização entre os dados geospaciais de bairros, arruamentos e lotes disponibilizados pela Prefeitura Municipal de Vitória com dados cadastrados de imóveis da União localizados neste município. Utilizando o endereço (logradouro, número e bairro) como chave de integração, o objetivo foi adicionar a componente geoespacial aos cadastros de imóveis da União e detectar possíveis registros que necessitam de revisão. Os resultados demonstram que 78,64% dos imóveis cadastrados foram compatibilizados com alta precisão, o que confere maior eficiência na gestão pública. A relevância prática dos resultados está no suporte aprimorado à administração pública, permitindo uma melhor gestão dos imóveis federais, facilitando processos como consultas de dominialidade, rotas de fiscalização e ações de regularização fundiária, além de otimizar o uso dos recursos públicos ao aumentar a confiabilidade e atualidade dos cadastros.

Palavras-chave: cadastro de imóveis federais; ETL (Extract, Transform, Load); geoespacialização, integração de dados; gestão pública

Abstract:

This article addresses the issue of outdated data on federal properties, stored in old and literal systems, and presents a proposal for an ETL process to harmonize geospatial data on neighborhoods, streets, and lots provided by the Municipality of Vitória with the federal property records for this municipality. Using the address (street name, number, and neighborhood) as an integration key, the aim was to add the geospatial component to the federal property records and identify possible entries that require review. The results show that 78.64% of the registered properties were successfully integrated with high precision, improving public administration efficiency. The practical relevance of these results lies in enhanced support for public administration, enabling better management of federal properties, facilitating processes such as ownership verification, inspection routes, and land regularization actions, while optimizing public resource use by increasing the reliability and accuracy of property records.

Keywords: federal property register; ETL (Extract, Transform, Load); geospatialization; data integration; public management

1. INTRODUÇÃO

A Secretaria do Patrimônio da União (SPU), vinculada ao Ministério da Gestão e da Inovação em Serviços Públicos (MGI), é responsável pela administração de todo o patrimônio imobiliário da União. Atualmente, a SPU opera com um conjunto de sistemas legados, entre eles o Sistema Integrado de Administração Patrimonial (SIAPA) e o Sistema de Gerenciamento dos Imóveis de Uso Especial da União (SPIUnet), que gerenciam mais de 700 mil imóveis, totalizando 3,31 trilhões de metros quadrados e com um valor contabilizado de 1,18 trilhões de reais.

Tais sistemas são literais e seus cadastros de imóveis apresentam um nível considerável de desatualização e/ou inconsistências (endereço, valor, localização, dimensões, matrícula, ocupante etc.), grande parte em razão da modelagem dos sistemas sem o mapeamento efetivo dos processos, da insuficiência de mecanismos de validação e da inexistência de integrações com bases de dados de outras instituições como os cartórios de registros de imóveis, a Receita Federal Brasileira, os Correios, bem como com as prefeituras municipais.

A integração de bases cadastrais com dados geoespaciais é uma questão recorrente no cenário global. O avanço de tecnologias de geoprocessamento e fluxos ETL (Extract, Transform, Load) tem permitido a modernização de sistemas de informação e a melhoria na qualidade dos registros públicos. Segundo Yildiz & Demir (2016), a integração de dados geoespaciais em registros cadastrais é essencial para aprimorar a precisão e a eficiência na administração pública, principalmente em países que lidam com sistemas fragmentados de registro de propriedades. Ainda, a utilização de fluxos ETL para a integração de dados provenientes de fontes heterogêneas foi amplamente explorada por Vassiliadis et al. (2005), que destacam a importância de um repositório centralizado de informações que permita a integração, limpeza e transformação de dados para uma tomada de decisão mais precisa.

No contexto da gestão do patrimônio público, vários estudos têm demonstrado a relevância da integração de dados cadastrais e geoespaciais. Furtado et al. (2018) afirmam que a interoperabilidade entre sistemas cadastrais pode otimizar a fiscalização, a cobrança de tributos e o planejamento urbano. Cerri & Carvalho (2020) ressaltam que a combinação de dados espaciais e cadastrais nos Sistemas de Informação Geográfica (SIG) melhora a eficiência na alocação de recursos e no monitoramento do uso do solo.

O presente trabalho se insere nesse contexto ao propor um fluxo ETL para compatibilizar os cadastros imobiliários da União com os dados geoespaciais disponibilizados pela Prefeitura Municipal de Vitória, utilizando o endereço (logradouro, número e bairro) como chave de integração. A originalidade desta proposta está na utilização de dados abertos municipais para geoespacializar os imóveis da SPU, permitindo um melhor aproveitamento dos recursos públicos e contribuindo para a modernização e atualização do cadastro imobiliário federal. Além disso, a adição da componente geoespacial possibilita uma gestão pública mais eficaz, com impacto direto em ações estratégicas do governo como a regularização fundiária, a fiscalização, e a cobrança de tributos.

2. REFERENCIAL TEÓRICO

2.1. Os dados geoespaciais e o geoprocessamento na Administração Pública

Segundo Bertini (2003), os dados geoespaciais, ou os também chamados dados geográficos, nada mais são do que dados que possuem uma dimensão espacial estabelecida por formas geométricas associadas à sua localização na superfície terrestre em um determinado tempo. Um conjunto de dados geoespaciais forma uma entidade geográfica, que por sua vez é concebida por objetos, fatos ou fenômenos identificáveis no mundo real que possuem relacionamentos espaciais com outras entidades geográficas.

Os dados geoespaciais agregam variáveis de lugar e tempo e identificação, respondendo às perguntas essenciais: “onde?”, “quando?” e “o que?”. A variável lugar traz a localização do objeto junto à sua geometria; o tempo traz o período no tempo que aquele dado se refere; e, a variável identificação traz os atributos não-espaciais que caracterizam os objetos. (CÂMARA et. al. 1996).

O geoprocessamento é um conjunto de ferramentas e técnicas voltadas para a coleta, tratamento, análise e visualização de dados geoespaciais. De acordo com Câmara et al. (1996), o geoprocessamento envolve a integração de tecnologias como sistemas de informação geográfica (SIG), sensoriamento remoto e banco de dados georreferenciados, com o objetivo de fornecer suporte a análises que envolvem componentes espaciais. Essa capacidade de lidar com dados espaciais é fundamental para melhorar o planejamento e a gestão territorial.

Para Cordovez (2002), é necessário convencer gestores e servidores, que, em sua grande maioria, desconhecem totalmente assuntos relacionados a cartografia e a sistemas de informação geográfica. Para o autor, um requisito para o êxito da introdução do geoprocessamento é difundir e incentivar a cultura GEO nas repartições públicas.

Xavier da Silva (2001) argumenta que o uso de geotecnologias pode aumentar a eficiência ao permitir que gestores visualizem espacialmente os ativos públicos, entendam melhor sua distribuição e tomem decisões com base em dados mais precisos. Nesse sentido, a adoção do geoprocessamento não representa apenas um avanço tecnológico, mas uma ferramenta estratégica para a tomada de decisões mais eficazes, beneficiando tanto gestores quanto cidadãos.

2.2. Extração, Transformação e Carga (Extract Transform Load - ETL)

O ETL (Extract, Transform, Load) é um processo fundamental na integração de dados que envolve três etapas principais: extração, transformação e carga. Na fase de extração, os dados são coletados de diversas fontes, como bancos de dados, arquivos e serviços, para consolidar informações dispersas. Em seguida, na fase de transformação, os dados extraídos são limpos e processados, assegurando consistência e qualidade, além de serem ajustados conforme as regras de negócio específicas. Por fim, na etapa de carga, os dados transformados são propagados para o repositório, banco de dados ou outro tipo de saída configurada, onde estarão disponíveis para análise e relatórios. Este processo é essencial para garantir que as informações integradas sejam úteis e confiáveis, possibilitando uma melhor tomada de decisão organizacional (VASSILIADIS et al., 2005).

Nesse sentido, Vassiliadis et al. (2005) demonstram que a aplicação de fluxos ETL permite a coleta de dados de fontes heterogêneas, sua limpeza e transformação, além de garantir que a informação final seja adequada para as necessidades dos gestores públicos.

No contexto da gestão pública, Furtado et al. (2018) realizaram um estudo focado

na integração de dados cadastrais com dados geoespaciais para otimizar a fiscalização de imóveis públicos. Eles exploraram como a implementação de um fluxo ETL poderia resolver problemas de desatualização e inconsistência nos cadastros imobiliários. A pesquisa destacou a importância de um repositório integrado que permita a visualização geográfica dos dados, o que facilita a identificação de irregularidades e a localização precisa dos imóveis. O estudo demonstrou que, ao aplicar técnicas de ETL, foi possível não apenas atualizar as informações, mas também integrar dados de diferentes órgãos públicos, melhorando significativamente a eficiência da fiscalização e a alocação de recursos.

De maneira similar, Cerri & Carvalho (2020) contribuíram para a discussão sobre a aplicação de ETL em sistemas de gestão pública, abordando como a integração de dados espaciais e cadastrais pode ser utilizada para otimizar a alocação de recursos e melhorar o planejamento urbano. Eles enfatizaram a necessidade de fluxos ETL robustos que possam lidar com a diversidade de formatos e fontes de dados, o que é comum em contextos administrativos. O trabalho ilustra que a implementação de ETL não só facilita a interoperabilidade entre sistemas, mas também permite a criação de um ambiente de dados mais acessível e utilizável para gestores públicos. Isso, por sua vez, contribui para a transparência e a eficiência nas operações governamentais.

2.3. Padrão de Registros de Endereço

O padrão de registro de endereços adotado é fundamental para garantir a interoperabilidade entre diferentes bases de dados. No Brasil, o Instituto Brasileiro de Geografia e Estatística (IBGE) é o responsável pela gestão do Cadastro Nacional de Endereços para Fins Estatísticos (CNEFE), que define um padrão de endereçamento que pode ser utilizado em várias esferas do governo. O CNEFE é um repositório nacional de endereços, criado a partir dos dados coletados pelo Censo 2000, cujo padrão pode ser adotado voluntariamente pela sociedade e em outras esferas de governo. (IBGE,2019)

De acordo com o documento "-Padrão de Registro de Endereços" (IBGE, 2019), o endereço é um texto que permite localizar e individualizar um bem imóvel dentro de um limite administrativo municipal. Com esse objetivo, o endereço é composto pelas seguintes partes integrantes: Unidade da Federação (UF); município; distrito; subdistrito; setor; localidade; quadra; face; logradouro (tipo, título e nome); número; modificador; complementos; ponto de referência; CEP; e coordenadas geográficas.

As Tabelas 1 e 2 apresentam exemplos de preenchimento de logradouros para os casos mais usuais e para os casos nos quais o logradouro não detenha um nome oficial, ou não seja possível identificá-lo (comuns em áreas de expansão urbana e rurais).

Tabela 1: Exemplos de preenchimento dos componentes do logradouro

Logradouro	Tipo	Título	Nome
Avenida Imperatriz Leopoldina	Avenida	Imperatriz	Leopoldina
Travessa Santa Lúcia	Travessa	Santa	Lúcia
Largo da Matriz	Largo		da Matriz
Rodovia BR-116	Rodovia		BR-116
Rua Imperador	Rua		Imperador

Fonte: IBGE (2019)

Tabela 2: Exemplos de logradouros não oficiais

Logradouro	Tipo	Título	Nome
Estrada para Represa de Gericinó	Estrada		para Represa de Gericinó
Acesso para Fazenda Santa Cruz	Acesso		para Fazenda Santa Cruz
Beco da Associação de Moradores	Beco		da Associação de Moradores
Vicinal da RJ-116 para São Martinho	Vicinal		da RJ-116 para São Martinho
Rua Sem Denominação	Rua		Sem Denominação

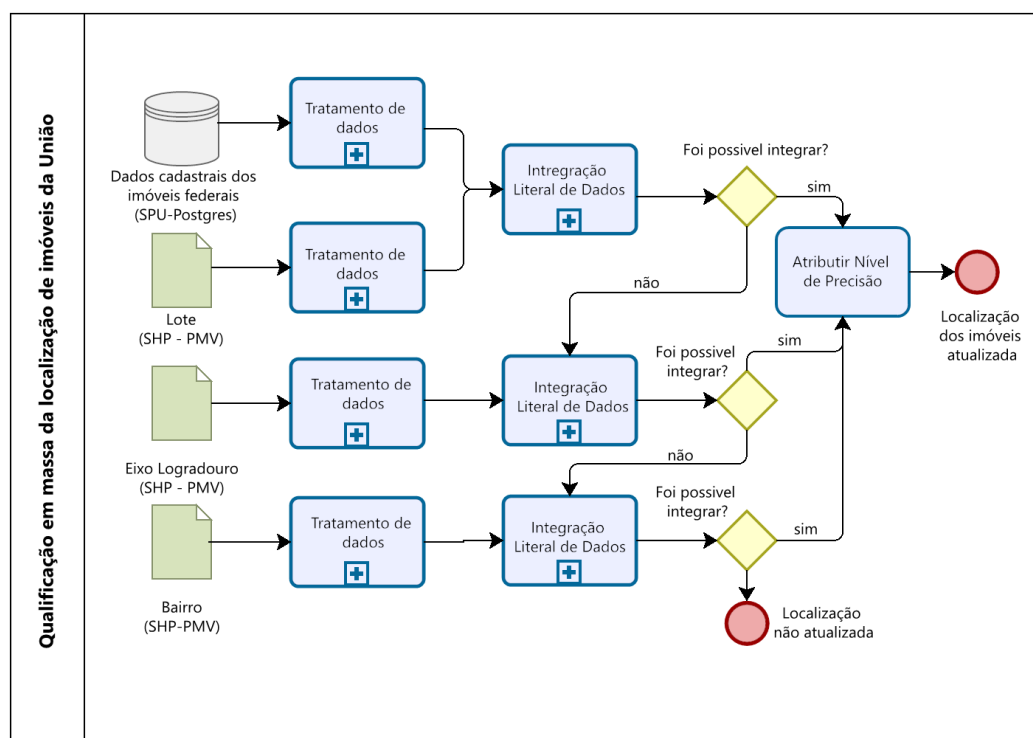
Fonte: IBGE (2018)

3. METODOLOGIA

Para a execução da metodologia foi escolhido o Software FME Desktop, tendo em vista a disponibilidade de licenças pela SPU e a sua alta capacidade de integrar e manipular uma ampla variedade de formatos de dados utilizados pela SPU, especialmente dados espaciais, facilitando a execução de operações complexas de geoprocessamento. Além disso, o FME permite a automação de fluxos de trabalho, alinhando-se perfeitamente com os objetivos deste estudo.

A metodologia se baseia num fluxo ETL composto por 3 grandes grupos de atividades de transformação de dados: o tratamento de dados, ocasião em que os dados coletados de fontes diversas são saneados e estruturados; a integração de dados, onde acontece a compatibilização entre o cadastro de imóveis federais (SPU) e as bases vetoriais disponibilizadas pela Prefeitura Municipal de Vitória (PMV), compondo um *dataset* único; e, por fim, a atribuição do nível de precisão, quando cada registro é classificado baseado na forma de unificação que foi possível na etapa anterior. A Figura 1 demonstra as macro-atividades apresentadas na presente metodologia.

Figura 1: Fluxo simplificado da compatibilização proposta



Fonte: elaborado pela autora

3.1 Fontes de Dados

Foram utilizadas três fontes de dados no estudo. A primeira é uma tabela desnormalizada proveniente de uma conexão Postgres SQL, que unifica mais de 750.000 cadastros dos dois principais sistemas da SPU: o SIAPA e o SPIUnet. Para a análise, foram aplicados filtros nessa tabela, selecionando apenas os dados de imóveis ativos cadastrados no Estado do Espírito Santo, especificamente no município de Vitória, resultando em aproximadamente 44.217 registros, sem geometria.

As demais fontes de dados consistem em arquivos vetoriais no formato "Shapefile", baixados gratuitamente do site "Geoweb" (<https://geoweb.vitoria.es.gov.br/#/shp>) da Prefeitura Municipal de Vitória- PMV. A base de lotes da cidade contém 51.567 registros do tipo polígono, enquanto a base de arruamentos possui 9.207 registros do tipo linha.

3.2. Tratamento de Dados

Tendo em vista a integração de dados com base em parâmetros predominantemente textuais, foram realizadas ações de limpeza e estruturação a partir das fontes configuradas. O processo utilizou como chave os dados literais de endereço presentes em ambas as bases, incluindo tipo de logradouro, nome do logradouro, número e bairro.

3.2.1. Tratamento dos dados cadastrais da SPU

Após os primeiros testes, percebeu-se que, mesmo o dado de input estando "limpo" (sem acentos, caracteres especiais e já em caixa alta), era necessária a inclusão de duas ações: sanear os logradouros que representavam esquinas e os logradouros que incluíam a informação de edifício dentro do atributo "logradouro" e não no atributo "complemento", como é o padrão.

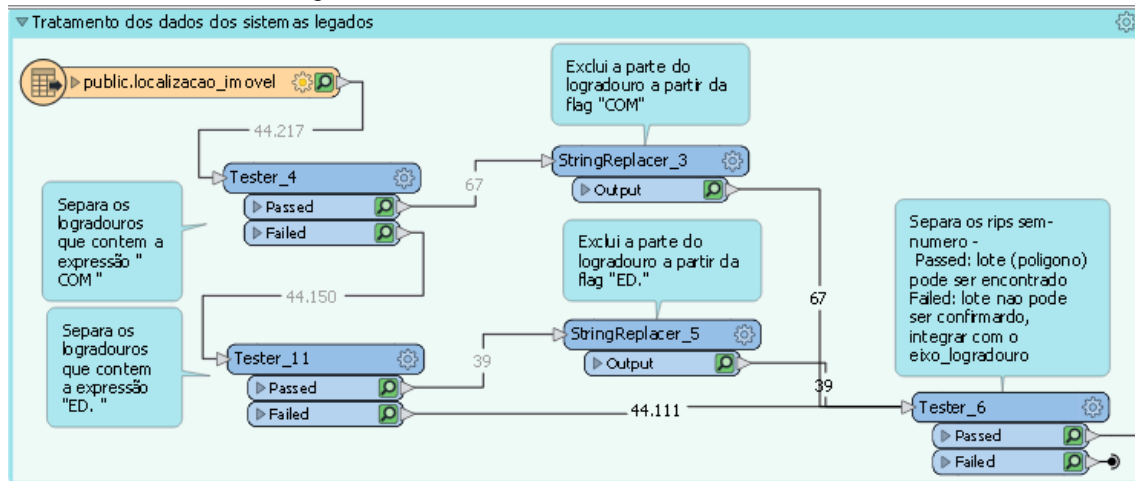
A Tabela 3 traz a lista de transformações realizadas no dado original da SPU, juntamente com qual condição foi aplicada. Cada transformação retrata um algoritmo pré-concebido pelo software utilizado, com o objetivo específico explicado no campo "Descrição". A Figura 2 ilustra a sequência de ações na etapa de tratamento de dados.

Tabela 3. Listagem de transformações e condições utilizadas nesta sub-etapa

Input	Atributo	Nome da transformação	Condição	Descrição
SPU	logradouro	tester	contains regex []COM[]	Filtra os logradouros que contenham expressão " COM "
		replacer	replace regex []COM[].+ com NULL	Exclui a parte do "logradouro" a partir da expressão "COM"
		tester	contains regex .+ED([.][])+	Filtra os logradouros que contenham expressão " ED "
		replacer	replace regex ED[.]?[]+. com NULL	Exclui a parte do "logradouro" a partir da expressão "ED"

número	tester	contains regex [0-9]+	filtra registros sem-número. Passed: lote (polígono) pode ser encontrado /Failed: lote não pode ser confirmado, integrar com o arruamento (linha)
--------	--------	--------------------------	--

Figura 2: Tratamento de dados na SPU no FME.



Fonte: elaborado pela autora

No primeiro teste, foram identificados e saneados 67 registros que possuíam a palavra “COM” dentro do texto do atributo “logradouro”. Já no segundo teste, foram saneados 39 registros. A Figura 3 exemplifica as inconsistências encontradas.

Figura 3: Representação do atributo “logradouro” antes e depois do saneamento proposto

logradouro	logradouro
1 X COM RUA UM	1 X
2 VITORIA COM R RAIMUNDO NONATO	2 VITORIA
3 ULISSES SARMENTO COM FERREIRA COELHO	3 ULISSES SARMENTO
4 PROJETADA COM SERVIDAO PUBLICA	4 PROJETADA
5 LEITAO DA SILVA ESQ COM D PEDRO I	5 LEITAO DA SILVA ESQ
6 JOSE DA S LOUREIRO COM RUA DJANIRA RIBEI	6 JOSE DA S LOUREIRO
7 JOAO DA CRUZ COM ALEXO NETO	7 JOAO DA CRUZ

Fonte: elaborado pela autora

3.2.2. Tratamento dos dados da Prefeitura Municipal de Vitória

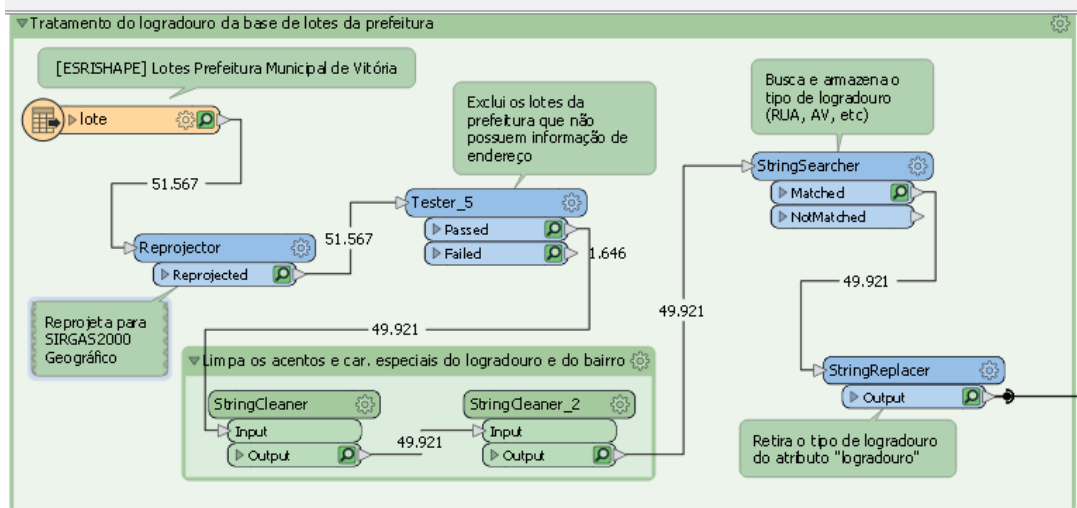
Para os *datasets* da Prefeitura, primeiramente foi realizada a reprojeção do Sistema de Referência de Coordenadas (SRC) do conjunto de dados para o mesmo SRC do Banco de Dados Geográfico da SPU. Tais dados estavam representados em coordenadas métricas e precisaram ser transformados para coordenadas geográficas, em graus, usualmente utilizadas em bases de maior abrangência.

Em seguida, excluiu-se da análise todos os registros que não apresentavam a informação principal a ser utilizada na integração: o nome do logradouro. Além disso, de forma a padronizar o nome do logradouro, foram retirados os acentos, caracteres especiais selecionados e o texto foi convertido para “caixa alta”.

Finalmente, percebeu-se que a base de lotes apresentava o tipo e o nome do logradouro unidos em um atributo único. Assim, foi feita uma busca textual para armazenar o primeiro termo como o tipo de logradouro e, em seguida, excluiu-se esse

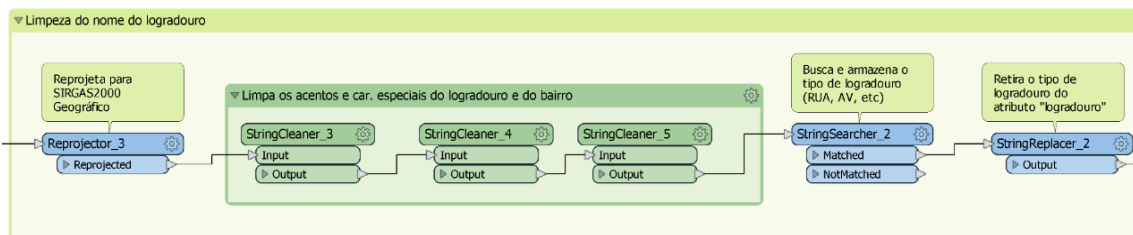
verbete do nome do logradouro original. Dessa forma, conseguiu-se separar o texto inicial conforme o padrão estabelecido no *dataset* da SPU. As Figuras 4 e 5 apresentam a sequência dos passos de tratamento realizados na base de lotes e arruamentos da PMV, respectivamente.

Figura 4: Tratamento de dados vetoriais de lotes da PMV no FME



Fonte: elaborado pela autora

Figura 5: Tratamento de dados vetoriais de arruamentos da PMV no FME



Fonte: elaborado pela autora

A Tabela 4 traz a lista de transformações realizadas no dado original de lotes e arruamentos da PMV, juntamente com qual condição foi aplicada. Cada transformação retrata um algoritmo pré-concebido pelo software utilizada, com o objetivo específico explicado no campo “Descrição”.

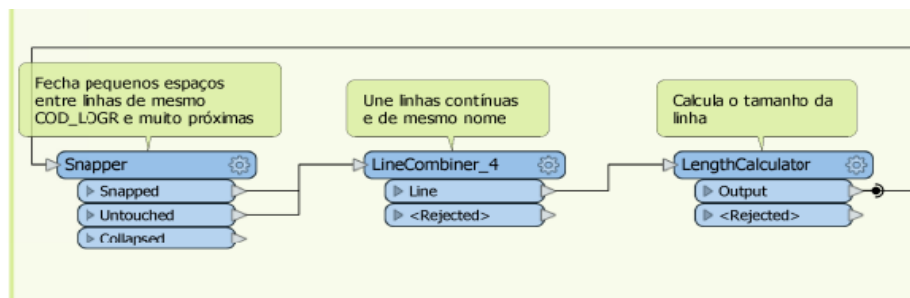
Tabela 4. Listagem de transformações e condições utilizadas nesta sub-etapa

Input	Nome da Transformação	Condição	Descrição
LOTE/ ARRUMAMENTOS	reprojector		Reprojeta os dados para o Sistema de Referência de Coordenadas utilizado pelo banco da SPU
PMV	tester	attribute has a value	Filtra apenas os registros com o nome do logradouro preenchidos
	string cleaner		Limpa acentos e caracteres especiais do logradouro
	string cleaner		Limpa acentos e caracteres especiais do bairro

string seacher	search re-gex [A-Z]+	Busca e armazena a primeira palavra do logradouro, armazena como o tipo do logradouro
string replacer	replace ^[A-Z]+\s+ com NULL	Limpa do logradouro o tipo do logradouro armazenado anteriormente

Já para a base de arruamentos, além das operações realizadas para a base de lotes, executaram-se operações adicionais. Na Figura 6, a primeira operação realizada é do tipo topológica e tem a função de unificar geograficamente as geometrias do tipo linha muito próximas, mas descontínuas. A transformação seguinte unifica registros que agora se tocam e possuem os mesmos atributos e, em seguida, o comprimento da linha que representa um arruamento é calculado.

Figura 6: Tratamento de dados vetoriais de arruamentos da PMV no FME.

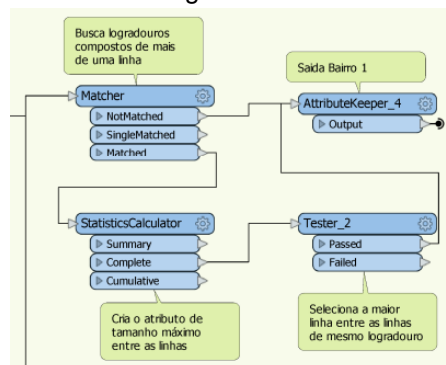


Fonte: elaborado pela autora

Com o comprimento do arruamento calculado, é feita uma análise de registros repetidos considerando os parâmetros “nome do logradouro” e “bairro”. O valor de comprimento máximo associado a esse agrupamento é adicionado a um atributo provisório, visando, na sequência, testar se o comprimento da linha contida em cada registro é igual ao máximo calculado. Caso seja, o registro segue no fluxo e será usado na fase seguinte. (Figura 7)

Tudo isso é feito para termos apenas um segmento de linha representando um logradouro em um determinado bairro, com o objetivo de não se multiplicar os registros resultantes da integração proposta na próxima fase. O mesmo procedimento é repetido alterando e/ou retirando os parâmetros de bairro, formando 3 agrupamentos dos maiores segmentos de arruamentos relativos ao bairro 1, bairro 2 e aos dois juntos.

Figura 7: Agrupando os maiores segmentos de arruamentos relativos ao bairro



Fonte: Elaborado pela autora

3.3. Integração Literal de Dados

É importante pontuar que a junção ideal procurada é com a base vetorial de lotes, uma vez que ela traz a informação na primitiva geométrica mais completa, que é o polígono com as delimitações exatas de cada imóvel.

Caso não seja possível, a segunda junção buscada é com a base vetorial de arruamentos. A ideia é que, mesmo que não seja conhecida a localização exata do lote, o segundo melhor nível de precisão possível seria o posicionamento o imóvel na rua correta.

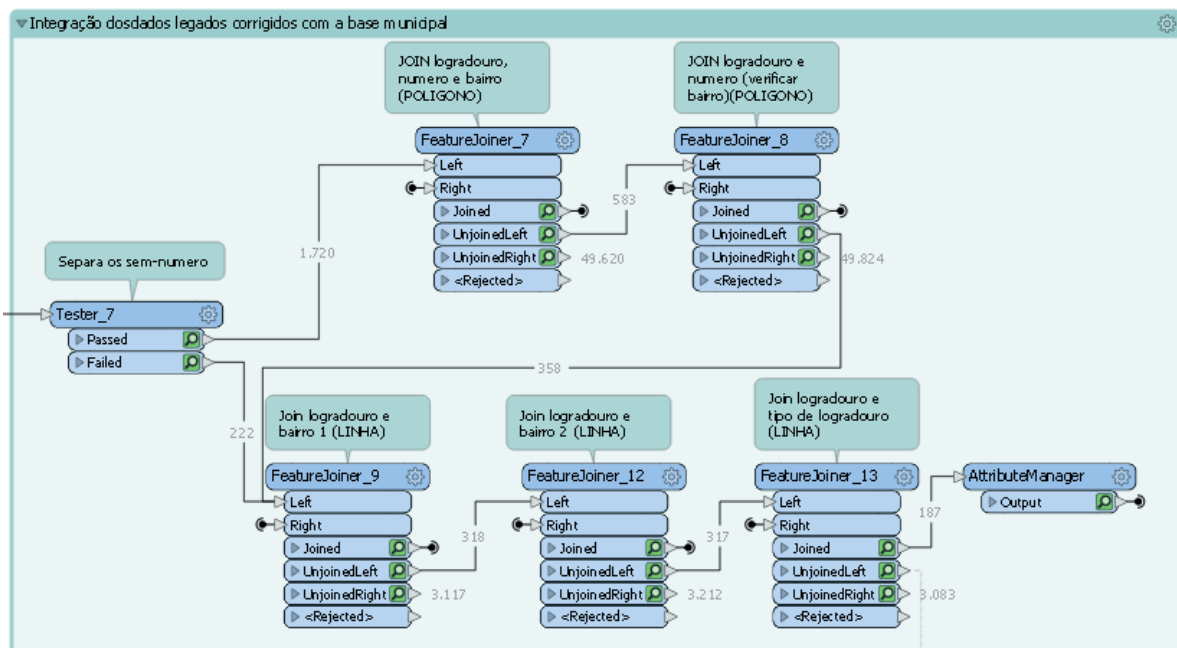
Caso não seja possível, visto que os logradouros podem ter mudado de nome no tempo ou serem logradouros não-oficiais, os registros serão separados e classificados como não encontrados.

Para o arquivo vetorial de arruamentos, foi verificado que o arquivo continha mais de um atributo de bairro: “Bairro_1” e “Bairro_2”. Isso se explica no fato de que um mesmo logradouro pode perpassar por mais de um bairro. Assim, como o bairro também é um atributo da base cadastral da SPU, na etapa anterior replicou-se a base de logradouros para cada bairro, com o objetivo de testar mais de uma integração em série, utilizando uma chave composta (logradouro+bairro_1) e em seguida (logradouro+bairro_2) para localizar um segmento de rua em um determinado bairro, conforme Figura 8.

Após a primeira integração, foi utilizado um algoritmo de similaridade “Fuzzy” para tratar os logradouros que não encontraram compatibilidade com nenhuma das bases da PMV. Utilizou-se um parâmetro empírico de 60% de similaridade para se tomar a decisão de substituição do logradouro da SPU pelo logradouro selecionado na base da PMV. Após essa ação, todas as ações de integração foram refeitas em busca de novas compatibilizações.

Para a compatibilização dos imóveis com os arruamentos, o resultado será do tipo linha, mas deverá ser substituído por um ponto posicionado na mediana da linha como forma de representação gráfica do imóvel como um ponto no mapa.

Figura 8: Integração de dados vetoriais da PMV com dados da SPU no FME.



Fonte: Elaborado pela autora

3.4. Atribuição do Nível de Precisão

Após a fase de integração, os registros são agrupados e organizados para as diferentes saídas do fluxo. São elas: recorte da base da SPU adicionada da informação vetorial de polígono extraída da PMV, recorte da base da SPU adicionada da informação vetorial dos pontos posicionados no centro dos logradouros extraídos, um recorte da base da SPU contendo todos os registros que não puderam ser compatibilizados e uma versão da base da SPU contendo os pontos de todos os imóveis.

Os registros da base de pontos da gerada pelo fluxo foram classificados em três níveis de precisão: “no centro do imóvel”, “na rua do imóvel” e “não localizado”.

Todos os imóveis que conseguiram ser compatibilizados com a base vetorial dos polígonos dos lotes foram classificados como “no centro do imóvel”, posto que o ponto do imóvel foi obtido a partir do cálculo e extração do centroide do polígono do lote.

Todos os imóveis que não conseguiram capturar a geometria do tipo polígono, mas tiveram sucesso na compatibilização com a base de arruamentos foram classificados como “na rua do imóvel”, uma vez que ponto foi obtido a partir da extração do ponto posicionado no meio do trecho de arruamento.

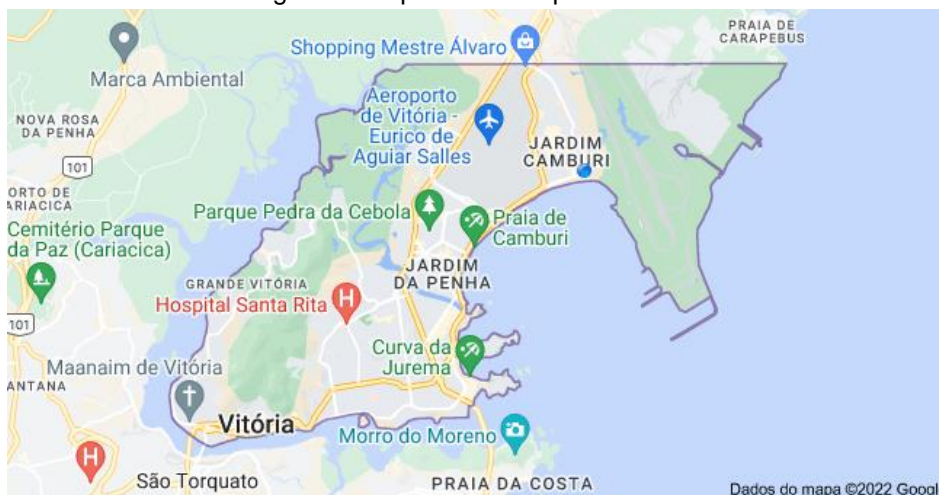
O restante dos imóveis cujos dados de endereço não foram compatíveis com a base municipal foram classificados como “não localizados” e necessitarão de análise posterior para o seu preciso posicionamento.

4. RESULTADOS

4.1 Área de estudo

Neste trabalho, a área de estudo compreende o município de Vitória, capital do Estado do Espírito Santo e formado por um arquipélago de 33 ilhas e uma parte continental, totalizando 97.123 km² de área e com uma população estimada de 369.534 pessoas. (IBGE, 2022)

Figura 9. Mapa do Município de Vitória



Fonte: Google Maps, 2024

4.2 Resultados Obtidos

Utilizando o software FME Desktop para a execução das etapas relatadas na metodologia, chegou-se ao resultado apresentado na Tabela 5 para a primeira tentativa de integração.

Tabela 5. Resultados da Primeira Tentativa de Integração

Input	Chaves de integração	Nº de Registros Integrados	% de Integração	Flags	Nível de Precisão
SPU+LOTE	logradouro, número e bairro	32426	73,33%	-	No centro do imóvel
	logradouro, número	985	2,23%	verificar bairro	
SPU+RUA	logradouro, bairro_1	7046	15,94%	-	Na rua do imóvel
	logradouro, bairro_1	81	0,18%	-	
	logradouro, tipo logradouro	1591	3,60%	verificar bairro	
	logradouro	76	0,17%	verificar bairro e tipo logradouro	
		2012	4,55%	-	Não localizado
Total de Registros		44217	100,00%		

Após a primeira tentativa, 2.012 registros da base da SPU não foram compatíveis com nenhum logradouro cadastrados nas bases municipais comparadas. Assim, após a análise de similaridade, 1.942 dos 2.012 registros da base SPU tiveram o atributo do nome do logradouro substituídos pelos logradouros com maior similaridade encontrados na base de arruamentos. Em seguida, os testes de integração foram repetidos, trazendo os resultados demonstrados na Tabela 6:

Tabela 6. Resultados da Segunda Tentativa de Integração

Input	Chaves de integração	Nº de Registros Integrados	% de Integração	Flags	Nível de Precisão
SPU+LOTE	logradouro, número e bairro	1137	58,55%	verificar logradouro	No centro do imóvel
	logradouro, número	224	11,53%	verificar logradouro, bairro	
SPU+RUA	logradouro, bairro_1	262	13,49%	verificar logradouro	Na rua do imóvel
	logradouro, bairro_2	1	0,05%	verificar logradouro	

	logradouro, tipo	318	16,37%	verificar logradouro, bairro	
	logradouro	0	0,00%	-	
Total de Registros		1942			Não localizado

Após a segunda tentativa, os 1942 registros foram compatibilizados com a base de arruamentos como esperado. A Tabela 7 apresenta o resultado consolidado das duas tentativas de integração, relativos ao nível de precisão obtido com o fluxo.

Tabela 7. Resultados Consolidados

Nível de Precisão	Nº de Registros	% de Integração	Representações possíveis
No centro do imóvel	34772	78,64%	Polígono Ponto
Na rua do imóvel	9375	21,20%	Ponto
Não localizados	70	0,16%	-
Total de imóveis	44217	100,00	Ponto

No total, 0,16% dos registros da SPU foram classificados como “não localizados”. De forma geral, trata-se de logradouros cadastrados de forma precária, seja pela falta ou pelo excesso de informação.

Além do nível de precisão, ao longo do fluxo foram atribuídas flags aos registros que não conseguiram se integrar utilizando todas as chaves propostas. Todos os registros que passaram pela análise de similaridade foram marcados como “verificar logradouro”, conforme a Tabela 8.

Tabela 8. Flags Atribuídas

Flags	Nº de Registros	%
verificar logradouro	1942	4,39%
verificar tipo logradouro	76	0,17%
verificar bairro	3194	7,22%

Após a apresentação das Tabelas 5, 6, 7 e 8, pode-se afirmar que foi possível compatibilizar, por meio de um fluxo ETL, os endereços de duas fontes de dados distintas, absorvendo a potencialidade da mais precisa delas, incluindo os limites de lotes cadastrais, possivelmente para 78,64% dos imóveis federais cadastrados no município de Vitória.

Conforme a Tabela 8, 11,78% dos registros foram sinalizados com flags para verificação posterior, indicando ao servidor da SPU os pontos onde pode haver inconsistências, uma vez que os dados municipais foram considerados corretos, devido à sua frequente atualização para fins de cobrança do IPTU.

Nesse sentido, é importante destacar que a validação dos resultados está planejada para ser realizada em um momento posterior, por meio de uma apuração especial no sistema que receberá a atualização em massa dos cadastros compatibilizados, garantindo a conferência pelo servidor responsável.

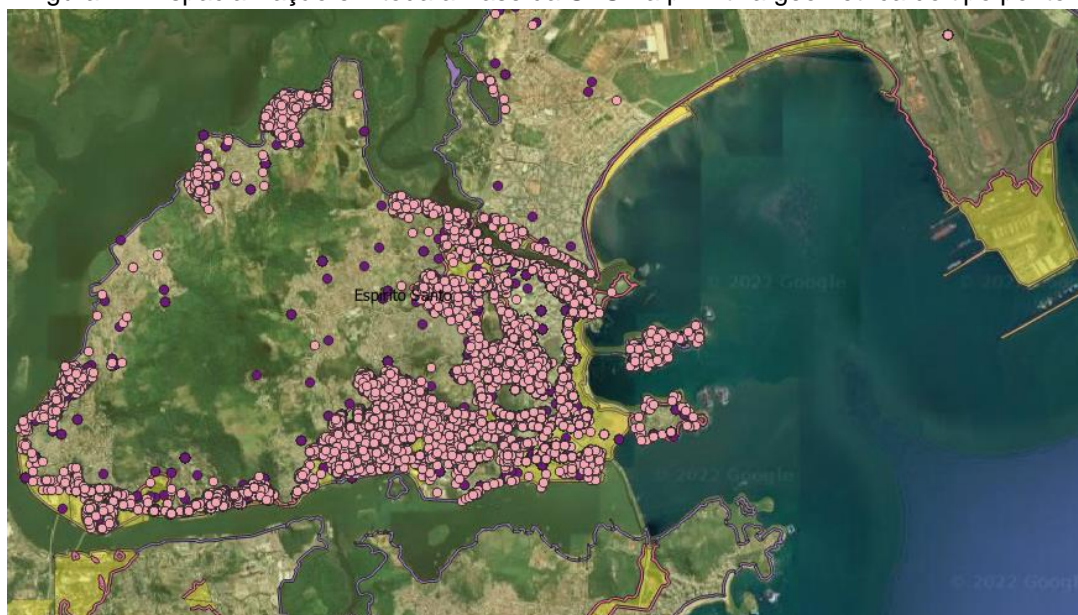
4.3 Mapas

Utilizando o software QGIS, que se apresenta como um Sistema de Informação Geográfica (SIG), foram elaborados os mapas apresentados nas Figuras 10 e 11.

Figura 10. Espacialização dos lotes junto aos pontos no centro dos eixos dos logradouros



Figura 11. Espacialização em toda a Base da SPU na primitiva geométrica do tipo ponto



5 CONCLUSÃO

Os resultados alcançados demonstram a viabilidade de compatibilizar os dados geoespaciais de bairros, arruamentos e lotes fornecidos pela Prefeitura Municipal de Vitória com os dados de endereços cadastrados de imóveis da União localizados

nesse município, utilizando um fluxo ETL.

Esse processo permitiu a adição da componente geoespacial aos cadastros de imóveis da SPU, o que orienta a tomada de decisões por gestores e auxilia os servidores em suas demandas diárias. Entre os benefícios práticos estão o suporte em ações de usucapião, consultas de dominialidade, definição de rotas de fiscalização, cálculo de áreas ocupadas e a detecção de “buracos de arrecadação”.

O trabalho também revelou que a maioria dos cadastros da SPU em Vitória está atualizada e alinhada com o padrão de endereços adotado pela prefeitura. Isso resultou em uma integração eficaz, com 78% dos registros apresentando um nível de precisão “no centro do imóvel”.

Uma recomendação prática de médio prazo é a atualização em massa dos cadastros junto à migração dos sistemas legados para o novo sistema **SPUnet**, garantindo que os cadastros de imóveis federais sejam atualizados e georreferenciados, o que otimizará as operações diárias da SPU e fortalecerá a gestão do patrimônio público. A adoção de fluxos ETL contínuos permitirá que essa atualização se torne um processo recorrente, prevenindo a obsolescência dos dados.

Além disso, recomenda-se que a SPU explore a replicação deste fluxo em outros municípios, priorizando aqueles com grandes volumes de imóveis da União, e invista em capacitação técnica para seus servidores. A ampliação do uso de tecnologias de geoprocessamento e automação de dados fortalecerá a eficiência e a transparência da gestão pública.

Para os cadastrados com flags, recomenda-se a revisita dos cadastros para realizar os ajustes no endereçamento e nos cadastros “não localizados”, os quais necessitarão de busca por maiores informações nos processos administrativos, mas em apenas 0,76% dos casos.

Essas análises e o fluxo descrito reforçam o potencial do geoprocessamento e da ciência de dados para promover maior eficiência na atualização de dados, por meio de rotinas de baixo custo implementadas internamente por servidores. Isso garante alta confiabilidade nas bases de dados institucionais e fortalece a gestão pública federal.

6 REFERÊNCIAS

BERTINI, G. C. *Uma modelagem orientada a objeto para o mapa urbano básico de Belo Horizonte (MUB/BH)*. Belo Horizonte: Pontifícia Universidade Católica de Minas Gerais, 2003.

DIOUF, Papa Senghane; BOLY, Aliou; NDIAYE, Samba. Variety of data in the ETL processes in the cloud: State of the art. In: **2018 IEEE International Conference on Innovative Research and Development (ICIRD)**. IEEE, 2018. p. 1-5.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. *Panorama da cidade de Vitória/ES*. Disponível em: <https://cidades.ibge.gov.br/brasil/es/vitoria/panorama>. Acesso em: 10 jul. 2022.

CÂMARA, G.; CASANOVA, M. A.; HEMERLY, A. S.; MAGALHÃES, G. C.; MEDEIROS, C. B. *Anatomia de Sistemas de Informação Geográfica*. Campinas: Ed. Unicamp, 1996. 193p.

CERRI, R.; CARVALHO, M. Geoinformação na administração pública: desafios e oportunidades para a gestão de imóveis. *Revista Brasileira de Cartografia*, v. 72, n. 1, p. 85-101, 2020.

CORDOVEZ, Juan Carlos Gortaire. Geoprocessamento como ferramenta de gestão urbana. In: **Simpósio Regional de Geoprocessamento e Sensoriamento Remoto**, v. 1, 2002.

FURTADO, R.; LOPES, L. M.; SOUZA, M. Integração de bases de dados cadastrais com geoprocessamento para otimização da fiscalização de imóveis públicos. *Revista de Gestão Pública*, v. 7, n. 3, p. 45-58, 2018.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. *Padrão de Registro de Endereços: Definições e orientações de uso*. Rio de Janeiro: IBGE, 2019. Disponível em: <https://biblioteca.ibge.gov.br/visualizacao/livros/liv101639.pdf>. Acesso em: 14 jul. 2022.

XAVIER-DA-SILVA, Jorge. *Geoprocessamento para análise ambiental*. Rio de Janeiro: 2001. 228 p.

VASSILIADIS, Panos et al. A generic and customizable framework for the design of ETL scenarios. *Information Systems*, v. 30, n. 7, p. 492-525, 2005. Disponível em: <http://www.dblab.ntua.gr/~mter/papers/is2005.pdf>. Acesso em: 27 jul. 2022.

YILDIZ, I.; DEMIR, H. Geospatial data integration for effective property management in cadastral systems. *International Journal of Geographical Information Science*, v. 30, n. 5, p. 1-12.