

JOAO LUIZ SALISSO FILHO

**AVALIAÇÃO DA QUALIDADE DO DADO ESPACIAL DIGITAL DE
ACORDO COM PARÂMETROS ESTABELECIDOS POR USUÁRIOS**

Dissertação de Mestrado apresentada à
Escola Politécnica da Universidade de
São Paulo para obtenção do título de
Mestre em Engenharia de Transportes.

**São Paulo
2013**

JOAO LUIZ SALISSO FILHO

**AVALIAÇÃO DA QUALIDADE DO DADO ESPACIAL DIGITAL DE
ACORDO COM PARÂMETROS ESTABELECIDOS POR USUÁRIOS**

Dissertação de Mestrado apresentada à
Escola Politécnica da Universidade de
São Paulo para obtenção do título de
Mestre em Engenharia de Transportes.

Área de concentração: Engenharia de
Transportes.

Sub-área: Informações Espaciais –
Geoprocessamento.

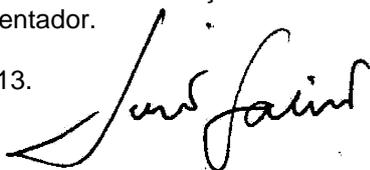
Orientadores: Professor Livre Docente
José Alberto Quintanilha

**São Paulo
2013**

Este exemplar foi revisado e alterado em relação à versão original, sob responsabilidade única do autor e com a anuência do orientador.

São Paulo, 01 de Julho de 2013.

Assinatura do autor



Assinatura do Orientador

I. FICHA CATALOGRÁFICA

Salisso Filho, João Luiz

Avaliação da qualidade do dado espacial digital de acordo com parâmetros estabelecidos por usuários / J.L. Salisso Filho. – ed. rev. -- São Paulo, 2013.

250 p.

Dissertação (Mestrado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Transportes.

1. Metadados 2. Geoprocessamento (Qualidade) I. Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Transportes II. t.

AGRADECIMENTOS

À minha família pelo apoio incondicional à continuidade dos meus estudos.

À minha esposa, pelo apoio sincero e ajuda em todos os momentos e principalmente por compreender as minhas ausências devido a este trabalho.

As colegas Mariana Giannotti e Claudia Machado pelo apoio incontestável na entrega desta dissertação. Muito obrigado!

Ao colega e amigo Rafael Walter de Albuquerque pela amizade e ajuda em vários momentos durante as atividades de nosso mestrado.

À Escola Politécnica da Universidade de São Paulo (EPUSP), pelo apoio institucional no desenvolvimento acadêmico-científico.

Ao Laboratório de Geoprocessamento da Escola Politécnica, pela estrutura e ambiente de trabalho que possibilitaram o desenvolvimento deste trabalho.

Ao professor José Alberto Quintanilha, meu orientador *de facto*, pelo direcionamento, pelas conversas, pela orientação paciente e objetiva. Com toda certeza este trabalho não teria sido concluído sem a sua ajuda.

RESUMO

Informações espaciais estão cada vez mais disseminadas no cotidiano do cidadão comum, de empresas e de instituições governamentais. Aplicações como o Google Earth, Bing Maps, aplicativos de localização por GPS, entre outros apresentam a informação espacial como uma *commodity*.

Cada vez mais empresas públicas e privadas incorporam o dado espacial em seu processo decisório, tornando ainda mais crítico a questão da qualidade deste tipo de dado.

Dada a natureza multidisciplinar e, principalmente, o volume de informações disponibilizadas para os usuários, faz-se necessário apresentar um método de avaliação de dados apoiado por processos computacionais, que permita ao usuário avaliar a verdadeira adequação que tais dados têm frente ao uso pretendido.

Nesta Dissertação de Mestrado propõe-se uma metodologia estruturada de avaliação de dados espaciais apoiada por computador. A metodologia utilizada, baseada em normas apresentadas pela International Standards Organization (ISO), permite ao usuário de dados espaciais avaliar sua qualidade comparando a qualidade do dado de acordo com os parâmetros estabelecidos pelo próprio usuário. Também permite ao usuário comparar a qualidade apresentada pelo dado espacial com a informação de qualidade provida pelo produtor do dado. Desta forma, o método apresentado, ajuda o usuário a determinar a real adequação do dado espacial ao seu uso pretendido.

Palavras-chave: Qualidade de Dados Espaciais, Avaliação de Dados Espaciais, Metadados.

ABSTRACT

Spatial information is increasingly widespread in everyday life of ordinary people, businesses and government institutions. Applications like Google Earth, Bing Maps, GPS location applications, among others present spatial data as a commodity.

More and more public and private companies incorporate the usage of spatial data into their decision process, increasing the importance of spatial quality issues.

Given the multidisciplinary nature and, especially, the volume of information available to all users, it is necessary to introduce a data quality evaluation method supported by computational processes, enabling the end user to evaluate the real fitness for use that such data have for an intended use.

This dissertation aims to present a structure methodology for spatial data evaluation supported by computational process. The methodology, based on standards provided by the International Standards Organization (ISO), allows users of spatial information evaluating the quality of spatial data comparing the quality of information against users' own quality parameters. It will also allow the user to compare the quality presented by the given spatial data with quality information provided by the data producer. Thus, the presented method will support the end user in determining the real fitness for use for the spatial data.

Keywords: Spatial Data Quality, Spatial Data Assessment, Metadata.

LISTA DE ILUSTRAÇÕES

Figura 1 - Classificação dos métodos de avaliação de qualidade (ISO, 2003).....	33
Figura 2 - Localização do Município de Itapeva no Estado de São Paulo.....	Erro! Indicador não definido.
Figura 3 - Fluxo metodológico	38
Figura 4 - Núcleo principal de avaliação do fluxo metodológico.....	40
Figura 5 - Fluxo de execução do processo de avaliação de qualidade	41
Figura 6 - Avaliação de erros de comissão de esquema de dados na tabela da feição Eixos de Logradouro.	48
Figura 7 - Rejeição da feição de eixos de logradouro na base digital da cidade de São Paulo.....	50
Figura 8 - Rejeição da feição de ruas em critério de omissão de registros.	52
Figura 9 - Avaliação de consistência de domínio na feição de RUAS.....	53
Figura 10 - Busca por registros nulos no campo TIPOPREF da feição RUAS.....	55
Figura 11 - Rejeição de consistência conceitual da base de hidrografias	56
Figura 12 - Avaliação de atributos não quantitativos no layer de hidrografias.	57
Figura 13 - Diagrama de classes - ISO 19113 - Fonte: o próprio autor.....	70
Figura 14 - Diagrama de classes - ISO 19114 - Fonte: o próprio autor.....	71
Figura 15 - Diagrama de classes - ISO 19115 - Fonte: o próprio Autor	72
Figura 16 - Diagrama de classes de Negócio - Fonte: o próprio Autor.....	73
Figura 17 - Fluxo geral de utilização do programa	75
Figura 18 - Tela principal do aplicativo de avaliação de qualidade de dados.....	77
Figura 19 - Detalhe do conjunto de dados em avaliação	77
Figura 20 – Tela inicial do aplicativo	78
Figura 21 - Relatório de qualidade	78
Figura 22 - Delimitação espacial da área de atuação da avaliação espacial	79
Figura 23 - Seção da tela <i>Quality Report</i>	79
Figura 24 - Descrição da medida de qualidade	80
Figura 25 - Tipos e descrição de avaliação	80
Figura 26 - Fonte da informação, qualidade informada e qualidade esperada.	81
Figura 27 - Qualidade do dado reportado pelo provedor.....	81
Figura 28 - Qualidade esperada pelo usuário	82
Figura 29 - Identificação do processo de avaliação	82

Figura 30 - Configuração do processo de avaliação finalizado	83
Figura 31 - Configuração de busca por inconsistências.....	83
Figura 32 - Fluxograma para avaliação de completeza de dados.....	87
Figura 33 - Início do procedimento de avaliação do dado	88
Figura 34 - Seleção do escopo de qualidade para atributos	88
Figura 35 - Seleção do Elemento e Sub Elemento de qualidade de dados	88
Figura 36 - Descrição da avaliação de qualidade a ser aplicada	88
Figura 37 - Tipo e descrição da avaliação de qualidade.	89
Figura 38 - Definição da fonte de informação e qualidade esperada	89
Figura 39 - Avaliação por completeza de esquema de dados.....	90
Figura 40 - Avaliação por completeza de registros	90
Figura 41 - Critério de aceitação / rejeição para avaliação de completeza por esquema de dados.....	90
Figura 42 - Critério de aceitação para completeza por registros do conjunto de dados.	91
Figura 43 - Identificação do processo de avaliação	91
Figura 44 - Definição dos critérios de busca por inconsistências no dado espacial ..	92
Figura 45 - Definição do esquema e atributos existentes.....	93
Figura 46 - Criação do esquema ideal ao usuário.....	93
Figura 47 - Definição dos campos de interesse do usuário.....	94
Figura 48 - Relacionando campos disponíveis com o esquema de dados esperado.	94
Figura 49 - Rejeição do dado	94
Figura 50 - Configuração do relatório de qualidade por omissão de registros	96
Figura 51 - Informação de qualidade apresentada pelo provedor do dado espacial. 97	
Figura 52 - Critério de aceitação configurado para rejeitar o dado espacial caso mais de 15% dos registros presentes no banco possuam valores nulos.....	97
Figura 53 - Configuração de busca por omissão/comissão de registros	98
Figura 54 - Configuração do critério de busca por campos nulos.	98
Figura 55 - Início do processo de avaliação	99
Figura 56 - Relatório final de avaliação do dado espacial.	99
Figura 57 - Configuração para avaliação de consistência de domínio.	100
Figura 58 - Critério de aceitação/rejeição da avaliação de consistência de domínio	101

Figura 59 - Qualidade informada pelo provedor a respeito da consistência de domínio do banco de dados em avaliação.	101
Figura 60 - Configuração do critério de busca por inconsistências de domínio	102
Figura 61 - Relatório de resultados com avaliação de conformidade.....	103
Figura 62 - Relatório de qualidade para avaliação de consistência de formato.	104
Figura 63 - Critério de aceitação do dado espacial caso seu formato seja o DGN.	105
Figura 64 - Qualidade aceita pelo aplicativo para avaliação de consistência de formato.....	105
Figura 65 - Configuração do relatório de qualidade para avaliação de consistência conceitual.....	106
Figura 66 - critério de aceitação para avaliação de inconsistência conceitual.	107
Figura 67 - Configuração de busca por elementos não lineares na feição de RUAS.	107
Figura 68 - Resultado da avaliação da feição de vias públicas em análise de consistência conceitual.	108
Figura 69 - Configuração do relatório de qualidade para avaliação da acurácia de atributos não quantitativos	109
Figura 70 - Configuração do critério de aceitação da avaliação de atributos não quantitativos.	110
Figura 71 - Critério de buscas por inconsistências de atributos não quantitativos.	110
Figura 72 - Resultado da avaliação de acurácia de atributos não quantitativos na feição hidrografia.....	111
Figura 73 - Tela de informação da linhagem do dado espacial	112
Figura 75 - Gravação e recuperação de metadados de qualidade no padrão ISO 19115	113
Figura 76 - Seleção do arquivo XML de metadados de qualidade.....	113

LISTA DE QUADROS

Quadro 1 - Elementos e Subelementos de Qualidade de Dados	28
Quadro 2 - Padrões de Metadados Geoespaciais (WEBER, <i>et al.</i> , 1999)	31
Quadro 3 - Temas e seus tipos utilizados para testes do método e aplicativo de avaliação de dados.....	35
Quadro 4 - Temas da base digital de Itapeva utilizados por esta dissertação.....	35
Quadro 6 - Atributos informados pelo provedor de dados cuja existência no banco foi comprovada pela avaliação.....	47
Quadro 6 - Configuração dos parâmetros de busca de inconsistências	84

LISTA DE ABREVIATURAS E SIGLAS

ANZLIC	Australia New Zealand Information Council
CADD	Computer Aided Draft and Design / Desenho e projeto assistido por computador (tradução livre)
CEN TC287	Comitê Europeu de Padronização
CETESB	Companhia Ambiental do Estado de São Paulo
DGN	Formato de arquivo de desenho de propriedade da Bentley Systems.
DWG	Formato de arquivo de desenho de propriedade da Autodesk.
FGDC	Federal Geographic Data Committee
GCMD/NASA	Global Change Master Directory/National Aero Spatial Agency
ISO	International Standards Organization
NCDCDS	National Committee on Digital Cartographic Data Standards
NSSDA	National Standard for Spatial Data Accuracy
PCC	Porcentual de Classificação Correta
PEC	Padrão de Exatidão Cartográfica
SDTS	Spatial Data Transfer Standard
SIG	Sistemas de Informações Geográficas
UML	Unified Modeling Language
XML	Extensible Markup Language
SIG	Sistemas de Informações Geográficas

SUMÁRIO

1. INTRODUÇÃO	14
1.1. OBJETIVO	15
1.2. JUSTIFICATIVA.....	15
1.3. HIPÓTESE.....	17
1.4. ESCOPO	17
2. QUALIDADE DE DADOS	18
2.1. PADRÕES EM QUALIDADE DE DADOS	23
2.2. ELEMENTOS DE QUALIDADE DE DADOS	24
2.2.1. Linhagem	24
2.2.2. Acurácia Posicional	25
2.2.3. Acurácia de Atributos	26
2.2.4. Completeza	27
2.2.5. Consistência Lógica	27
2.2.6. Sumário de Elementos de Qualidade	28
2.3. METADADOS	29
2.4. PADRÕES DE METADADOS GEOESPACIAIS.....	30
2.5. AVALIAÇÃO DE QUALIDADE DE DADOS ESPACIAIS	32
3. MATERIAIS E MÉTODOS	34
3.1. MATERIAIS.....	34
3.1.1. Área de Estudo	34
3.1.1.1. Base de Dados da Cidade de São Paulo	34
3.1.1.2. Base Digital da cidade de Itapeva	35
3.1.1.3. Considerações Sobre os Formatos de Dados Utilizados	36
3.1.2. Recursos de software	37
3.2. MÉTODOS.....	37
3.2.1. Premissas	37
3.2.2. Avaliação da qualidade de dados espaciais	38
3.2.2.1. Descrição do fluxo metodológico.....	39
4. RESULTADOS E DISCUSSÃO	46

4.1. AVALIAÇÃO DE COMPLETEZA DE DADOS.....	46
4.1.1. Erros de Comissão e Omissão em Esquema de Dados.....	47
4.1.2. Erros de Omissão em Esquema de Dados.....	49
4.1.3. Erros de Omissão em Registros	51
4.2. AVALIAÇÃO DE CONSISTÊNCIA LÓGICA.....	52
4.2.1. Consistência de Domínio.....	52
4.2.2. Consistência de Formato.....	55
4.2.3. Consistência Conceitual.....	55
4.3. AVALIAÇÃO DE ACURÁCIA DE ATRIBUTOS	56
4.4. DISCUSSÕES À RESPEITO DOS CRITÉRIOS DE ACEITAÇÃO OU REJEIÇÃO.....	57
5. CONCLUSÕES.....	59
5.1. AVALIAÇÃO DE QUALIDADE DE DADOS ESPACIAIS	60
5.2. COMUNICAÇÃO DA QUALIDADE DO DADO ESPACIAL	61
5.3. CONCLUSÕES FINAIS.....	62
6. REFERÊNCIAS BIBLIOGRÁFICAS.....	63
ANEXOS	67

1. INTRODUÇÃO

O desenvolvimento tecnológico ocorrido nas tecnologias de levantamento de dados, programas de geoprocessamento, banco de dados e internet nas últimas três décadas modificou a forma como instituições públicas e privadas e a sociedade em geral utilizam a informação espacial.

Dispositivos de posicionamento estão cada vez mais diluídos no cotidiano do cidadão através do uso de aplicativos de localização por satélite. Aplicativos voltados à análise de informações geográficas podem ser baixados gratuitamente de sítios na internet (GOODCHILD, 2006). Outros aplicativos como o Google Maps, Bing Maps e Yahoo Maps popularizaram o uso do dado espacial através da internet.

A informação espacial é matéria prima para atividades de planejamento urbano, segurança e serviços públicos, lazer e gerenciamento de emergências. Governos e instituições privadas reconhecem o valor da informação espacial em seu processo decisório.

Apesar de todo aspecto positivo que o aumento na demanda e na produção de informações espaciais apresenta, é cada vez maior a quantidade de usuários que fazem uso da informação sem determinar a sua adequação à sua finalidade. Conseqüentemente, o usuário daquele dado está exposto ao risco de utilizar uma informação inadequada à sua aplicação (GOODCHILD, 1995), (FRANK, 2008) e (DÉVILLERS, *et al.*, 2007) o que pode comprometer o resultado de suas análises e seu processo decisório.

Atualmente, o volume de dados, sua complexidade e principalmente a forma como provedores de informação e usuários comunicam a qualidade da informação, torna o processo de avaliação da qualidade do dado espacial um processo complexo, demorado e suscetível a erros. Isto impossibilita o usuário decidir com segurança se um determinado dado está adequado a um determinado propósito.

1.1. OBJETIVO

Nesta dissertação tem-se o seguinte objetivo:

- Apresentar uma metodologia de avaliação estruturada da qualidade de dados espaciais que permita ao usuário:
 - Avaliar a qualidade do dado via um procedimento automatizado.
 - Criar metadados sobre a qualidade de bases de dados espaciais, caso não existam.

Esta metodologia será aplicada na avaliação da qualidade da base de dados da cidade de São Paulo produzida para fins comerciais e bases de dados espaciais disponibilizadas pela prefeitura municipal de Itapeva-SP.

O protótipo de software também permite:

- Apresentar relatórios de conformidade do dado espacial, com a indicação de sua aderência à especificação do usuário.
- Criar metadados de qualidade em padrão ISO.

1.2. JUSTIFICATIVA

A qualidade da informação é uma questão crítica em qualquer sistema de informações. A eficácia de serviços públicos, ordem social, segurança pública, financeira, ambiental e distribuição de recursos têm na qualidade da informação um fator determinante (BAIRD, 1998). A informação é a base da sociedade moderna, do bem estar econômico e das instituições que governam nossas ações. Se a sociedade moderna não pode contar com informação correta e com acurácia, então muitas de suas tarefas se tornarão impossíveis (STEIN; SHI; BIJKER, 2008). A qualidade da informação espacial segue o mesmo princípio, traz em si, dada a sua natureza, outros complicadores, tais como ciclos de atualização mais longos, custos de produção e manutenção maiores e complexidade no processo de sua manutenção.

Como o dado espacial está enraizado em vários pilares da sociedade, não é aceitável ou possível ignorar a importância de sua qualidade. Ao contrário, pesquisas em certos aspectos da qualidade do dado espacial vêm sendo realizadas

por vários anos. Chrisman (1994) apresentou um método de determinação de adequação de dados para análises ambientais; Couclelis (2003) apresentou estudos sobre incertezas em dados espaciais e Frank (2008) apresentou um método de avaliação de qualidade de dados baseado em um modelo de decisão. Boin e Hunter (2008) apresentaram métodos mais eficazes de comunicação da qualidade do dado ao usuário; Wu e Bittenfield (1994) argumentaram sobre a avaliação da qualidade do dado espacial e o uso de metadados neste processo.

Boin e Hunter (2008), Agumya e Hunter (1997), Dévillers e Jeansoulin (2006), Chrisman (1994) e Zargar (2009) argumentaram que a comunicação da qualidade do dado espacial e o seu papel na determinação da adequação ao uso deste dado em um propósito específico ainda é um problema crítico. Processos decisórios baseados na combinação de dados oriundos de diversas fontes, em diferentes escalas e conseqüentemente com níveis de qualidade diferentes, tornam a atividade de avaliação da adequação da informação espacial a um propósito específico uma tarefa ainda mais complexa (TSOULOS; JAKOBSSON, 2005). Beard (1989), Dévillers (2004) e Dévillers, *et al.*, (2004) sugeriram que uma maneira de solucionar o problema de avaliação da adequação da qualidade de um dado para um determinado uso, consiste em apoiar o usuário em descrever de forma explícita e estruturada a qualidade do dado espacial que sua aplicação requer.

Na literatura estudada apresentou-se o problema comum relacionado ao modo de comunicação e ao processo de avaliação da qualidade do dado espacial: O volume de dados, variedade de temas e de fontes de dados. Tais problemas tornam o processo de avaliação da qualidade do dado uma atividade complexa de ser executada e comunicada sem apoio de metodologia e ferramentas especializadas.

Além disso, literatura mostra que a popularização da informação espacial trouxe problemas que inexistiam no início da utilização de dados espaciais pelas empresas, tais como: desconhecimento da qualidade da informação em uso pelo usuário, desconhecimento da adequação da informação, necessidade de comunicação eficiente da qualidade da informação e do nível de esforço em adequá-la para

determinado uso e a necessidade de uma ferramenta de apoio na avaliação da qualidade do dado.

1.3. HIPÓTESE

A hipótese desta dissertação é: É possível dar suporte a um usuário de dados espaciais no processo de avaliação da qualidade da informação com base nos requisitos de aplicação definidos pelo usuário.

1.4. ESCOPO

No capítulo 0 desta dissertação apresenta-se a revisão bibliográfica realizada pela elaboração deste documento. Apresentando o conceito de qualidade de dados, elementos e subelementos de qualidade. Descreve também métodos de avaliação de qualidade e os instrumentos utilizados na comunicação da qualidade, tais como metadados e seus padrões.

O capítulo 3 descreve os materiais e métodos a serem utilizados na construção desta dissertação, destacando o método proposto para avaliação de alguns elementos de qualidade. O capítulo relaciona as bases de dados utilizadas para validação do método e do protótipo de software desenvolvido. O capítulo apresenta também os aplicativos que apoiarão o processo de avaliação.

O capítulo 4 apresenta os resultados alcançados com a aplicação do método de avaliação apoiado pelo aplicativo desenvolvido nesta Dissertação.

O capítulo 0 apresenta as conclusões relacionadas ao desenvolvimento desta dissertação.

O capítulo 6 apresenta a referência bibliográfica consultada durante a elaboração desta dissertação.

2. QUALIDADE DE DADOS

Neste capítulo serão apresentados conceitos e definições sobre a qualidade de dados espaciais, os padrões utilizados mundialmente na descrição e na avaliação da qualidade do dado e os elementos descritores da qualidade. Serão apresentados também os conceitos de metadados e seus padrões de metadados relacionados à descrição da qualidade do dado espacial.

Antes de iniciar a fundamentação teórica sobre qualidade de dados espaciais, faz-se necessário caracterizar os tipos de usuário existentes atualmente no mercado no contexto desta dissertação:

Quadro 1 - Caracterização dos usuários de dados espaciais

Tipo de usuário	Descrição
Provedor de dados	O provedor do dado espacial representa o produtor do dado, o indivíduo com conhecimento em levantamento de dados, compilação, produção e especificação do dado espacial. Este usuário representa o "vendedor" de dados espaciais no mercado.
"Usuário final" ou apenas "Usuário"	O usuário final é o usuário "consumidor" do dado espacial, representa o indivíduo final que aplica o dado espacial em alguma finalidade específica. Geralmente este usuário tem pouco ou nenhum conhecimento em produção de dados espaciais ou avaliação do dado em si. Este usuário por outro lado, possui conhecimento específico em sua área de estudo. Por exemplo, um biólogo que procura relacionar espacialmente um biota em um mapa pode representar o usuário final deste tipo de dado.
Avaliador de dados	O avaliador do dado espacial possui características únicas: ele compreende e identifica as necessidades do usuário final e consegue traduzir tais necessidades específicas em requerimentos de dados espaciais; ele compreende e identifica os quesitos de qualidade apresentados pelo fabricante; e também consegue avaliar o dado espacial de acordo com ambas perspectivas, comparando-as e qualificando o dado espacial de acordo com ambos critérios (provedor e usuário final).

O Quadro 1 apresenta as caracterizações de usuário sob o contexto desta dissertação. Existe uma grande distância em termos de conhecimento do dado espacial entre o Provedor de dados e o Usuário final. Enquanto o primeiro busca criar um dado espacial que atenda a maior variedade possível de usuários, o segundo busca um dado espacial que atenda à sua necessidade específica. Em

muitas situações, o Provedor simplesmente desconhece as necessidades da aplicação do usuário final; o Usuário final, por sua vez, frequentemente não é capaz de prover uma especificação para produção dos dados que atenderão à sua aplicação.

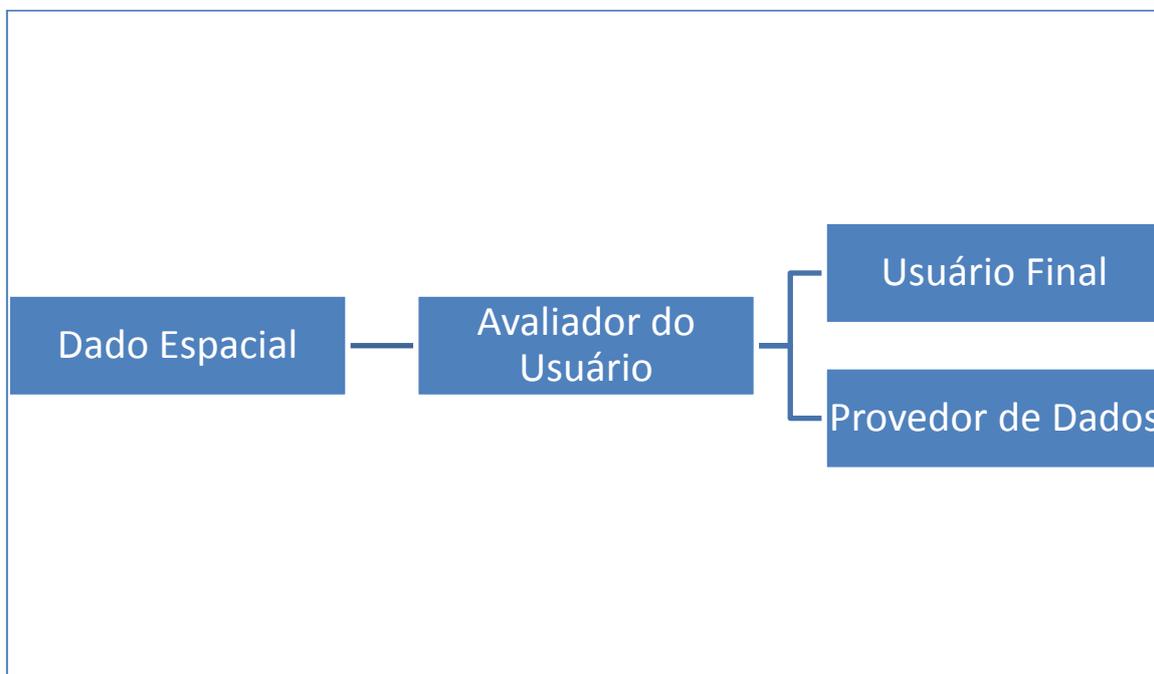


Figura 1 - Relacionamento ideal entre os diferentes tipos de usuário no contexto desta dissertação. Fonte, o autor.

O Avaliador do dado espacial, por outro lado, pode atuar como *expert* com conhecimento a respeito da necessidade do usuário final e capacidade para traduzi-la em uma especificação para o provedor, ou, consegue determinar se o dado espacial disponível atende às expectativas de sua aplicação.

Uma vez caracterizados os usuários do dado espacial no contexto desta aplicação, passa-se para a revisão bibliográfica necessária ao entendimento dos conceitos abordados nesta dissertação.

Duckham (1999) afirmou que, apesar de muitas pesquisas sobre a qualidade de dados espaciais, os softwares comerciais de SIG incluíam muito pouco ou nenhuma capacidade de gerenciar a qualidade de dados.

Qiu and Hunter (2002) comentou sobre a falta de ferramentas para armazenar, manipular, recuperar, atualizar e mostrar informações sobre a qualidade dos dados.

Zargar (2009) discutiu e testou vários aspectos para melhorar o gerenciamento da qualidade de dados espaciais em softwares SIG. Um desses aspectos é o uso de indicadores para comunicar a qualidade do dado espacial aos usuários e se utiliza do Quality Information Management Model (QIMM) desenvolvido por Dévillers et al. (2005) para gerenciar a informação da qualidade especial nos diferentes níveis de detalhes dos dados.

Ainda segundo Zargar *op. cit.*, SIGs são usados sem ter uma maneira de comunicar a qualidade dos dados e, segundo (Devillers *et al.*, 2005; 2007), muitos dos atuais usuários de SIG não têm experiência em usar informação geográfica.

Boin e Hunter (2006) afirmaram que, à época, exceto por poucas ferramentas acadêmicas usadas como extensões aos softwares SIGs, esses softwares não permitiam a comunicação da qualidade do dado espacial. Ressaltaram que metadados são ferramentas importantes para descrever a qualidade de dados aos consumidores e apresentaram uma estratégia para comunicar a qualidade da informação espacial que permite ao consumidor decidir se essa informação é ou não possível de se usar um particular conjunto de dados. A pesquisa explora a experiência do consumidor de dados e fornece uma comunicação da qualidade da informação via web, usando a terminologia do usuário (e não a da indústria), a solução proposta foca em descrever a adequação e a confiabilidade do produto ao invés do método de produção.

Dévillers *et al.* (2010), fizeram um artigo de revisão sobre os últimos 30 anos de pesquisas acadêmicas sobre qualidade de dados espaciais e tentaram identificar os sucessos, fracassos e oportunidades para futuras pesquisas. Ressaltaram também, o aspecto do despreparo dos usuários e afirmaram que a publicação dos metadados é o principal aspecto que tem sido usado pelos produtores de dados, nas últimas duas décadas, para comunicar a qualidade do dado espacial.

As pesquisas desenvolvidas por Lush *et al.* (2012) mostraram que conjuntos de dados com arquivos de metadados completos, boa reputação do provedor e dos

dados, são vistos como de “alta qualidade” pelos usuários, que confiam nessas informações.

Harding (2006) afirmou que, tradicionalmente, o produtor dos dados é o único responsável por definir e prover sua qualidade. Entretanto, outros trabalhos como os de Devillers e Jeansoulin (2006) e Vasseur *et al.* (2006), mostram a necessidade de se considerar o ponto de vista do usuário para se determinar se algum dado é apropriado para o uso previsto, conforme postulado por Chrisman (1991):

“A qualidade de uma base de dados digital para GIS está associada à sua adequação a um determinado uso”.

Essa visão está apresentada em Quintanilha (2002) e uma metodologia para fazer o controle sob essa óptica está em Ho *et al.* (2007) e Quintanilha *et al.* (2008), advogando que há uma hierarquia entre os aspectos a serem testados a ser seguida quando se controla a qualidade especificada pelo usuário.

Pierkot *et al.* (2011), David e Fasquel (1997) foram além e especificaram que a definição de qualidade sob a óptica do produtor ou do usuário segue duas qualificações distintas:

- qualidade interna: é o conjunto de propriedades ou características de um produto ou serviço que confere a eles a capacidade de satisfazer as especificações do seu conteúdo. É medida pela diferença entre o que deveria ter sido produzido e o que efetivamente foi feito.

- qualidade externa: é a adequação às especificações requeridas pelo usuário.

Guptill e Morisson (1995) incluíram a linhagem, a geometria, a semântica e a exatidão temporal, a completeza e a consistência lógica como aspectos a serem observados na consistência interna e foram utilizados para se estabelecer a ISO 19113.

Nesta dissertação, embora a literatura reconheça que, em geral, os usuários não tem conhecimento específico para avaliação da qualidade de dados espaciais, parte-se do princípio de que o usuário tem habilidades suficientes para determinar os requisitos mínimos que sua aplicação requer. Apropria-se da visão tradicional na

literatura que divide o controle da qualidade de dados espaciais em qualidade interna e externa e apresenta protótipo de software para o controle de algumas dos objetos mais comumente encontrados em bases de dados urbanos.

Apesar de muitos usuários associarem o termo “qualidade” somente com a acurácia espacial (posicional) do dado coletado, o conceito de qualidade engloba um espectro muito maior. Ele afeta todo o processo de aquisição, gerenciamento, comunicação e uso do dado espacial (KUMI-BOATENG; YAKUBU, 2010).

Um dado espacial poderá ser considerado adequado ou não a um determinado propósito depois de submetido a uma avaliação de sua qualidade (NOGUEIRA JUNIOR, 2003).

Uma avaliação de qualidade de um dado espacial para ser considerada eficiente deverá ser coerente com o uso ao qual este dado será aplicado. Isto implica em habilitar o usuário em definir quais critérios melhor avaliam o dado espacial em relação à aplicação a qual ele será submetido. Por exemplo, se o usuário precisa verificar a completeza de um dado em relação ao seu esquema de dados¹, é necessário que o usuário consiga descrever de maneira estruturada quais atributos são importantes para sua aplicação. Além disso, um método de avaliação de dados espaciais também deve permitir que o resultado da avaliação seja comparado com a qualidade informada pelo provedor do dado e com a qualidade avaliada por outro usuário. Desta forma, torna-se possível determinar a conformidade do dado espacial.

Uma avaliação da qualidade de dados espaciais que produza resultados que possam ser comparáveis com os resultados de outras avaliações deve ser realizada sob as mesmas normas e padrões de qualidade (DÉVILLERS, *et al.*, 2007).

Padrões em qualidade de dados apresentam uma base comum de conceituação de qualidade e possibilitam a comunicação estruturada da qualidade da informação espacial (CHRISMAN, 1994).

¹ Esquema de Dados: descrição da estrutura física das tabelas que compõe a informação espacial. Por exemplo, nome das colunas, tipo de dado de cada coluna (inteiro, real, caractere, etc.).

2.1. PADRÕES EM QUALIDADE DE DADOS

O uso generalizado da informação espacial para tomada de decisões, navegação, planejamento, desenvolvimento e serviços de emergência, levam a considerações relacionadas ao risco de se utilizar informação inadequada. O fácil acesso a fontes de dados secundários e a rápida evolução e uso de tecnologias de comunicação como a internet, potencializam este risco (GOODCHILD, 1995), (SHI; FISHER; GOODCHILD, 2002) e (GUPTILL; MORRISON, 1995).

Padrões proveem um método comum para adquirir, gerenciar e apresentar uma informação. Eles são necessários para manter o conceito de um sistema aberto e uma aplicação extensível (CAPRIOLI, *et al.*, 2003). Para qualquer organização é possível maximizar o uso e aplicação de dados digitais pela utilização de padrões internacionalmente aceitos, ou formatos proprietários largamente utilizados (*de facto Standards*).

O primeiro padrão de qualidade de dados espaciais importante foi publicado em 1994 nos Estados Unidos, requerendo todas as agências americanas a coletar e produzir dados espaciais para incluir informação a respeito da qualidade do dado em seus produtos (FORESMAN; WIGGINS; PORTER, 1996). O relatório "Content Standard for Digital Geospatial Metadata" publicado pela Federal Geographic Data Committee-FGDC (FEDERAL GEOGRAPHIC DATA COMMITTEE, 1998) contém diretrizes para organizações incluírem metadados descritivos com seus dados espaciais. Outro padrão largamente utilizado é o publicado pela International Organization for Standardization - ISO (ISO, 2003) e foi adotado nos Estados Unidos como um substituto para o FGDC.

Os elementos de qualidade descritos no FGDC (1998) e na ISO 19113 (2002) compartilham cinco elementos em comum: linhagem, acurácia posicional, acurácia temática (de atributos), consistência lógica e completude.

Tais elementos de qualidade, conforme apresentado pela ISO 19113, definem os critérios avaliação de qualidade interna do dado espacial. Estes critérios avaliam o nível de similaridade entre o dado produzido e o dado *perfeito* ou, em outras palavras, o dado produzido sem erros (DÉVILLERS; JEANSOULIN, 2006).

Em contraste, Dévillers e Jeansoulin (2006) neste mesmo trabalho, apresentam a definição de qualidade externa como a "totalidade de características de um produto que busca satisfazer necessidades implícitas e explícitas", desta forma, a qualidade externa está preocupada com a adequação ao uso, ou adequação ao propósito de um produto (CHRISMAN, 1994).

Uma questão pertinente ao padrão de qualidade de dados é a escolha de qual padrão deve-se utilizar para avaliar a qualidade do dado. Sua escolha deve ter como objetivo a comunicação clara e a troca de informações entre empresas e agências de produção de dados e usuários. Se as empresas utilizam vários padrões diferentes a comunicação e troca de informações entre elas será complexa e provavelmente inviável.

2.2. ELEMENTOS DE QUALIDADE DE DADOS

Em 1982 foi estabelecido nos Estados Unidos estabeleceram o National Committee on Digital Cartographic Data Standards (NCDCCDS), Comitê Nacional em Padrões para Cartografia Digital em uma tradução livre, que elaborou a primeira versão de um relatório denominado "A Draft Proposed Standard for Digital Cartographic Data" (GUPTILL; MORRISON, 1995). Este relatório designou cinco componentes de qualidade para bancos de dados espaciais: Linhagem, Acurácia Posicional, Acurácia de Atributos, Completude e Consistência Lógica. Estes componentes definem o esquema básico² de comunicação da qualidade do dado, definindo-a e documentando-a.

A acurácia temporal foi posteriormente adicionada à lista de elementos de qualidade reconhecidos pela ISO (ISO, 2002). A acurácia semântica foi reconhecida pelo ICA (International Cartographic Association) em 1995 (GUPTILL; MORRISON, 1995).

2.2.1. Linhagem

Linhagem do dado inclui descrições das fontes de dados, métodos que foram utilizados para criação da base de dados, incluindo todas as transformações de

² Conjunto mínimo aceitável de informações necessárias para comunicar ao usuário sobre a qualidade do dado espacial que irá utilizar.

dados e transações utilizadas neste processo (CHRISMAN, 1994). Este componente deve conter todos os dados importantes para as fontes de dados e para os processos de atualização (ISO, 2002).

O propósito final do relatório de linhagem é manter o registro histórico do dado para seus usuários (ISO, 2002, p. 48).

Usualmente os metadados de uma base de dados espacial comunicam sua linhagem, apresentando por parte do produtor do dado para seu consumidor, quais foram as transformações que este dado sofreu ao longo de seu processo de construção.

A linhagem, em alguns casos, pode ser uma consideração importante na escolha de um conjunto de dados para uma aplicação específica (ROCHA SANTOS, *et al.*, 2010).

2.2.2. Acurácia Posicional

Acurácia posicional é a acurácia da posição das feições (ISO, 2002, p. 53).

Para feições pontuais a acurácia posicional é representada pela discrepância entre a localização do dado na realidade e a localização em sua representação. Nas ciências de mapeamento a posição de uma entidade do mundo real é descrita por valores em um sistema de coordenadas apropriado. A acurácia posicional representa a proximidade entre tais valores com seu homólogo no “mundo real”.

Métricas para acurácia espacial dependem da entidade dimensional que é considerada (VEREGIN; HARGITAI, 1995). No caso de feições pontuais, é representado por meios estatísticos como erro médio quadrático, desvio padrão ou intervalo de confiança. A acurácia posicional de um ponto pode ser apresentada pela equação (1).

$$PA_p = \sqrt{\sigma_x^2 + \sigma_y^2} \quad (1)$$

Em que σ_x representa o desvio padrão na coordenada X, e σ_y representa o desvio na coordenada Y.

Para feições lineares sua métrica é mais complexa. Acurácia posicional linear é o resultado da acurácia posicional de pontos que definem as linhas ou seus segmentos utilizados como formas generalizadas da realidade. A acurácia de elementos lineares como estradas, curvas de nível, etc., são frequentemente apresentadas por ε -bands. Bandas Epsilon (ε -bands) definem uma zona de incerteza em torno da linha medida dentro da qual a linha existe com alguma probabilidade (JOKSIC; BAJAT, 2004).

No caso de feições lineares a acurácia posicional pode ser representada pela equação (2) a seguir.

$$PA_L = \sqrt[2]{n(\sigma_X^2 + \sigma_Y^2)} \quad (2)$$

Em que σ_X representa o desvio padrão na coordenada X, e σ_Y representa o desvio na coordenada Y e n representa o número de vértices da feição.

Feições poligonais são representadas como elementos lineares fechados, e sua acurácia posicional é também representada pela equação (2). Alguns modelos de representação e mensuração da acurácia poligonal foram propostos como, por exemplo, a distância de Hausdorff (STEIN; VAN OORT, 2006).

2.2.3. Acurácia de Atributos

Acurácia de atributos é definida como a confiabilidade dos valores associados a feições (camadas) em um conjunto de dados em relação aos seus possíveis valores no mundo real (ISO, 2002). A avaliação da acurácia de atributos busca indicar o quão confiável, correto ou livre de tendências o atributo está.

A acurácia qualitativa de atributos se refere ao quanto as variáveis nominais estão corretas (GOODCHILD, 1995). Por exemplo, associando um tipo de classe de uso de solo A, onde na verdade deveria ser B.

A acurácia quantitativa de atributos se refere ao nível de tendência na estimativa de valores associados às feições (VEREGIN; HARGITAI, 1995). Por exemplo, nas estimativas de valores de pH em um mapa de solos.

2.2.4. Completeza

Completeness ou Completeza é a presença ou ausência de feições, seus atributos e seus relacionamentos (ISO, 2002, pág. 51). A completeness é dividida em completeness de dados e completeness de modelo.

A completeza de dados descreve a comissão de entidades em um banco de dados espacial relacionado com sua quantidade no mundo real.

A completeza de modelo indica o nível de credibilidade de quão acurada está a base de dados. Completeness de modelo está relacionada com a acurácia semântica. É dependente de aplicação e, portanto é um aspecto de adequação ao uso (VEREGIN; LANTER, 1995).

As medidas de completeza de dados indicam o quão corretamente a informação reflete o universo definido pela sua especificação. Levando esta abordagem em consideração, duas medidas possíveis são sugeridas para quantificar a completeza de dados (CEN/TC287/WG02, 1995). Estas medidas são respectivamente:

- Omissão – Percentual de dados faltantes em relação a especificação
- Comissão – Percentual de dados presentes que não existem na especificação do conjunto de dados ou sua extração.

2.2.5. Consistência Lógica

Consistência lógica representa o grau de confiabilidade alcançado das regras lógicas e conexões em uma estrutura de dados. Consistência lógica é definida como a ausência de contradições em uma base de dados. Refere-se às regras lógicas da estrutura (topologia) e regras de atributos para dados espaciais (KAINZ, 1995). Regras derivam de teorias matemáticas e testes formais que podem ser construídos para verificar se existe consistência em diferentes camadas (feições) de um conjunto de dados. Regras topológicas para objetos espaciais não devem ser violadas (por exemplo, uma edificação não ocupa dois terrenos diferentes ao mesmo tempo) e regras de atributos (por exemplo, não existem dados censitários de uma doença em uma quadra sem população).

A consistência lógica pode ser classificada em cinco categorias diferentes:

- Consistência Conceitual
- Consistência de Domínio
- Consistência de Formato
- Consistência Topológica

2.2.6. Sumário de Elementos de Qualidade

Os elementos de qualidade de dados e seus respectivos subelementos, podem ser agrupados de acordo com o Quadro 2 a seguir:

Quadro 2 - Elementos e Subelementos de Qualidade de Dados

Elemento de qualidade	Subelemento de qualidade	de	Descrição do subelemento de qualidade
Completeza	Comissão		Excesso de dados presentes em um conjunto de dados
	Omissão		Ausência de dados em um conjunto de dados
Consistência Lógica	Consistência Conceitual		Aderência às regras do esquema conceitual
	Consistência Domínio	de	Aderência de valores com os valores de domínio
	Consistência Formato	de	Nível no qual o dado está armazenado em concordância com a estrutura física do conjunto de dados.
	Consistência Topológica		Correção das características topológicas explicitamente codificadas no conjunto de dados
Acurácia Posicional	Acurácia externa ou absoluta	ou	Proximidade dos valores de coordenadas reportados com os valores aceitos como verdadeiros
	Acurácia interna ou relativa	ou	Proximidade das posições relativas das feições em um conjunto de dados em relação as suas respectivas posições aceitas como verdadeiras.
	Acurácia de dados posicionados e grade		Proximidade dos valores de posição de dados em grade com os valores aceitos como sendo verdadeiros
Acurácia Temporal	Acurácia de uma medida de tempo		Correção da referência temporal de um item (relato de erro em medida de tempo)
	Consistência temporal		Correção de eventos ordenados ou seqüências, se relatado.

	Validade Temporal	Validade do dado em relação ao tempo
Acurácia de Atributos	Correção de classificação	Comparação de classes associadas as feições ou seus atributos com o universo em discurso (isto é, com a verdade ou dados de referência)
	Correção de atributos não quantitativos	Correção de atributos não quantitativos
	Acurácia de atributos quantitativos	Acurácia de atributos quantitativos

O Quadro 2 apresenta a compilação dos elementos e subelementos de qualidade conforme apresentados pelas ISO 19113 e 19114.

2.3. METADADOS

Metadados são dados a respeito de dados. Metadado é uma documentação de dados. Tem sido primariamente utilizado como informação para permitir gestores de dados ou usuários em entender, comparar e trocar o conteúdo do conjunto de dados descrito pelo Metadado (ISO, 2003).

Com a introdução de padrões em metadados, um crescente número de produtores de dados os tem utilizado para descrever a qualidade de seus produtos (DÉVILLERS, *et al.*, 2002). No Brasil, a utilização de metadados contendo descritores de qualidade de dados tem sido adotada principalmente por organizações públicas (GT1/CEMG, 2009).

Entretanto, prover informação de qualidade de dados ainda é opcional na maioria dos padrões disponíveis tais como a norma ISO 19115, FGDC e ANZLIC, o que pode limitar a documentação da qualidade do dado de forma prática (DÉVILLERS, *et al.*, 2007), (DÉVILLERS, *et al.*, 2002), (BOIN; HUNTER, 2006).

Metadados podem ser publicados em arquivos em formato de texto simples, bancos de dados ou arquivos XML³. Este último foi adotado como estrutura de armazenamento dos dados de qualidade. A ISO 19139 (ISO, 2009) recomenda o formato XML como formato de gravação para metadados no padrão ISO 19115.

³ XML – eXtensible Markup Language (<http://www.iee.org>)

Boin e Hunter, (2006), entretanto, argumentam que enquanto os elementos de qualidade, listados anteriormente neste mesmo capítulo, podem ser apropriados para registrar dados, ainda permanece a questão de quão apropriados os metadados são para a comunicação da qualidade do dado ao usuário. Dévillers, *et al.*, (2007) argumentam que a maioria dos elementos de qualidade está concentrada nas características relacionadas ao produtor de dados (sob a perspectiva do produtor de dados) e que é necessário incluir mais informação para alcançar a definição de "adequação ao uso".

2.4. PADRÕES DE METADADOS GEOESPACIAIS

Vários padrões de metadados geoespaciais digitais estão disponíveis atualmente no mercado. Muitos deles estão operacionais e são suportados por ferramentas de software para permitir o cadastramento de seus metadados e esquemas de metadados para sua disseminação.

O documento Perfil de Metadados Geoespaciais do Brasil (CEMG-CONCAR, COMITÊ DE ESTRUTURAÇÃO DE METADADOS GEOSPACIAIS, 2009), estabelece que o padrão de metadados em implementação na Infraestrutura Nacional de Dados Espaciais seja baseado na norma ISO 19115:2003. O perfil de metadados geográficos do Brasil passa a ser então, uma especialização⁴ desta norma.

Os padrões de metadados são importantes, pois determinam a estrutura de comunicação da qualidade dos dados espaciais. Tais padrões determinam como serão registradas as informações sobre a qualidade do dado e como tais informações serão transmitidas para outros usuários.

No Quadro 3 a seguir apresentam-se os principais padrões de metadados internacionais propostos pelas organizações e grupos de estudos e desenvolvimento (WEBER, *et al.*, 1999).

⁴ Entende-se por especialização da ISO 19115, a utilização desta norma como um marco referencial, criando à partir dela, especializações da norma para atendimento de necessidades regionais.

Quadro 3 - Padrões de Metadados Geoespaciais (WEBER, et al., 1999)

Spatial Data Transfer Standard - SDTS	Desenvolvido pelo Federal Geographic Data Committee (FGDC) nos Estados Unidos. Foi desenvolvido para permitir às agências federais dos Estados Unidos o compartilhamento de dados espaciais entre aplicações.
Content Standards for Digital Geospatial Metadata	Desenvolvido pelo Federal Geographic Data Committee (FGDC) dos Estados Unidos. Foi desenvolvido em 1994 e entrou em vigor em 1995. O padrão foi desenvolvido para prover um conjunto comum de terminologias e definições para a documentação ligada a esses metadados.
Spatial Archive and Interchange Format - SAIF	Foi desenvolvido pela Divisão de Levantamento e Mapeamento de Recursos do Ministério do Ambiente Território e Parques da Columbia Britânica (Canadá). O padrão SAIF foi concebido para facilitar a interoperabilidade, particularmente no contexto da troca de dados.
CEN TC287	Desenvolvido pelo Comitê Técnico 287 (TC287) do Comitê Europeu de Padronização (CEN). O padrão CEN define um conjunto mínimo de metadados que devem ser providos pelos produtores/fornecedores de dados.
ANZLIC	O Padrão foi proposto pelo Australia New Zealand Land Information Council, baseado no padrão americano CSDGM, mas utiliza um subconjunto menor dos elementos de metadados.
GCMD/NASA	O Global Change Master Directory (GCMD) foi desenvolvido pela NASA, como parte do National Space Science Data Center (NSSDC) para promover a troca de conjuntos de dados científicos através do projeto Catalog Interoperability (CI)
Dublin Core	Especificação de metadados para catálogo de documentos eletrônicos. Atualmente, a intenção deste padrão é definir um conjunto pequeno e universalmente aceito de elementos de metadados que permita a autores e provedores de informação espacial a descrever seu trabalho.
ISO 19115	Especificado pelo Comitê Técnico 211 (TC 211) da ISSO, a ISO 19115:2003 (Geographic Information – Metadata) faz parte de uma família de várias normas para informação geográfica e suporta o referenciamento espacial. Utiliza a modelagem UML para representar suas seções, entidades e elementos de metadados. Combinando aspectos de vários metadados, esta norma visa tornar-se um padrão universal para o

	armazenamento e distribuição de metadados geoespaciais.
--	---------------------------------------------------------

Padrões de metadados geoespaciais, tais como representados no Quadro 3, são importantes, pois apresentam as diretrizes de comunicação da qualidade do dado espacial conforme determinadas pelos Padrões em Qualidade de Dados, conforme descritos neste mesmo capítulo, no item 2.1. Tais padrões detalham como cada elemento de qualidade será armazenado, descrito e transmitido nos metadados.

2.5. AVALIAÇÃO DE QUALIDADE DE DADOS ESPACIAIS

Este capítulo apresenta métodos descritos na literatura para avaliação da qualidade do dado espacial. Apesar da natureza subjetiva do termo qualidade do dado espacial, os métodos descritos buscam apresentar um enquadramento que permita quantificar a qualidade do dado espacial.

Uma avaliação da qualidade, mesmo que realizada em sua plenitude, não eliminaria todos os riscos decorrentes da existência de dados imprecisos (FRANK, 2008). Mas é possível que tanto o produtor de dados, quanto o seu usuário estabeleçam um limite aceitável para os riscos relacionados ao uso de uma informação imprecisa ou incorreta. O usuário poderá estabelecer, por exemplo, um limite de 10% de registros com erro de omissão como um limite aceitável para aceitação de uma feição em uma base de dados espaciais.

Um procedimento de avaliação de qualidade é realizado através da aplicação de um ou mais métodos de avaliação. Os métodos de avaliação são divididos em duas classes principais (ISO, 2003): Métodos Diretos e Métodos Indiretos.

Métodos Diretos determinam a qualidade através da comparação do dado com uma informação de referência interna ou externa.

Métodos Indiretos inferem ou estimam a qualidade do dado utilizando informação sobre o dado, por exemplo, a informação de linhagem. A Figura 2 representa esta estrutura de classificação.

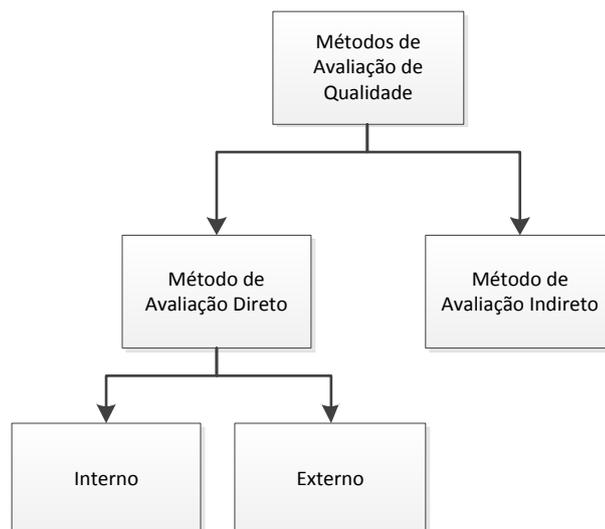


Figura 2 - Classificação dos métodos de avaliação de qualidade (ISO, 2003).

No método de avaliação direto, utilizando dados de qualidade interno do conjunto de dados analisado, toda informação necessária para realizar a avaliação está interna a este conjunto de dados. Como exemplo pode-se citar que toda informação necessária para realizar um teste de topológico de fechamento de polígonos reside na estrutura topológica do próprio conjunto de dados.

Avaliação Direta utilizando dados externos necessita de informação de referência externa ao conjunto de dados sendo testado. Por exemplo, para verificação da completeza de conjunto de dados de nomes de ruas é necessário utilizar outra fonte de informações de segmento de logradouros.

No método de avaliação indireta da qualidade do conjunto de dados, a avaliação em si é baseada em conhecimento ou informação externo. Tal conhecimento pode incluir, mas não estar limitado, a uma avaliação da qualidade do dado por outros meios (avaliação de peritos ou avaliadores, por exemplo). Um exemplo de avaliação por método indireto é a avaliação da acurácia posicional do dado. Nesta avaliação, a posição espacial do dado é comparada com uma referência externa, ou seja, com o resultado de um levantamento topográfico, um levantamento por GPS ou pela comparação com outro dado com qualidade reconhecida como superior.

3. MATERIAIS E MÉTODOS

Este capítulo identifica os materiais e métodos que foram utilizados nas atividades relacionadas a esta dissertação.

3.1. MATERIAIS

Este capítulo identifica os materiais utilizados nos testes e avaliações realizados pelo aplicativo desenvolvido para esta dissertação.

3.1.1. Área de Estudo

Duas bases de dados espaciais distintas foram utilizadas nesta dissertação:

- Base de dados da cidade de São Paulo
- Base digital da cidade de Itapeva

3.1.1.1. Base de Dados da Cidade de São Paulo

A base de dados da cidade de São Paulo⁵ foi gerada pela empresa provedora a partir da vetorização de imagens Ikonos e datam do ano de 2008.

O formato utilizado pelo provedor de dados é o ESRI Shapefile⁶ e MapInfo TAB file⁷ e suas feições possuem, além das geometrias, atributos alfanuméricos.

A generalização cartográfica aplicada a esta base de dados utilizou como referência a escala 1:150.000⁸.

O Quadro 4 apresenta os temas e o tipo de suas feições existentes na base de dados.

⁵ A base de dados da cidade de São Paulo foi fornecida pela empresa Multispectral, pelo Engenheiro Cartógrafo Elivagner Barros. O uso desta base de dados é restrito para fins acadêmicos.

⁶ ESRI Shapefile é um formato aberto de dados espaciais, contendo geometrias e atributos alfanuméricos desenvolvidos pela empresa americana ESRI (Environmental Systems Research Institute).

⁷ MapInfo TAB file é um formato vetorial proprietário de dados espaciais, capaz de armazenar geometrias e atributos. Este formato foi desenvolvido e regulamentado pela empresa MapInfo.

⁸ Estas informações foram providas verbalmente pelo provedor. Não existem metadados de qualidade para tais bases.

Quadro 4 - Temas e seus tipos utilizados para testes do método e aplicativo de avaliação de dados

Tema	Geometria
Logradouros	Linear
Avenidas	Linear
Hidrografia	Linear / Poligonal
Estações de metrô	Pontual
Bairros	Poligonal
Limite político	Poligonal

3.1.1.2. Base Digital da cidade de Itapeva

A base digital da cidade de Itapeva é disponibilizada livremente pelo site de sua prefeitura municipal.

A cidade de Itapeva está localizada a aproximadamente 300 km da cidade de São Paulo-SP.

O mapa de Itapeva-SP foi disponibilizado no formato DWG⁹ e pode ser baixado de <http://www.itapeva.sp.gov.br/downloads-iss/#servicos>.

As camadas de informação disponibilizadas pelos dados digitais da cidade, e utilizadas por esta dissertação estão apresentadas no Quadro 5 abaixo.

Quadro 5 - Temas existentes na base digital de Itapeva.

Tema	Geometria
Uso do solo	Poligonal
Logradouros	Linear
Quadras	Poligonal
Perímetros de bairros	Poligonal
Áreas de Proteção	Poligonal
Edificações	Poligonal
Bairros	Poligonal
Hidrografia	Linear

Não há informações de linhagem ou metadados de qualidade informados no site da prefeitura de Itapeva.

⁹ DWG é o formato do arquivo de desenho 2D e 3D do fabricante de software AutoDESK.

3.1.1.3. Considerações Sobre os Dados e Formatos de Dados Utilizados

Esta dissertação utilizou os formatos descritos neste documento (Esri shapefile, MapInfo TAB, DWG e DGN) pelos seguintes motivos:

- Tais formatos são utilizados pelas principais agências de mapeamento no país e são largamente utilizados por empresas privadas de mapeamento.
- Cada formato de arquivo citado possui uma estruturação diferente para armazenamento de seus dados.
 - O arquivo shapefile, por exemplo, admite apenas um tipo de geometria para suas feições.
 - O arquivo DWG e DGN por serem formatos CADD armazenam atributos de banco de dados de maneira diferente dos formatos shapefile e TAB

Para realização dos testes de avaliação de dados apresentados nesta dissertação, não foram utilizados todos os temas existentes nas bases de dados.

Os temas utilizados foram:

- Base de São Paulo
 - Eixos de Logradouro
- Base de Itapeva
 - Hidrografia

Tais temas foram utilizados por serem as informações mais relevantes em ambas as bases, e foram suficientes para realização dos testes de laboratório necessários para uma primeira validação do aplicativo.

3.1.2. Recursos de software

Para desenvolvimento dos trabalhos descritos nesta dissertação foram utilizados os seguintes softwares:

Para avaliação e manuseio de arquivos shapefile foi utilizado o software MapWindow¹⁰ e suas bibliotecas de programação.

Para acesso e manuseio dos arquivos foi utilizado o software Bentley Map Enterprise¹¹ e suas bibliotecas de programação.

Para codificação do protótipo de software para suporte a avaliação do dado espacial foi utilizada a linguagem de desenvolvimento C# no ambiente de programação provido pelo Microsoft Visual C# 2010 Express¹².

3.2. MÉTODOS

Neste capítulo é detalhada a metodologia aplicada no processo de avaliação da qualidade do dado espacial. Serão sugeridos métodos de avaliação para os elementos de qualidade utilizados nesta dissertação.

3.2.1. Premissas

A metodologia terá as seguintes premissas:

- O avaliador da qualidade do dado possui experiência no reconhecimento de elementos de qualidade.
- Os dados espaciais não precisam ter metadados de qualidade *a priori*.
- Os dados espaciais deverão estar no formato vetorial e podem possuir atributos alfanuméricos.
- Os dados espaciais poderão estar em formato CADD (DWG e DGN) ou SIG (ESRI shapefile ou MapInfo TAB). Para ambos os formatos, quando existirem

¹⁰ O software MapWindow pode ser baixado gratuitamente do site <http://www.mapwindow.org/>

¹¹ Software de mapeamento de propriedade da Bentley Systems Inc. A licença de uso do software Bentley Map Enterprise foi cedida gratuitamente pela empresa Bentley Systems ao aluno Joao Salisso pelo período de um ano.

¹² O Visual C# 2010 Express é gratuito e pode ser baixado do site <http://www.microsoft.com/visualstudio/en-us/products/2010-editions/visual-csharp-express>

atributos alfanuméricos, eles deverão estar armazenados em uma base de dados estruturada.

3.2.2. Avaliação da qualidade de dados espaciais

O método de avaliação de qualidade de dados espaciais proposto por esta dissertação é apresentado na Figura 3.

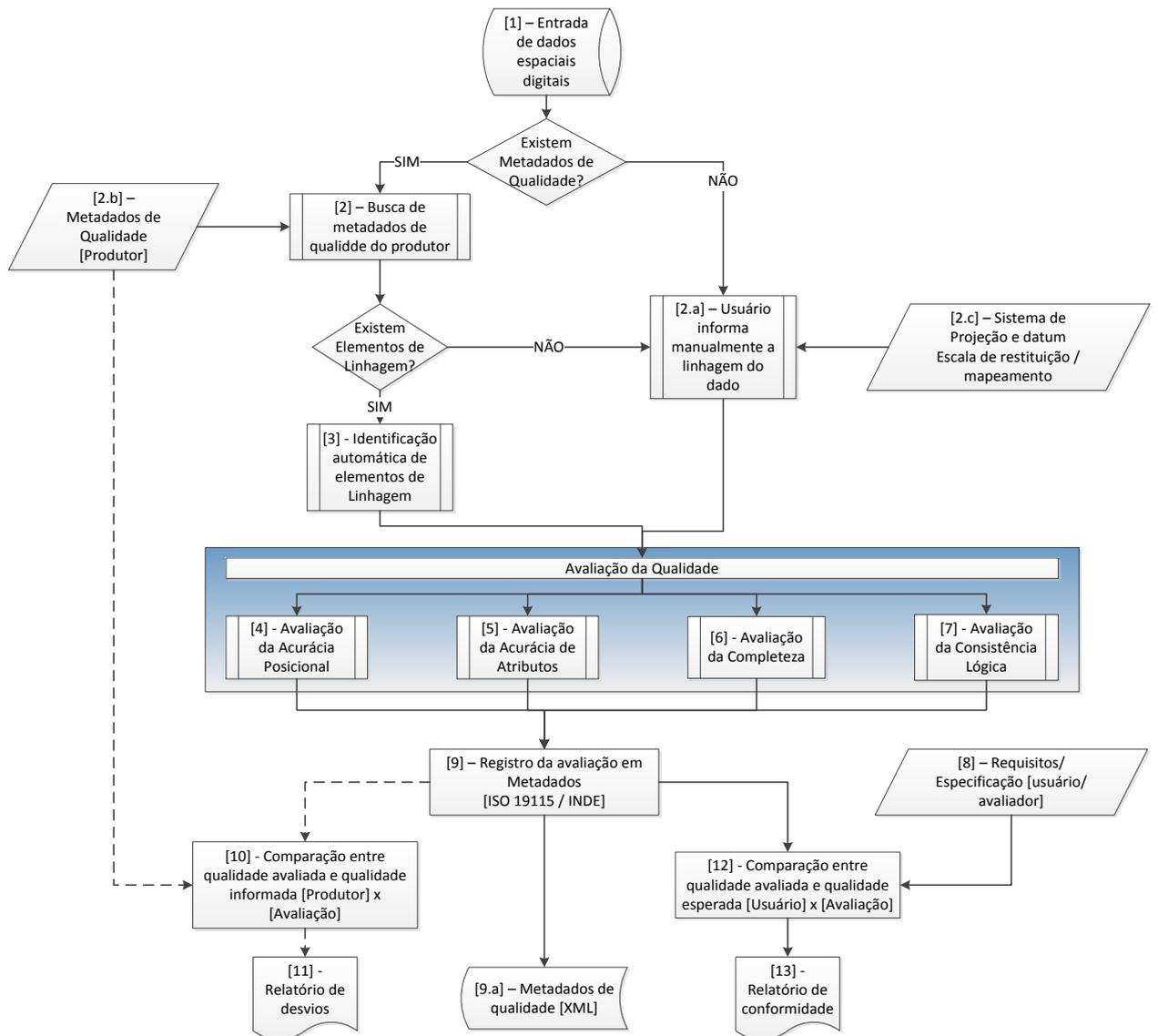


Figura 3 - Fluxo metodológico

3.2.2.1. Descrição do fluxo metodológico

O fluxo metodológico (Figura 3) apresentado a seguir será apoiado por protótipo de programa computacional. Este protótipo armazenará as informações de qualidade de dados levantadas pelo usuário durante o processo de avaliação.

[1] Entrada de dados espaciais digitais

O processo de avaliação tem início com a seleção do dado espacial. O processo de entrada de dados espaciais irá verificar se existem ou não metadados vinculados à base de dados em avaliação.

[2] Busca de Metadados de Qualidade [Produtor]

Caso existam metadados vinculados à base de dados digital em avaliação, este processo irá identificar os elementos de qualidade que são essenciais a avaliação da acurácia posicional. Os elementos necessários para esta avaliação são o sistema de projeção, datum e escala de mapeamento.

[2.b] Metadados de Qualidade [Produtor]

Este processo apresenta a leitura dos metadados existentes pelo protótipo de software. Em geral, o arquivo de metadados está armazenado em formato XML, na mesma pasta onde o dado espacial está armazenado.

O protótipo de software irá apoiar este processo devido à quantidade de informações existentes em um arquivo de metadados e devido à sua estrutura de leitura complexa, que tornaria impossível sua compreensão sem o apoio desta ferramenta.

[3] – Identificação automática de elementos de linhagem

Neste processo, o protótipo identifica os metadados de linhagem que acompanha a base de dados espacial. Existindo tais metadados, o aplicativo automaticamente os recupera e apresenta ao usuário. Os elementos que o aplicativo identifica são: sistema de projeção, datum e escala de mapeamento.

[2.a] – Usuário informa manualmente a linhagem do dado

Caso não existam metadados sobre qualidade ou, caso não existam informações sobre a linhagem nos metadados existentes, o avaliador do dado deverá obtê-las

por outros meios. O protótipo de software irá apoiar o usuário na entrada manual destas informações. Estas informações serão salvas pelo protótipo para uso posterior.

[2.c] – Sistema de Projeção e datum, Escala de restituição / mapeamento

Este processo representa a entrada manual da informação de sistema de projeção, datum e escala de mapeamento pelo usuário.

Se a informação a respeito do sistema de projeção e datum não estiverem disponíveis, a avaliação da acurácia posicional estará comprometida.

Avaliação da Qualidade

O núcleo principal de avaliação de qualidade do dado espacial, apresentado no fluxo metodológico, irá considerar três dos quatro elementos apresentados.

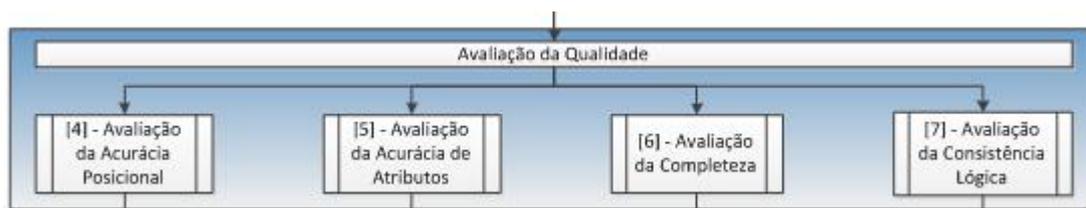


Figura 4 - Núcleo principal de avaliação do fluxo metodológico

No processo de avaliação da qualidade, o avaliador identificará os elementos de qualidade que melhor avaliam o dado de acordo com o propósito que lhe será atribuído. Nesta dissertação os elementos de qualidade de dados serão limitados a quatro, a saber:

- [4] Avaliação de Acurácia Posicional. Esta avaliação realiza a verificação da acurácia posicional dos elementos gráficos da base de dados.
 - o Esta avaliação não foi implementada no aplicativo, pois esta avaliação demandaria a implementação de algoritmos que fogem ao objetivo desta dissertação.
- [5] Avaliação da Acurácia de Atributos. Esta avaliação realizou a consistência dos atributos físicos das geometrias e suas definições de armazenamento na base de dados.

- [6] Avaliação da Completeza. Esta avaliação realizou a verificação de casos omissão e comissão de dados.
- [7] Avaliação da Consistência Lógica. Esta avaliação realizou a verificação de inconsistências de conceituação e inconsistências de formato.

Todos os processos de avaliação seguiram o mesmo fluxo de execução, conforme mostrado na Figura 5.

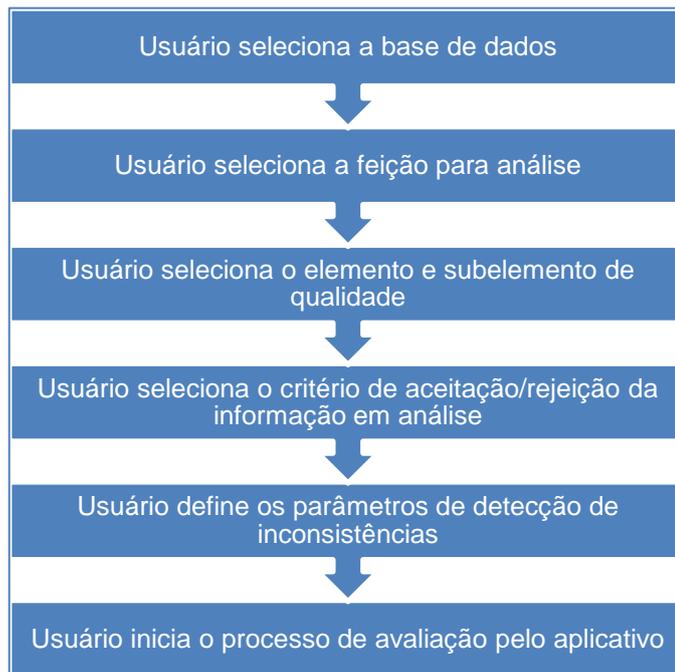


Figura 5 - Fluxo de execução do processo de avaliação de qualidade

Conforme a Figura 5, o usuário selecionará a base de dados e a feição para avaliação de sua qualidade. Em seguida, o usuário definirá os critérios de avaliação selecionando elemento e subelemento de qualidade que ele deseja aplicar. Por exemplo, o usuário poderá selecionar **COMPLETEZA** como elemento e, **OMISSÃO** como subelemento de qualidade.

O próximo passo no fluxo de execução é a definição do critério que o usuário deseja utilizar para aceitar ou rejeitar o dado em avaliação. Ainda no exemplo anterior, caso 10% dos registros em análise, possuam erros de **omissão**, a feição será rejeitada pelo processo.

Em seguida, o usuário definirá quais critérios serão utilizados para detectar as inconsistências em sua base de dados. No exemplo anterior, na análise de erros de comissão de uma base de dados, o usuário poderá definir que o atributo **nome de rua**, da feição **eixos de logradouro**, será auditado em busca de registros nulos.

Uma vez concluídas as definições, o dado estará pronto para ser avaliado.

[5] Avaliação da Acurácia de Atributos

O processo de avaliação da acurácia de atributos irá realizar avaliações não quantitativas de acurácia de atributos.

Esta avaliação irá buscar elementos cujos atributos físicos não estão em conformidade com a especificação da feição em análise. Por exemplo, os elementos qualificados como pertencentes a feição **HIDROGRAFIA** devem possuir a cor azul, estilo de linha contínua, espessura 0 e devem estar armazenadas na camada **RIOS**.

O processo de avaliação irá então buscar todos os elementos qualificados como hidrografia, mas que não cumpram com aquelas definições.

[6] Avaliação da Completeza

A avaliação da completeza realizará avaliações em busca de erros de omissão e comissão.

Esta avaliação será limitada a avaliações nos atributos da base de dados e seu esquema. Não serão avaliadas comissões ou omissões em geometrias.

As seguintes avaliações de completeza foram implementadas:

- Completeza
 - De Esquema de Dados
 - Omissão
 - Comissão
 - De Registros
 - Omissão

A avaliação de omissão de esquema de dados detecta a ausência de atributos definidos pelo usuário como necessários na base de dados. Por exemplo, o usuário

pode definir que o atributo CEP da rua é importante para sua aplicação, caso este atributo não exista na tabela associada a feição de ruas, a feição será rejeitada.

A avaliação de omissão de registros de dados detecta a ausência de valores nos atributos da feição. Por exemplo, o usuário define que o atributo CEP não pode conter mais de 10% de registros sem esta informação, se o processo detectar mais que 10% de registros com este problema a feição será rejeitada.

A avaliação de comissão de esquema de dados detecta o excesso de atributos na base de dados. Por exemplo, o usuário define que apenas os atributos NOME DE RUA, CEP, NUMERO INICIAL e NUMERO FINAL devem existir na tabela associada a feição **eixos de logradouro**. Caso outros atributos existam na tabela, a feição será rejeitada.

[7] Avaliação da Consistência Lógica

A avaliação da consistência lógica utilizará os seguintes critérios de avaliação: Consistência conceitual, consistência de domínio e consistência de formato.

Os critérios de avaliação de consistência conceitual, consistência de domínio e consistência de formato poderão ser aplicados em bases de dados SIG (ESRI shapefile ou MapInfo TAB file), o critério de avaliação de consistência de domínio não pode ser aplicada ao formato CADD se não houver uma tabela de dados associada às suas feições.

Consistência Conceitual

A avaliação de consistência conceitual busca elementos na base de dados que não estão em conformidade com a definição da feição em análise. Por exemplo, a feição **eixos de logradouro** deve possuir elementos lineares apenas. Caso elementos pontuais ou poligonais sejam encontrados a feição será rejeitada.

Consistência de Domínio

A avaliação de consistência de domínio busca registros na base de dados onde seus valores devam seguir um esquema definido. Por exemplo, o atributo **material**, da feição **tubulação** deverá possuir os seguintes valores: **PE, CONCRETO, FERRO**

FUNDIDO. Caso a incidência de registros cujos valores não estejam previstos nesta lista de materiais extrapole o limite estabelecido pelo usuário, a feição será rejeitada.

Consistência de Formato

Esta avaliação irá verificar se o formato do dado segue a especificação definida pelo usuário. Por exemplo, caso a aplicação do usuário necessite receber o dado em formato DWG, e o dado em avaliação estiver no formato MAPInfo TAB file, a base de dados será rejeitada.

[9] – Registro da avaliação em metadados [ISO 19115]

Uma vez finalizado o processo de **Avaliação de Qualidade**, o protótipo de software irá armazenar o resultado da avaliação de cada elemento de qualidade segundo a norma ISO 19115:2003.

[9.a] – Metadados de Qualidade [XML]

Este processo representa o armazenamento dos metadados de qualidade em formato XML nos padrões ISO 19115:2003.

[10] – Comparação entre qualidade avaliada e qualidade informada [Produtor] x [Avaliação]

Neste processo, o protótipo de software irá apresentar as informações de qualidade de dados recuperadas pelos processos [2] ou [2.c] caso existam e, as informações do resultado do processo de **Avaliação da Qualidade**, comparando a qualidade do dado no ponto de vista do produtor de dados contra a qualidade do dado de acordo com a avaliação executada pelo avaliador.

[11] – Relatório de desvios

O processo de comparação dos metadados de qualidade executados pelo processo [10], resultarão em um relatório de desvios, identificando os pontos de divergência entre a qualidade informada pelo produtor do dado e a qualidade avaliada pelo usuário.

[8] – Requisitos/Especificação [usuário/avaliador]

Neste processo, o avaliador do dado irá apresentar os níveis de qualidade, na forma dos elementos de qualidade, que são requeridos pela sua aplicação. Por exemplo, o usuário poderá informar que, para sua aplicação é necessário que a acurácia posicional da base de dados esteja na ordem de ± 50 cm, e que pelo menos 95% dos registros existentes na base de dados não possuam problemas de acurácia de atributos. Os outros elementos, para esta aplicação, são desnecessários.

Estas informações formam o nível de conformidade requerido pela aplicação explicitado pelo usuário.

Obrigatoriamente, os elementos de qualidade que o avaliador irá informar serão os mesmos que foram utilizados no processo de **Avaliação da Qualidade**.

[12] – Comparação entre a qualidade avaliada e a qualidade esperada [Usuário] x [Avaliação]

De maneira análoga ao processo **[10]**, o protótipo de software irá apresentar ao avaliador da qualidade do dado espacial os resultados do processo de **Avaliação da Qualidade** comparando-os com as informações referentes ao nível de conformidade definida pelo usuário, conforme descrito no processo **[8]**. Permitindo ao avaliador comparar o resultado da avaliação da qualidade de dados, contra a qualidade requerida pelo usuário.

[13] – Relatório de conformidade

O processo de comparação entre a qualidade avaliada e a qualidade esperada pelo usuário, conforme descrito no processo **[12]**, irá resultar em um relatório de conformidade. Este relatório apontará as divergências entre a qualidade avaliada do dado espacial e a qualidade esperada pelo usuário para sua aplicação.

4. RESULTADOS E DISCUSSÃO

Este capítulo detalha os resultados obtidos nos processos de avaliação que foram aplicados contra a base de dados, seus tempos de processamento e comunicação dos resultados.

Nem todas as feições existentes nas bases de dados foram avaliadas nesta dissertação. Uma vez que o propósito desta dissertação foi apresentar um método automatizado de avaliação e não apresentar a qualidade das bases em utilização, aplicou-se o método de avaliação nas feições mais significativas das bases, tais como:

- Base de São Paulo
 - Eixos de Logradouro
- Base de Itapeva
 - Hidrografia

Ainda com relação à avaliação das bases de dados utilizadas por esta dissertação, é importante ressaltar que:

- Os critérios de avaliação para determinar a qualidade do dado espacial contido nas bases de dados são subjetivos e estão vinculadas à necessidade do usuário.
- Uma mesma base de dados pode ser considerada como de qualidade por um usuário e de não qualidade por outro, pois os critérios de avaliação de cada usuário são totalmente dependentes do propósito ao qual o dado será aplicado.

Logo, esta dissertação não apresentará o dado como sendo de qualidade ou de não qualidade, mas adequado ou não à aplicação de um determinado usuário.

4.1. AVALIAÇÃO DE COMPLETEZA DE DADOS

A avaliação da Completeza de Dados utilizou a feição de ***eixos de logradouro*** existente na base de dados de São Paulo. Esta feição foi selecionada por conter aproximadamente 200.000 registros (atributos e elementos geométricos)

O procedimento de avaliação verificou a existência de erros de omissão em registros, omissão em esquema de dados e comissão em esquema de dados na feição citada.

4.1.1. Erros de Comissão e Omissão em Esquema de Dados

A avaliação de erros de comissão no esquema de dados buscou detectar atributos na feição ***eixos de logradouro*** além daqueles apontados pelo usuário como necessários para sua aplicação.

O aplicativo computacional comparou os atributos apontados pelo usuário como necessários ao seu propósito contra aqueles existentes no esquema de dados da feição ***eixos de logradouro***.

O quadro abaixo apresenta os atributos existentes no esquema de dados da feição citada.

Quadro 6 - Atributos informados pelo provedor de dados cuja existência no banco foi comprovada pela avaliação.

Atributo
RUA
CEPPAR
CEPIMPAR
OBSNUM
CODLOGR
BAIRROPAR
BAIRROIMPA
TIPOVIA
NAVEG
ESTADO
MUNICIPIO
DIRECAO
CONV
DIVISA
MI_PRINX
TIPOPREF
PREPPREF
TITULOPREF
NOME PREF
TIPOCORR
PREPCORR
TITULOCORR
NOMECORR
ASSOCIADO
OBS_PRODUC

O Quadro 6 lista os atributos existentes no esquema de dados da feição **eixo de logradouro**. O procedimento de avaliação consiste em apoiar o usuário em apontar quais atributos existentes no esquema de dados são necessários para sua aplicação. Se houver atributos não identificados pelo usuário como necessários, o aplicativo computacional irá caracterizar a existência de atributos em comissão.

Comissão de atributos não é determinante para a rejeição do dado, uma vez que a aplicação do usuário pode tolerar excesso de atributos na base de dados.

O teste de avaliação aplicado aceitou a base de dados caso existam comissões no esquema de atributos da feição.

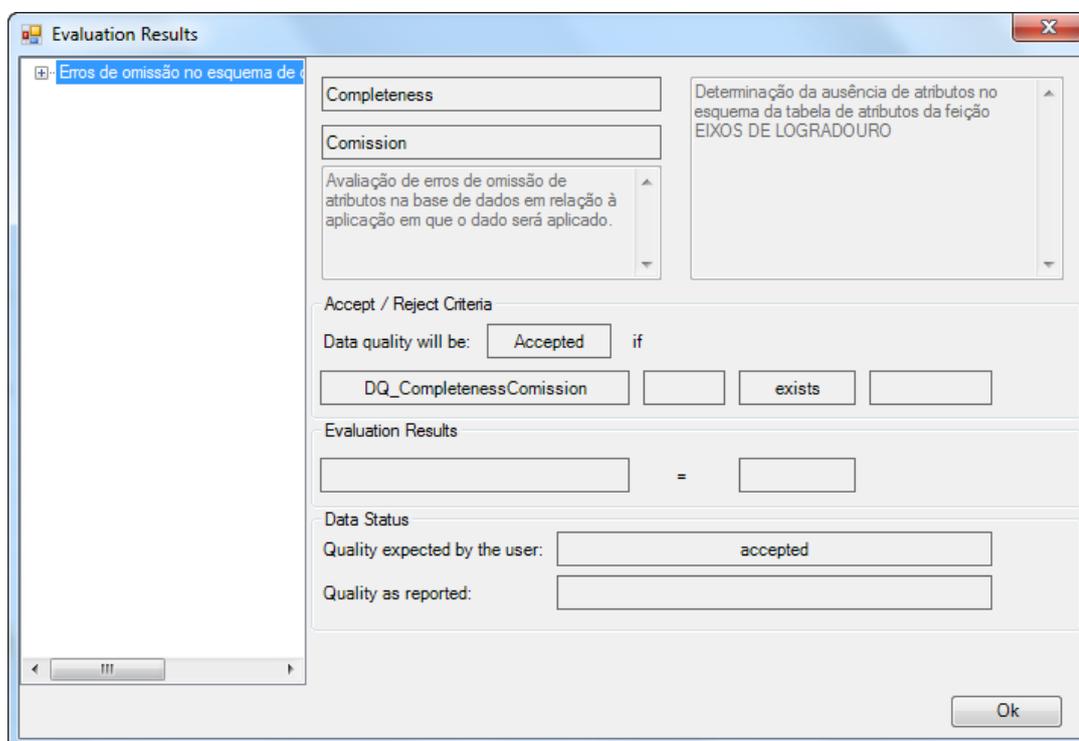


Figura 6 - Avaliação de erros de comissão de esquema de dados na tabela da feição Eixos de Logradouro.

A Figura 6 apresenta o resultado da avaliação de completeza para erros de comissão na base de dados. A tela apresenta ao usuário o elemento, o subelemento de qualidade, a descrição do avaliação e seu método de avaliação, conforme Figura 7 a seguir.

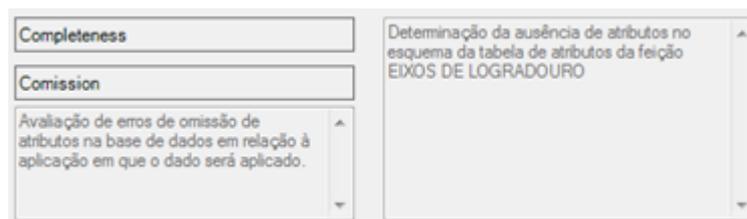


Figura 7 - Resultado da avaliação ao usuário: Elemento, Subelemento, Descrição e Método de avaliação.

A tela de resultados apresentada na Figura 6, detalha ao usuário o critério de aceitação ou rejeição do dado.

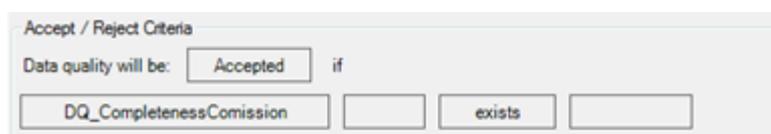


Figura 8 - Critério de aceitação para avaliação de Completeza.

Na Figura 8, o aplicativo irá aceitar o dado espacial em avaliação mesmo se existir comissão na base de dados.

A base foi aceita no processo de avaliação, pois, caso o usuário decida pela utilização da base de dados em sua aplicação mesmo com o a comissão de dados, o excesso de atributos pode ser facilmente removido através de programas computacionais para edição de bancos de dados. O contrário não é possível.

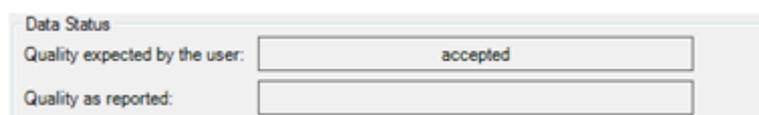


Figura 9 - Status da avaliação do teste de Completeza

Ainda na Figura 6, o aplicativo apresenta ao usuário se o dado foi aceito ou rejeitado pelo critério de avaliação. Na Figura 9, o aplicativo **aceita** o dado avaliado.

Esta avaliação foi executada pelo aplicativo em menos de 5 segundos.

4.1.2. Erros de Omissão em Esquema de Dados

A feição de **ruas** existente na base de dados digital da cidade de São Paulo foi rejeitada (Figura 10) na avaliação de erros de omissão de dados.

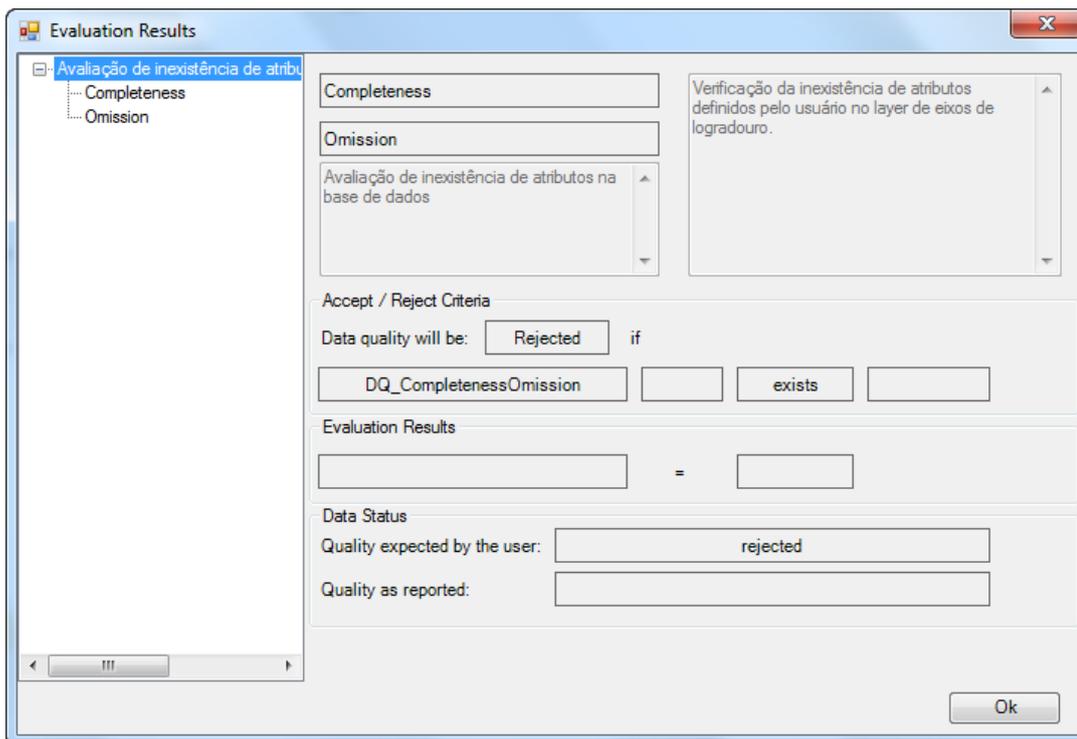


Figura 10 - Rejeição da feição de eixos de logradouro na base digital da cidade de São Paulo.

A tela apresentada na Figura 10, apresenta o relatório de avaliação do dado para erros de Omissão. O relatório apresenta o Elemento, Subelemento de qualidade utilizados, descrição da avaliação e seu método, conforme detalhe apresentado na figura a seguir.

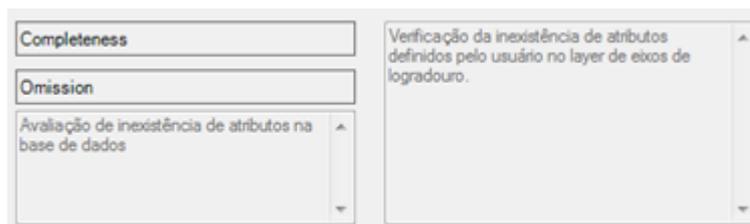


Figura 11 - Elemento, subelemento, descrição e método de avaliação para teste de Omissão.

A tela apresentada na Figura 10, apresenta também o critério de aceitação ou rejeição da base de ruas de São Paulo para o teste de omissão de dados.

Accept / Reject Criteria
 Data quality will be: if
 exists

Figura 12 - Critério de aceitação/rejeição da camada de ruas para teste de Omissão de dados.

Na Figura 12, o aplicativo apresenta o critério de aceitação, onde o dado será **rejeitado** caso exista erro de Omissão no dado em avaliação.

No teste de avaliação de erros de Omissão, o aplicativo detectou que nem todos atributos marcados como necessários pelo usuário estavam presentes na base de dados, o que levou a sua rejeição, conforme a Figura 13 a seguir.

Accept / Reject Criteria
 Data quality will be: if
 exists

Evaluation Results
 -

Data Status
 Quality expected by the user:
 Quality as reported:

Figura 13 - Critério e Status da avaliação do eixo de ruas de São Paulo.

Apesar disso, outros usuários para outras aplicações podem considerar a base de dados útil para suas atividades. Por exemplo, a avaliação aplicada rejeita a base de dados caso o código municipal da via (ou código do logradouro) não exista no esquema de dados. Outras aplicações que não necessitem deste atributo podem aceitar sua utilização.

Esta avaliação foi executada pelo aplicativo em menos de 5 segundos.

4.1.3. Erros de Omissão em Registros

A avaliação de erros de omissão em registros na base de dados digital da cidade de São Paulo rejeitou a feição de ruas. A feição avaliada possui mais de 10% de ocorrências da feição **nome de rua** com valores nulos.

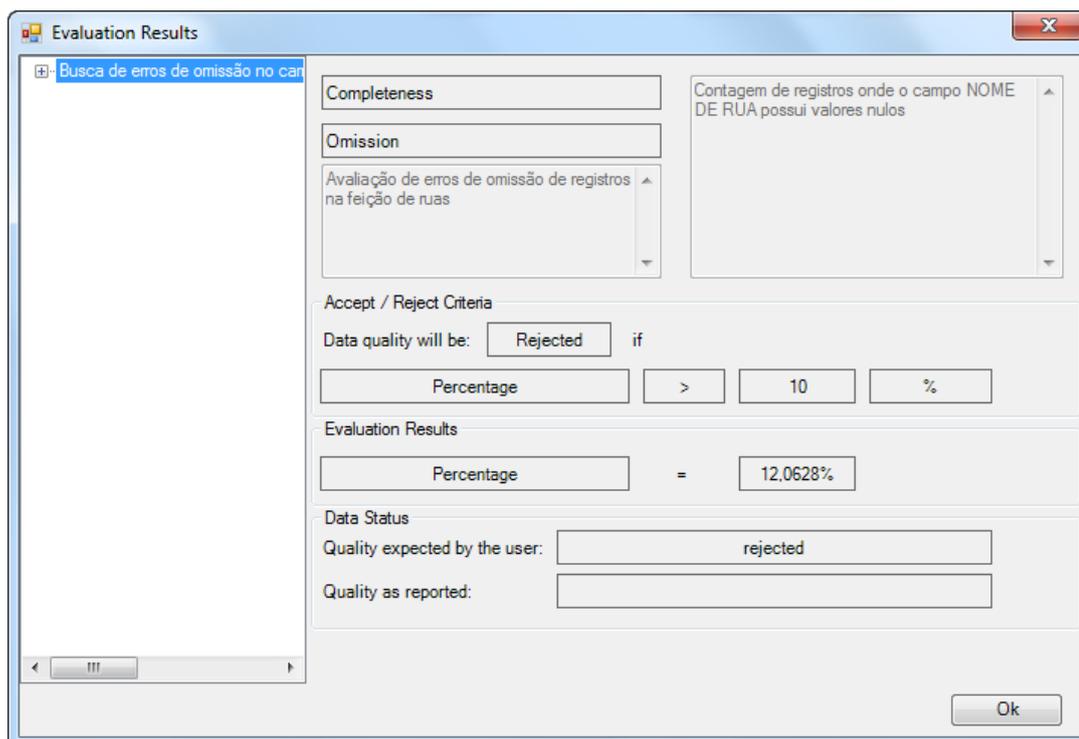


Figura 14 - Rejeição da feição de ruas em critério de omissão de registros.

O critério de avaliação utilizou um parâmetro de rejeição de 10% de registros com erros de omissão. A feição avaliada possui 12,063% de registros com valores nulos. Logo, sob estes critérios, a feição de eixos de logradouro na base de dados digital da cidade de São Paulo foi rejeitada.

Esta avaliação foi executada pelo aplicativo em aproximadamente 45 segundos.

4.2. AVALIAÇÃO DE CONSISTÊNCIA LÓGICA

A avaliação de consistência lógica foi aplicada em três diferentes critérios: Consistência de Domínio, Consistência de Formato e Consistência Conceitual.

4.2.1. Consistência de Domínio

A avaliação de consistência de domínio foi aplicada contra o atributo TIPOPREF da feição **RUAS** da base digital da cidade de São Paulo. O atributo possui informações sobre o **TIPO** da via pública: RUA, ESTRADA, AVENIDA, ALAMEDA, etc.

A avaliação buscou neste campo valores diferentes de RUA, ESTRADA, TRAVESSA e AVENIDA.

A avaliação **rejeitou** a base de dados pois mais de 69,83% dos registros possuíam valores fora da lista de domínio definida.

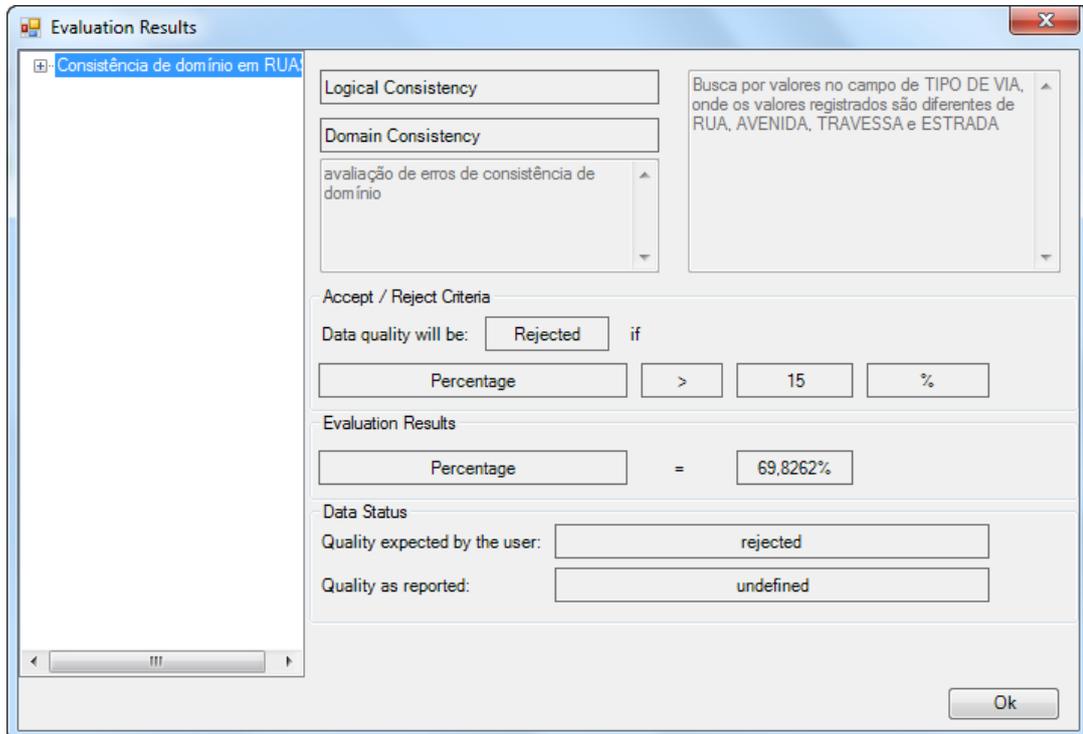


Figura 15 - Avaliação de consistência de domínio na feição de RUAS.

A Figura 15 apresenta a tela de resultados da avaliação da qualidade do dado. Nesta tela, é apresentado ao usuário o elemento e subelemento de qualidade contra o qual o dado espacial será avaliado:

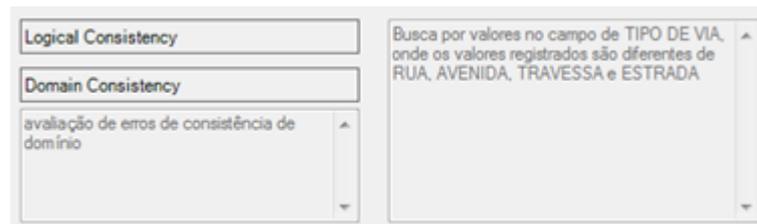


Figura 16 - Elemento, Subelemento, descrição e Método de avaliação para consistência de domínio na feição de ruas.

O critério de aceitação/rejeição da qualidade do dado:

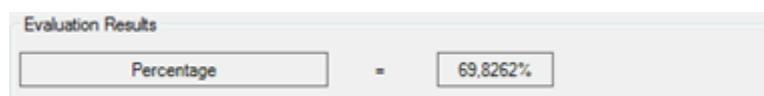


Accept / Reject Criteria
Data quality will be: if

Figura 17- Critério de aceitação ou rejeição da feição de ruas para avaliação de consistência de domínio.

No caso da Figura 15 o dado espacial será rejeitado se o percentual de elementos não coerentes com o critério de avaliação for maior que 15% dos elementos existentes na base de dados.

A Figura 15 também apresenta o resultado da avaliação:



Evaluation Results
 =

Figura 18 - Resultado da avaliação de consistência de domínio.

Neste caso o percentual de elementos sem conformidade com os requisitos estabelecidos pelo usuário foi de **69,82%**.

Como o valor ultrapassou em muito o critério aplicado para rejeição, procedeu-se uma análise dos registros desta feição, onde se descobriu que o atributo TIPOPREF utilizado nesta avaliação também possui muitos registros em branco.

Finalmente, a Figura 15 apresenta o resultado da avaliação:



Data Status
Quality expected by the user:
Quality as reported:

Figura 19 - Status da avaliação do dado espacial.

O aplicativo então, baseado nos requisitos do usuário, rejeitou a base de dados em avaliação. Como não havia informação de qualidade do provedor o aplicativo não pôde concluir sobre a aceitação/rejeição do dado.

Este atributo em questão possui mais de 131.000 registros (Figura 20) com valores nulos no campo TIPOPREF. Caso o usuário precise utilizar esta informação de qualquer maneira, um grande esforço terá que ser demandado para corrigir esta falha na feição avaliada.

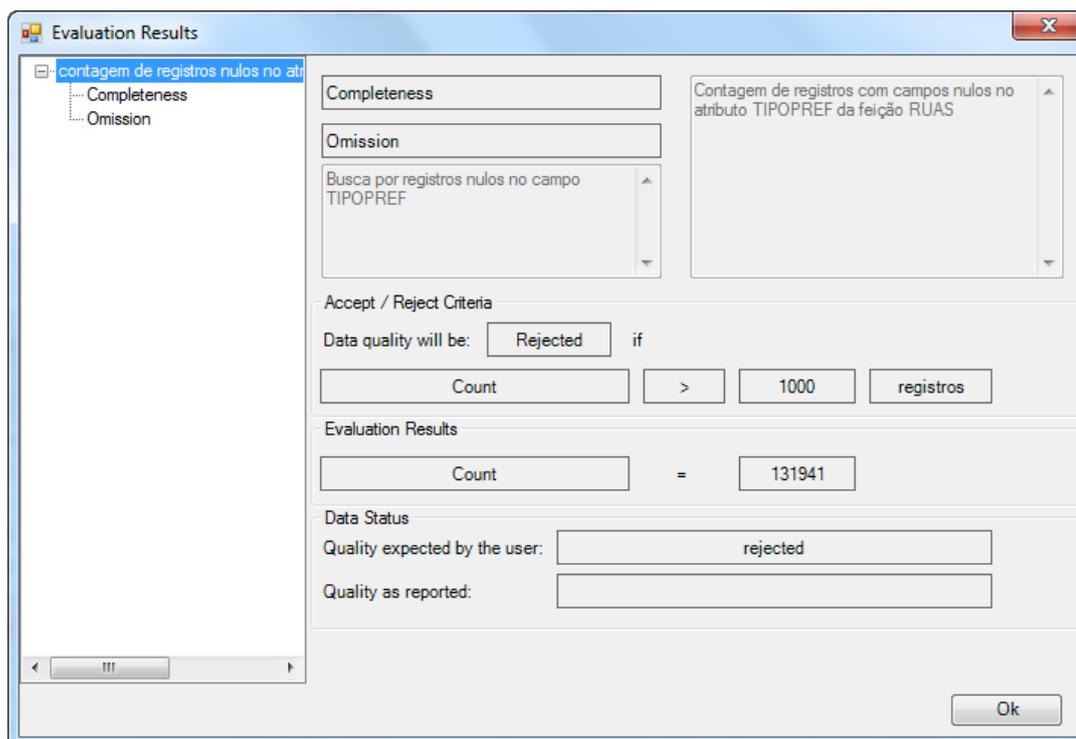


Figura 20 - Busca por registros nulos no campo TIPOPREF da feição RUAS.

Para avaliação de Omissão de registros na feição de *ruas*, o aplicativo foi configurado para *rejeitar* o dado caso mais de 1000 elementos na base de dados possuam valores *nulos* ou *vazios* no campo *TIPOREF*, conforme descrito na seção **Accept/Reject Criteria** apresentado na Figura 20.

Nesta mesma tela o aplicativo rejeita o dado, uma vez que foram detectados **131.941** elementos que atendem este requisito.

4.2.2. Consistência de Formato

A avaliação de consistência de formato confirmou que os dados em avaliação estão realmente no formato comunicado pelo provedor (ESRI Shapefile, MAPInfo TAB file, DWG e DGN).

4.2.3. Consistência Conceitual

A avaliação de consistência conceitual foi aplicada na feição HIDROGRAFIA registrada na base de dados da cidade de Itapeva.

Esta avaliação foi aplicada na base CADD em lugar da base SIG, pois os formatos ESRI Shapefile e MapInfo TAB file não permitem que uma feição possua mais de um

tipo de geometria. Por exemplo, se a base shapefile define uma feição linear, o formato simplesmente não permite que elementos pontuais ou poligonais sejam registrados em seu banco de dados. Já o mesmo não ocorre com dados CADD que aceitam praticamente qualquer informação que o usuário cadastrar, estando mais sujeita a este tipo de erro.

O critério de rejeição foi definido para rejeitar a feição caso mais de 10% dos elementos classificados como hidrografias não estejam em conformidade com o tipo da feição. A feição de **hidrografias** foi rejeitada pela avaliação por possuir mais de 14% de elementos não lineares registrados como hidrografia.

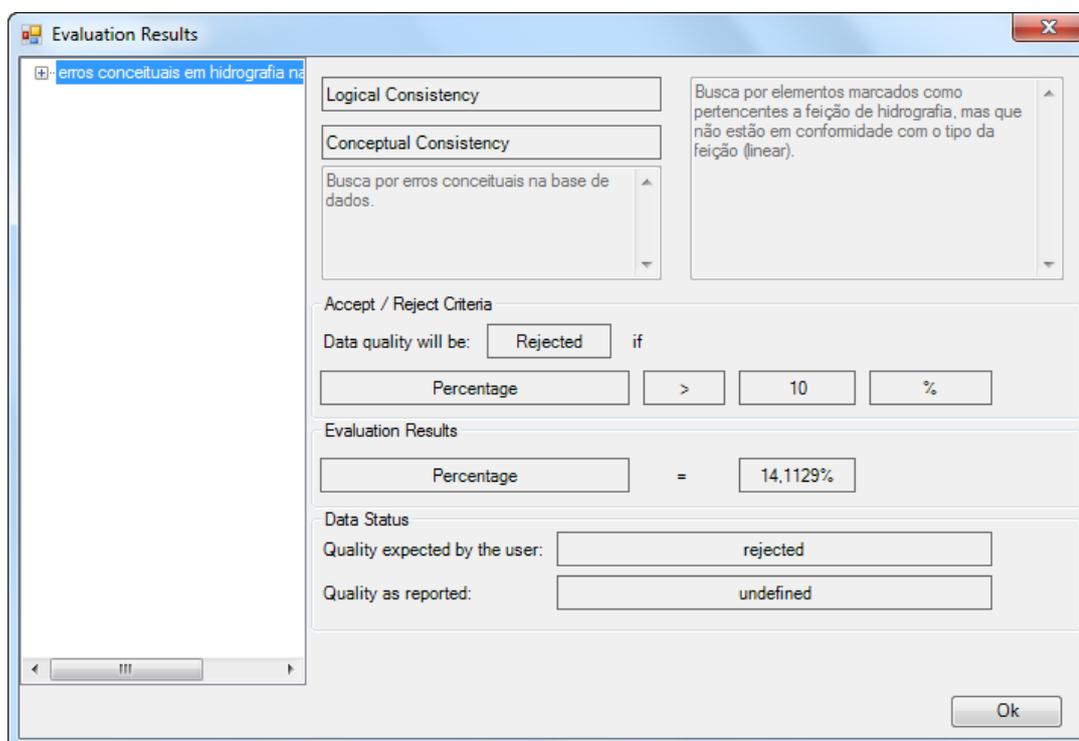


Figura 21 - Rejeição de consistência conceitual da base de hidrografias

Esta avaliação foi executada em menos de 10 segundos.

4.3. AVALIAÇÃO DE ACURÁCIA DE ATRIBUTOS

A avaliação da acurácia de atributos foi aplicada contra a base digital de Itapeva, onde foi avaliada a ocorrência de erros de atributos não quantitativos da feição. Ou seja, foram verificados todos os elementos classificados como **hidrografia**, mas nos

casos em que seus atributos físicos como cor, espessura de linha e estilo de linha não estejam em conformidade com a definição da feição.

A avaliação **aceitou** a feição de acordo com o seguinte critério: A feição será aceita caso menos de 5 elementos classificados como hidrografias possuam erros em seus atributos físicos.

A avaliação não encontrou nenhum elemento naquela feição fora dos padrões de conformidade estabelecidos.

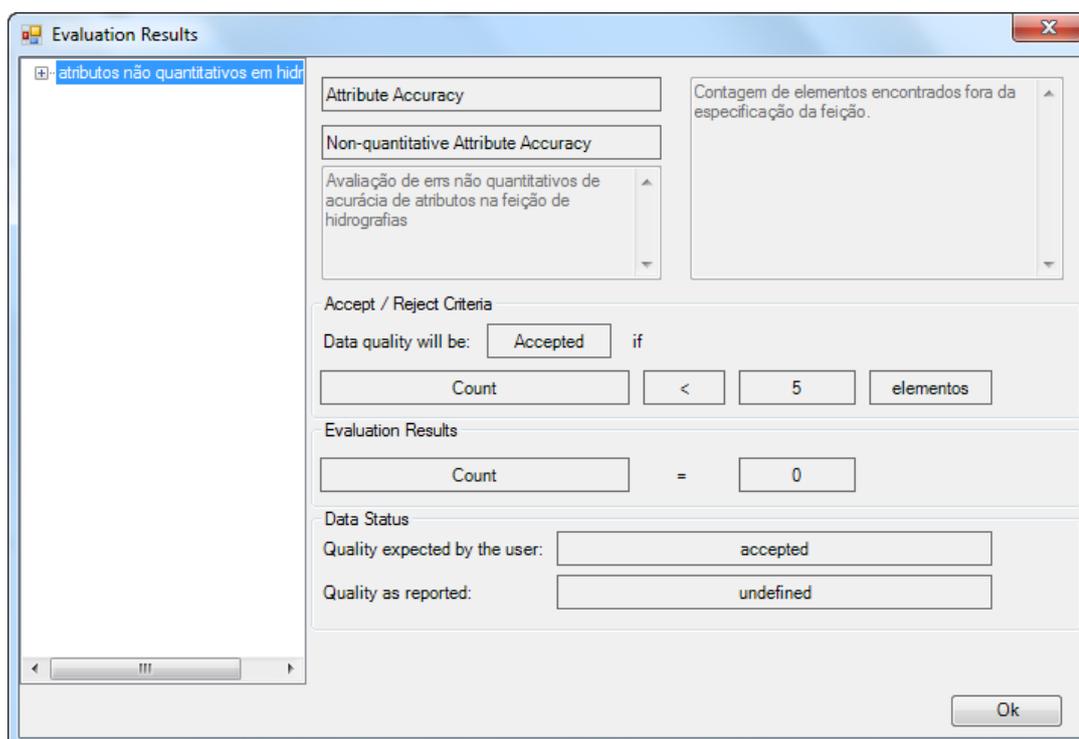


Figura 22 - Avaliação de atributos não quantitativos na camada de hidrografias.

Este procedimento foi executado em menos de 10 segundos pelo aplicativo.

4.4. DISCUSSÕES À RESPEITO DOS CRITÉRIOS DE ACEITAÇÃO OU REJEIÇÃO

Os critérios de rejeição e aceitação de uma feição em uma base de dados digital podem variar de acordo com o usuário que está promovendo a avaliação. Usuários diferentes podem ter critérios de avaliação diferentes.

As avaliações aplicadas nesta dissertação de um modo geral, não invalidam nenhuma das bases utilizadas, pois, novamente, cabe ao usuário final aplicar o critério de aceitação/rejeição mais adequados à sua aplicação.

4.5. DISCUSSÕES FINAIS À RESPEITO DOS RESULTADOS

Nas avaliações apresentadas neste capítulo, o avaliador do dado apresentou para cada dado espacial seus critérios de avaliação.

Quadro 7 - Parâmetros de avaliação descritos pelo usuário

Camada	Elemento	Subelemento	Critério de aceitação
Eixos de logradouro	Completeza	Comissão	O dado será aceito mesmo na presença de erros de comissão de dados
Eixos de logradouro	Completeza	Omissão (esquema de dados)	O dado será rejeitado caso faltem atributos na base de dados
Eixos de logradouro	Completeza	Omissão (registros)	O dado será rejeitado caso mais que 10% dos registros apresentem valores nulos ou vazios)
Eixos de logradouro	Consistência Lógica	Consistência de Domínio	O dado será rejeitado caso mais que 15% dos registros apresentem valores diferentes daqueles definidos em sua lista de domínio)
Eixos de logradouro	Consistência Lógica	Consistência Conceitual	O dado será rejeitado caso mais que 10% dos elementos não obedeçam a estruturação descrita pelo usuário como correta
Hidrografia	Acurácia de Atributos	Acurácia de atributos não quantitativos	O dado será rejeitado caso mais que 5 elementos na base de dados não obedeçam ao critério apresentado.

O aplicativo por sua vez, com base em cada critério apresentado pelo usuário, avaliou os elementos existentes na base de dados. Para cada critério o aplicativo apresentou duas informações distintas: O resultado da avaliação e o status do dado. Com tais informações, o usuário pode decidir se o dado está aderente aos seus requisitos, o esforço necessário à corrigi-lo ou adequá-lo ao seu uso, bem como o risco em utilizá-lo da forma em que se encontra.

A Figura 84, descrita no Anexo 2 desta dissertação, apresenta a opção do usuário em gravar os metadados de qualidade em um arquivo XML, permitindo que a informação de qualidade seja compartilhada com outros usuários.

5. CONCLUSÕES

A automação do processo de avaliação de qualidade de dados espacial pode aumentar o conhecimento do usuário a respeito da qualidade do dado que está utilizando. Esta percepção advém de três informações básicas que o usuário recebe do aplicativo após o processo de avaliação ser concluído:

- Resultado da avaliação: onde o aplicativo informa se o dado espacial foi aprovado ou rejeitado pelo processo;
- A condição que o aplicativo utilizou para aceitar/rejeitar o dado espacial; por exemplo, se o aplicativo detectou mais elementos com problemas de qualidade do que o critério de aceitação permitiria.
- A quantidade, quando cabível, de elementos com problemas no dado avaliado.

Esta informação transmite ao usuário uma noção do esforço necessário para corrigir o dado, caso ele opte por utilizá-lo em sua aplicação. A informação da quantidade de elementos que não estão em conformidade com os requisitos de qualidade do usuário apresentam a ele o nível de risco associado à utilização desta informação sem corrigi-la.

Procedimentos envolvendo grande volume de dados também podem ser beneficiados por este método automatizado de avaliação, onde tais dados podem ser processados e avaliados em um intervalo de tempo curto, conforme os tempos apresentados durante os testes.

Uma limitação no procedimento de avaliação, que foi detectada ao longo dos testes realizados, é que não é possível avaliar se a escala do mapeamento ou, se o nível de generalização¹³ cartográfica da base de dados, poderá inviabilizar o uso de um determinado dado espacial para uma determinada aplicação.

Procedimentos de avaliação para acurácia de atributos e acurácia posicional só podem ser avaliados contra relatórios de levantamento de campo ou contra outras bases de dados cuja qualidade é reconhecidamente superior. No caso de acurácia

¹³ Generalização cartográfica é o nível de simplificação aplicado no dado espacial em relação à escala que este dado está sendo representado.

de atributos, deve-se levar em conta que o critério de classificação do avaliador também possui um fator de erro associado.

5.1. AVALIAÇÃO DE QUALIDADE DE DADOS ESPACIAIS

Conforme a literatura estudada na elaboração desta dissertação, foi observado que as maiores dificuldades encontradas no processo de avaliação de qualidade do dado espacial são: o volume de informações para processamento e a dificuldade em permitir ao usuário final descrever, de forma estruturada, qual elemento ou, elementos de qualidade, melhor avaliam o dado de acordo com o seu propósito.

Esta dissertação mostrou que a descrição estruturada a respeito da avaliação de qualidade que o usuário deseja aplicar em seu conjunto de dados depende dos seguintes fatores a priori:

- Conhecimento por parte do usuário sobre dados e dados espaciais
- Utilização de normas e padrões de comunicação de elementos de qualidade;

Tais requisitos foram necessários para garantir a eficiência da comunicação estruturada por parte do usuário na definição dos seus critérios de avaliação.

A avaliação da qualidade de dados espaciais, principalmente aqueles advindos de fontes diferentes, apresentou-se como uma tarefa complexa e trabalhosa. A complexidade deste processo é advinda da natureza da avaliação de alguns elementos de qualidade, tais como Completeza e Consistência Conceitual, onde foi necessário realizar várias iterações na base de dados. Cada elemento de qualidade demandou uma comunicação estruturada de seu critério de avaliação, tal estruturação permitiu a automação do processo de avaliação de toda a base de dados. Desta forma, esta dissertação apresentou uma solução viável para o processo de avaliação da qualidade do dado espacial.

Avaliações de completeza de registros ou completeza de esquema de dados requerem que o usuário possua conhecimentos relacionados à configuração de bancos de dados.

A avaliação de consistência lógica requer do usuário conhecimento em relação à tipologia da feição de interesse, por exemplo, se uma feição deve ser pontual, linear ou poligonal.

Por outro lado, só foi possível automatizar o procedimento de avaliação de uma base de dados espacial, uma vez que se especializaram os metadados de qualidade definidos pela ISO 19115 de forma que permitisse a documentação estruturada da avaliação de qualidade ideal para os anseios do usuário.

5.2. COMUNICAÇÃO DA QUALIDADE DO DADO ESPACIAL

As ISO 19113, 19114 e 19115 sozinhas não podem prover os mecanismos de comunicação necessários para este tipo de comunicação, pois são abrangentes em não detalhar procedimentos como a avaliação de completeza para os registros de uma tabela, ou para o modelo de dados adotado pela feição. Apesar disso, as definições e metadados apresentados por tais normas servem perfeitamente como um framework inicial de criação de um padrão voltado para automação do processo de avaliação de qualidade de dados espaciais.

Comprovou-se nesta dissertação que, apesar dos metadados serem apropriados para registrar informações sobre os elementos de qualidade, ainda permanece a questão de quão apropriados os metadados são para a comunicação da qualidade do dado ao usuário.

Esta afirmação se mostrou verdadeira no contexto em produtores de dados espaciais buscam comunicar¹⁴ a qualidade do dado espacial através de metadados. O formato XML comumente utilizado para o registro de metadados não é eficiente em comunicar uma informação, mas se apresenta como um excelente meio de transporte de uma informação qualificada, onde através de um dispositivo de tradução deste metadado (por exemplo, um aplicativo computacional), a comunicação pode se tornar efetiva.

¹⁴ Comunicar no sentido de se fazer entender um assunto ou tema.

Neste sentido, esta dissertação também apresentou uma solução, ainda que parcial, para o problema de comunicação da qualidade do dado ao usuário final. Considera-se parcialmente resolvido, pois questões relacionadas a comunicação da qualidade das entidades da base de dados (geometrias ou atributos) que não atendam aos critérios de qualidade estabelecidos pelo usuário não foram implementadas.

5.3. CONCLUSÕES FINAIS

Esta dissertação foi capaz de provar a hipótese de que é possível apoiar o usuário do dado espacial em avaliar a adequação do uso de um determinado dado espacial, utilizando seus próprios parâmetros de avaliação. O objetivo desta dissertação, em apresentar uma metodologia de avaliação estruturada da qualidade do dado espacial foi cumprido, de forma que o usuário é capaz de informar de maneira estruturada os elementos e subelementos de qualidade que melhor descrevem seu propósito de utilização.

O método de avaliação automatizado apresenta ao usuário a conformidade do dado espacial em relação aos seus critérios de qualidade, e em relação a avaliação da qualidade informada pelo provedor, uma vez que o processo de avaliação quantifica os desvios existentes entre a qualidade do dado espacial verificada no processo de avaliação contra a qualidade informada pelo provedor e os critérios de aceitação do dado pelo usuário.

Ainda que nem todos os processos de avaliação fossem resolvidos pelo programa computacional desenvolvido durante o decorrer deste trabalho (acurácia posicional, acurácia de atributos, acurácia temporal, etc.), esta dissertação provou que é possível prover ferramentas ao usuário que lhe permitem: comunicar qual o Elemento de Qualidade melhor atende os requisitos de sua aplicação, comunicar de maneira estruturada seus critérios de aceitação e busca de maneira eficiente.

Como sugestão futura, a avaliação da eficiência da comunicação da qualidade apresentada pelo aplicativo, poderá ser feita submetendo-o para testes por uma comunidade de usuários, coletando suas impressões e eficiência percebida.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- BAIRD, M. **The Role of Evaluation**. World Bank. Washington, DC, p. 7-12. 1998.
- BEARD, K. **Use Error: The Neglected Error Component**. Proceeding of AUTOCARTO 9. Baltimore, Maryland: [s.n.]. 1989. p. pp. 808-817.
- BOIN, A. T.; HUNTER, G. J. **Do Spatial Data Consumers Really Understand Data Quality Information?** 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences. [S.I.]: [s.n.]. 2006.
- BOIN, A. T.; HUNTER, G. J. What Communicates Quality to the Spatial Data Consumer. In: STEIN, A.; SHI, W.; BIJKER, W. **Quality Aspects in Spatial Data Mining**. Boca Ratón: CRC PRESS, 2008. p. 285-296.
- BUCKLEY, D. J. The Nature of Geographic Information. In: BUCKLEY, D. J. **The GIS Primer. An Introduction to Geographic Information Systems**. Fort Colins: GIS Solutions Inc., 1997. p. 7-21.
- CAPRIOLI, M. et al. **Rules and Standards for Spatial Data Quality in GIS Environments**. International Cartographic Conference. Durban, South Africa: ICA. 2003. p. 1740-1747. 21st International Cartographic Conference, Cartographic Renaissance, ICC 2003.
- CEMG-CONCAR, COMITÊ DE ESTRUTURAÇÃO DE METADADOS GEOSPACIAIS. **Perfil de Metadados Geoespaciais do Brasil (Perfil MGB) - Conteúdo de Metadados Geoespaciais em Conformidade com a Norma ISO 19115:2003**. Comissão Nacional de Cartografia. [S.I.], p. 194. 2009.
- CEN/TC287/WG02. **Geographic Information - Data Description - Quality. Draft for discussion produced by project team 5 (PT05)**. CEN Central Secretariat. Brussels, p. 36. 1995.
- CHRISMAN, N. R. Metadata Required to Determine the Fitness of Use for Use in Environmental Analysis. In: CHRISMAN, N. R. **Environmental Information Management and Analysis: Ecosystems to Global Scale**. Philadelphia: Taylor & Francis, 1994.
- COUCLELIS, H. The Certainty of Uncertainty: GIS And The Limits of Geographic Knowledge. **Transactions in GIS**, v. 7, n. No 2, p. 165-175, 2003.
- DÉVILLERS, R. **Conception d'Un Système Multidimensionnel d'Information Sur La Qualité des Données Géospatiales**. Université Laval, Canada. [S.I.]. 2004.
- DÉVILLERS, R. et al. **Spatial Data Quality: From Metadata to Quality Indicators and Contextual End-User Manual**. OEEPE/ISPRS Joint Workshop on Spatial Data Quality Management. Istanbul: [s.n.]. 2002.
- DÉVILLERS, R. et al. Towards Spatial Data Quality Information Analysis Tools For Experts Assessing the Fitness For Use of Spatial Data. **International Journal of Geographical Information Science**, p. 261-282, 2007.

DÉVILLERS, R.; BÉDARD, Y.; GERVAIS, Y. Indicateurs de Qualité Pour Réduire Les Risques de Mauvaise Utilisation des Données Géospatiales. **Revue Internationale de Géomatique**, 2004. 35-57.

DÉVILLERS, R.; JEANSOULIN, R. **Fundamentals of Spatial Data Quality**. Londres: ISTE Ltda, 2006.

DUCKHAM, M. Implementing an object-oriented error-sensitive GIS. In: LOWELL, K.; JATON, A. **Spatial Accuracy Assessment: Land Information Uncertainty in Natural Resources**. Michigan: Ann Arbor Press, 1999. p. 209-215.

FEDERAL GEOGRAPHIC DATA COMMITTEE. National Standard for Spatial Data Accuracy. In: INFRASTRUCTURE, N. S. D. **Geospatial Positioning Accuracy Standards**. Reston, Virginia: Federal GeoGraphic Data Committee Secretariat, 1998.

FORESMAN, T. W.; WIGGINS, H. V.; PORTER, D. L. **Metadata Myth: Misunderstanding the Implications of Federal Metadata Standards**. Proceedings of the First IEEE Metadata Conference. Silverspring: [s.n.]. 1996.

FRANK, A. Assessing the Quality of Data with a Decision Model. In: STEIN, A.; SHI, W.; BIJKER, W. **Quality Aspects in Spatial Data Mining**. Boca Raton, Florida: CRC Press - Taylor & Francis Group, 2008.

GOODCHILD, M. F. Attribute Accuracy. In: GUPTILL, S. C.; MORRISON, J. L. **Elements of Spatial Data Quality**. 1. ed. Oxford: Pergamon, v. 1, 1995. Cap. 4, p. 59-79.

GOODCHILD, M. F. Sharing Imperfect Data. **Sharing GeoGraphic Information**, New Jersey, p. Pages 413-425, 1995.

GOODCHILD, M. F. Forward. In: DÉVILLERS, R.; JEANSOULIN, R. **Fundamentals of Spatial Data Quality**. London: ISTE, 2006. p. pp. 13-16.

GUPTILL, C.; MORRISON, J. L. **Elements of Spatial Data Quality**. New York: Elsevier Science Inc., 1995.

HARDING, J. Vector data quality: A data provider's perspective. In: DÉVILLERS, R.; JEANSOULIN, R. **Spatial data quality: Concepts**. [S.l.]: ISTE, 2006. Cap. 8.

HO, L. L.; QUINTANILHA, J. A.; QUININO, R. C. The design of a zero-defect sampling with rectification procedure in the presence of classification errors: an application to database with spatial data. **Brazilian Journal of Operations and Production Management.**, 3, 2007. 87-99.

ISO. **ISO 19113 - Geographic Information - Quality Principles**. Switzerland. 2002.

ISO. **ISO 19114 - Geographic Information - Quality Evaluation Process**. Switzerland. 2003.

ISO. **ISO 19115 - Geographic Information - Metadata**. Switzerland. 2003.

ISO. **ISO 19139 - Geographic Information - Metadata - XML schema implementation**. International Standards Organization. Switzerland. 2009.

JEANSOULIN, R. (Ed.). **Fundamentals of spatial data quality**. [S.l.]: ISTE - Geographic Information Systems Series, 2006. 141-160 p.

- JOKSIC, D.; BAJAT, B. Elements of Spatial Data Quality as Information Technology Support for Sustainable Development Planning. **Spatium**, 2004. 77-83.
- KAINZ, W. Logical Consistency. In: GUPTILL, S.; MORRISON, J. L. **Elements of Spatial Data Quality**. Tarrytown, NY: International Cartographic Association, 1995. p. 109-137.
- KUMI-BOATENG, B.; YAKUBU, I. Assessing the Quality of Spatial Data. **European Journal of Scientific Research**, 43, 2010. 507-515.
- LIA, D.; ZHANGB, J.; WUA, H. Spatial data quality and beyond. **International Journal of Geographical Information Science**, 26, 2012. 2277-2290.
- LUSH, V.; BASTIN, L.; LUMSDEN, J. **Geospatial Data Quality Indicators**. 10th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environment Sciences. Florianópolis: [s.n.]. 5-7 July 2012.
- MOSTAFAVI, M. A.; EDWARDS, G.; JEANSOULIN, R. **An Ontology-Based Method for Quality Assessment of Spatial Databases**. Proceeding of the 4th International Symposium on Spatial Data Quality. Viena: Geo-Info. 2004. p. pp. 49-66.
- NOGUEIRA JUNIOR, J. B. **Controle de Qualidade de Produtos Cartográficos, Dissertação de Mestrado**. Universidade Estadual Paulista. Presidente Prudente. 2003.
- PIERKOT, C. et al. **Advocacy for External Quality in GIS**. [S.l.]: Springer-Verlag Berlin Heidelberg, 2011. 151-165 p.
- QIU, J.; HUNTER, G. J. A GIS with the Capacity for Managing Data Quality Information. In: SHI, W.; FISHER, P. F.; GOODCHILD, M. F. **Spatial Data Quality**. London: Taylor & Francis, 2002. p. 230-250.
- QUINTANILHA, J. A. Qualidade em bases de dados para sistemas de informações geográficas. **Dados geográficos: aspectos e perspectivas**, Salvador, Bahia, 2002. 33-42.
- QUINTANILHA, J. A.; HO, L. L.; QUININO, R. C. Controle de qualidade de bases de dados espaciais através de uma amostragem de zero-defeitos com retificação. **Revista Brasileira de Cartografia**, v. 60, p. 9-15, 2008.
- ROCHA SANTOS, S. D. et al. **Considerações Sobre a Utilização do PEC (Padrão de Exatidão Cartográfica) Nos dias Atuais**. III Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação. Recife, PE: [s.n.]. 2010. p. 001-005.
- SHI, W.; FISHER, P.; GOODCHILD, M. F. **Spatial Data Quality**. London: Taylor & Francis, 2002.
- STEIN, A.; SHI, W.; BIJKER, W. **Quality Aspects in Spatial Data Mining**. Boca Raton, FL: CRC Press, 2008.
- STEIN, A.; VAN OORT, P. The Impact of Positional Accuracy on the Computation of Cost Functions. In: DÉVILLERS, R.; JEANSOULIN, R. **Fundamentals Of Spatial Data Quality**. Newport Beach: ISTE, 2006. Cap. 6, p. 108 - 122.

TSOULOS, L.; JAKOBSSON, A. **The Role of Quality in Spatial Data Infrastructures**. Athens. 2005.

VASSEUR, B. et al. External quality evaluation of geographical applications: An ontological approach. In: DÉVILLERS, R.; JEANSOULIN, R. **Spatial Data Quality**. [S.I.]: ISTE - Geographical Information Systems Series, 2006. Cap. 13, p. 255-270.

VEREGIN, H.; HARGITAI, P. An Evaluation Matrix for Geographical Data Quality. In: GUPTILL, S. C.; MORRISON, J. L. **Elements of Spatial Data Quality**. Oxford: Elsevier, 1995. Cap. 9, p. 167-188.

VEREGIN, H.; LANTER, D. P. Data-Quality Enhancement Techniques in Layer-Based Geographic Information Systems. **Computers, Environment and Urban Systems**, Kent, v. 19, n. No. 1, p. 23-36, 1995.

WEBER, E. et al. **Qualidade de Dados Geoespaciais**. Universidade Federal do Rio Grande do Sul. Porto Alegre-RS. 1999.

WU, C. V.; BUTTENFIELD, B. P. Spatial Data Quality and Its Evaluation. **Computers, Environment and Urban Systems**, Buffalo, v. 18, n. No. 3, p. 153-165, 1994.

ZARGAR, A. **An Operation-Based Approach To The Communication of Spatial Data Quality in GIS**. Memorial University of Newfoundland. St. John's. 2009. Thesis submitted to the School of Graduate Studies in partial fulfillment of the requirements for a degree of Master of Science.

ANEXOS

Anexo A – Amostra de metadados de qualidade

A seguir, é mostrado um exemplo de arquivo de metadados de qualidade gerado pelo aplicativo desenvolvido nesta dissertação.

```
<?xml version="1.0" encoding="utf-8"?>
<DQ_DataQuality>
  <lineage>
    <source scaleDenominator="1:100.000" sourceReferenceSystem="UTM" sourceDatum="SAD 69"
sourceEllipsoid="HAYFFORD">
      <processSteps>
        <ProcessStep dateTime="03/12/2012" description="Reprojeção do sistema de projeção e datum de
UTM/SAD-69 para UTM/WGS-84" />
      </processSteps>
    </source>
  </lineage>
  <report>
    <dataQualityInfo InfoSource="2">
      <scope level="Feature">
        <extent westBoundLongitude="" eastBoundLongitude="" southBoundLatitude="" northBoundLatitude=""
/>
        <levelDescription description="" feature="" />
      </scope>
      <DQ_Element nameOfMeasure="avaliação do formato do dado" Name="Logical Consistency">
        <DQ_Subelement Name="Format Consistency" DQ_Date="">
          <DQ_Measure DQ_MeasureDesc="" DQ_MeasureID="1">
            <DQ_EvalMethod DQ_EvalMethodType="1" DQ_EvalMethodDesc="" />
            <DQ_QualityResult MeasureType="" DQ_ValueType="" DQ_Value="" DQ_ValueUnit="" />
          </DQ_Measure>
        </DQ_Subelement>
      </DQ_Element>
      <ExpectedQuality AcceptReject="1" FileFormat="1" />
    </dataQualityInfo>
  </report>
</DQ_DataQuality>
```

Anexo B – O Aplicativo de Apoio a Avaliação de Qualidade Espacial

Este anexo descreve o aplicativo que foi construído para apoiar a metodologia de avaliação de dados espaciais. Este anexo apresenta as telas principais e descreve as funcionalidades que foram codificadas para alcançar os resultados demonstrados nesta dissertação.

Propósito do aplicativo

O aplicativo desenvolvido para esta dissertação, nomeado *ISOSpatialQuality* foi criado para:

1. Permitir a comunicação estruturada da qualidade do dado espacial tal como determinada pelo seu produtor.
2. Permitir a descrição estruturada da qualidade do dado espacial de acordo com o propósito que o usuário pretende aplica-la.
3. Registrar metadados de qualidade e assim permitir sua comunicação para outros usuários.
4. Automatizar o processo de avaliação do dado espacial.
5. Apresentar um relatório de avaliação e conformidade do dado espacial.

A descrição estruturada da qualidade do dado espacial foi desenvolvida de acordo com os Elementos e Sub-elementos de qualidade tais como descritos no capítulo 2.

O aplicativo foi desenvolvido em linguagem de programação C#, no ambiente de programação provido pelo Microsoft Visual Studio C# Express.

Classes Principais e Componentes

O aplicativo desenvolvido para esta dissertação possui dois grupos principais de classes: Classes ISO e Classes de Negócio.

Todo o aplicativo foi desenvolvido pelo autor desta dissertação. As classes ISO foram codificadas a partir das definições da ISO 19113, 19114 e 19115.

As classes ISO codificaram as definições ISO pertinentes a definição dos elementos de qualidade de dados espaciais, conforme descrito nas ISO 19113, 19114 e 19115.

A Figura 23 a seguir, apresenta o diagrama de classes desenvolvido pelo autor a partir da ISO 19113.

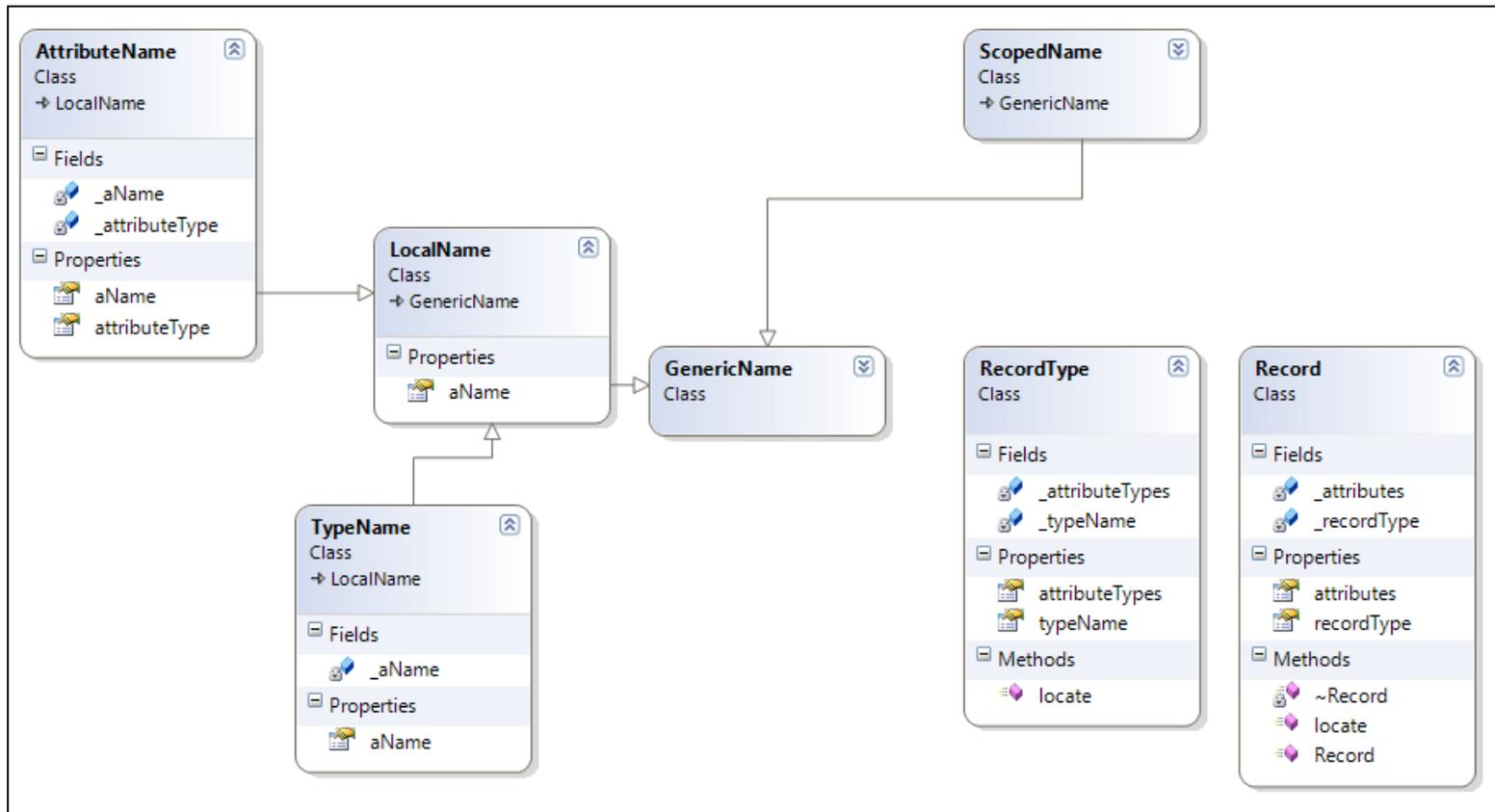


Figura 23 - Diagrama de classes - ISO 19113 - Fonte: o próprio autor.

A Figura 24 a seguir apresenta o diagrama de classes desenvolvido pelo autor a partir da ISO 19114.

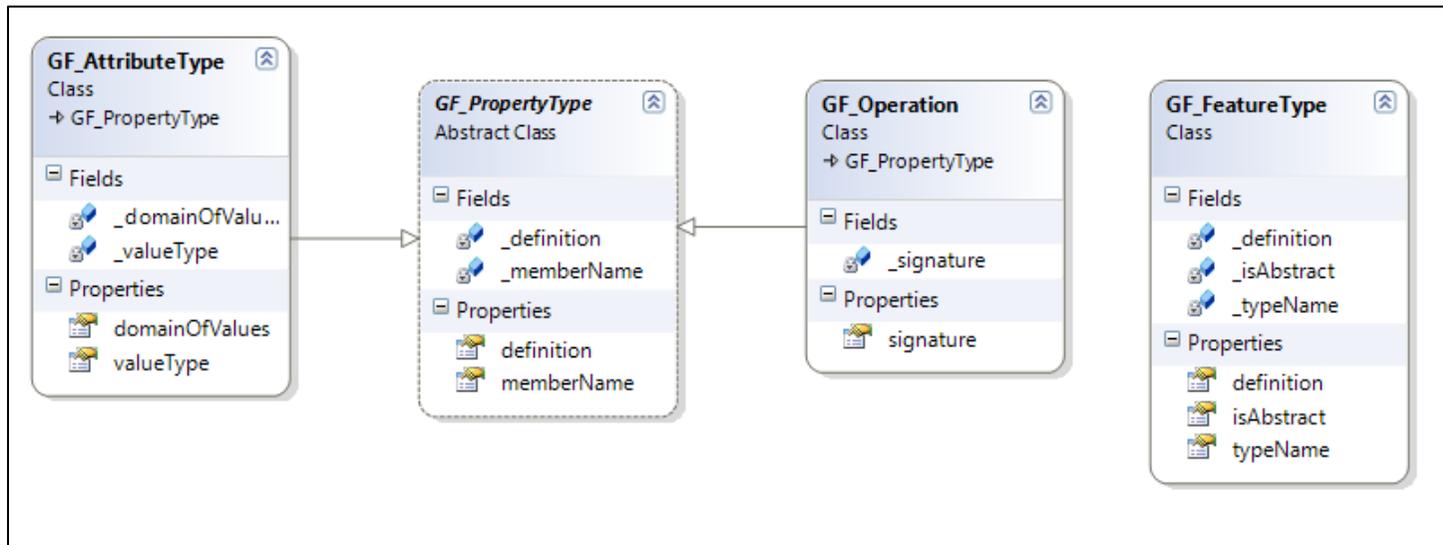


Figura 24 - Diagrama de classes - ISO 19114 - Fonte: o próprio autor

A Figura 25 a seguir apresenta o diagrama de classes desenvolvido pelo autor a partir da ISO 19115.

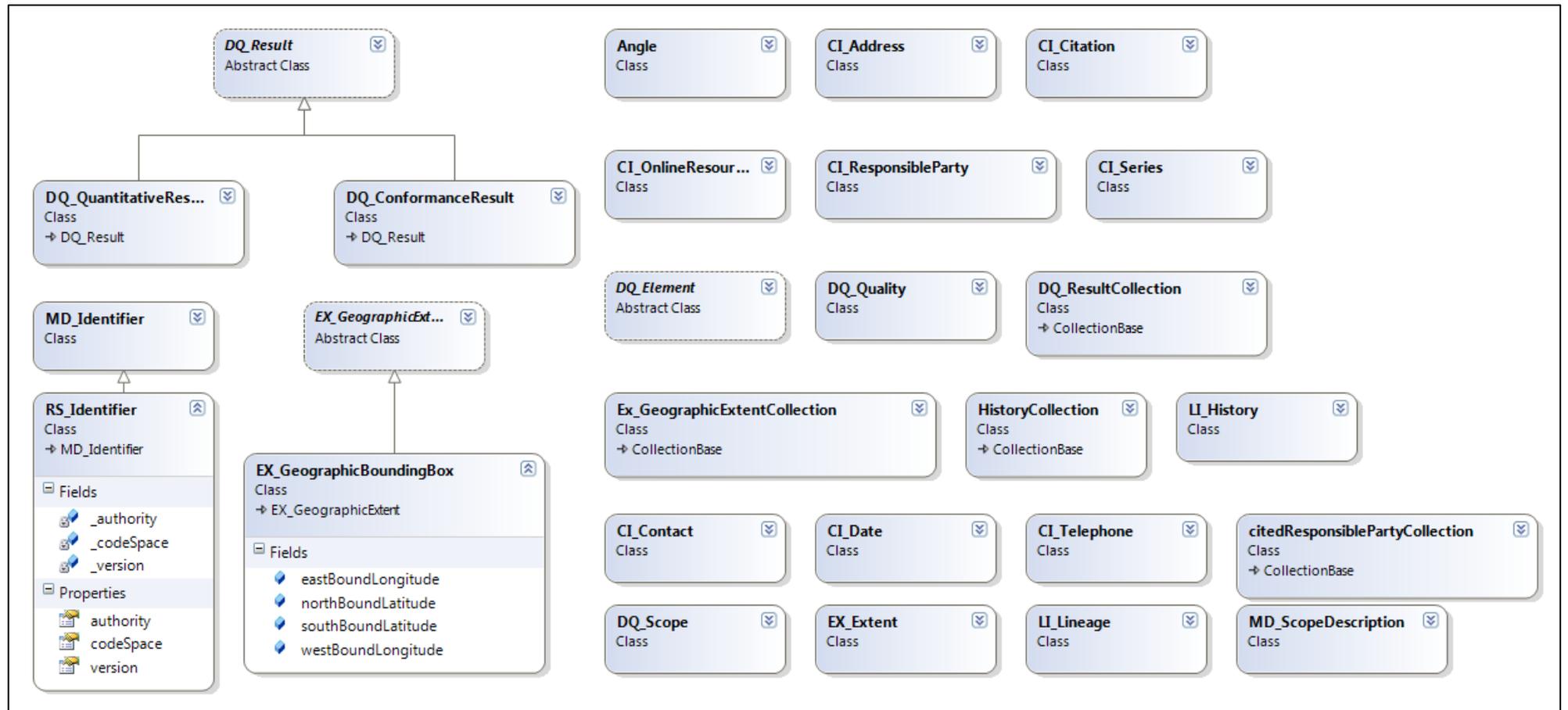


Figura 25 - Diagrama de classes - ISO 19115 - Fonte: o próprio Autor

A Figura 26 apresenta o diagrama de classes desenvolvido pelo autor apresentando as classes de negócio necessárias para a conclusão do programa.

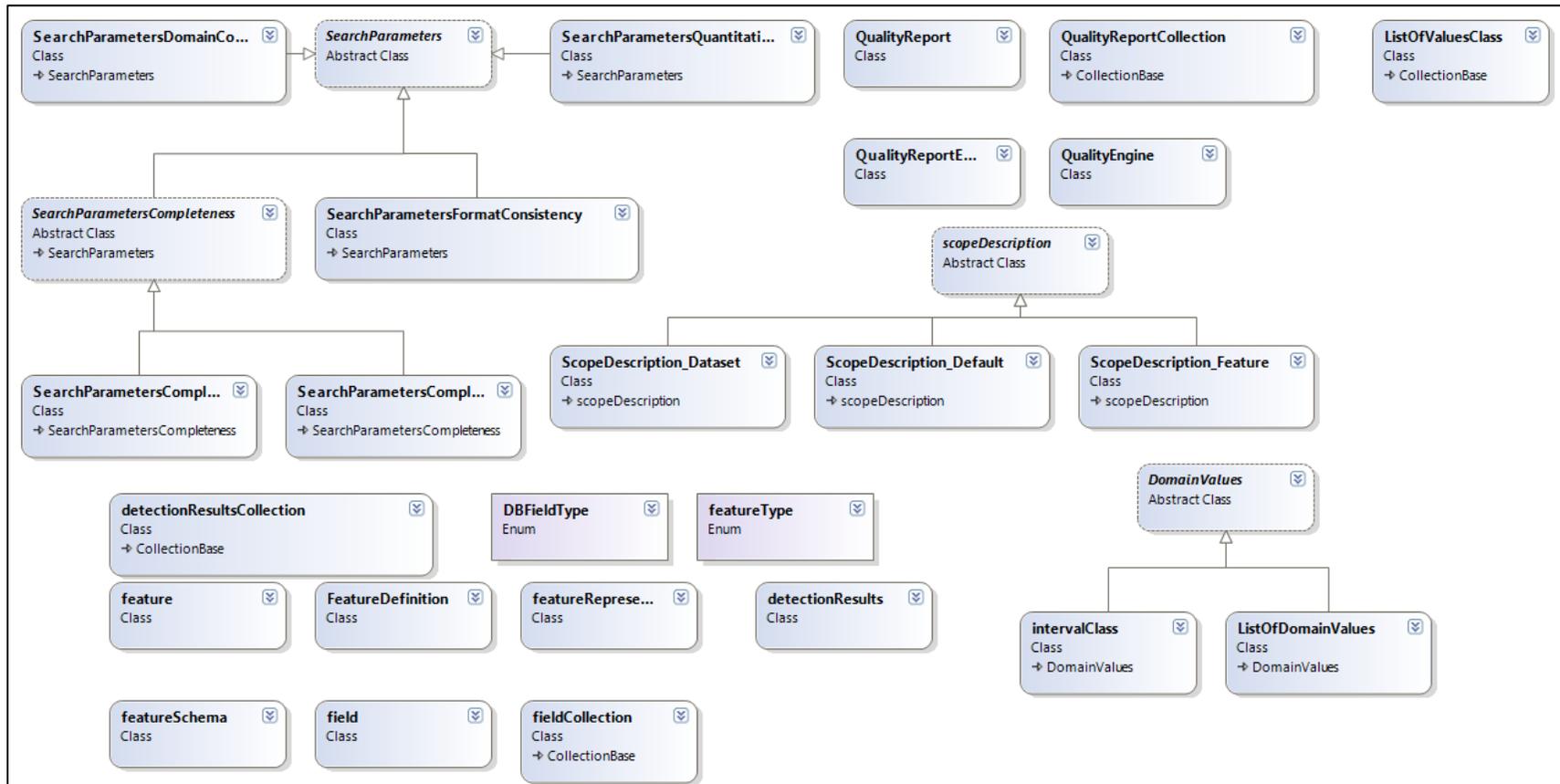


Figura 26 - Diagrama de classes de Negócio - Fonte: o próprio Autor

As classes de Negócio descrevem a lógica de comunicação estruturada da qualidade por parte do usuário e sua aplicação nas bases de dados de interesse.

O aplicativo construído pode ser descrito em 3 componentes principais: Dados espaciais, Padrões ISO e Metadados.

Dados Espaciais

O primeiro componente necessário para construção do aplicativo foi elaborado para ler e interpretar o conteúdo dos arquivos de dados espaciais.

As rotinas de leitura de dados se concentraram em duas entidades principais:

- Leitura de geometrias
 - Reconhecendo feições pontuais, lineares ou poligonais
- Leitura de atributos
 - Reconhecendo os atributos existentes na feição, e seus tipos. Por exemplo, o atributo NOME DE RUA, contém o tipo STRING; o atributo COMPRIMENTO contém o tipo REAL.

Cada formato demandou rotinas diferentes de leitura e reconhecimento de suas feições.

Padrões ISO

Existem vários padrões de avaliação e de comunicação de qualidade, conforme descritos no capítulo 0 desta dissertação. As definições contidas na ISO 19113 (ISO, 2002) tem sido uma das mais utilizadas (ZARGAR, 2009).

Metadados

O formato XML é o formato comumente aceito para armazenamento de metadados (CEMG-CONCAR, COMITÊ DE ESTRUTURAÇÃO DE METADADOS GEOSPACIAIS, 2009), e foi utilizado como formato padrão para armazenamento e processamento de dados pelo aplicativo.

Desenvolvimento do Aplicativo

As normas ISO 19113, 19114 e 19115, proveram as definições e esquema de elementos necessário para desenvolvimento do aplicativo de avaliação de dados espaciais. Por outro lado, tais normas e padrões não foram criados originalmente para uso exclusivo em aplicativos computacionais. Por exemplo, na descrição de completeza de dados, pode-se tratar de completeza de dados gráficos, do esquema de dados ou de registros. Tal limitação na ISO levou a criação de uma extensão à norma ou, a sua especialização¹⁵.

O aplicativo foi desenvolvido inicialmente para atuar como um Add-In do software Bentley Map Enterprise, e pode ser iniciado a qualquer momento pelo usuário através linhas de comando dentro do software.

A Figura 27 a seguir apresenta o fluxo geral de utilização do programa desenvolvido nesta dissertação.

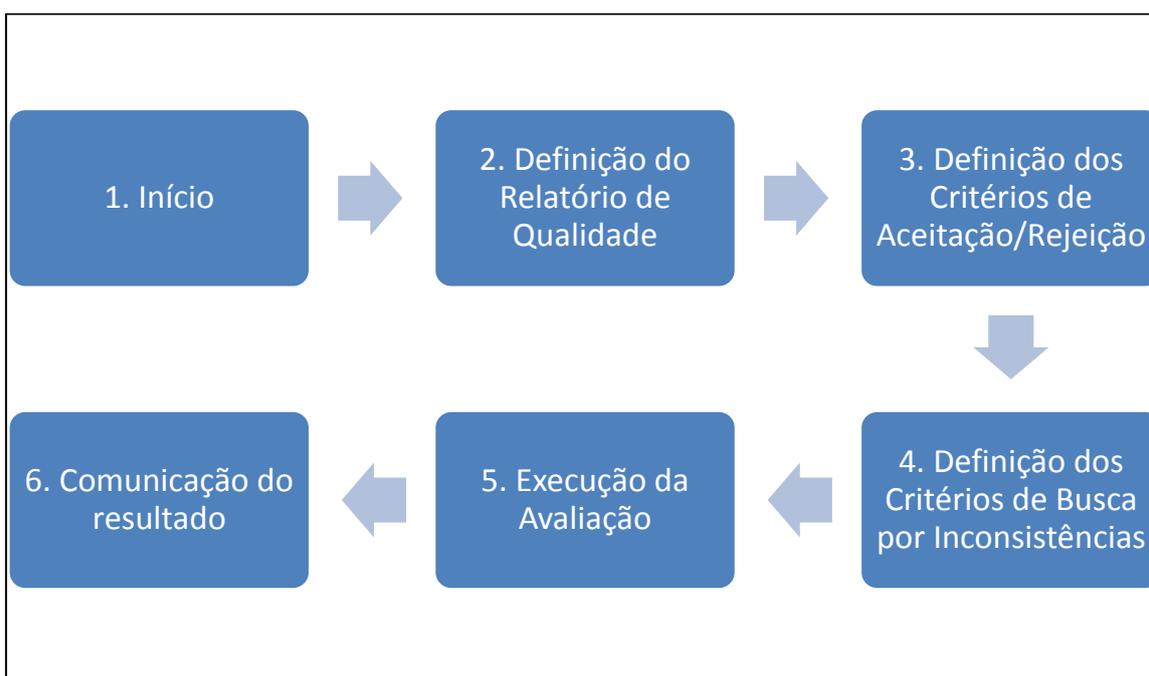


Figura 27 - Fluxo geral de utilização do programa

¹⁵ Uma especialização de uma norma, busca criar especificações única e voltadas para um determinado assunto ou área de conhecimento, à partir de uma norma previamente existente.

1. Início

O aplicativo inicia com uma chamada explícita do usuário, a qualquer momento após a abertura do arquivo de dados para avaliação.

2. Definição do Relatório de Qualidade

O usuário, através de interfaces específicas, apresentará os elementos e sub elementos de qualidade que melhor avaliam o dado para seu uso.

3. Definição dos Critérios de Aceitação / Rejeição do dado

O usuário irá definir qual será o critério de eliminação ou aceitação da qualidade do dado, de forma a permitir que o aplicativo usará este critério para decidir quando aceitar ou rejeitar a qualidade do dado em avaliação durante seu processamento. Por exemplo, o usuário poderá definir que o aplicativo irá rejeitar se mais de 10% dos elementos pertencentes a feição de **eixos de logradouro** se o atributo **nome de rua** possuir valores nulos.

4. Definição dos Critérios de Buscas Por Inconsistências

Nesta fase, o usuário irá definir o critério de busca por elementos cujas primitivas (geometrias ou atributos) serão avaliadas. Por exemplo, o usuário poderá definir que a busca por completeza será executada contra o atributo **material** contido na feição **postes**.

5. Execução da Avaliação

Ocorre o processo de avaliação do dado, onde este será auditado pelos critérios definidos pelo usuário do aplicativo.

6. Comunicação dos Resultados.

Uma vez completo o processo de avaliação, o aplicativo irá comunicar ao seu usuário o resultado final se o dado foi aceito ou não, e a explicação do motivo de tal decisão.

Interfaces de Usuário

Ao todo, foram desenvolvidas 22 telas de acesso ao usuário, de forma a permitir que ele comunique de maneira estruturada o método de qualidade que ele deseja submeter seu dado espacial.

A Figura 28 apresenta a tela principal do aplicativo desenvolvido durante a evolução desta dissertação.

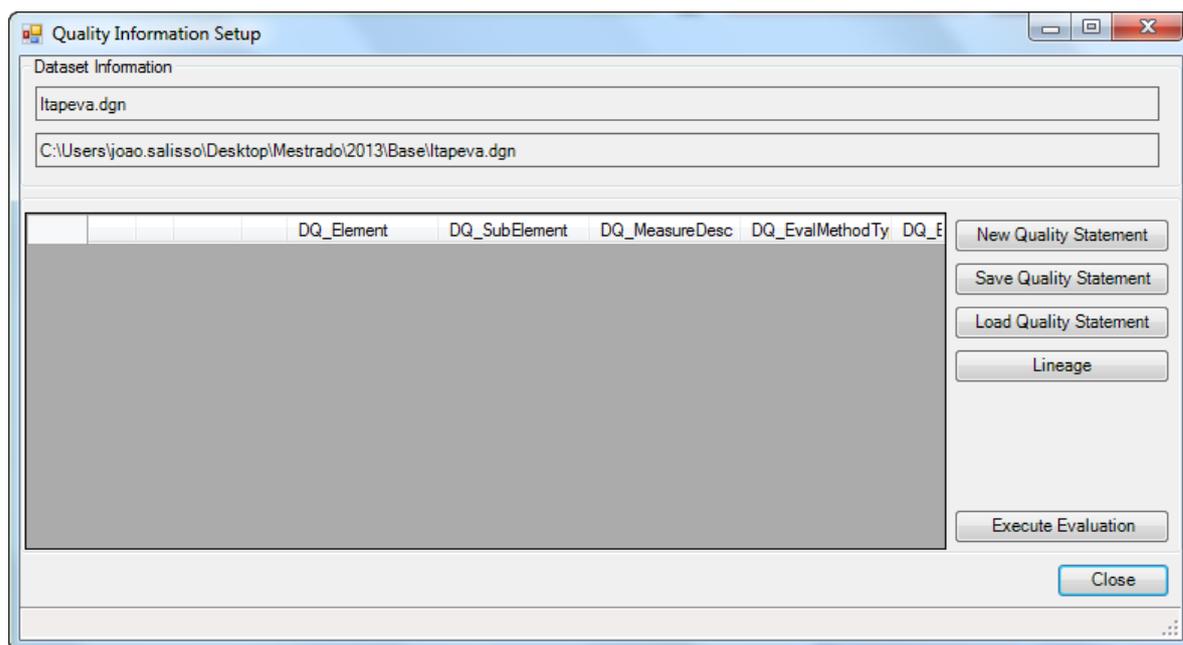


Figura 28 - Tela principal do aplicativo de avaliação de qualidade de dados

A tela principal apresenta ao usuário qual o conjunto de dados que será avaliado e o diretório onde este dado se localiza.



Figura 29 - Detalhe do conjunto de dados em avaliação

A descrição da qualidade do dado é realizada através da configuração do relatório de qualidade (QUALITY REPORT), que é apresentada ao usuário através do botão "New Quality Statement", conforme apresentado na Figura 30 a seguir.

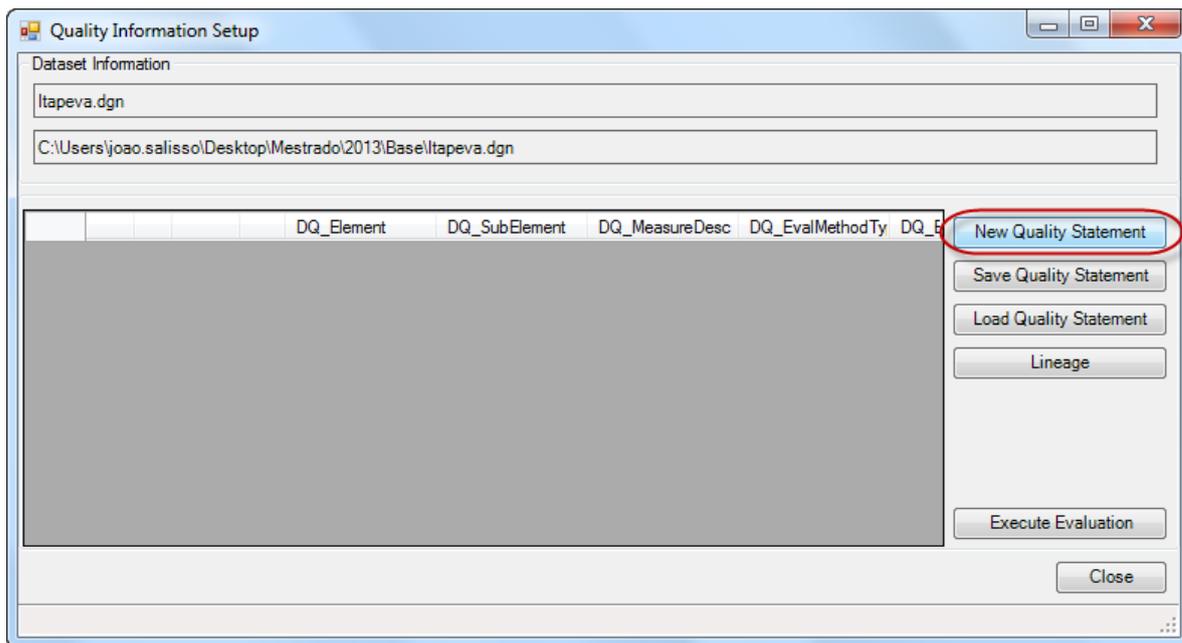


Figura 30 – Tela inicial do aplicativo

O relatório de qualidade, apresentado a seguir na Figura 31, acompanha o usuário na descrição da qualidade a ser executada em seu conjunto de dados espacial.

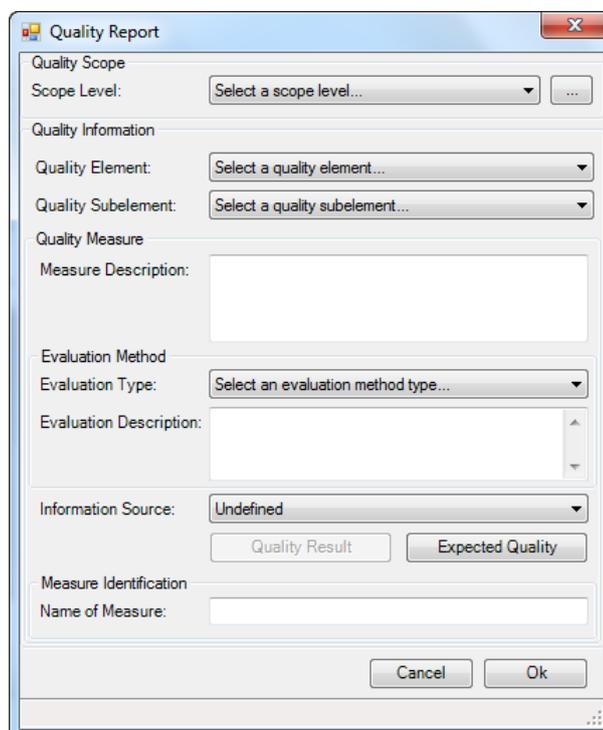


Figura 31 - Relatório de qualidade

Escopo da Qualidade

O nível de escopo identifica os limites do processo de avaliação da qualidade do dado. Por exemplo, o processo de avaliação pode ser realizado em todo o conjunto de dados; ou pode estar limitado a uma região geográfica; pode ser aplicado apenas ao banco de dados, ou a uma feição específica.

Ao selecionar **Feature**, o usuário pode querer definir uma limitação na área geográfica de atuação da avaliação espacial. A Figura 32 apresenta a tela onde esta configuração será feita pelo usuário.

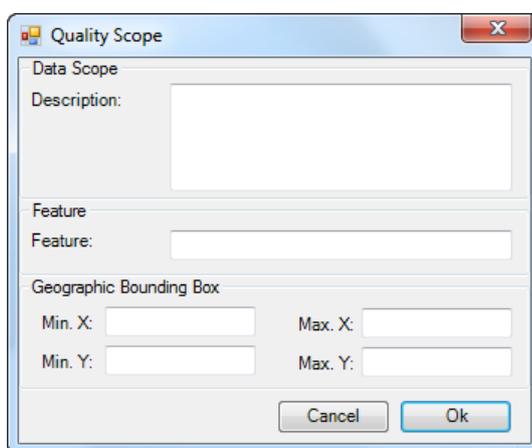


Figura 32 - Delimitação espacial da área de atuação da avaliação espacial

Informação da Qualidade / Quality Information

Em seguida, o usuário deve definir o elemento e subelemento de qualidade que melhor avaliam o dado de acordo com seu propósito. Ainda na tela de relatório de qualidade (Quality Report), o usuário pode fazer tais seleções, conforme demonstrado na Figura 33, a seguir.

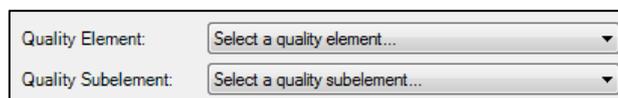
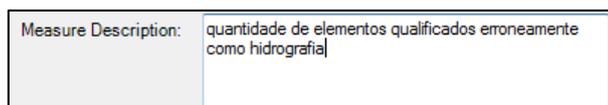


Figura 33 - Seção da tela Quality Report

Cada elemento de qualidade possui seus próprios subelementos de qualidade, conforme descrito no capítulo 2 deste documento.

Medida de Qualidade / Quality Measure

Ainda na tela de relatório de qualidade, o usuário pode fazer uma descrição da medida de qualidade que pretende aplicar ao dado. Mantendo conformidade com a ISO 19114, esta informação pode ser inserida conforme seção da tela principal, mostrada na Figura 34, a seguir.



The image shows a rectangular input field with a light gray background. On the left side, the text 'Measure Description:' is displayed in a small font. To the right of this label, the text 'quantidade de elementos qualificados erroneamente como hidrografia' is entered into the field.

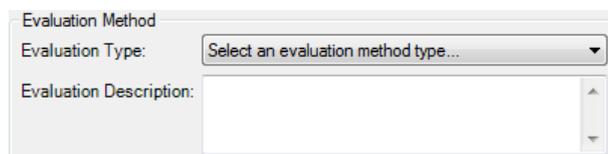
Figura 34 - Descrição da medida de qualidade

Esta informação é meramente descritiva e não afeta a automação do procedimento. Mas cabe lembrar que é importante informar aos próximos usuários desta informação, ou ao seu provedor, que medida de avaliação será aplicada ao dado espacial.

Descrição do Método de Avaliação

Conforme descrito no capítulo 2, o método de avaliação pode ser **direto interno**, **direto externo** ou **indireto**. Esta informação pode ser apresentada pelo usuário, bem como uma descrição de como será aplicada a avaliação. Por exemplo: contagem dos elementos na camada de hidrografia que não possuem as características físicas de elementos de hidrografia (elemento linear, cor azul, estilo de linha contínuo, etc.).

A Figura 35 a seguir, apresenta a tela do aplicativo onde o usuário pode informar o tipo de avaliação que será aplicado ao dado espacial.



The image shows a dialog box titled 'Evaluation Method'. It contains three main components: a label 'Evaluation Type:' followed by a dropdown menu with the text 'Select an evaluation method type...'; a label 'Evaluation Description:' followed by a text area with a vertical scrollbar on the right side.

Figura 35 - Tipos e descrição de avaliação

Fonte da Informação / Information Source, Qualidade Informada e Qualidade Esperada

A avaliação da qualidade do dado espacial pode seguir dois caminhos distintos: avaliar o dado conforme os requisitos de aplicação do usuário ou avaliar o dado espacial conforme a qualidade informada pelo seu provedor. Não existe na ISO 19113, 19114 ou 19115 nenhuma definição para este tipo de informação, mas esta se mostrou necessária para permitir ao usuário informar quando for necessário realizar a validação da informação de qualidade informada pelo provedor.

Conforme a Figura 36, o campo **information source**, permite ao usuário informar se o dado possui informação sobre a qualidade do dado informada pelo seu provedor.

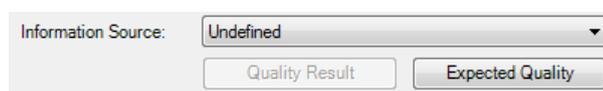


Figura 36 - Fonte da informação, qualidade informada e qualidade esperada.

Na Figura 36, a fonte da informação pode ser: **provider** ou **user**. Uma vez que o usuário marque a opção **information source** com o valor **provider**, o aplicativo irá habilitar o botão **Quality Result**, onde a informação da qualidade informada pelo usuário poderá ser registrada no banco de dados do aplicativo.

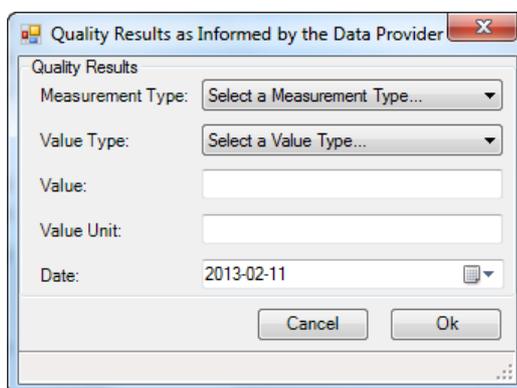


Figura 37 - Qualidade do dado reportado pelo provedor

Se houver informação de qualidade relatada pelo provedor, será possível aplicar uma avaliação de conformidade do dado.

Nesta fase, o aplicativo requer que o usuário apresente o critério de aceitação (qualidade mínima) para determinar se o dado espacial possui qualidade para seu propósito.

Tais parâmetros podem ser informados na tela apresentada pela Figura 38.

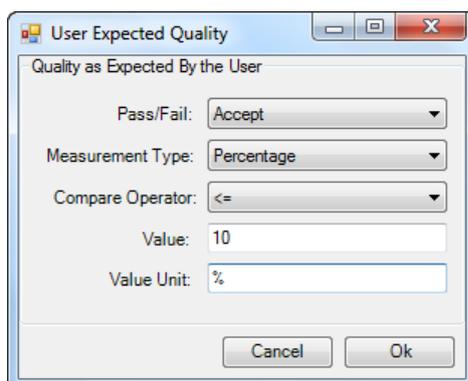


Figura 38 - Qualidade esperada pelo usuário

A tela apresentada na Figura 38, permite que o usuário defina o seu critério de aceitação/rejeição do dado espacial.

Ainda na tela de relatório de qualidade (*quality report*), é necessário apresentar um nome, um identificador único para o processo de avaliação que foi configurado. Desta forma é possível identifica-lo em uma lista de outros processos.

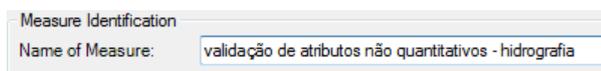


Figura 39 - Identificação do processo de avaliação

A Figura 39 apresenta a tela onde é feita a identificação do processo de avaliação. Esta identificação é apresentada em forma de texto livre e não afeta o processo de avaliação.

Uma vez concluída a configuração da avaliação da qualidade que o usuário deseja conduzir, este processo é listado na tela principal do aplicativo, conforme apresentado na Figura 40.

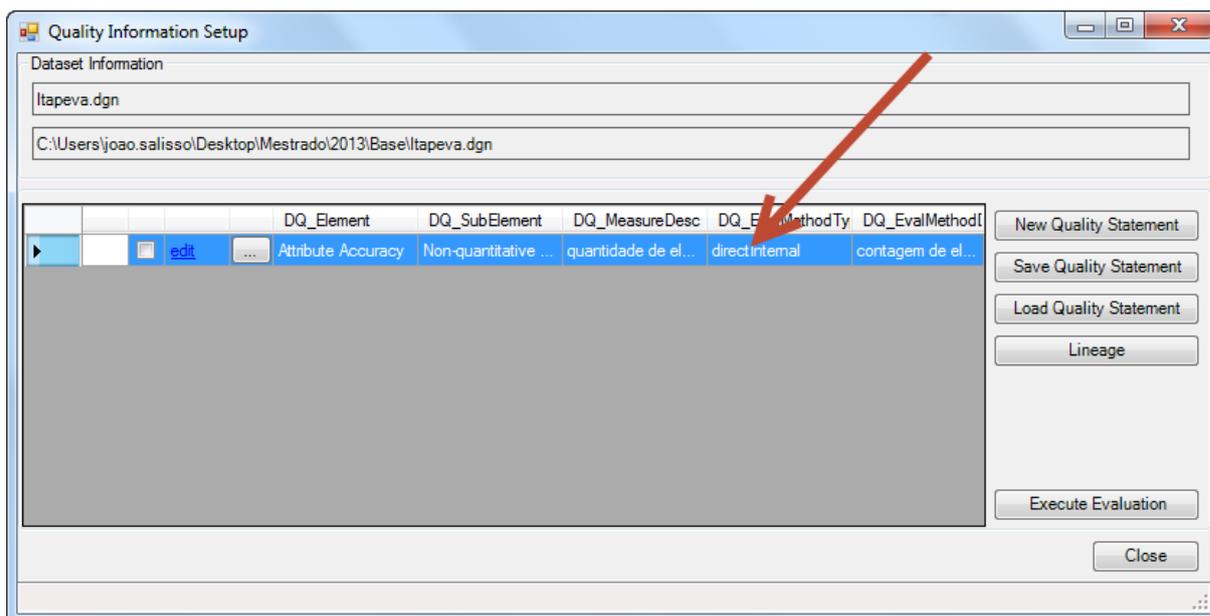


Figura 40 - Configuração do processo de avaliação finalizado

Configuração de Buscas por Inconsistências

Uma vez concluída a configuração da avaliação de qualidade, é necessário identificar os parâmetros de busca por inconsistências no dado espacial.

Esta configuração é feita acessando o botão "... " na tela principal.

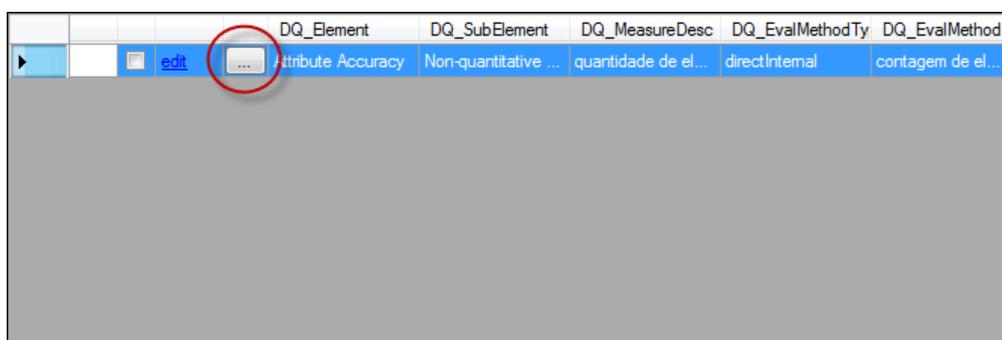
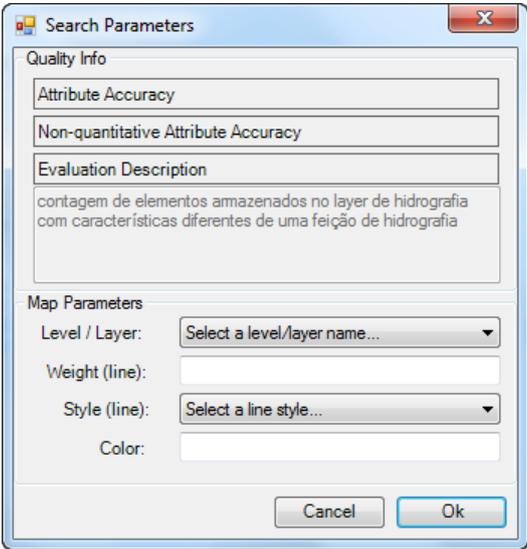
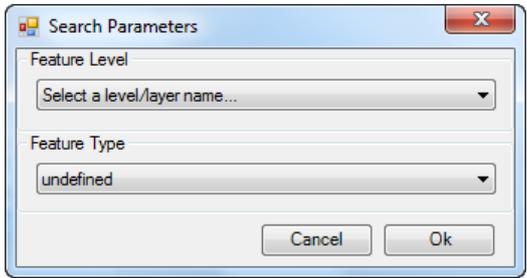


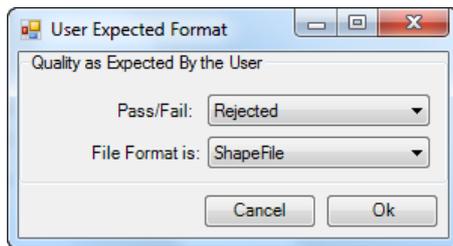
Figura 41 - Configuração de busca por inconsistências.

Antes de prosseguir com as próximas telas do aplicativo, é necessário informar que, as o termo inconsistências utilizado nesta seção, se refere a características informadas pelo usuário, que identificam os elementos (geometria e atributos alfanuméricos) existentes no dado espacial, como adequados ao seu propósito. Desta maneira, elementos que não estejam em conformidade com tais parâmetros serão considerados desvios ou inconsistências do dado.

As inconsistências que o aplicativo irá buscar no dado espacial podem variar de acordo com os elementos de qualidade que o usuário selecionar na tela de avaliação. Sendo assim o Quadro 8 a seguir apresenta as telas de configuração dos parâmetros de busca por inconsistências nos dados que foram criadas cada elemento e sub elemento de qualidade.

Quadro 8 - Configuração dos parâmetros de busca de inconsistências

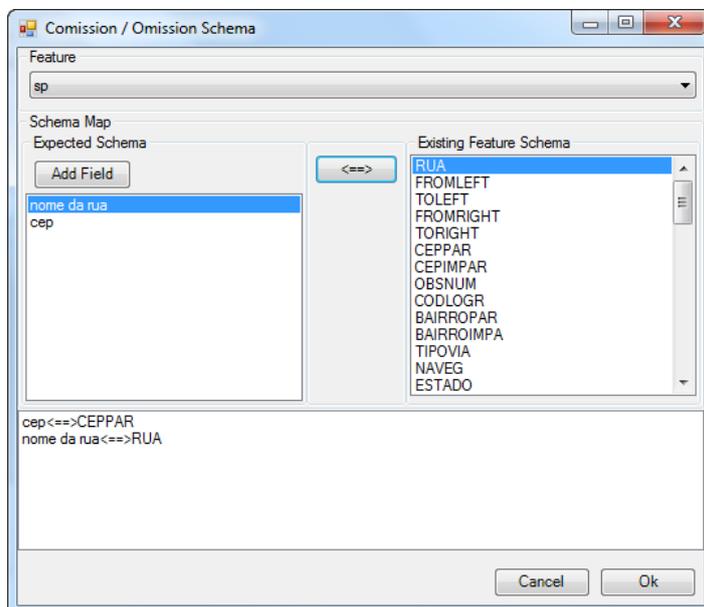
Tela do aplicativo	Descrição da inconsistência a ser configurada
	<p>Busca por inconsistências de atributos não quantitativos.</p> <p>Os parâmetros de busca são:</p> <ul style="list-style-type: none"> - Nível/Camada - Peso (espessura) do elemento linear - Estilo (tracejado, contínuo, etc.). - Cor (do elemento)
	<p>Busca por inconsistências conceituais, por exemplo, uma camada de eixos de logradouro, contendo elementos poligonais.</p> <p>Os parâmetros de busca são:</p> <ul style="list-style-type: none"> - Nível / Camada - Tipo da feição <ul style="list-style-type: none"> o Pontual o Linear o Poligonal



Busca por inconsistências de formato

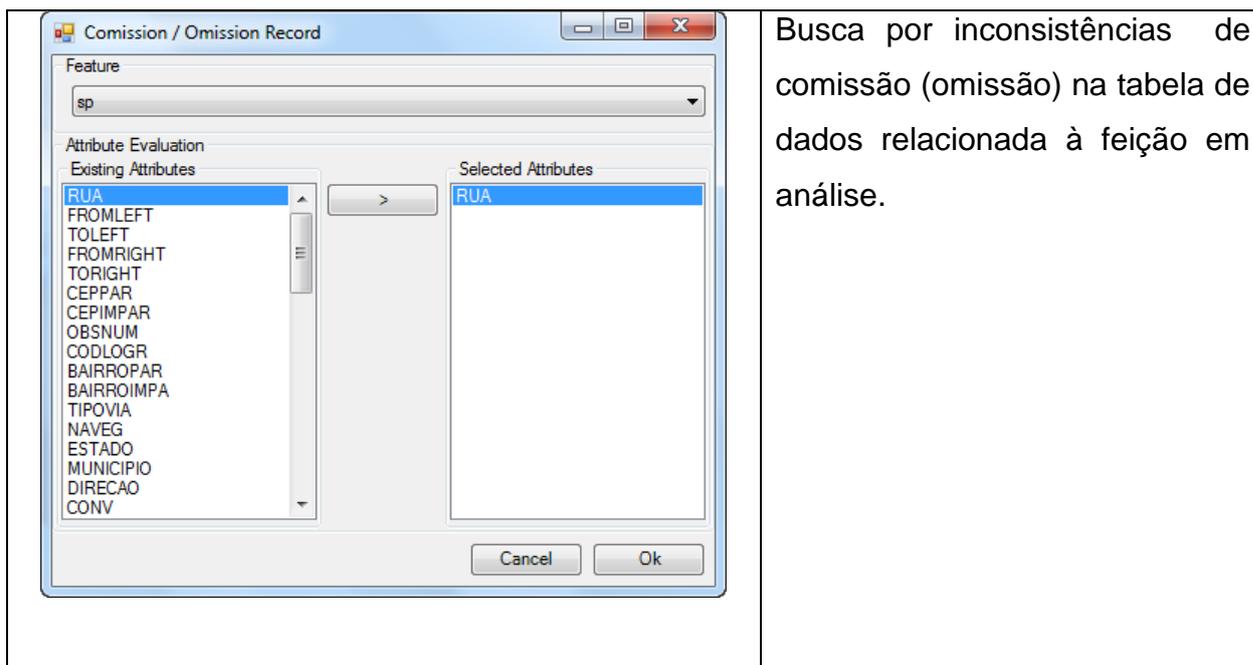
Os parâmetros de busca são:

- Aceita/Rejeita
- Formato do arquivo pode ser:
 - o DGN
 - o DWG
 - o Shapefile



Busca por inconsistências de comissão (ou omissão) no esquema de dados.

- Comissão: excesso de informação no dado espacial.
- Omissão: falta de informação no dado espacial.



Busca por inconsistências de comissão (omissão) na tabela de dados relacionada à feição em análise.

Utilização do Aplicativo Para Avaliação da Qualidade do Dado Espacial

Esta seção do anexo apresenta o modo de utilização do aplicativo desenvolvido para realização da avaliação da qualidade do dado espacial. Os exemplos aqui apresentados utilizaram as bases de dados descritas no capítulo 3.

Avaliação de Completeza de Dados

Comissão

O procedimento de avaliação de completeza de dados desenvolvido nesta dissertação verifica a existência de erros de omissão ou comissão de dados concentrando-se em seus atributos. Desta forma, para aplicação bem sucedida deste procedimento, é necessário que a feição a ser avaliada possua uma tabela de dados ou um banco de dados vinculado a ela.

O procedimento de avaliação de completeza seguiu o fluxograma apresentado na Figura 42, a seguir.

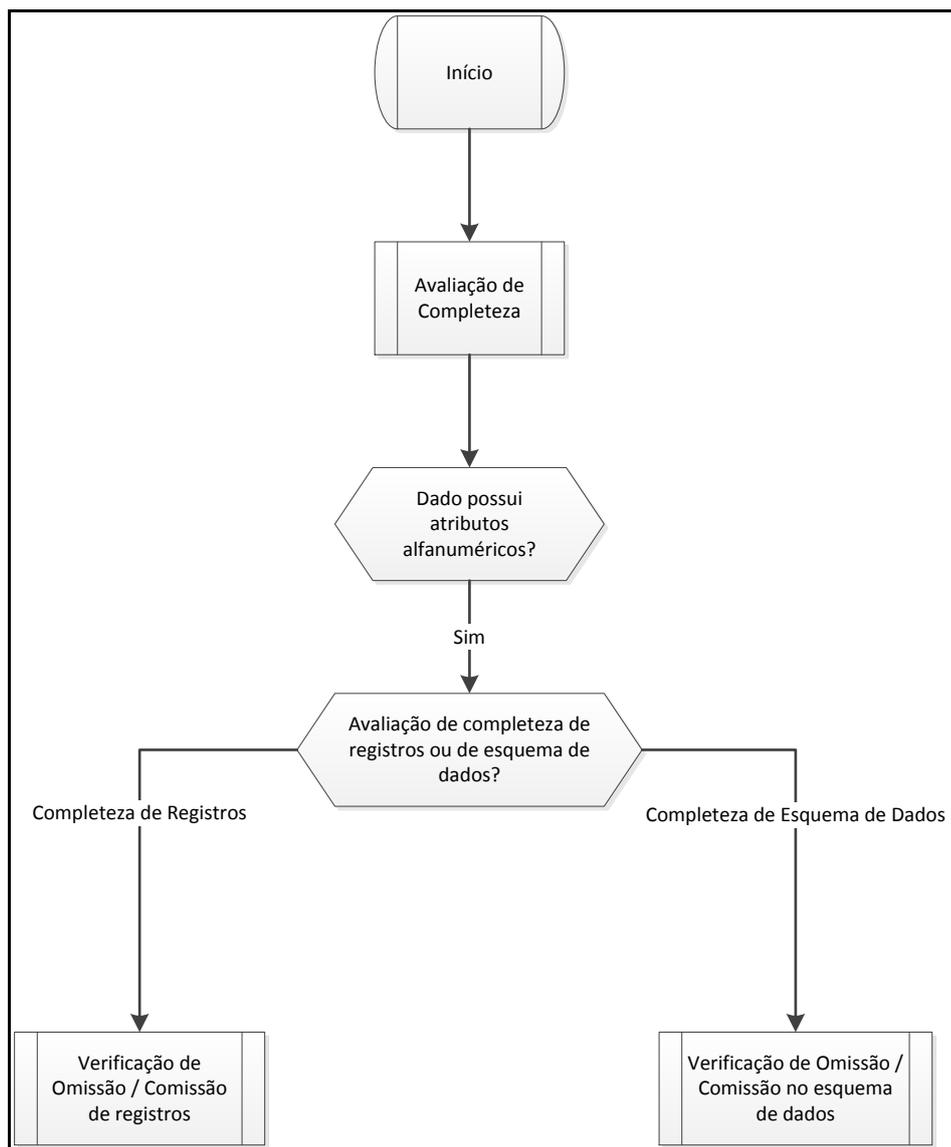


Figura 42 - Fluxograma para avaliação de completeza de dados

O usuário ao iniciar o aplicativo de apoio à avaliação da qualidade do dado espacial, não conseguirá prosseguir com o procedimento se o dado espacial em análise não possuir um banco de dados ou uma tabela de dados vinculado a sua feição.

O procedimento iniciou com a criação do relatório de qualidade através do botão **New Quality Statement** na tela principal do aplicativo (Figura 43).

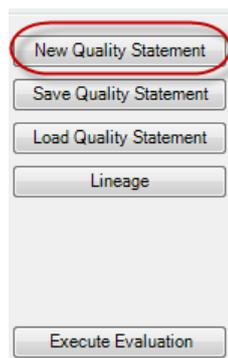


Figura 43 - Início do procedimento de avaliação do dado

A tela **Quality Report** surge e podem-se descrever os elementos de qualidade que melhor refletem a utilização que fará do dado espacial.

O primeiro dado informado foi o escopo (**quality scope**) do processo de avaliação.



Figura 44 - Seleção do escopo de qualidade para atributos

O procedimento em questão foi aplicado no banco de dados, especificamente na tabela de atributos vinculada a feição, portanto, o nível do escopo definido foi **Attribute** (atributo, em tradução livre) conforme a Figura 44 acima.

Em seguida, foi definido o Elemento e Sub Elemento de qualidade (Figura 45) como critérios de avaliação.

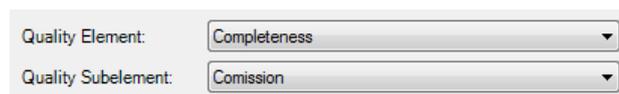


Figura 45 - Seleção do Elemento e Sub Elemento de qualidade de dados

No próximo passo, foi descrita a Medida de Qualidade que foi aplicada ao dado. Esta descrição é um texto livre, conforme mostrado na Figura 46.

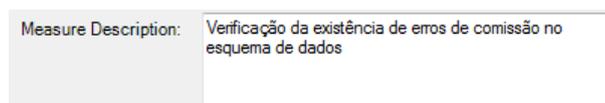


Figura 46 - Descrição da avaliação de qualidade a ser aplicada

Ainda na descrição da qualidade, foi necessário descrever o método de avaliação, onde foi informado o tipo de avaliação (direto interno, direto externo ou indireto) conforme Figura 47 indica. Apesar deste campo de descrição ser de livre preenchimento, sua importância está na documentação para outros usuários sobre o procedimento utilizado para avaliar aquele elemento e sub elemento de qualidade.

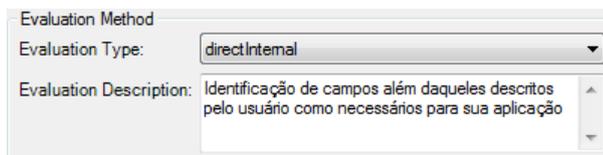


Figura 47 - Tipo e descrição da avaliação de qualidade.

Também foi necessário informar sobre a fonte da informação de qualidade, se existe informação de qualidade fornecida pelo provedor, pelo usuário, ou de ambos.

Caso a informação esteja sendo avaliada apenas sob a perspectiva do usuário, o campo **information source** poderá ser preenchido como **User**. Se a avaliação deverá levar em conta também a informação de qualidade informada pelo provedor do dado, então deverá ser selecionado **Provider**. Conforme a Figura 48 apresenta, caso a opção seja **User**, o botão **Quality Result** ficará desabilitado.

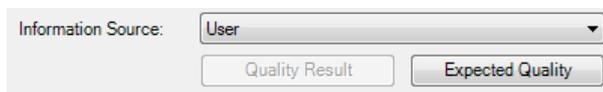


Figura 48 - Definição da fonte de informação e qualidade esperada

Além destas informações, foi necessário informar quais parâmetros de aceitação / rejeição foi utilizada pelo aplicativo.

Por se tratar de avaliação de completeza, o aplicativo conduziu para uma definição mais específica: Para avaliar a completeza do dado em relação ao esquema de dados (Figura 49) ou em relação aos registros vinculados a feição espacial (Figura 50).

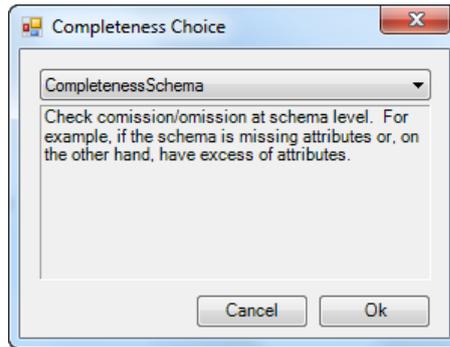


Figura 49 - Avaliação por completeza de esquema de dados

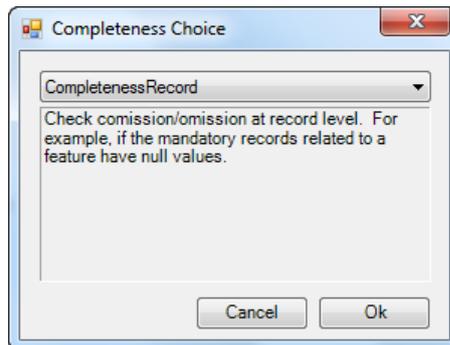


Figura 50 - Avaliação por completeza de registros

Para avaliação da completeza do dado em relação ao seu esquema de dados, pode-se definir seu critério de aceitação aceitando ou rejeitando a feição, conforme Figura 51, a seguir.

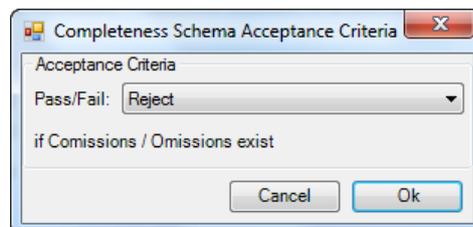


Figura 51 - Critério de aceitação / rejeição para avaliação de completeza por esquema de dados.

Para avaliação da completeza do dado em relação aos registros existentes no banco, a avaliação assume uma característica estatística, onde se pode estabelecer um limite de aceitação (percentual, contagem, existência, etc.) para os problemas encontrados no dado espacial.

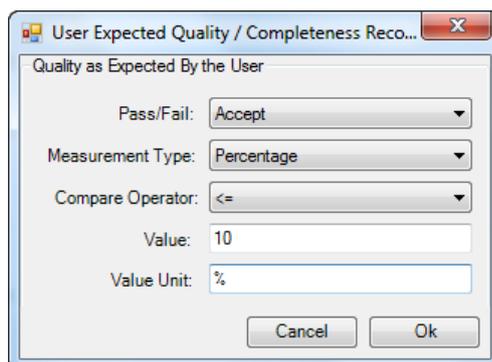


Figura 52 - Critério de aceitação para completudeza por registros do conjunto de dados.

A Figura 52 exemplifica a tela de critério de aceitação para completudeza de dados por registros, onde foi definido que o dado será aceito se o percentual de registros com erro de comissão for menor ou igual a 10% do total de registros no banco de dados.

Definiu-se também um nome para o processo de avaliação, uma maneira única de identificar este processo, conforme descrito na Figura 53.

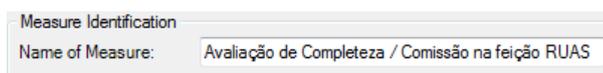


Figura 53 - Identificação do processo de avaliação

Uma vez definido o processo de criação do relatório de qualidade, definiu-se o critério de buscas, em conformidade com o método apresentado no capítulo 3.2.2.1.

Este critério instruiu o aplicativo sobre como identificar uma inconsistência nos dados espaciais. Esta fase do processo é iniciada clicando no botão "...", conforme Figura 54 a seguir.

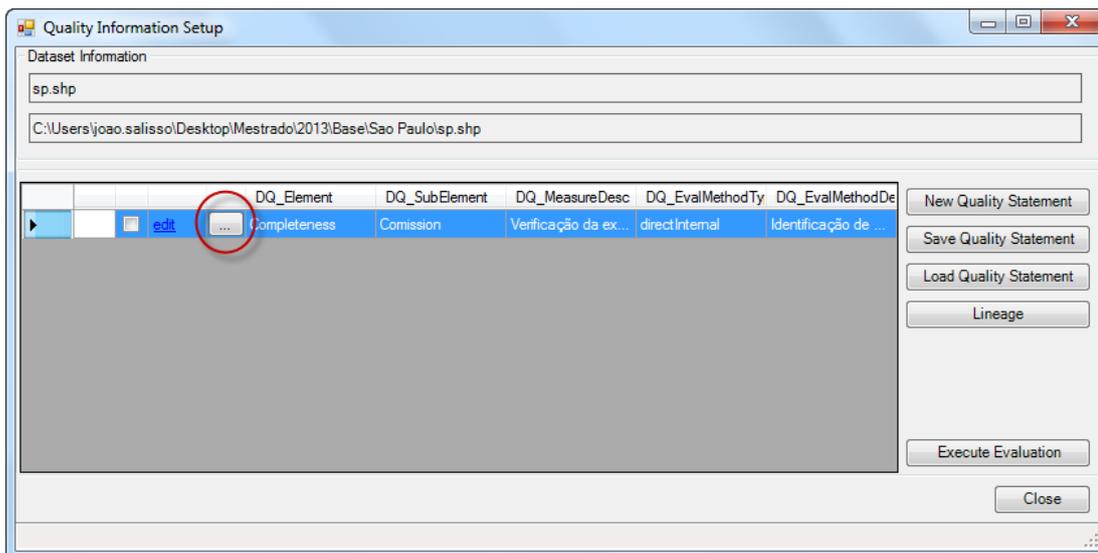


Figura 54 - Definição dos critérios de busca por inconsistências no dado espacial

Esta descrição é feita pela tela **Comission / Omission Schema** apresentado pela Figura 55 a seguir.

Esta seção do aplicativo auxilia o usuário no processo de descrição do esquema de dados que atende ao seu propósito. No exemplo da Figura 55, a feição escolhida se chama **SP**, e esta feição contém todos os eixos de logradouro de São Paulo bem como seus atributos (nome de rua, cep, numeração de início e de fim, etc.). Uma vez que a feição a ser avaliada é selecionada no campo **Feature**, todos os atributos definidos para esta feição são listados na área **Existing Feature Schema**, onde é possível visualizar todos os campos disponíveis para aquela feição.

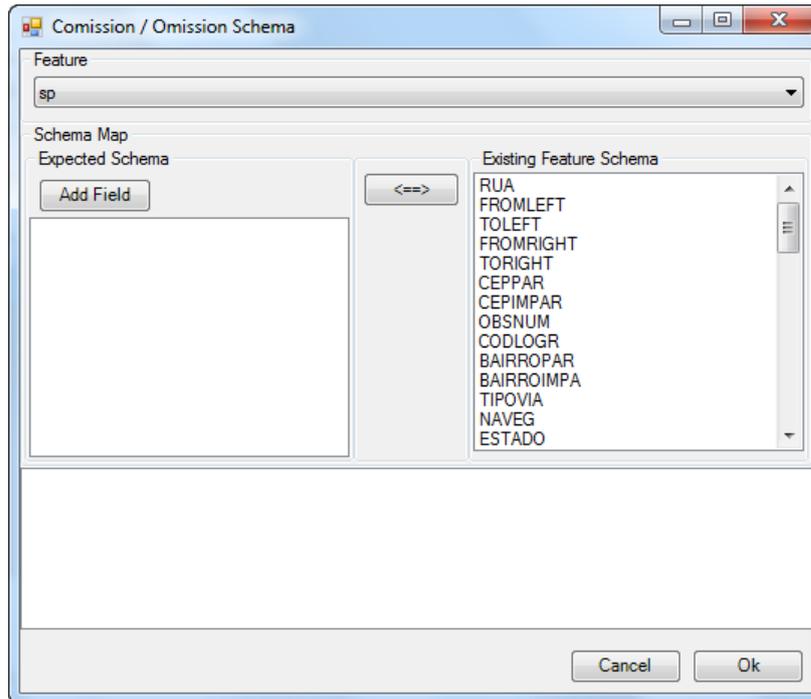


Figura 55 - Definição do esquema e atributos existentes

No passo seguinte, foram descritos quais campos deverão obrigatoriamente existir no esquema de dados e mapeá-lo com o esquema de dados da feição.

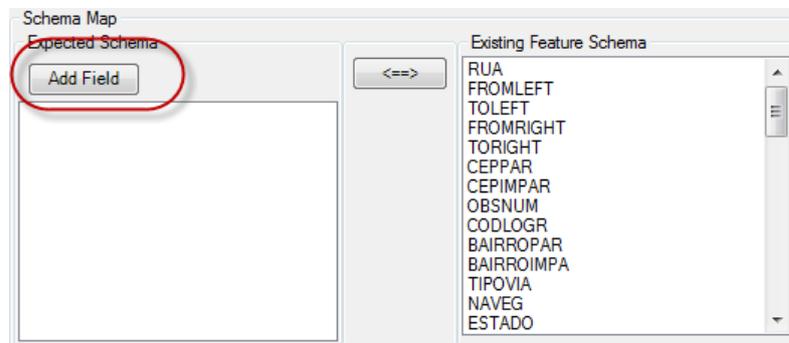


Figura 56 - Criação do esquema ideal ao usuário

Para isso, seleciona-se o botão **Add Field** (Figura 56), onde foi informada a descrição dos campos de interesse.

Nesta tela (Figura 57) foi definido o nome do campo de interesse (NOME DE RUA), e seu tipo (CARACTERE).

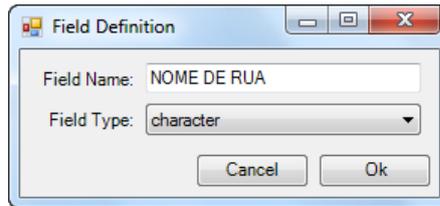


Figura 57 - Definição dos campos de interesse do usuário

Finalmente, relacionou-se o campo recém-criado, com o campo disponível na feição. Isto foi feito através da interface apresentada na Figura 58, que permite realizar o mapeamento entre os campos de interesse do usuário e os campos existentes na base de dados.

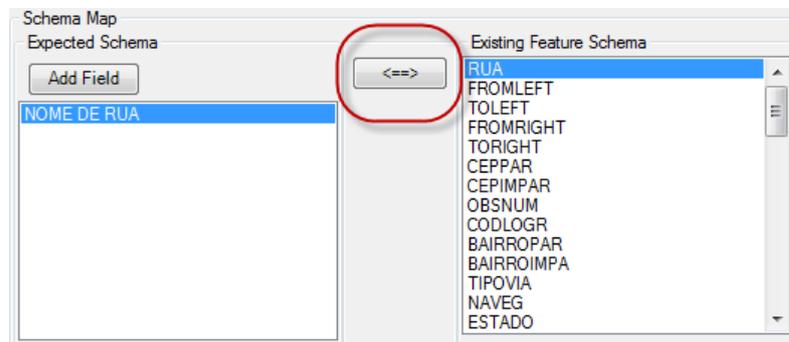


Figura 58 - Relacionando campos disponíveis com o esquema de dados esperado.

Voltando a tela principal do aplicativo (Figura 59), pode-se ver que o esquema de dados já efetuou a validação do modelo de dados existente contra o modelo de dados esperado. Como a instrução do usuário era a de rejeitar o dado caso exista comissão no esquema de dados, ou seja, rejeitar o dado caso exista mais atributos do que aqueles que o usuário deseja, o aplicativo sinalizou imediatamente a rejeição na tela principal.

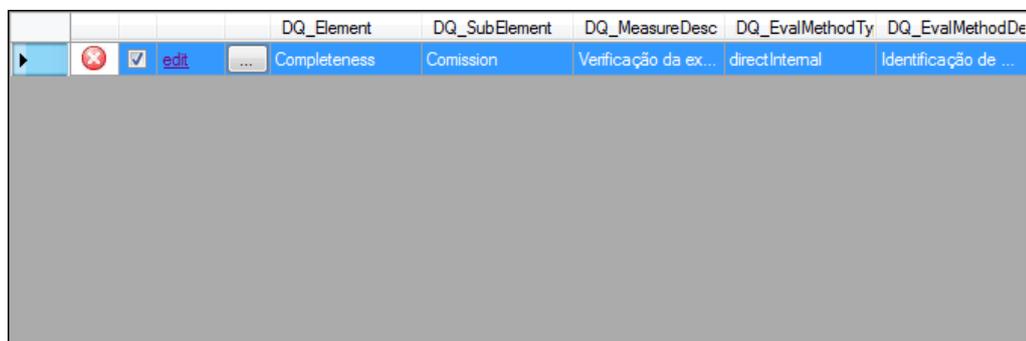


Figura 59 - Rejeição do dado

Omissão

Neste exemplo, foi utilizado o mesmo conjunto de dados do exemplo anterior. Desta vez, a avaliação de completeza levou em consideração a omissão de dados ao nível do registro. Desta forma, o aplicativo varreu o banco de dados em busca dos atributos definidos pelo usuário cujos valores eram nulos.

A Figura 60 a seguir, apresenta a tela de relatório de qualidade que foi configurada com os dados para esta avaliação.

The image shows a 'Quality Report' dialog box with the following configuration:

- Quality Scope:** Scope Level: Attribute
- Quality Information:** Quality Element: Completeness; Quality Subelement: Omission
- Quality Measure:** Measure Description: Verificação de ocorrências de registros com nome de rua com valores nulos.
- Evaluation Method:** Evaluation Type: directInternal; Evaluation Description: contagem de registros do atributo NOME_RUA onde seus valores são nulos.
- Information Source:** Provider; Buttons: Quality Result, Expected Quality
- Measure Identification:** Name of Measure: Avaliação de completeza / comissão por registros

Buttons: Cancel, Ok

Figura 60 - Configuração do relatório de qualidade por omissão de registros

A qualidade deste dado, conforme apresentada pelo seu provedor era de que menos de 15% dos registros de seu banco de dados estava incompleta (nulos), segundo consta na documentação do dado, esta informação de qualidade foi aferida pelo provedor em 14 de Junho de 2011. O usuário deverá apresentar este dado para o aplicativo, através do botão **Quality Result**

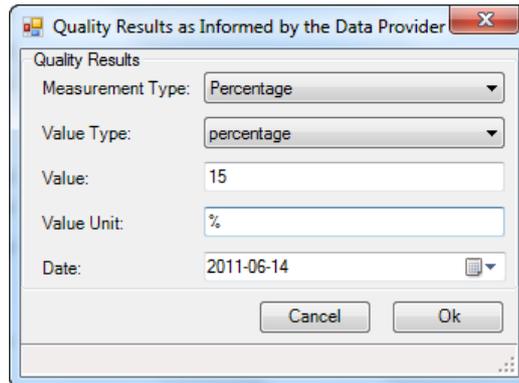


Figura 61 - Informação de qualidade apresentada pelo provedor do dado espacial.

O usuário deve agora informar o critério de aceitação/rejeição do dado em avaliação. Isto pode ser feito, conforme visto no exemplo anterior, através do botão ***Expected Quality***.

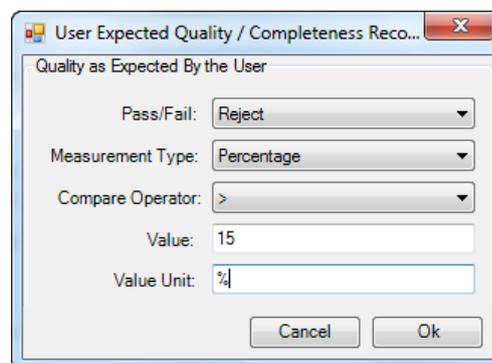


Figura 62 - Critério de aceitação configurado para rejeitar o dado espacial caso mais de 15% dos registros presentes no banco possuam valores nulos.

Nesta avaliação, o aplicativo deverá rejeitar o dado espacial se mais de 15% dos registros no banco de dados possuírem valores nulos.

O critério de buscas neste exemplo irá descrever quais campos dentre os disponíveis no modelo de dados da feição em análise não poderão ter valores nulos. Esta informação é inserida no aplicativo pela tela ***Comission / Omission Record***, conforme apresentado na Figura 63.

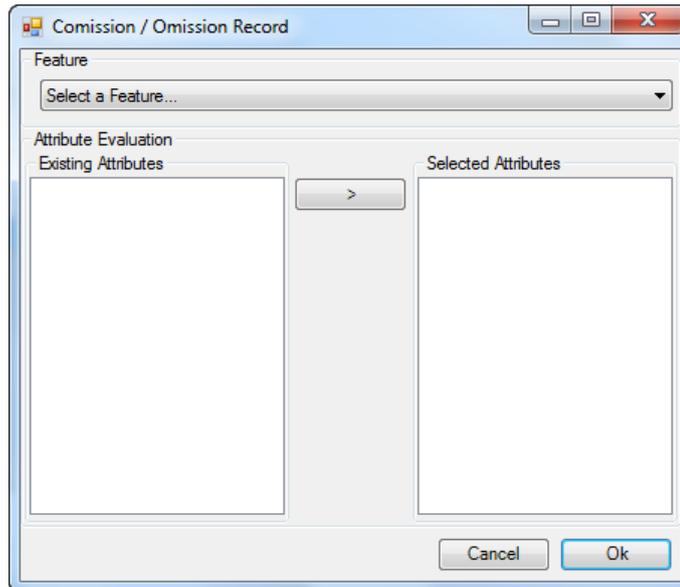


Figura 63 - Configuração de busca por omissão/comissão de registros

Ao selecionar a feição em análise, neste exemplo a feição de eixos de logradouros chamada **SP**, o aplicativo novamente irá apresentar ao usuário todos os atributos disponíveis em seu modelo de dados (Figura 64).

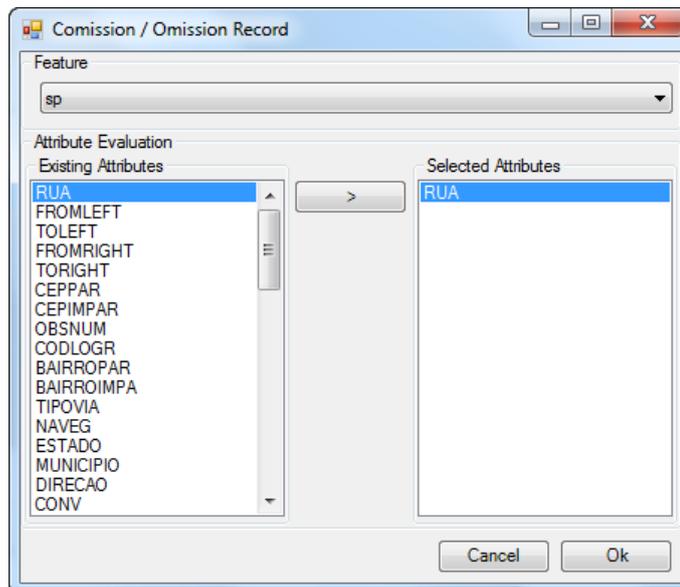


Figura 64 - Configuração do critério de busca por campos nulos.

Uma vez concluído esta etapa do processo, o aplicativo está pronto para avaliar o dado espacial.

Na tela principal do aplicativo, o usuário deverá selecionar o botão **Execute Evaluation**, este botão irá iniciar o processo de avaliação da qualidade do dado espacial de acordo com os parâmetros de aceitação e critérios de busca definidos pelo usuário.

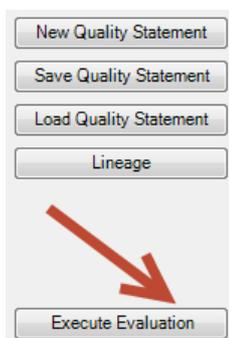


Figura 65 - Início do processo de avaliação

Ao final do processo, o aplicativo irá apresentar um relatório de avaliação (Figura 66), onde os relatórios de avaliação da qualidade do dado são detalhados ao usuário.

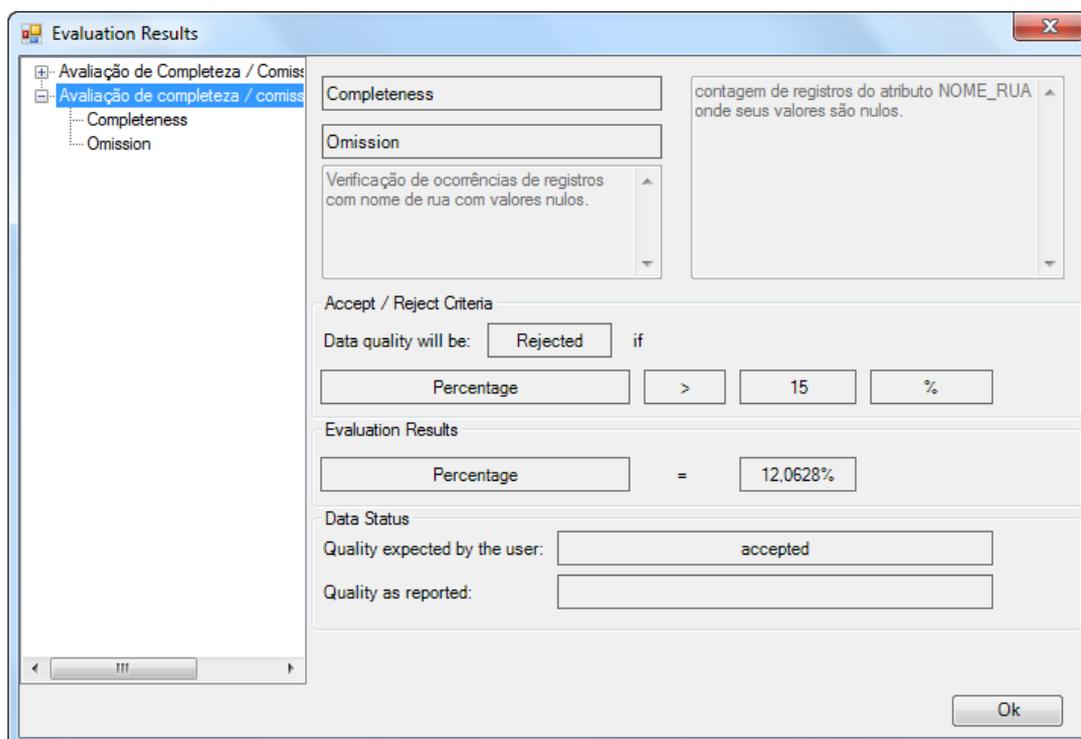


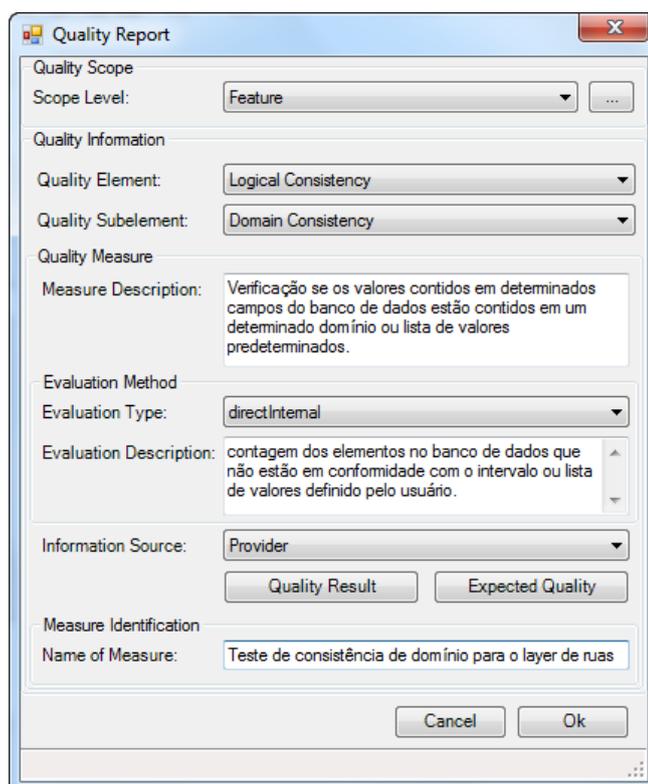
Figura 66 - Relatório final de avaliação do dado espacial.

Avaliação de Consistência Lógica

Consistência de Domínio

A avaliação de Consistência de Domínio irá realizar buscas nos dados espaciais, contabilizando atributos definidos pelo usuário cujos valores não obedeçam a uma lista de domínio ou intervalo de valores também definidos pelo usuário.

A configuração do relatório de qualidade segue o mesmo padrão das avaliações anteriores, conforme podemos ver na Figura 67 a seguir.



The image shows a software dialog box titled "Quality Report". It contains several sections for configuration:

- Quality Scope:** "Scope Level:" is set to "Feature".
- Quality Information:** "Quality Element:" is set to "Logical Consistency" and "Quality Subelement:" is set to "Domain Consistency".
- Quality Measure:** "Measure Description:" contains the text: "Verificação se os valores contidos em determinados campos do banco de dados estão contidos em um determinado domínio ou lista de valores predeterminados."
- Evaluation Method:** "Evaluation Type:" is set to "directInternal". "Evaluation Description:" contains the text: "contagem dos elementos no banco de dados que não estão em conformidade com o intervalo ou lista de valores definido pelo usuário."
- Information Source:** "Information Source:" is set to "Provider". Below this are two buttons: "Quality Result" and "Expected Quality".
- Measure Identification:** "Name of Measure:" is set to "Teste de consistência de domínio para o layer de ruas".

At the bottom of the dialog are "Cancel" and "Ok" buttons.

Figura 67 - Configuração para avaliação de consistência de domínio.

As configurações específicas para esta varredura são o Elemento (**Logical Consistency**) e o Sub Elemento de qualidade (**Domain Consistency**).

Além destas informações referentes ao tipo de avaliação que se deseja realizar, também é necessário estabelecer o critério de aceitação da avaliação. A Figura 68 apresenta a definição criada para o exemplo em questão: O aplicativo irá **aceitar** o dado espacial, se o **percentual** de ocorrências for **menor ou igual a 15%** do total de elementos avaliados no banco de dados.

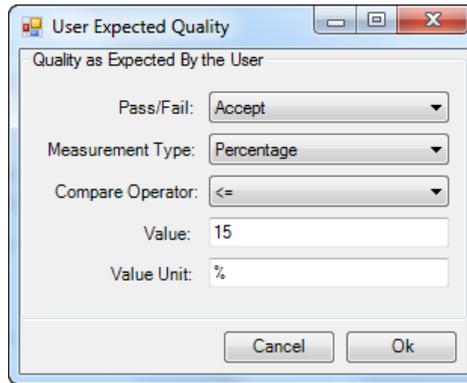


Figura 68 - Critério de aceitação/rejeição da avaliação de consistência de domínio

Além desta configuração, também apresentamos a qualidade do dado conforme avaliada anteriormente pelo próprio provedor.

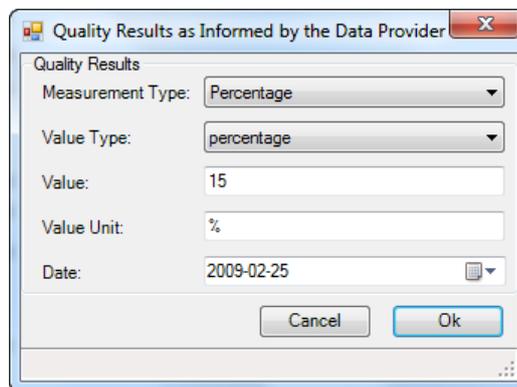


Figura 69 - Qualidade informada pelo provedor a respeito da consistência de domínio do banco de dados em avaliação.

O provedor do dado informou que até 15% dos elementos existentes no banco de dados apresentavam problemas de consistência de domínio, e que a data da última avaliação submetida ao banco é de 25-Fevereiro-2009.

Uma vez configurado o relatório de qualidade, é necessário configurar o critério de busca por inconsistências. O usuário deverá informar ao sistema de maneira estruturada, quais campos deverão ser varridos em busca de valores que estejam fora de um intervalo ou lista pré-determinada.

O aplicativo desenvolvido auxiliará o usuário a definir a feição que será submetida à avaliação, e qual atributo desta feição será varrido.

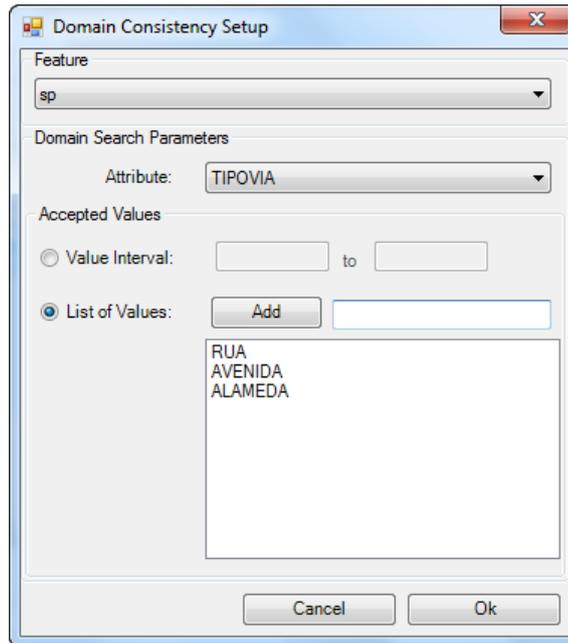


Figura 70 - Configuração do critério de busca por inconsistências de domínio

A Figura 70 apresenta a configuração que foi construída para o teste em execução, onde a feição de eixos de logradouro (**sp**) terá o atributo TIPO DE VIA (**TIPOVIA**) varrido em busca de valores diferentes daqueles definidos na lista definida pelo usuário: **RUA, AVENIDA e ALAMEDA**.

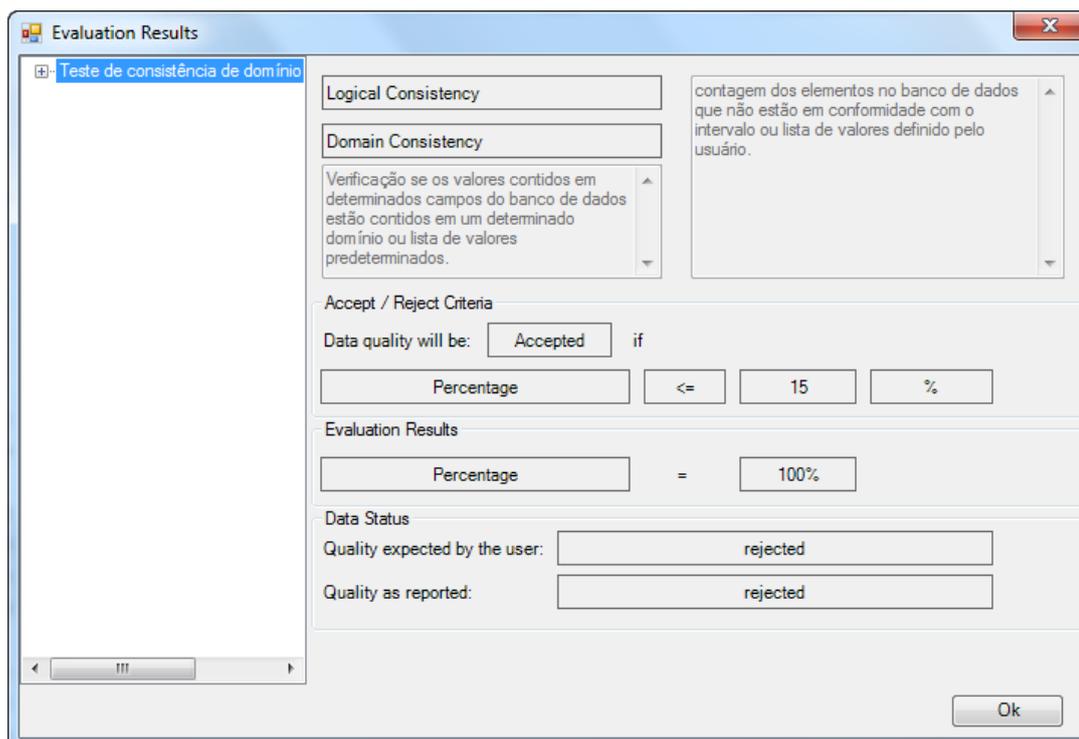


Figura 71 - Relatório de resultados com avaliação de conformidade.

O relatório de conformidade apresenta ao usuário que o banco de dados em análise foi rejeitado tanto em relação aos parâmetros de qualidade definidos pelo usuário, quanto em comparação com a qualidade informada pelo provedor.

Na área **Data Status** o dado apresenta status **rejected** tanto para a qualidade conforme esperada pelo usuário (**Quality expected by the user**) quanto pela qualidade conforme relatada pelo provedor (**Quality as reported**).

Consistência de Formato

Para realização dos testes deste capítulo, foi utilizada a base de dados digital da cidade de Itapeva-SP. Não há informação de Linhagem deste dado disponibilizado pela prefeitura. Esta base digital está disponibilizada em formato DWG.

Para a avaliação da consistência do formato do dado, o usuário deverá preencher o relatório de qualidade da mesma forma que nos testes anteriores.

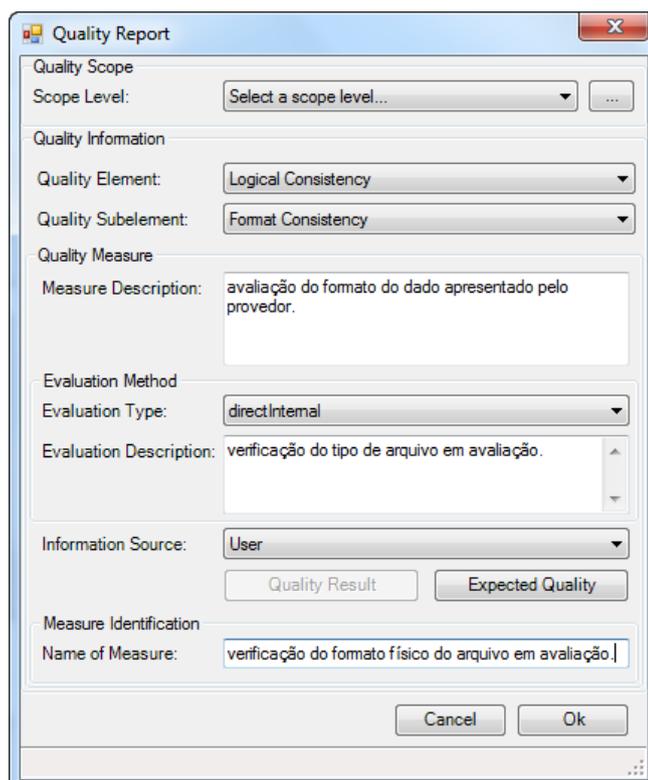


Figura 72 - Relatório de qualidade para avaliação de consistência de formato.

Conforme a Figura 72, o relatório será preenchido em conformidade com a avaliação que o usuário deseja aplicar ao dado.

A qualidade esperada, ou seu critério de aceitação deverá ser informado de acordo com o formato físico do dado que ele deseja receber de seu provedor ou de sua IDE. Os formatos suportados pelo aplicativo desenvolvido nesta dissertação são: **DGN**, **DWG** e **ESRI Shapefile**.

Neste teste em questão, o usuário irá aceitar o dado espacial se o seu formato de armazenamento for o DGN (Figura 73).

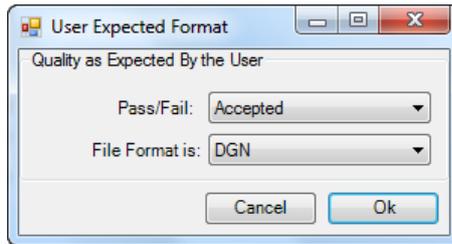


Figura 73 - Critério de aceitação do dado espacial caso seu formato seja o DGN.

O relatório de resultados pode ser acessado imediatamente após a conclusão desta etapa, apresentando o dado espacial como aceito pelo aplicativo por estar em conformidade com o formato especificado pelo usuário (Figura 74).

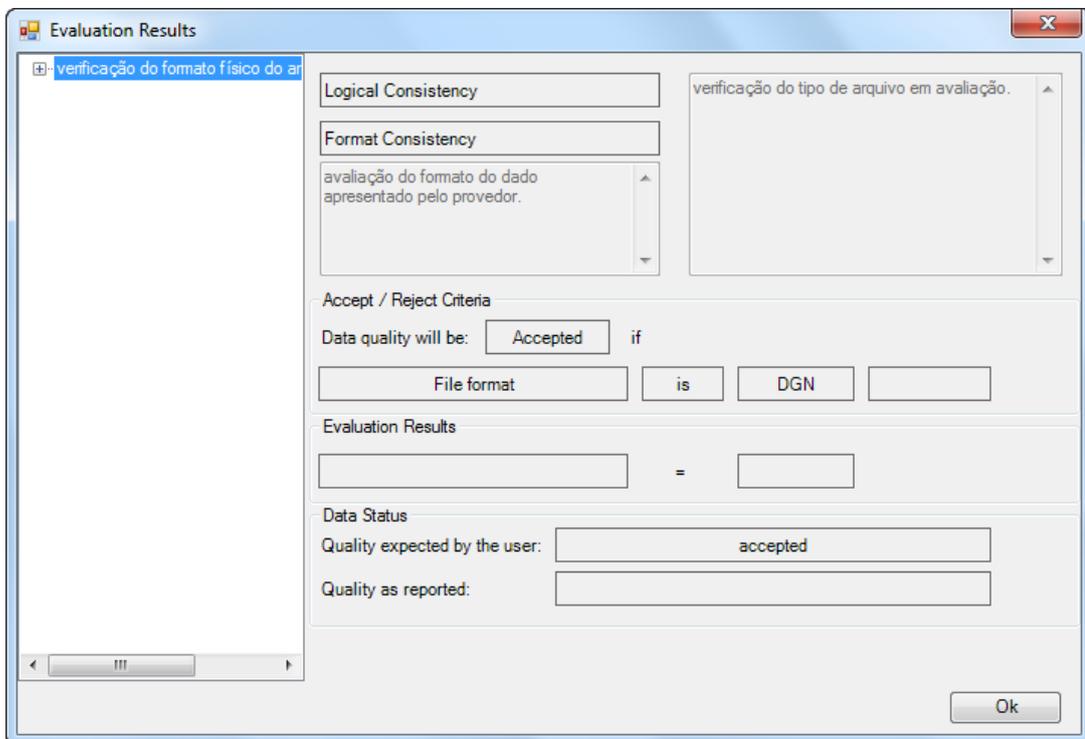


Figura 74 - Qualidade aceita pelo aplicativo para avaliação de consistência de formato.

Consistência Conceitual

Este teste foi realizado utilizando a mesma base digital do município de Itapeva-SP.

A avaliação de consistência conceitual irá avaliar o dado em busca de elementos que não estejam em conformidade com o modelo conceitual da feição em análise. Por exemplo, a feição eixos de logradouro foi definida como uma feição **linear** logo não poderá existir nesta mesma feição elementos pontuais ou poligonais.

Para realização deste teste, foi preenchido o relatório de qualidade conforme a Figura 75 a seguir.

The image shows a 'Quality Report' dialog box with the following configuration:

- Quality Scope:** Scope Level: Feature
- Quality Information:** Quality Element: Logical Consistency; Quality Subelement: Conceptual Consistency
- Quality Measure:** Measure Description: verificação de elementos em não conformidade com o modelo conceitual da feição em avaliação.
- Evaluation Method:** Evaluation Type: directInternal; Evaluation Description: contagem de elementos em não conformidade com o modelo conceitual da feição.
- Information Source:** User
- Buttons:** Quality Result, Expected Quality
- Measure Identification:** Name of Measure: avaliação de consistência conceitual do layer de vias p...
- Bottom Buttons:** Cancel, Ok

Figura 75 - Configuração do relatório de qualidade para avaliação de consistência conceitual.

O critério de aceitação neste teste foi definido de forma que o usuário aceita a qualidade do dado espacial caso exista menos de 50 elementos em não conformidade com o modelo conceitual da feição em análise (Figura 76).

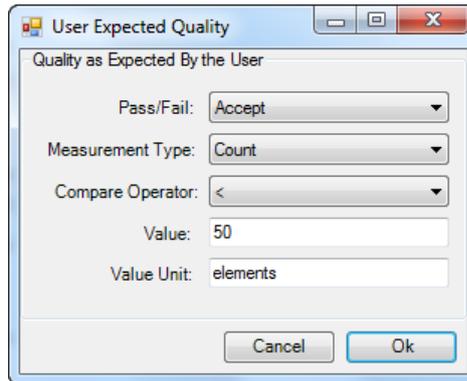


Figura 76 - critério de aceitação para avaliação de inconsistência conceitual.

O critério de busca por inconsistências deve ser preenchido de forma que, irá contar todos os elementos existentes no layer de **RUAS** e que não sejam elementos **lineares**, conforme configurado na tela apresentada na Figura 77.

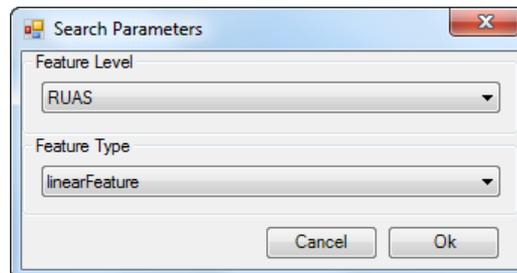


Figura 77 - Configuração de busca por elementos não lineares na feição de RUAS.

No teste aplicado, o programa aceitou a qualidade espacial da feição RUAS, pois apenas um elemento (Figura 78) não está em conformidade com o modelo conceitual de vias públicas, conforme descrito pelo usuário.

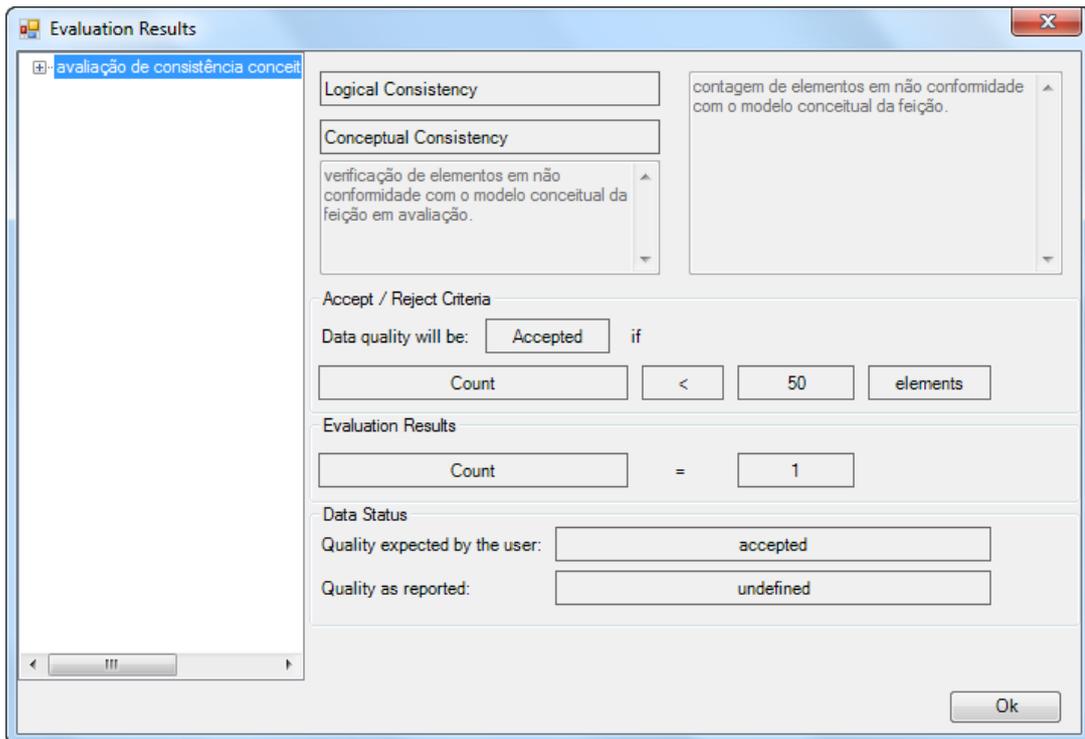
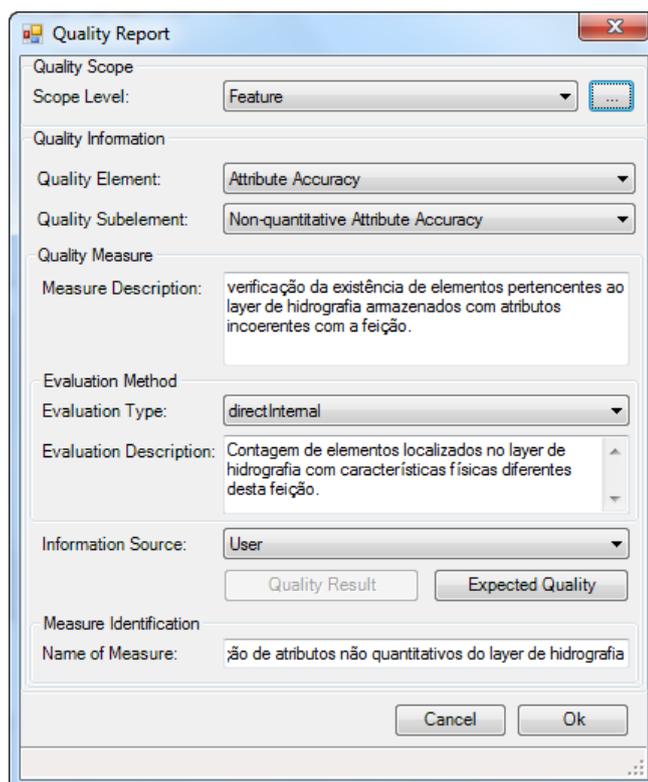


Figura 78 - Resultado da avaliação da feição de vias públicas em análise de consistência conceitual.

Avaliação de Acurácia de Atributos

Acurácia de Atributos Não-Quantitativos

De forma análoga aos testes apresentados nos capítulos anteriores, o preenchimento do relatório de qualidade segue o mesmo procedimento. A Figura 79 a seguir, apresenta a tela de configuração do relatório de qualidade para esta avaliação.



The image shows a software dialog box titled "Quality Report". It is organized into several sections:

- Quality Scope:** "Scope Level" is set to "Feature".
- Quality Information:** "Quality Element" is "Attribute Accuracy" and "Quality Subelement" is "Non-quantitative Attribute Accuracy".
- Quality Measure:** "Measure Description" contains the text: "verificação da existência de elementos pertencentes ao layer de hidrografia armazenados com atributos incoerentes com a feição."
- Evaluation Method:** "Evaluation Type" is set to "direct:Internal". "Evaluation Description" contains: "Contagem de elementos localizados no layer de hidrografia com características físicas diferentes desta feição."
- Information Source:** Set to "User". Below this are two buttons: "Quality Result" and "Expected Quality".
- Measure Identification:** "Name of Measure" is "ção de atributos não quantitativos do layer de hidrografia".

At the bottom of the dialog are "Cancel" and "Ok" buttons.

Figura 79 - Configuração do relatório de qualidade para avaliação da acurácia de atributos não quantitativos

A configuração do critério de aceitação/rejeição é feita da mesma maneira que nos procedimentos anteriores.

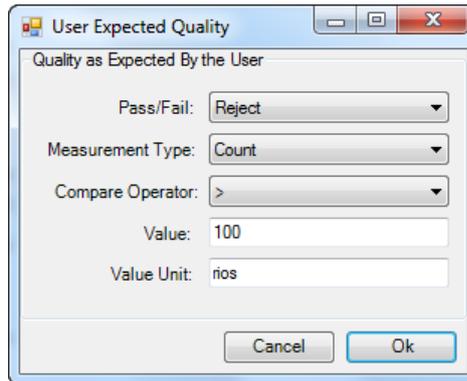


Figura 80 - Configuração do critério de aceitação da avaliação de atributos não quantitativos.

Neste teste conforme a Figura 80, o dado será rejeitado se existir mais de 100 elementos categorizados como pertencentes ao layer de RIOS, mas armazenados com características¹⁶ diferentes.

O critério de busca por inconsistências deverá ser também configurado pelo usuário.

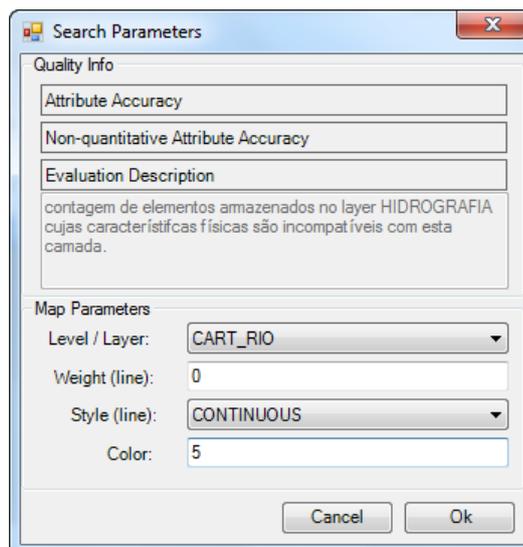


Figura 81 - Critério de buscas por inconsistências de atributos não quantitativos.

Na Figura 81, o parâmetro de busca por inconsistências instruiu o aplicativo a buscar por elementos que estejam armazenados na feição **CART_RIO**, onde a espessura

¹⁶ As características utilizadas neste teste são as características físicas da geometria armazenada no arquivo de desenho. Logo são seu nível (layer), cor, estilo de linha e espessura de linha.

da linha seja diferente de **0**, o estilo da linha seja diferente de **CONTINUOUS** e a cor do elemento gráfico seja diferente de 5.

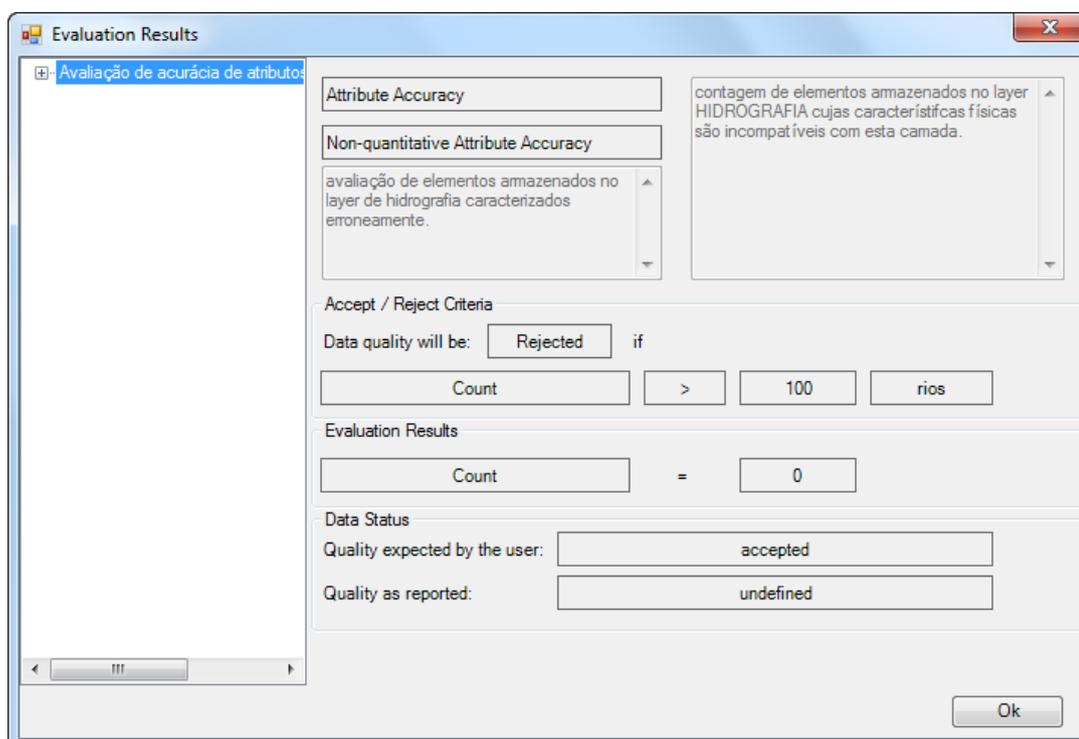


Figura 82 - Resultado da avaliação de acurácia de atributos não quantitativos na feição hidrografia.

Neste exemplo, conforme demonstrado pela Figura 82, o aplicativo aprovou o dado avaliado pois não existem elementos na feição de hidrografias cujas características não obedecem a definição daquela feição.

Linhagem

A Linhagem, conforme apresentado no capítulo 2, informa ao usuário do dado espacial toda a "história" passada do dado espacial em avaliação, todas as transformações que a informação sofreu até estar disponível aquele usuário.

O aplicativo também permite o armazenamento desta informação em um formato de arquivo e padrões de armazenamento coerentes com a ISO 19115.

A tela apresentada na Figura 83, apresenta a informação do dado espacial, refletindo seu sistema de projeção e datum, elipsoide e escala de mapeamento. O

aplicativo permite também a entrada de dado histórico documentando todas as transformações sofridas pelo dado até a sua disponibilização para consumo.

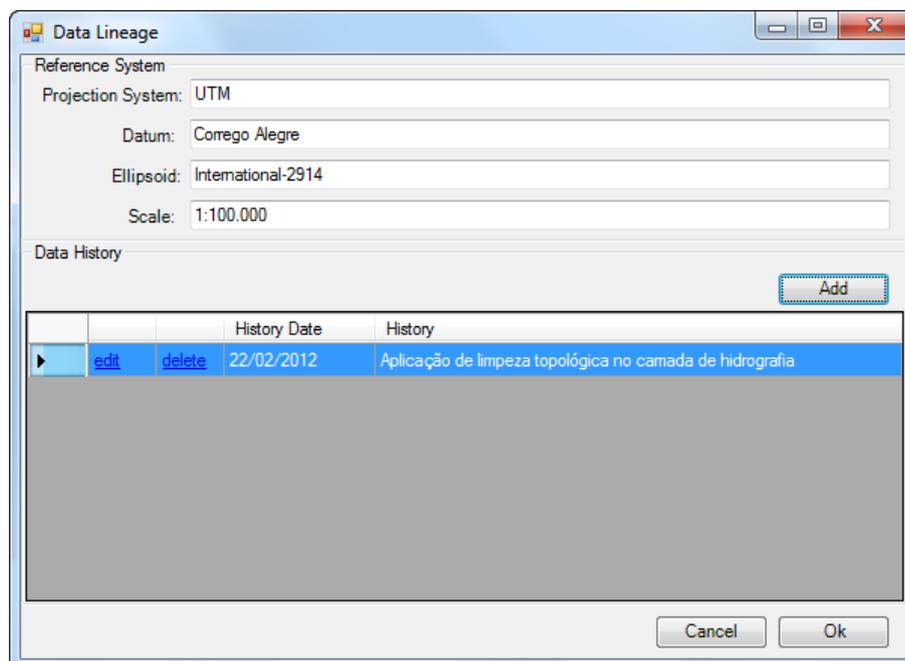


Figura 83 - Tela de informação da linhagem do dado espacial

Metadados de Qualidade

Uma avaliação da qualidade de dados espaciais que produza resultados que possam ser comparáveis com os resultados de outras avaliações deve ser realizada sob as mesmas normas e padrões de qualidade (DÉVILLERS, *et al.*, 2007).

Assim sendo, o aplicativo de avaliação de qualidade desenvolvido para esta dissertação, permite a criação e o reaproveitamento de metadados de qualidade.

Foi desenvolvida uma interface (Figura 84) de acesso aos metadados de qualidade, criados no padrão ISO 19115, onde a informação de qualidade pode ser salva e recuperada pelo aplicativo.

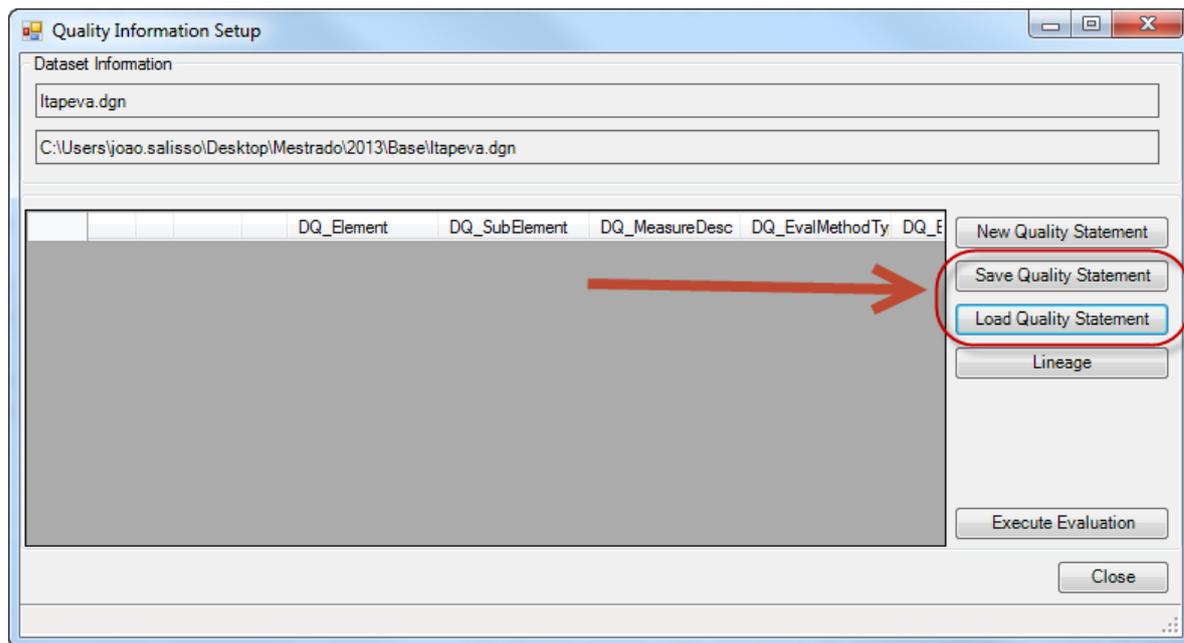


Figura 84 - Gravação e recuperação de metadados de qualidade no padrão ISO 19115

Através desta interface de acesso, o usuário poderá gravar e recuperar informações de qualidade de dados e aproveitá-los em suas avaliações.

Ao selecionar o botão **Load Quality Statement**, o usuário poderá carregar no aplicativo os metadados de qualidade que foram criados em uma utilização anterior do aplicativo.

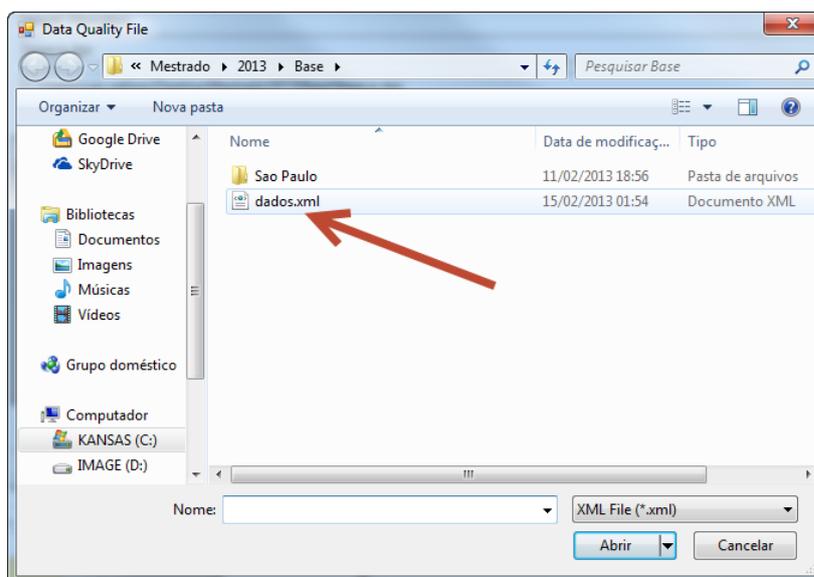


Figura 85 - Seleção do arquivo XML de metadados de qualidade.

O arquivo de metadados de qualidade, utilizado no exemplo apresentado na Figura 85, possui o formato descrito conforme apresentado no Anexo 1.