



UNIVERSIDADE ESTADUAL PAULISTA
Faculdade de Ciências e Tecnologia
Câmpus de Presidente Prudente

Programa de Pós-Graduação em Ciências Cartográficas

Glaucia Gabriel Sass

An abstract graphic featuring a central globe with a purple and blue color scheme. The globe is surrounded by several overlapping, hand-drawn style lines in a light gray color, creating a complex, swirling pattern that frames the central globe and the title text.

**Um Método de Análise de Dados
Temporais para o Cadastro Territorial
Multifinalitário Urbano**

Presidente Prudente

2013



UNIVERSIDADE ESTADUAL PAULISTA
Faculdade de Ciências e Tecnologia
Câmpus de Presidente Prudente

Programa de Pós-Graduação em Ciências Cartográficas

Glaucia Gabriel Sass

A background graphic featuring a central globe with a purple and blue color scheme. Overlaid on the globe are several thick, grey, hand-drawn style lines that form a complex, abstract pattern of overlapping loops and curves, resembling a stylized atomic model or a network diagram.

Um Método de Análise de Dados Temporais para o Cadastro Territorial Multifinalitário Urbano

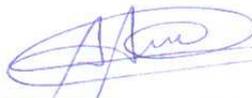
Tese de doutorado apresentada ao Programa de Pós-Graduação em Ciências Cartográficas da Universidade Estadual Paulista – Campus de Presidente Prudente, como requisito parcial para a obtenção do título de Doutora em Ciências Cartográficas. Área de Concentração: Aquisição, Análise e Representação da Informação Espacial.

Orientador: Prof. Dr. Amilton Amorim

Presidente Prudente

2013

BANCA EXAMINADORA



Prof. Dr. AMILTON AMORIM
ORIENTADOR



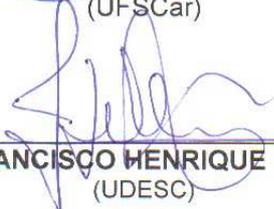
Prof. Dr. MILTON HIROKAZU SHIMABUKURO
(FCT/UNESP)



Prof. Dr. EDMUR AZEVEDO PUGLIESI
(FCT/UNESP)



Prof. Dr. RICARDO RODRIGUES CIFERRI
(UFSCar)



Prof. Dr. FRANCISCO HENRIQUE DE OLIVEIRA
(UDESC)



GLAUCIA GABRIEL SASS

Presidente Prudente (SP), 6 de dezembro de 2013.

Resultado: ARROVADO

DEDICATÓRIA

Ao meu esposo e companheiro

Sergio Ricardo

Aos meus filhos

Thales Augusto

João Ricardo

Á minha família

AGRADECIMENTOS

A Deus, em primeiro lugar.

Ao Sergio Ricardo, Thales Augusto e João Ricardo, esposo e filhos, pelo amor, carinho e paciência durante esse período.

À minha família, em especial minha mãe Hulda, meu avô Augusto (*in memoriam*) e minha avó Querubina, pelo apoio e incentivo ao estudo durante toda a minha vida.

Ao meu orientador Amilton Amorim, pela atenção, dedicação e orientação.

Aos amigos da Pós-Graduação e do GARDE, pelos momentos de discussões e confraternizações.

À banca de qualificação, professores Dr. Ricardo Rodrigues Ciferri (UFSCar), Dr. Nilton Nobuhiro Imai (FCT/UNESP) e Dr. Ronaldo Celso Messias Correa (FCT/UNESP), pelas correções e contribuições.

Ao programa de Pós-graduação em Ciências Cartográficas, pela estrutura e apoio fornecidos.

Aos funcionários da secretaria de Pós-graduação da UNESP, pelos atendimentos e orientações sobre as normas e resoluções.

À UEMS, por ter possibilitado a realização do curso de doutorado, concedendo afastamento integral.

À FUNDECT - Fundação de Apoio ao Desenvolvimento do Ensino, Ciência e Tecnologia do Estado de Mato Grosso do Sul, pela concessão da bolsa de doutorado.

A todos aqueles que direta ou indiretamente contribuíram para a realização deste trabalho.

A todos meu Muito Obrigada!!

EPÍGRAFE

⁹*Portanto, eu vos digo: pedi e recebereis; procurai e encontrareis; batei e vos será aberto.* ¹⁰*Pois quem pede recebe; quem procura encontra; e, para quem bate, se abrirá.*

Lucas 11, 9-10.

”Deus nos fez perfeitos e não escolhe os capacitados, capacita os escolhidos. Fazer ou não fazer algo só depende de nossa vontade e perseverança.”

Albert Einstein.

RESUMO

O Cadastro Territorial Multifinalitário é um instrumento usado para a gestão pública dos municípios e vem sendo cada vez mais utilizado. Com a evolução tecnológica alguns paradigmas vêm sendo alterados, como é o caso dos tradicionais formulários e mapas em papel que evoluíram para ambientes informatizados, otimizando e melhorando em vários aspectos o processo de implantação e manutenção do Cadastro Territorial Multifinalitário. Normalmente, os Sistemas de Informação Geográfica são utilizados nos Cadastros informatizados, gerando dados 2D, em banco de dados convencionais e geográficos, nos quais os dados são sobrescritos no momento da atualização. Todavia, os tradicionais Sistemas de Informação Geográfica, geralmente possuem poucos recursos para a geração de relatórios, análises e gráficos para dar suporte à tomada de decisão que possibilitem a definição de políticas públicas, comparado aos ambientes de *Business Intelligence*. O objetivo deste trabalho é contribuir com o desenvolvimento de procedimentos adequados para incorporar o aspecto temporal, principalmente, às análises pertinentes à gestão territorial a partir de dados do Cadastro Territorial Multifinalitário Urbano (Cadastro 2D acrescido do tempo). Segundo os modelos cadastrais e de Sistemas de Informação Geográfica os atributos espaciais e temporais fazem parte de seus conceitos. Entretanto, ao longo da história o atributo temporal não foi implantado. Isso ocorreu, em grande parte, pela falta de tecnologias adequadas para tratar a característica temporal. Com a evolução tecnológica ocorrida nas últimas décadas esse cenário vem se transformando. Com isso, conceitos discutidos há décadas estão tornando-se realidade. O Cadastro em sua essência precisa dizer onde e quando ocorreram as mudanças, para cumprir seu papel nas políticas de planejamento e gestão urbana. Nesse contexto, esta tese apresenta um método de análise de dados temporais para o Cadastro Territorial Multifinalitário. Esse método envolve o uso dos conceitos de *Data Warehousing* para a elaboração de um ambiente de apoio à tomada de decisão. No desenvolvimento deste trabalho foram utilizadas ferramentas que apoiam o processo de *Business Intelligence* e que permitem gerar informações espaço-temporais, a suíte Pentaho, o PostgreSQL/PostGIS e o gvSIG. Os resultados da modelagem e aplicação do método demonstraram ser eficazes para a geração de informações espaço-temporais. Finalmente, o método foi contextualizado por meio da elaboração do *Data Warehousing* com dados reais, tomando como base os dados cadastrais do município de Ribeirão dos Índios – SP. Essa aplicação foi desenvolvida com o objetivo de investigar formas de fornecer informações espaço-temporais aos gestores, capazes de responder questionamentos importantes às análises.

Palavras-chave: Cadastro Territorial Multifinalitário. Análise Temporal. *Data Warehouse* Espaço-Temporal.

ABSTRACT

The Multipurpose Cadastre is a tool used for the public management of the municipalities and it is becoming more widely used. With the technological evolution some paradigms are being modified, such as the traditional paper maps and forms which have evolved to computerized environments, optimizing and improving in many aspects the implementation and maintenance of the Multipurpose Cadastre. Normally, the Geographic Information Systems are used in the computerized Cadastres, generating 2D data, on conventional and geographic databases, in which the data are overwritten at the moment of the update. However, the traditional Geographic Information Systems, generally have few resources for the generation of reports, analysis and graphics to provide support for the decision making which enable the definition of public policies, compared to the environments of Business Intelligence. The aim of this essay is to contribute to the development of appropriate procedures to incorporate the temporal aspect, mainly, to the relevant analyzes to the territorial management, using data from the Urban Multipurpose Cadastre (2D Cadastre plus time) . According to the cadastral models and the Geographic Information Systems, the spatial and the temporal attributes are part of their concepts. Nevertheless, throughout the history, the temporal attribute was not implemented. This occurred, mostly, due to the lack of adequate technologies to treat the temporal characteristic. With the technological evolution over the past decades this scenario has been changing. Thus, concepts discussed decades ago are becoming reality. The Cadastre, in its essence, needs to tell where and when the changes occurred, to fulfill its role on the planning policies and the urban management. In this context, this thesis presents a method of analysis of the temporal data for the Multipurpose Cadastre. This method involves the use of the concepts of Data Warehousing for the elaboration of an environment to support the decision making. During the development of this study, there were used tools that support the process of Business Intelligence and that allow the generation of spatio-temporal information, the Pentaho Suite, the PostgreSQL/PostGIS and the gvSIG. The results of the modeling and implementation of the method have shown to be effective for the generation of spatio-temporal information. Finally, the method was contextualized through the elaboration of the Data Warehousing with real data, having as a basis the cadastral data of the municipality of Ribeirão do Índios – SP. This implementation was developed with the objective of investigating ways of providing spatio-temporal information to the managers, which are capable of answering important questionings to the analysis.

Keywords: Multipurpose Cadastre. Temporal Analysis. Spatio-Temporal Data Warehouse.

LISTA DE QUADROS

Quadro 1 : Questões de integração de dados espaciais.	42
Quadro 2 : Diferenças entre os Bancos de Dados Transacionais e os <i>DWs</i>	57
Quadro 3 : <i>MOLAP</i> versus <i>ROLAP</i>	62
Quadro 4 : Atributos para a dimensão.	65
Quadro 5 : Comparativo dos trabalhos relacionados.	75
Quadro 6 : Dicionário de Dados parcial.	84
Quadro 7 : “Processos de negócios” do <i>Data Warehousing</i>	86

LISTA DE FIGURAS

Figura 1 : Exemplo de <i>kudurrus</i> kassitas	18
Figura 2 : Evolução do conceito do Cadastro.....	21
Figura 3 : Informações do Cadastro.	23
Figura 4 : Etapas de execução do Cadastro.	23
Figura 5 : Pilares do Cadastro.....	25
Figura 6 : Pilares do Cadastro sob influência.	25
Figura 7 : Representação de dados espaciais e temporais.	28
Figura 8 : Representação da parcela.	30
Figura 9 : Edifício de apartamentos.....	30
Figura 10 : Construções complexas.....	31
Figura 11 : Representação 4D da parcela.	31
Figura 12 : Exemplo de integração e transformação dos dados.	35
Figura 13 : Hierarquia e interoperabilidade de dados espaciais.	37
Figura 14 : Estruturas organizacionais para gestão territorial.	46
Figura 15 : A hierarquia da Infraestrutura de Dados Espaciais.	46
Figura 16 : Componentes de uma Infraestrutura de Dados Espaciais.	47
Figura 17 : Relação temporal em 3-dimensões.	49
Figura 18 : Tipos de ordem temporal.	50
Figura 19 : Representações de rastrr em duas dimensões.	52
Figura 20 : Representações vetoriais em duas dimensões.....	52
Figura 21 : Representação do espaço absoluto e relativo.....	53
Figura 22 : Categorias de dimensões da informação geográfica.	55
Figura 23 : Tipos de consultas espaço-temporais.	56
Figura 24 : Fases de criação de um <i>DW</i>	58
Figura 25 : Arquitetura <i>Warehouse</i> para sistemas <i>MOLAP</i>	61
Figura 26 : Arquitetura <i>Warehouse</i> para sistemas <i>ROLAP</i>	62
Figura 27 : Cubo de três dimensões.....	63
Figura 28 : Exemplo de cubo esparso.....	64
Figura 29 : Esquema estrela.	65
Figura 30 : Hierarquia do atributo localização.	65
Figura 31 : Esquema floco de neve.	66
Figura 32 : Notação do modelo MultiDim.	69
Figura 33 : Metamodelo da extensão espacial do modelo MultiDim.....	70
Figura 34 : Metamodelo da temporalidade estendido do modelo MultiDim.	72
Figura 35 : Hierarquia básica do MADS para tipos de dados abstratos espaciais.....	74
Figura 36 : Mapa de localização do Município de Ribeirão dos Índios - SP.	78
Figura 37 : Etapas do Método de análise temporal para CTM.....	79
Figura 38 : Arquitetura para o Método Temporal do CTM.....	82
Figura 39 : Modelagem parcial do CTM 2012 – Parcela Cadastral.	83
Figura 40 : Diagrama de Entidade-Relacionamento de acordo com o MGD.....	85
Figura 41 : Processo de negócio, assunto Parcela.	86
Figura 42 : Transformação do processo de <i>ETL</i> para tabelas Dimensões.....	87
Figura 43 : Transformação do processo de <i>ETL</i> para tabela Fato.	87
Figura 44 : <i>Schema Workbench</i> apresentando os cubos.....	87
Figura 45 : <i>Schema Workbench</i> com componentes do cubo Parcela.	88
Figura 46 : Interface do <i>jPivot</i> com o cubo Parcela.	88
Figura 47 : <i>Metadata Editor</i> com o metamodelo da Fato Parcela.....	89

Figura 48 : Interface <i>WAQR</i> com o metamodelo Fato Parcela.....	89
Figura 49 : Relatório de Residentes desenvolvido no <i>Report Desing</i>	90
Figura 50 : Relatório de Programas Sociais desenvolvido no <i>Report Desing</i>	90
Figura 51 : Resultado da consulta, parcela <i>ssqllff</i> = '1010101', no <i>gvSIG</i>	91
Figura 52 : Resultado da consulta, vizinhos da parcela '01050401', no <i>gvSIG</i>	92
Figura 53 : Visualização da consulta <i>SQL</i> sobre aposentados e pensionistas, no <i>PostgreSQL</i> . 92	
Figura 54 : Visualização da consulta <i>SQL</i> sobre pensionistas, no <i>PostgreSQL</i>	93
Figura 55 : Mapa com a consulta <i>SQL</i> sobre aposentados e pensionistas, no <i>gvSIG</i>	94
Figura 56 : Mapa com a consulta <i>SQL</i> sobre aposentados, no <i>gvSIG</i>	95
Figura 57 : Mapa com a consulta <i>SQL</i> sobre aposentados em 2004 e 2010, no <i>gvSIG</i>	97
Figura 58 : Mapa com a consulta <i>SQL</i> sobre desmembramento entre 2010 e 2012, no <i>gvSIG</i>	98
Figura 59 : Mapa com a consulta <i>SQL</i> sobre remembramento entre 2010 e 2012, no <i>gvSIG</i> . 99	
Figura 60 : Relatório: Alteração da Área Construída nas Parcelas, 2004 e 2010.....	100
Figura 61 : Relatório: Alteração da Área Construída nas Parcelas, 2010 e 2012.....	100
Figura 62 : Relatório: Ocorrência de Cardiopatias, Hipertensão Arterial e Diabetes. Anos 1996, 2004, 2010 e 2012.	101
Figura 63 : Mapa sobre a alteração de área construída entre 2004 e 2010.....	102
Figura 64 : Mapa sobre a alteração de área construída entre 2010 e 2012.....	103
Figura 65 : Recorte do mapa de Ocorrência de Diabetes.	104
Figura 66 : Mapa da ocorrência de Cardiopatia em 1996, 2004, 2010 e 2012.....	105
Figura 67 : Mapa sobre a ocorrência de Hipertensão em 1996, 2004, 2010 e 2012.	106
Figura 68 : Mapa da ocorrência de Diabetes em 1996, 2004, 2010 e 2012.	107
Figura 69 : Visualização dos dados tabulados no <i>Saiku Analytics</i>	108
Figura 70 : Visualização dos dados em gráfico no <i>Saiku Analytics</i>	108
Figura 71 : Relatório: Alunos matriculados em creche e no ensino fundamental, nos anos de 2004, 2010 e 2012.	109
Figura 72 : Representação gráfica das parcelas com alunos matriculados no ensino fundamental e creche em 2004.	110
Figura 73 : Representação gráfica das parcelas com alunos matriculados no ensino fundamental e creche em 2010.	111
Figura 74 : Mapa das parcelas com alunos matriculados no ensino fundamental e creche em 2012.	112
Figura 75 : Mapa com as parcelas com alunos matriculados no ensino fundamental e creche em 2004, 2010 e 2012.	113
Figura 76 : Mapa com interseção das parcelas: 2004 - 2010 e 2010 - 2012.	114
Figura 77 : Visualização dos dados tabulados no <i>Saiku Analytics</i>	115
Figura 78 : Visualização dos dados em gráfico no <i>Saiku Analytics</i>	116

LISTA DE SIGLAS

2D	-	Duas Dimensões
3D	-	Três Dimensões
4D	-	Quatro Dimensões
a. C.	-	antes de Cristo
ABNT	-	Associação Brasileira de Normas Técnicas
ADT	-	<i>Abstract Data Types</i>
BDET	-	Banco de Dados Espaço-temporais
BDG	-	Banco de Dados Geográficos
BDT	-	Banco de Dados Temporais
BI	-	<i>Business Intelligence</i>
BIC	-	Boletim de Informações Cadastrais
CAD	-	<i>Computer-Aided Design</i>
CASE	-	<i>Computer Aided Software Engineering</i>
CTM	-	Cadastro Territorial Multifinalitário
DER	-	Diagrama de Entidade Relacionamento
DW	-	<i>Data Warehouse</i>
ETL	-	<i>Extraction, Transformation and Load</i>
FIG	-	<i>Fédération Internationale de Géomètres</i> (acrônimo francês de Federação Internacional de Agrimensores)
GARDE	-	Grupo de pesquisa em Aquisição e Representação de Dados Espaciais
GML	-	<i>Geographical Markup Language</i>
GSDI	-	<i>Global Spatial Data Infrastructure</i>
HOLAP	-	<i>Hybrid On-Line Analytical Processing</i>
IBGE	-	Instituto Brasileiro de Geografia e Estatística
INDE	-	Infraestrutura Nacional de Dados Espaciais
MADS	-	<i>Modeling of Application Data with Spatiotemporal features</i>
MGD	-	Modelo Global de Dados
MOLAP	-	<i>Multidimensional On-Line Analytical Processing</i>
OGC	-	<i>Open GIS Consortium</i>
OLAP	-	<i>On-Line Analytical Processing</i>
OQL	-	<i>Object Query Language</i>
PGV	-	Planta Genérica de Valores
ROLAP	-	<i>Relational On-Line Analytical Processing</i>
SADT	-	<i>Spatial ADTs</i>
SDI	-	<i>Spatial Data Infrastructures</i>
SDW	-	<i>Spatial Data Warehouse</i>
SGBD	-	Sistema de Gerenciamento de Banco de Dados
SI	-	Sistema de Informação
SICART	-	Sistema de Cadastro e Registro Territorial
SIG	-	Sistema de Informação Geográfica
SIT	-	Sistema de Informação Territorial
SOLAP	-	<i>Spatial On-Line Analytical Processing</i>
SQL	-	<i>Structure Query Language</i>
STDW	-	<i>Spatio-Temporal Data Warehouse</i>
TADT	-	<i>Temporal ADT</i>
TDW	-	<i>Temporal Data Warehouse</i>
UML	-	<i>Unified Modeling Language</i>
XML	-	<i>Extensible Markup Language</i>

SUMÁRIO

1	INTRODUÇÃO	13
1.1	Objetivo	14
1.2	Estrutura da Tese	17
2	REVISÃO DA LITERATURA	18
2.1	Cadastro Territorial Multifinalitário	18
2.1.1	Sistemas de informação.....	25
2.1.2	Cadastro 2D, 3D e 4D	29
2.2	Integração de Dados	33
2.2.1	Sistemas de integração de dados	37
2.2.2	Integração de dados espaciais.....	40
2.2.3	Integração de dados no Brasil	43
2.2.3.1	<i>Modelo Global de Dados</i>	43
2.2.3.2	<i>Infraestrutura Nacional de Dados Espaciais</i>	44
2.3	Banco de Dados.....	48
2.3.1	Dados temporais	48
2.3.2	Dados espaciais	52
2.3.3	Dados espaço-temporais.....	54
2.3.4	Banco de dados espaço-temporais.....	55
2.4	Data Warehouse.....	56
2.4.1	Arquitetura de <i>Data Warehouse</i>	60
2.4.2	Modelo Multidimensional	63
2.4.3	<i>Data Warehouse Espaço-Temporal</i>	67
2.4.3.1	<i>MADS</i>	73
2.5	Trabalhos Correlatos	75
3	MATERIAIS E MÉTODOS	78
4	RESULTADOS	83
4.1	Consultas Espaciais, Temporais e Espaço-temporais	91
4.1.1	Consultas espaciais	91
4.1.2	Consultas Temporais.....	92
4.1.3	Consultas espaço-temporais	93
4.2	Análise espaço temporal: Área construída.....	100
4.3	Análise espaço temporal: Evolução das patologias cardiopatia, hipertensão arterial e diabetes.....	101
4.4	Análise espaço temporal: Ensino fundamental e creche.....	108
5	DISCUSSÃO	117
6	CONCLUSÃO E TRABALHOS FUTUROS.....	123
6.1	Conclusão	123
6.2	Trabalhos futuros	124
	REFERÊNCIAS	126
	APÊNDICE A – MODELAGEM DO BANCO DE DADOS ESPAÇO-TEMPORAIS	135
	APÊNDICE B – MODELO GLOBAL DE DADOS.....	140
	APÊNDICE C – SPATIO-TEMPORAL DATA WAREHOUSE	147

1 INTRODUÇÃO

O espaço urbano, nas últimas décadas, sofreu grandes transformações. Houve um período em que a população rural migrou em grande quantidade para as cidades, causando problemas no ambiente urbano com o crescimento desordenado. Isso ocorreu pela falta de planejamento adequado para suportar as mudanças. Para auxiliar o planejamento do espaço urbano, muitas discussões nas esferas administrativas públicas aconteceram para identificar recursos e ferramentas que apoiassem a gestão. Neste contexto, o Cadastro Territorial Multifinalitário (CTM), que visa gerenciar os dados das parcelas territoriais, tem se destacado como importante instrumento de suporte ao planejamento e à gestão.

Representar o espaço urbano é um desafio, principalmente no que se refere à garantia dos direitos de propriedade e uso dos imóveis pela população. A inovação tecnológica tem contribuído significativamente para o desenvolvimento dos chamados Cadastros Territoriais Urbanos, principalmente na aquisição, armazenamento e visualização do espaço, permitindo a criação de políticas de planejamento coerentes com a realidade.

Entanto, um dos problemas do planejamento urbano é a falta de informações que apoiem a tomada de decisão. No caso do Cadastro os dados existem nas prefeituras, em papel ou digitalizados, mas muitas vezes não são transformados em informações úteis.

As instituições públicas e privadas também se transformaram ao longo dos anos, e sua relação com a informação também mudou. Atualmente as organizações possuem muitos dados, porém pouca informação. A informação está entre os principais recursos de uma organização, a qual armazena seus dados em Sistemas de Informação Transacional¹, que apoiam as operações do dia-a-dia e a gerência operacional. Contudo, as gerências superiores das organizações necessitam de informações sumarizadas e históricas que, geralmente, não são contempladas pelos sistemas transacionais.

O Cadastro, administrado pelas prefeituras no Brasil, além dos dados atualizados sobre as parcelas, precisa ter histórico para gerenciar o território. Mesmo não existindo padrões e modelos para o Cadastro no país, muitos municípios possuem pelo menos o Cadastro Econômico, e em alguns casos informatizados. Porém, quando informatizados, estão em Sistemas de Informação Transacional, ou seja, quando o banco de dados é atualizado perde a informação anterior, com dados geográficos ou não. Nesse contexto, como gerar informações históricas do Cadastro?

¹ Sistemas de Informação Transacional ou Sistemas de Processamento de Transações: executam e registram transações solicitadas por usuários de um banco de dados (O'BRIEN, 2004).

As pesquisas sobre o CTM obtiveram muitos avanços na última década, grande parte apoiado por novos recursos tecnológicos. O Cadastro passou do conceito 2D para o 3D, e atualmente a pesquisa trabalha com conceito de Cadastro 4D (Cadastro 3D + Tempo). Além de ter passado de um registro básico do território para ser considerado um inventário público do território e todos os objetos nele contido.

Em outras áreas do conhecimento, as pesquisas têm buscado métodos de armazenamento de informações históricas, um exemplo são os Bancos de Dados Temporais (BDTs). Entretanto, a tecnologia não acompanhou os conceitos sobre os dados temporais, e os Sistemas Gerenciadores de Banco de Dados (SGBDs) não possuem os recursos necessários para a implantação dos conceitos por completo.

A realidade atual do CTM no Brasil apresenta um sistema geralmente voltado para a tributação, com pouca atualização e baseado em Sistemas de Informação Transacional sem seguir nenhum tipo de padronização tecnológica. Nesse cenário, a geração de informações que apoiem a tomada de decisão se torna complexa. Com poucos recursos a serem aplicados pelo poder público, o conceito de Inteligência de Negócio (*Business Intelligence – BI*), um conjunto de tecnologias voltadas a disponibilizar informações e conhecimento em uma organização, e o uso de *Data Warehouse (DW)*, surgem como opção de integrar dados para gerar informações históricas do CTM.

1.1 Objetivo

Objetivo Geral

Definir um método para análise de dados temporais, aplicado ao Cadastro Territorial Multifinalitário urbano, a partir de um *Data Warehouse* para acessar fontes de dados heterogêneas.

Objetivos específicos

- Analisar as estruturas dos bancos de dados que dão suporte ao CTM, para projetar um *Data Warehousing*.
- Integrar dados distintos para disponibilizar informações de apoio à decisão.
- Avaliar o *DW* como recurso para a análise de dados temporais.

Hipótese

O CTM utiliza representações estáticas dos fenômenos espaciais, porém, modelos espaço-temporais melhoram a representação da variação visual dos fenômenos no espaço e no tempo, para apoiar a gestão pública.

Justificativa

A evolução tecnológica trouxe melhoria às atividades rotineiras na sociedade. Com o crescimento do volume de informações a serem armazenadas e manipuladas, surgiram também problemas de distribuição espacial e heterogeneidade dos dados. Essas transformações afetaram diversos domínios de aplicações, entre eles os Sistemas de Informação Territorial (SIT, termo proveniente do inglês *Land Information Systems*).

O crescente volume de fontes independentes de dados geográficos trouxe problemas de disseminação e entendimento desses dados. Por isso, as aplicações precisam ser capazes de interpretar e processar dados oriundos de diversas fontes, capazes de compartilhar e trocar informações e processos entre diferentes usuários.

O desafio do intercâmbio de dados é enfrentar a diversidade de modelos conceituais dos Sistemas de Informação Geográfica (SIGs) e SITs, onde cada instituição tem sua maneira de organizar a informação espacial.

No contexto tecnológico, um recurso capaz de integrar os dados mantendo a rotina de cada instituição é o *DW*, podendo agregar diferentes bancos de dados (relacional, objeto relacional, geográfico) com diferentes estruturas, que em algumas situações foram desenvolvidos sem nenhuma técnica de modelagem apropriada. Além de ter a característica de ser temporal, permitindo o armazenamento do histórico dos dados.

Para Kimball e Caserta (2004) um *DW* é um sistema que extrai, limpa, adapta e fornece dados de uma (ou várias) origem em um repositório de dados dimensional e, em seguida, apoia e implementa consultas e análises para fins de tomada de decisão.

O armazenamento dos dados que contenham a evolução histórica permitirá, por exemplo, rastrear as informações de um terreno, ao longo do tempo, podendo ser interessante para um usuário, mostrando quando as edificações foram construídas ou modificadas em um determinado terreno. Para realizar esse rastreamento é preciso definir um método de análise temporal para acompanhar, analisar e orientar a produção do CTM.

No CTM o uso do *DW* possibilitará ao administrador público pensar sobre questões estratégicas e projetar o futuro por meio de cenários fundamentados em dados precisos do presente e do passado. Na maioria dos projetos é considerada uma solução onerosa, mas propicia agilidade e segurança a tomada de decisões. Essa desvantagem pode ser suprida, em parte, com ferramentas livres, porém a capacitação de recursos humanos é fundamental.

Neste contexto tecnológico aplicado ao CTM, procura-se responder questões do dia-a-dia das administrações públicas como por exemplo:

- Qual a variação quantitativa dos casos de dengue na população da cidade, segmentada por setores e por anos?
- Após 10 anos, quais parcelas poderão ter residentes com mais de 60 anos?
- Após 5 anos, quais parcelas poderão ter residentes em idade escolar, ingressantes na rede pública?
- Quais parcelas foram desmembradas no ano anterior e tiveram mudança de proprietário?
- Quais parcelas foram unificadas no ano anterior e tiveram mudança de proprietário?
- Existem tendências aparentes?
- Qual é o padrão construtivo predominante em determinada região da cidade? E, quais os equipamentos comunitários existem na região?
- Em determinado período de tempo, quais parcelas tiveram mudanças de proprietário? E, quais equipamentos comunitários estavam disponíveis na região dessas parcelas?

O CTM temporal, respondendo essas questões, pode ser utilizado como um inventário para armazenar uma descrição completa da parcela, representar as mudanças armazenadas em computador, analisar, explicar, explorar ou prever elementos contidos nas atividades de uma região. Também pode ser utilizado para apoiar a atualização dos dados com controle de qualidade, avaliando se os novos dados são logicamente consistentes com as versões e estados anteriores.

Além disso, pode permitir o agendamento de tarefas, identificando ou antecipando o começo de um estado no banco de dados, que desencadeia respostas pré-definidas, e ainda exibir as informações, gerando um mapa estático ou dinâmico, ou um relatório alfanumérico, dos processos temporais das atividades de uma região.

Neste trabalho, não são tratados os processos de aquisição, modelagem e representação das parcelas e os dados sobre o Registro de Imóveis. Pretende-se contribuir nas discussões de como utilizar os CTMs existentes, integrá-los aos dados antigos e gerar informações históricas úteis para os usuários a partir de recursos tecnológicos de baixo custo, porém com recursos humanos capacitados.

1.2 Estrutura da Tese

O Capítulo 1, Introdução, apresenta uma visão geral do trabalho, objetivos, hipótese, justificativas e a estrutura da tese.

O Capítulo 2, Revisão da Literatura, com as seguintes Seções: Cadastro Territorial Multifinalitário, apresenta um histórico do Cadastro no mundo e no Brasil, define sua função e estrutura. Discute as características dos Sistemas de Informação Cadastral e a evolução do Cadastro 2D (duas Dimensões) ao 4D (quatro dimensões); Integração de Dados, discute a integração dos dados geográficos, os padrões e os esforços que vem sendo despendidos para o estabelecimento da padronização e conseqüentemente da integração dos dados; Banco de Dados, apresenta os conceitos e as tecnologias de banco de dados não convencionais para o armazenamento de dados temporais, espaciais (geográficos) e espaço-temporais; *DW* e *Spatio-Temporal Data Warehouse (STDW)*; e Trabalhos Correlatos, apresenta alguns trabalhos relacionados ao tema.

O Capítulo 3, Materiais e Métodos, apresenta as etapas do método de análise temporal para CTM, a arquitetura e as ferramentas utilizadas.

O Capítulo 4, Resultados, apresenta os resultados sobre os estudos realizados.

O Capítulo 5, Discussão dos resultados apresentados no Capítulo 4.

O Capítulo 6, Conclusão e Trabalhos Futuros. Apresenta as conclusões sobre este trabalho e questões para sua continuidade.

O Apêndice A - Modelagem do Banco de Dados Espaço-Temporais.

O Apêndice B – Modelagem Global de Dados.

O Apêndice C – *Spatio-Temporal Data Warehouse*.

2 REVISÃO DA LITERATURA

2.1 Cadastro Territorial Multifinalitário

O termo Cadastro surgiu há muito tempo. Tradicionalmente ele foi projetado para apoiar a tributação territorial, transferência de imóveis e redistribuição de terra. Fornecendo informações pertinentes a terra, que melhoram a eficiência e segurança das transações em relação à sua posse. Assim, com informações sobre a propriedade, apoiando a tributação e a regulamentação territorial (FIG, 1995).

Como evidência de quão antigo é o Cadastro, foram encontradas documentações de taxação de terras e outras contribuições para o estado, 3000 anos a.C. (antes de Cristo – a.C.), no Egito antigo. Nesse período em função das cheias periódicas do rio Nilo, havia um inventário descritivo de terras dentro de um sistema de coordenadas (LARSSON, 1996).

Outro registro histórico, datado entre 1600 a.C. e 1200 a.C., são dos kassitas da antiga Mesopotâmia que utilizavam os *kudurrus*, ou *kassita boundary stone*, para o registro de terras, que serviam como registro de doação de terras feita pelo rei a uma pessoa (KUDURRU, 2011). As informações sobre a terra eram escritas duas vezes em um mesmo *kudurru*, que era quebrado, uma parte ficava com o rei e a outra com o proprietário, a Figura 1 mostra alguns exemplares.



Figura 1: Exemplo de *kudurrus* kassitas

Fonte: (a) <http://www.larshammaren.se/bildgal.htm>,

(b) <https://feldman.berkeley.edu:453/?course=nes120a&do=detail&imgid=00321>,

Ainda há relatos que os Caldeus, em aproximadamente 400 a.C., teriam um registro das parcelas de terra, no qual eram descritas geometricamente, possibilitando o conhecimento da estrutura fundiária para fins de tributação (LARSSON, 1996).

Esses são alguns dos relatos mais antigos, desde então os governantes sempre dispunham de algum recurso para administração e distribuição do uso da terra, até que

Napoleão, em 1807, tornou-se um marco para o Cadastro contemporâneo. Após a Revolução Francesa, Napoleão decretou um completo levantamento cadastral de todo território nacional francês e também das terras ocupadas, com a finalidade de mapear as áreas estratégicas, estimular a cidadania e a tributação justa dos imóveis (LARSSON, 1996).

Esse ato consistia em uma simples descrição verbal e um mapa, no qual se detalhava a localização e as fronteiras das terras em questão. Essa herança, deixada por Napoleão no século XIX, foi desenvolvida na tentativa de tornar a Europa um único Império Francês, e assim os países colonizados herdaram a semelhança desse sistema.

No Brasil as primeiras medições territoriais sugeriram com o sistema de Sesmarias, quando o governo português distribuiu terras para desenvolver a agricultura, a criação de gado, e posteriormente o extrativismo vegetal. Paralelamente, as Sesmarias ajudaram a povoar o território e recompensar nobres, navegadores ou militares por serviços prestados à Coroa Portuguesa. Esse sistema era semelhante ao sistema jurídico português, definido pela lei de 26 de maio de 1375, e foi encerrado em 1822, após essa data não havia uma política de ocupação territorial, por isso o único modo de aquisição de terras na época era a ocupação (LOCH, 2007).

Para regularizar a demarcação das terras foi criada a primeira lei de terras brasileiras, Lei nº 601, de 18 de setembro de 1850, que discriminou os bens de domínio público do particular, criando o registro paroquial das terras do Império e obrigando os proprietários rurais a registrarem suas terras. Em 1854 a lei foi regulamentada pelo Governo Imperial, tentando promover a colonização nacional. Todavia, essa lei não conseguiu legalizar e cadastrar as terras particulares e públicas. Essa lei não foi relacionada ao Cadastro Territorial, essa relação somente aconteceu com a criação do Estatuto da Terra, Lei 4.504 de novembro de 1964², que regulamenta os direitos e obrigações relacionados aos bens e imóveis rurais com o objetivo de desenvolver a reforma agrária e promoção de políticas agrícolas (LOCH, 2007).

Seguindo os registros históricos, em 1973 foi sancionada a Lei 6015/1973³, também chamada de Lei de Registros Públicos, incluindo o “Título V: Do Registro de Imóveis”. Nessa lei, cada imóvel devia possuir matrícula individualizada, ou seja, nenhum imóvel possuía mais de uma matrícula e nenhuma matrícula mais de um imóvel, entre outras orientações.

Mais recentemente, em 2001 foi instituída a Lei 10.267/2001 de georreferenciamento dos imóveis rurais padronizando os procedimentos, desde a caracterização do imóvel até a sua

² Disponível em: http://www.planalto.gov.br/ccivil_03/leis/L4504.htm.

³ http://www.planalto.gov.br/ccivil_03/leis/L6015.htm

localização por meio das coordenadas dos vértices definidores dos limites, exigindo que a precisão posicional dessas coordenadas fosse igual ou superior a 50 centímetros (AMORIM et al., 2007).

Nenhuma das leis citadas anteriormente dispunha sobre os imóveis urbanos. Somente em 1998, uma norma que tratava da área urbana foi publicada. A Associação Brasileira de Normas Técnicas (ABNT) publicou a norma NBR – 14166/1998 estabelecendo normas para a implantação e manutenção da Rede de Referência Cadastral Municipal (AMORIM et al., 2007).

No entanto, em 2009, o Ministério das Cidades publicou a Portaria Nº 511 estabelecendo as diretrizes para o CTM. Esta portaria descreve que o CTM “será o inventário territorial oficial e sistemático do município e será embasado no levantamento dos limites de cada parcela, que recebe uma identificação numérica inequívoca.” (BRASIL, 2009) A partir dessa portaria algumas ações começaram a ser desenvolvidas para a efetivação do CTM brasileiro.

Desde o surgimento, o CTM passou por várias fases, associando novas características conforme a necessidade da administração pública, assim, vários tipos de sistemas cadastrais surgiram ao longo do tempo. Dale e Mclaughlin (1990) identificam três categorias de Cadastro, segundo sua aplicação e evolução: Cadastro Jurídico, o qual serve como um registro legal de posse da terra; Cadastro Econômico, o qual foi desenvolvido, inicialmente, para a tributação do imóvel; e Cadastro Multifinalitário, o qual reúne ambos o econômico e o jurídico adicionando outras informações da parcela, no Brasil denominado CTM.

Para WILLIAMSON et al. (2010) o Cadastro passou por estágios ao longo do tempo, são eles: Cadastro como ferramenta econômica; Cadastro como ferramenta para transações imobiliárias; Cadastro como ferramenta para o planejamento; e por último Cadastro como ferramenta para o gerenciamento territorial, o CTM .

As transformações do Cadastro ocorreram por fatores que influenciaram o formato e sua administração, incluindo (FIG, 1995): história, cultura e acordos tradicionais de posse da terra; área; geografia física e econômica; distribuição da população; nível de tecnologia; planos tradicionais de administração pública; planos de lei sobre a propriedade e a terra; prioridades políticas para a jurisdição, em relação a terra.

Acompanhando a história do Cadastro, a Figura 2 apresenta uma visão da evolução do Cadastro ao longo do tempo, denominados de ondas, parafraseando Alvin Toffler⁴.

⁴Alvin Toffler: autor futurista norte americano, conhecido pelas ideias sobre a revolução digital, revolução das comunicações e da singularidade tecnológica.

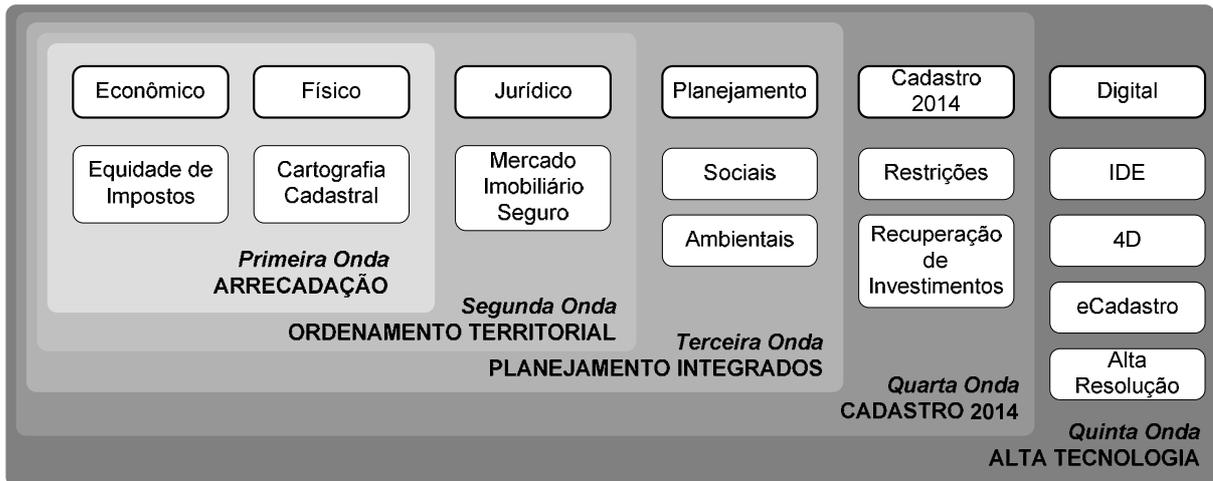


Figura 2: Evolução do conceito do Cadastro.
Fonte: ÁGUILA; ERBA (2007).

Na primeira onda, o Cadastro tinha por finalidade a arrecadação de impostos, na segunda o Cadastro agregava o conceito de direito de propriedade, na terceira a multifinalidade do Cadastro foi reconhecida para apoiar a gestão pública. A partir da quarta onda, uma visão futurista do Cadastro, amplia o registro de dados e o transforma em um inventário público de todos os objetos territoriais legais, mensurado por seus limites, na quinta um Cadastro que faz uso dos mais modernos recursos tecnológicos, mencionando pela primeira vez a temporalidade (4D). Na implantação do Cadastro, seus gestores devem identificar em qual onda eles se encontram e onde querem chegar.

Independente do formato e administração, o Cadastro pode ajudar no monitoramento e no controle de alguns aspectos territoriais, como (FIG, 1995):

- o tamanho das parcelas, máximo e mínimo, para prevenir fragmentação excessiva;
- a forma da parcela, evitando projetos de subdivisão antieconômica ou estradas e sistemas de água ineficientes, etc.;
- realocação de parcelas para melhorar políticas sociais e econômicas por subdivisão, consolidação de terra, etc.;
- uso da terra, como na agricultura ou assegurar baixo custo de edificação pública para alocação de grupos de pessoas;
- controles e medidas escolhidas para implementar programas sociais, buscando melhorar o acesso a propriedade territorial pelas mulheres e grupos de minoria;
- avaliação da terra para a arrecadação de impostos e taxas do governo;
- arrecadação de contribuições para melhorar instalações comuns, como sistemas de água;
- o valor da terra como resultado do desenvolvimento;

- aquisição de terra para fins públicos ou comuns.

Alguns desses aspectos podem ser melhorados com a disponibilização do histórico da parcela. Por exemplo, para apoiar os programas sociais as informações sobre a idade e renda mensal dos residentes podem indicar domicílios que precisam do acompanhamento do setor público. Outra situação é a alteração da área construída, em alguns casos não são regularizadas pelos proprietários, levando a arrecadação injusta de impostos e taxas.

Os principais objetivos do CTM urbano, de acordo com Loch (2005) são: a) coletar e armazenar informações descritivas do espaço urbano; b) manter atualizado o sistema descritivo das características dos imóveis; c) implantar e manter atualizado o sistema cartográfico; d) fornecer dados físicos e descritivos para o planejamento urbano, informações que estão sempre relacionadas ao sistema cartográfico, respeitando o nível de detalhamento da escala da carta; e) fazer com que o sistema cartográfico e o descritivo gerem as informações necessárias à execução de planos de desenvolvimento integrado da área urbana; f) tornar as transações imobiliárias mais confiáveis, através da definição precisa da propriedade imobiliária; g) colocar os resultados do CTM urbano à disposição dos diversos órgãos públicos envolvidos com a terra, do cidadão e do contribuinte; h) facilitar o acesso rápido, atualizado e confiável aos dados gerados pelo CTM a todos os usuários que precisam desse tipo de informações.

Para a FIG (acrônimo francês de Federação Internacional de Agrimensores) (1995), o Cadastro é um sistema, normalmente baseado em parcelas, que registra interesses sobre a terra, como direitos, restrições e responsabilidades, que pode ser estabelecido para arrecadação, legal e, ou, de apoio ao planejamento. Buscando sempre o desenvolvimento social e econômico, mas, sem a necessidade de um CTM uniforme para todos os países.

Para atender esses preceitos, o CTM, geralmente, deve conter textos e mapas, relacionados por um identificador único, que pode ser o código postal, as coordenadas do “centroide” ou um número sequencial, como mostra a Figura 3. Os dados podem ser organizados baseados na parcela territorial, coletados, armazenados e referenciados (DALE; MCLAUGHLIN, 1990).

A Portaria 511, de 07 de dezembro de 2009, no Art. 7^a, diz que o CTM deve ser constituído de: “I - Arquivo de documentos originais de levantamento cadastral de campo; II - Arquivo dos dados literais (alfanuméricos) referentes às parcelas cadastrais; III - Carta Cadastral.” (BRASIL, 2009).

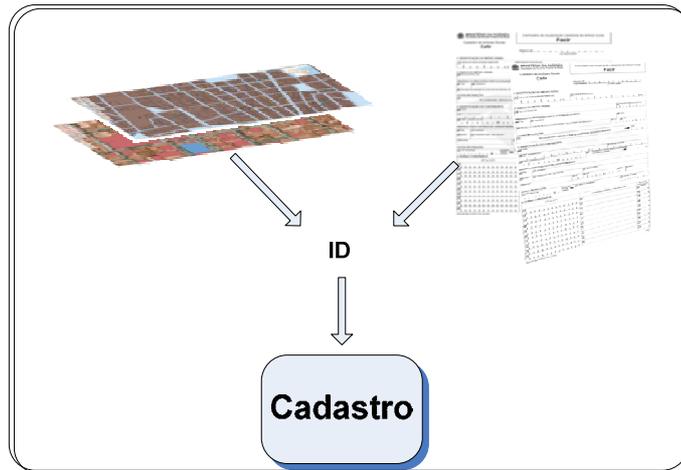


Figura 3: Informações do Cadastro.

Os documentos originais de levantamento cadastral de campo são documentos que descrevem a origem das informações sobre as parcelas cadastrais, como, croquis, planilhas de cálculos e referências de equipamentos. Os arquivos literais, atualmente guardados em bancos de dados, contêm dados descritivos das parcelas e das pessoas a elas relacionadas, por exemplo, identificador da parcela, nome do proprietário, área e uso. A carta cadastral contém a representação cartográfica do levantamento sistemático das parcelas territoriais, em grande escala, com elementos físicos e naturais (OLIVEIRA, 2010).

Com esses arquivos, o CTM torna-se um recurso importante para a gestão pública, pois disponibiliza informações importantes sobre a ocupação territorial, no Brasil, é mantido e administrado pelas prefeituras. Para a criação de um sistema cadastral, vários fatores precisam ser considerados para que o CTM não se torne oneroso e possa disponibilizar informações em um prazo aceitável. De forma geral, as etapas de execução do CTM, para a geração dos documentos definidos pela Portaria 511/2009, seriam a de planejamento, trabalho de campo e trabalho de escritório, conforme a Figura 4.

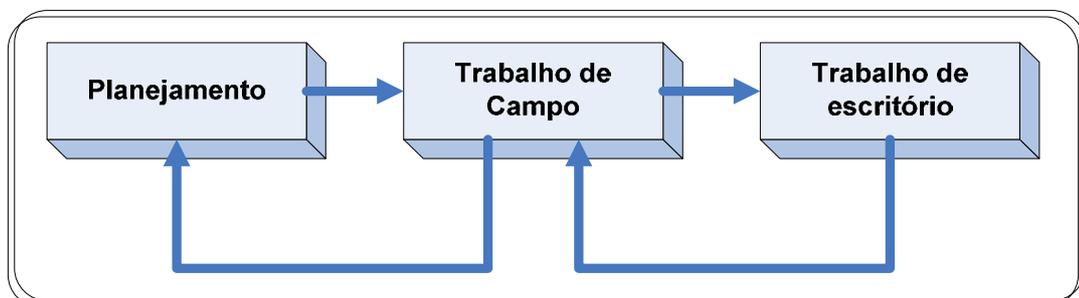


Figura 4: Etapas de execução do Cadastro.

Na etapa de planejamento, se faz um diagnóstico da situação atual, verificando como o Cadastro se insere na estrutura administrativa da prefeitura e quais as políticas que o sustenta. A partir desses dados, definem-se quais dados estarão contidos no CTM, geralmente, no

início, é voltado para a tributação (Cadastro Econômico). Entretanto, com as necessidades atuais dos gestores públicos, dados temáticos são inseridos no CTM, como dados sobre educação, saúde, meio ambiente, entre outros. Também faz parte dessa etapa um estudo sobre a viabilidade técnica e financeira da implantação do Cadastro.

Na etapa de campo se faz a coleta dos dados sobre as parcelas, seus proprietários e residentes. A cartografia topográfica, normalmente, é feita por empresas privadas, que utilizam o aerolevanteamento ou levantamento topográfico. A coleta de dados ocorre *in loco*⁵, para o preenchimento do Boletim de Informação Cadastral (BIC). No BIC coletam-se os dados temáticos e de caracterização do imóvel.

A etapa de escritório é responsável pelo processamento e integração das informações da cartografia topográfica e da coleta de dados. Processos fotogramétricos, com restituidores analógicos, analíticos e digitais, computadores, *plotters* e impressoras, softwares de sensoriamento remoto, SIG, banco de dados entre outros recursos, são utilizados para tratar os dados de campo. Nessa etapa, tanto os recursos tecnológicos são importantes, como os recursos humanos, pois a forma de utilização dos recursos tecnológicos vai refletir diretamente no CTM (LOCH; ERBA, 2007).

As três etapas de execução do CTM são desenvolvidas em torno da parcela cadastral, é a unidade fundamental, são contíguas e não se sobrepõem umas às outras, nem existem lacunas entre si (PHILIPS, 2010). Essas unidades são definidas pelos limites formais ou informais que as demarcam (FIG, 1995). A Portaria 511/2009 define a parcela como a menor unidade do Cadastro, representando toda e qualquer porção da superfície do município, com um código único e estável. Outras unidades são modeladas por uma ou mais parcelas, por exemplo, lotes, glebas, vias públicas, praças, lagos, rios e outras (BRASIL, 2009).

As etapas de execução (apresentadas na Figura 4) envolvem muitas atividades e especialistas de diversas áreas. Souza (2011) chama a atenção para uma questão importante do Cadastro, a multidisciplinaridade. O referido autor afirma que o Cadastro está estabelecido sobre alguns pilares que permitem entender a sua complexidade e a sua efetiva aplicação. Os pilares são, como mostra a Figura 5: legal, que envolve os aspectos da seguridade jurídica do Cadastro, como leis, decretos e normas; econômico, que envolve os recursos financeiros gerados pelo Cadastro e os que o financiam; técnico, referente ao aspecto técnico, como equipamentos de levantamento de campo, computadores, aplicativos, banco de dados, imagens, entre outros; político, envolve as negociações com a administração pública para a

⁵ *in loco*: do Latin, significa “no local”.

viabilidade do Cadastro, como esclarecimentos a população e o convencimento do Poder Legislativo.

A Figura 5 representa o Cadastro com os pilares nas mesmas dimensões, trabalhando juntos para a sua efetivação. Nota-se que Souza (2011) não considera os recursos humanos como um dos pilares do Cadastro, para ele os recursos humanos influenciam em todos os pilares. Caso os recursos humanos não sejam equilibrados nas suas decisões influenciarão na base de apoio do Cadastro, deixando-o desequilibrado como mostra a Figura 6, com isso ele não cumpre seu objetivo de forma adequada.

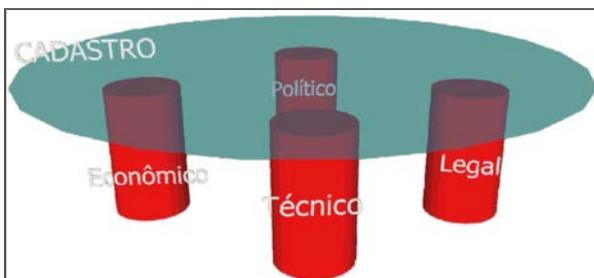


Figura 5: Pilares do Cadastro.
Fonte: SOUZA (2011).

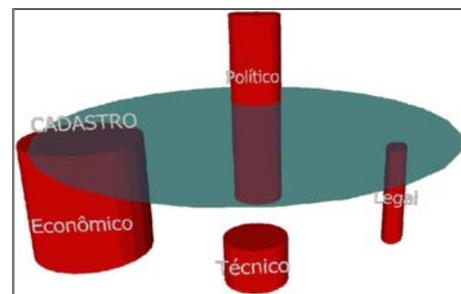


Figura 6: Pilares do Cadastro sob influência.
Fonte: SOUZA (2011).

Amorim, Souza, e Yamashita (2008) ressaltam que o caráter multidisciplinar do Cadastro apresenta-se como um conjunto de informações organizadas para diversos departamentos de setores da administração municipal. No entanto, a maioria dos projetos de Cadastro apoiados por SIG implantado nas prefeituras, não foram modelados com a participação de equipes multidisciplinares de usuários internos, implicando na falta de funcionalidade operacional do Cadastro (AMORIM; SOUZA; DALAQUA, 2004). Ou seja, a multidisciplinaridade do Cadastro ainda não é tratada com a devida atenção.

2.1.1 Sistemas de informação

Os Sistemas de Informação (SI) tornaram-se parte do cotidiano de trabalho da maioria das organizações representando um conjunto organizado de pessoas, hardware, software, redes de comunicação e recursos de dados que coleta, transforma e difunde informação, fornecendo apoio às operações, à tomada de decisão e à obtenção de vantagem estratégica (O'BRIEN, 2004).

Desde a década de 1950, os SIs foram recebendo diferentes denominações conforme o tipo de informação que eles processavam e para qual finalidade eram utilizados, como Sistemas de Informação Gerencial, Sistemas de Apoio à Decisão, Sistemas de Informação

Estratégica, Sistemas Especialistas, Sistemas de Informação Executiva, Sistemas de Computação do Usuário Final até os Sistemas de *e-business* e *e-commerce*.

O CTM precisa de um SI que trate de dados alfanuméricos e geográficos para apoio à operação e tomada de decisão. Nesse contexto se enquadram os SIGs, que utilizam banco de dados geográficos para desenhar e exibir mapas e outros demonstrativos gráficos que apóiam decisões relativas à distribuição geográfica de pessoas e outros recursos (Ó'BRIEN, 2004).

O SIG, utilizado para a gestão, está relacionado aos Sistemas de Suporte à Decisão (do termo inglês *Decision Support System*) e inclusive em alguns casos, integrado para formar os Sistemas Espaciais de Suporte à Decisão (*Spatial Decision Support System*). São caracterizados pelo desenvolvimento de complexas análises e modelagens de situações, a partir da utilização do conjunto de funções e ferramentas disponíveis nesses sistemas. Algo muito importante deste tipo de aplicação é que são elas que têm um maior impacto dentro das organizações em termos de resultados e retorno de investimentos (PIUMETTO; ERBA, 2007).

Na literatura são muitas as definições para SIG, Piumetto e Erba (2007) apresentam uma série de citações sobre o assunto, e fazem a seguinte ressalva, consideram como SIG todo sistema que permite modelar o espaço geográfico, estruturar CTMs e realizar análises espaciais com a finalidade de dar suporte à tomada de decisão na definição de políticas para o território urbano.

Os SIGs ajudam a solucionar problemas complexos da realidade urbana e fornecem informações geográficas para a tomada de decisão em inúmeras áreas, entre elas citam-se: planejamento, CTM, mercado imobiliário, uso do solo, redes de infraestrutura, telecomunicação, trânsito, logística, estudos sociais, demografia, educação, saúde, segurança, desenvolvimento econômico, entre outros. Por tanta diversidade de aplicações, os SIGs recebem diferentes denominações de acordo com a área de aplicação e o tipo de dados que administram. Os SIGs, orientados para o CTM, com base na parcela territorial são comumente denominados de SIT (PIUMETTO; ERBA, 2007) (FIG, 1995). Nesse contexto, o SIT é um sistema para a aquisição, processamento, armazenamento e distribuição de informações sobre o território (WILLIAMSON et al., 2010).

O SIT faz parte dos Sistemas de Administração Territorial (do inglês *Land Administration System*), que são uma infraestrutura para a implementação de políticas fundiárias e de estratégias de gestão territorial em prol do desenvolvimento sustentável. A infraestrutura inclui acordos institucionais, uma estrutura legal, processos, normas, informações territoriais, sistemas de gestão e divulgação, e as tecnologias necessárias para

apoiar a alocação, comercialização de terras, avaliação, controle do uso e desenvolvimento de interesses pelas terras. (WILLIAMSON et al., 2010). Segundo Williamson et al. (2010), o CTM é o núcleo dos Sistemas de Administração Territorial que fornece integridade espacial e a identificação única de cada parcela territorial.

Nesses sistemas os CTMs são representações de grande escala de como a comunidade divide suas terras em partes utilizáveis, as parcelas. A integridade espacial dentro do CTM é fornecido por um mapa cadastral, atualizado por levantamentos cadastrais. A identificação única da parcela fornece a ligação entre o mapa cadastral, o registro territorial e os servidores, como a base do Sistema de Administração Territorial e suas informações territoriais. O CTM deverá ainda incluir todas as parcelas de uma jurisdição: espaço público, privado, comunitário e aberto.

Assim, no contexto dos Sistemas de Administração Territorial, o SIT, baseado na parcela e atualizado, contém um registro de interesse e uma descrição geométrica das parcelas territoriais, ligados a outros registros que descrevem a natureza desses interesses, e muitas vezes o valor da parcela e suas melhorias. Pode ser estabelecido para fins econômicos (por exemplo, avaliação e tributação igualitária), para finalidade legal (como transferência de propriedade), para auxiliar na gestão da terra e de seu uso (por exemplo, para outros propósitos administrativos e de planejamento), e para permitir o desenvolvimento sustentável e a proteção ambiental (FIG, 1995) (WILLIAMSON et al., 2010).

No Brasil, o termo SIT não é muito utilizado, em outros países ele é bastante conhecido quando integra o Cadastro ao Registro de Imóveis (AMORIM; SOUZA; YAMASHITA, 2008). Todavia, os Art. 4º e 5º, da Portaria 511/2009, diz que, quando os dados do CTM estão relacionados aos dados do Registro de Imóveis, formam um Sistema de Cadastro e Registro Territorial (SICART), e acrescentando dados dos cadastros temáticos cria-se o SIT (BRASIL, 2009).

Nesse é importante estabelecer a necessidade da coleta sistemática, atualização, processamento e distribuição dos dados espaciais para apoio à tomada de decisão administrativa, econômica e legal. Também dever ser aplicado no planejamento do desenvolvimento do território e para avaliar as consequências de diferentes ações alternativas (LARSSON, 1996).

Por essas características, o SIT pode conter dados de diferentes categorias de informação, permitindo ser projetado para fornecer: informação ambiental, cujo foco é a delimitação de zonas ambientais associadas a alguns fenômenos físicos, químicos ou bióticos únicos, como áreas sujeitas a inundações; informação de infraestrutura, que se concentra

principalmente em estruturas de engenharia e serviços públicos, tais como serviços subterrâneos e oleodutos; informação cadastral, relacionada a áreas onde direitos, responsabilidades ou restrições sobre o território são reconhecidos, como áreas sujeitas a restrições de planejamento particular; e, informação socioeconômica, que inclui tipos de dados censitários e estatístico (DALE; MCLAUGHLIN, 1990).

O SIT permite fornecer estes dados na forma de um produto (como mapas ou títulos certificados) ou em forma de serviços (como consultoria profissional), composto de: dados de atributos, que podem ser apresentados de forma textual ou numérica; dados espaciais, que podem ser apresentados em mapas; e dados temporais, que indicam sua ocorrência (DALE; MCLAUGHLIN, 1990). A Figura 7 apresenta uma forma de representar os dados espaciais e temporais das parcelas territoriais em um SIT (SASS; AMORIM, 2013).

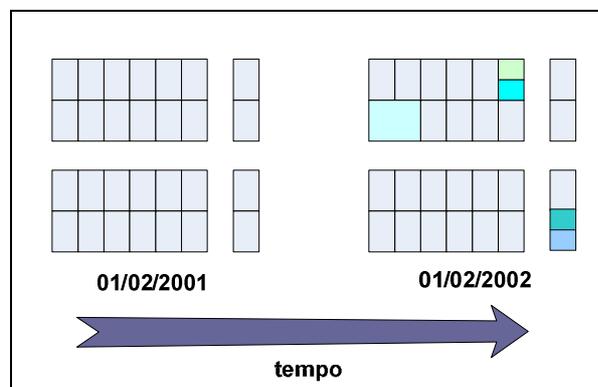


Figura 7: Representação de dados espaciais e temporais.
Fonte: SASS; AMORIM (2013).

Os componentes de localização dos dados podem ser representados no mapa como pontos, linhas ou polígonos. Eles podem incorporar três tipos de relações espaciais: geométrica, onde os dados são referenciados em um sistema espacial preciso, tais como coordenadas; cartográfica, onde os dados são generalizados e simbolizados; e topológica, onde há localização relativa e não métrica, e definido o relacionamento de vários elementos de dados (DALE; MCLAUGHLIN, 1990). O ponto pode representar um medidor de água, um poste ou uma árvore. A linha pode representar uma rede de serviço, como, esgoto, água e eletricidade. Por fim, o polígono, representa uma parcela, um quarteirão, uma construção ou um lago, com um número determinado de vértices.

Pode-se dizer, que os SITs são ferramentas para apoiar o processo de tomada de decisão, seja em um projeto ou em uma instituição, e que agrega os dados geográficos as bases de dados alfanuméricos para gerar uma visão mais completa e adequada da realidade analisada (PIUMETTO; ERBA, 2007). Porém, os dados temporais não têm sido contemplados pelos SITs desenvolvidos.

As novas tecnologias têm aumentado muito o potencial do desenvolvimento de sistemas, mas também impõe restrições e condições. Weir (1984⁶ apud LARSSON, 1996) relata que informações relacionadas ao território têm uma inestimável importância para o uso ordenado, justo e inteligente e para o desenvolvimento do território. No passado essas informações estavam em papel, com as novas tecnologias, estas atividades são, agora, computadorizadas e automatizadas. Esta transição foi, e é importante para agências governamentais, privadas e indústrias. Por isso, é fundamental na construção dos SIT: a existência de um modelo de referência de fácil acesso; ações construtivas por parte do governo na coordenação do uso do território; e padronização de procedimentos e terminologias.

2.1.2 Cadastro 2D, 3D e 4D

A utilização dos SIs nos CTM trouxe benefícios para a gestão e controle dos dados cadastrais. Tradicionalmente, os registros cadastrais consistiam em um conjunto de mapas que continham as parcelas, com uma identificação única da parcela e um arquivo de papel onde suas informações eram mantidas. Desde o final do século passado as inscrições cadastrais, em países desenvolvidos, começaram a ser convertidas de registros cadastrais analógicos para registros digitais. As informações espaciais sobre a parcela não são mais mantidas em mapas de papel, mas em SIG e *Computer-Aided Design (CAD)* (STOTER; VAN OOSTEROM, 2006).

Nos registros digitais, uma ligação é feita entre o mapa cadastral digital e o banco de dados cadastrais. A ligação fornece a possibilidade de examinar a parte espacial e a administrativa do registro cadastral e combinar os resultados. Nos SIGs é possível consultar a parte espacial e administrativa do registro cadastral em um ambiente integrado. Assim, os bancos de dados cadastrais contêm os dados obtidos por meio do levantamento cadastral, dados alfanuméricos referentes aos imóveis e seu uso e dados geográficos referentes à sua localização (AMORIM; SOUZA; DALAQUA, 2004).

A maioria dos CTMs possuem dados 2D que representam a parcela no plano (X,Y), como mostra a Figura 8 (a), porém, essa representação já não atende às necessidades de informação sobre uma determinada área nos dias de hoje. Pois, sua característica básica é não permitir espaços ou sobreposições entre áreas na superfície terrestre, isso gera limitações para

⁶ WEIR, C. H. **The decision maker and land information systems**. Canadian Institute of Surveying, Ottawa, 1984.

registrar o aumento das características complexas (sobreposições) e restrições do mundo atual.

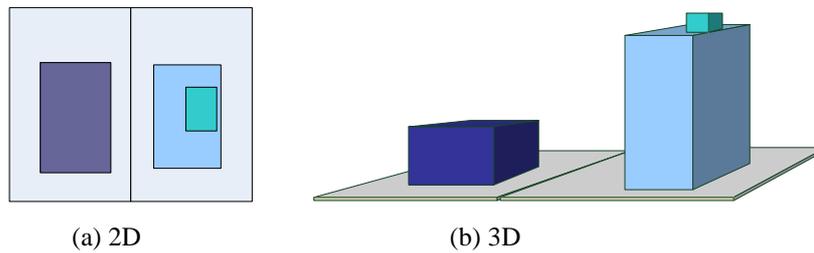


Figura 8: Representação da parcela.
Fonte: SASS; AMORIM (2012).

Como solução para resolver o problema de sobreposições, surgiu o conceito de Cadastro 3D (três dimensões), como mostra a Figura 8(b). O conceito 2D não é traduzido diretamente para o Cadastro 3D e sim estendido, pois o 3D é baseado em uma estrutura topológica complexa com volumes, faces, extremidades e nodos.

Esse conceito surgiu pela demanda por espaço físico, que em grandes cidades, tem gerado a sobreposição e a conexão de construções. Neste contexto, surge o desafio de registrar construções que se sobrepõem e se conectam quando projetadas na superfície em uma inscrição cadastral que registra informação sobre parcelas 2D. Embora a propriedade tenha ficado, por muitos anos, situada uma em cima da outra, como mostra a Figura 9, novas sobreposições vêm surgindo no ambiente urbano.



Figura 9 : Edifício de apartamentos⁷

As situações mostradas nas Figura 10(a) e 10(b) apontam para a necessidade de se buscar soluções para a representação cadastral na forma tridimensional (3D).

⁷ Imagem disponível em: <http://www.sinomar.com.br/portal/conteudo.asp?codigo=4191&page=17>. Acesso em 05 dez. 2011.

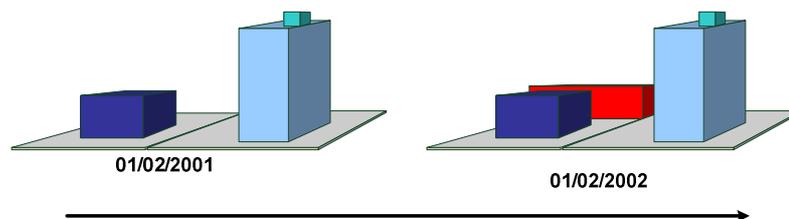
(a)⁸ Viaduto Otávio Rocha (Porto Alegre - RS).(b)⁹ Radisson Montevidéo Victoria Plaza Hotel – Montevideú – URU.

Figura 10: Construções complexas.

Souza (2011) ressalta que acrescentar a informação tridimensional ao CTM não é uma tarefa fácil, pois não basta acrescentar um dado sobre a altimetria das parcelas ou edificações. É preciso que os objetos tridimensionais estejam inseridos no modelo do Cadastro, no qual o sistema de referência utilizado esteja claro. No Brasil, e em muitos outros países, não existe um modelo de Cadastro urbano bem definido.

Complementando o conceito do CTM, a variável tempo sempre teve um papel importante em sistemas cadastrais, mas, o aspecto temporal tem sido tratado independentemente do aspecto espaço, 2D ou 3D. O registro do tempo permite identificar, por exemplo, quem tem e quais direitos em um determinado momento, para aquele espaço e para aquele período. Por isso uma abordagem integrada dos aspectos temporais e espaciais deve ser investigada buscando identificar melhorias para os Cadastros (VAN OOSTEROM et al., 2006).

Considerando o Cadastro 3D acrescido do tempo, cria-se o Cadastro 4D. O Cadastro 4D, segundo Van Oosterom et al. (2006), introduz o atributo tempo para descrever várias parcelas e suas trajetórias históricas, como mostra a Figura 11. O cadastro 4D é usado quando o registro histórico de um Cadastro 3D é importante para entender a evolução do uso da terra durante um espaço de tempo apoiando políticas públicas. Assim, uma nova dimensão é adicionada ao Cadastro, nesse caso o tempo, criando a quarta dimensão.

Figura 11: Representação 4D da parcela.
Fonte: SASS; AMORIM (2013).

⁸ Imagem disponível em: http://proweb.procempa.com.br/pmpa/prefpoa/vivaocentro/default.php?p_secao=130
Acesso em: 18 mai. 2010

⁹ Fonte: Foto de Gláucia Gabriel Sass. 29 nov. 2012.

O conceito do Cadastro 4D mantém o conceito do 2D e 3D, não deixando espaços vazios, contudo considerando em paralelo a dimensão de tempo. Esta dimensão pode ser integrada pela dimensão espacial ou com um atributo. Uma maneira de implementar esse cadastro é utilizando uma estrutura topológica no espaço-tempo 4D como fundamento. No entanto, essa estrutura ainda não está disponível nos softwares, uma alternativa para a implantação do Cadastro 4D é usar as técnicas do Cadastro 3D e definir atributos temporais, com restrições, para se obter o Cadastro 4D (DONER et al., 2008).

Assim, para o Cadastro 4D mantém-se a estrutura 3D do objeto no banco de dados, acrescentando o “local no tempo” que a informação da parcela foi utilizada pelo Cadastro. Neste contexto, o aspecto temporal é um aspecto fundamental no qual um registro está recorrendo a parcela já existente, porém em um tempo diferente. Uma parcela pode mudar com o passar do tempo, ocorrendo desmembramento, remembramento ou até mesmo ser apagada do CTM, mantendo-se as referências corretas e a consistência no sistema, uma versão específica da parcela em um determinado tempo pode ser recuperada (DONER et al., 2008).

Para Langran (1993) as funções que a componente temporal poderia apoiar o CTM e o registro de terra seriam:

- inventário (descrição completa): esse poderia ser um registro inicial que compreende longo espaço de tempo. Assim o 'ponto de partida' de um Cadastro pode ser diferente em uma região, isto também é válido para um 'ponto final', por exemplo, no caso de um novo regime legal;
- analisar (explicar, explorar, prever): isso é pertinente no caso do uso dos dados cadastrais para planejamento do espaço e consolidação da terra, para análises de tendências e imóveis recém inscritos em hipotecas, para indexação de bens imóveis comercializados por um período de tempo mais longo, para identificação de transações durante um certo tempo, para requisição de prognóstico para propósitos de planejamento de recursos humanos;
- atualizações (atualizar informações antigas por novas): esse é um ponto fundamental, também pode incluir a geração e entrega de arquivos de atualizações referentes a um certo período para os cidadãos, para uso próprio;
- controle de qualidade (monitorar e avaliar novos dados, verificando a consistência em relação aos dados antigos): isso pode ser importante em muitos casos, por exemplo, em uma atualização representa a geração de uma nova versão completa do mapa cadastral com os impactos em uma determinada área. O mesmo acontece

no caso de novo levantamento de parcelas ou no caso da conversão de mapas cadastrais analógicos para digitais;

- agendamento (identificando estados limites que ativam ações predefinidas): algumas ações de manutenção e processos de atualização precisam ser executados em ordem sequencial;
- apresentação (geração de mapas ou tabelas de um processo temporal): esse poderia ser um processo de consolidação da terra no caso de Cadastro e Registro de terra.

Para Langran (1993) um SIG temporal deveria rastrear a mudança de estado da área de estudo, armazenar seu histórico e antecipar estados geográficos. Essa definição pode ser aplicada diretamente ao CTM.

2.2 Integração de Dados

A integração de dados é um desafio para aplicações que precisam realizar consultas em várias fontes de dados independentes e heterogêneas. Ela é importante para as grandes organizações que possuem um grande número de fontes de dados, para o progresso em projetos científicos de grande escala, para uma melhor cooperação entre as agências governamentais, e para permitir a pesquisa de dados nas milhões de fontes de dados da *World Wide Web* (HALEVY; RAJARAMAN; ORDILLE, 2006).

Essa atividade representa o problema da combinação de dados residentes em diferentes fontes, que proporciona ao usuário uma visão unificada dos dados. O projeto de sistemas de integração de dados é importante em aplicações do mundo real, e é caracterizada por uma série de problemas, entre eles a heterogeneidade de dados.

Para Hull (1997) as fontes de dados são heterogêneas devido a diferentes plataformas de *hardware* e *software* (nível físico), modelos de dados (nível lógico), e esquemas e conceitos (nível conceitual). Diferentes representações de um mesmo conceito podem acontecer, ou porque o modelo permite diferentes representações ou porque projetistas têm diferentes percepções da realidade. A multiplicidade de possíveis representações de um conceito do mundo real é chamada de relativismo semântico.

Para Wiederhold (2000) a heterogeneidade existe em vários níveis. O referido autor apresenta algumas questões sobre os níveis de heterogeneidade:

- A heterogeneidade de modelos organizacionais tem sido uma área ativa de pesquisa. Os primeiros trabalhos foram em mapeamento de modelos de banco de dados em rede, hierárquico, relacional e universal. A informação semântica formal pode ajudar na transformação de dados relacionais em objetos estruturados;

- Heterogeneidade na representação causa problemas nos limites. Em campos de comprimentos diferentes, o escopo das variáveis é susceptível de variar, assim, telefones poderiam ter 10 dígitos (99 9999-9999) ou 12 dígitos (99 99 9999-9999);
- A heterogeneidade do escopo ocorre quando variáveis com o mesmo nome e representação, em bases de dados diferentes, capturam diferentes conjuntos e subconjuntos de objetos do mundo real. Os registros, muitas vezes, se sobrepõem parcialmente: uma pessoa é empregada, é estudante, é remunerada, tem plano de saúde. Por exemplo, um novo funcionário, em período de experiência, não entra no plano de saúde, nem por isso ele estará fora da folha de pagamento. O conhecimento sobre o escopo do banco de dados é necessário para processar corretamente esses dados em conjunto;
- Frequentemente há heterogeneidade do nível de abstração. Por exemplo, os estoques podem ser agregados pelo fornecedor, por material, ou por data para efeitos fiscais. Assim como, locais podem ser identificados por CEP em um domínio e pelo nome da cidade em outro. A não ser que os dados atômicos sejam mantidos por toda parte, no nível mais detalhado de agregação, este tipo de heterogeneidade causará resultados parciais quando os dados forem unidos;
- Ainda mais dependentes do contexto de origem é a heterogeneidade de significado, ou seja, a designação dos valores dos atributos para atributos do mundo real. Enquanto o peso do automóvel pode ser objetivo, a cor não é. Da mesma forma, as notas de um curso diferem entre as escolas e os membros do corpo docente, e podem ser insuficientes para realmente classificar indivíduos;
- Quando os bancos de dados coletarem mais *snapshots*¹⁰ atuais, problemas da heterogeneidade de validade temporal surgirão. Valores têm, muitas vezes, implicitamente, diferentes faixas temporais: o salário de um empregado, benefícios de pensão e um bônus não podem ser simplesmente agregados. Além disso, os bancos de dados independentes não serão atualizados de forma síncrona. A provisão de históricos permite voltar atrás no tempo e obter um *snapshot* que é temporalmente coerente, mas um resumo para a última hora pode ser bastante obsoleto.

Para Sheth (1999), a heterogeneidade pode ser classificada em: **sintática**, ocorre quando existe um mesmo dado com nomes diferentes em diferentes sistemas, ou sintaxes

¹⁰ *Snapshots*: “foto instantânea” dos dados. Mostram os estados dos dados em um instante no tempo.

(escrita) diferentes são atribuídas a conceitos correspondentes; de **estrutura**, diferentes formatos de representação de um dado, ou seja, os dados são organizados segundo esquemas conceituais diferentes; e de **semântica**, diferentes significados dos dados, como em dados que são parecidos, considerando diferentes interpretações para os dados.

A classificação, apresentada por Sheth (1999), é simplificada e reduzida, comparando a Wiederhold (2000) citado anteriormente, no entanto em alguns pontos traz semelhanças. Wiederhold (2000) considera toda a organização na qual o dado existe, ou seja, o domínio do problema interfere diretamente no modelo de dados e demonstra as características da organização, implicando na heterogeneidade dos dados. Enquanto Sheth (1999) dá ênfase na estrutura e definição dos dados, tratando a heterogeneidade de forma mais direta. A classificação apresentada por Sheth (1999) esta inserida nos níveis de Wilderhold (2000).

Um exemplo de heterogeneidade semântica e estrutural pode ser visto na Figura 12, que mostra os dados sobre a disponibilidade de água nas parcelas em três bancos de dados do CTM de uma mesma cidade. No CTM de 1996, o projetista convencionou que a resposta, registrada no banco de dados, seria representada por “S”, se o serviço público estivesse disponível na parcela e “N”, se o serviço público não estivesse disponível. Já no CTM de 2004, desenvolvido por outra equipe, essas mesmas informações foram registradas como “Sim” e “Não”. Por último, no CTM de 2010, foi definido “1” para a resposta sim, e “0” para a resposta não. Nesse exemplo os dados foram mapeado e transformados para uma única representação.

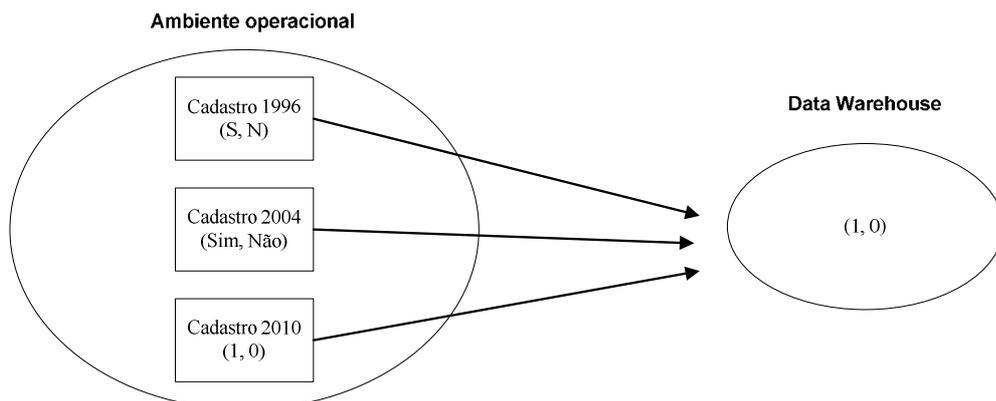


Figura 12: Exemplo de integração e transformação dos dados.

Além da integração de dados, em muitos casos, é necessária à integração de sistemas, denominada Interoperabilidade. A Interoperabilidade em SI é a capacidade de diferentes tipos de computadores, redes, sistemas operacionais e aplicações de trabalhar em conjunto de forma

eficaz, sem comunicação prévia, a fim de trocar informações de forma útil e significativa (INPROTEO, 2005¹¹ apud WILLIAMSON et al., 2010).

A interoperabilidade assume a capacidade de comunicar, executar programas ou transferir dados entre vários computadores de forma que o usuário tenha pouco ou nenhum conhecimento das características específicas dessas unidades (RAWAT, 2003).

No domínio da informação espacial, a interoperabilidade permite a compatibilidade entre SIs, de diferentes organizações, para executar, manipular, trocar e compartilhar dados relacionados com a informação espacial acima e abaixo da superfície da Terra (RAWAT, 2003). O conceito de interoperabilidade foi aplicado em empresas e organizações, além da administração pública, para: melhorar a colaboração e a produtividade em geral; aumentar a flexibilidade; melhorar a eficiência do serviço; aumentar a produtividade; e, reduzir custos (WILLIAMSON et al., 2010).

A interoperabilidade possui três níveis, semelhante à heterogeneidade de dados: sintático, estrutural e semântico (MOHAMMADI; RAJABIFARD; WILLIAMSON, 2010), como mostra a Figura 13.

A **interoperabilidade sintática** ajuda na reutilização, visualização, consulta e análise dos conjuntos de dados incluídos na integração. A necessidade de interoperabilidade sintática de várias origens e de várias arquiteturas de software distribuídas é tratada com padrões de serviços baseados na *Web* e formatos padrões como *Geographical Markup Language (GML - Linguagem de Marcação Geográfica)*. No entanto, este é apenas o primeiro passo para a integração real da informação.

A **interoperabilidade estrutural** permite um nível básico de conversão entre um esquema e outro, e que os dados espaciais sejam estruturados de acordo com a visão do fenômeno correspondente. Existem várias diferenças entre os conjuntos de dados provenientes de diferentes fontes, uma vez que essa estrutura é muito influenciada pelas organizações. A interoperabilidade estrutural está relacionada com incompatibilidades, como nomes, especificações, nomes de atributos, granularidade e valores de domínio.

A **Interoperabilidade semântica** é a habilidade do entendimento do significado dos dados entre vários sistemas. Além das diferenças estruturais, as diferenças de semântica das informações causam problemas na integração de dados provenientes de diversas fontes. Os usuários dos dados de uma organização produtora de dados não tem dificuldades para entender seus dados, mas quando esses dados são reutilizados por outra organização, esse

¹¹ INPROTEO. Definition for interoperability. <http://www.inproteomics.com>. (accessed August 1, 2009).

conhecimento prévio dos dados não existe, e aparecem as dificuldades de entendimento desses dados.

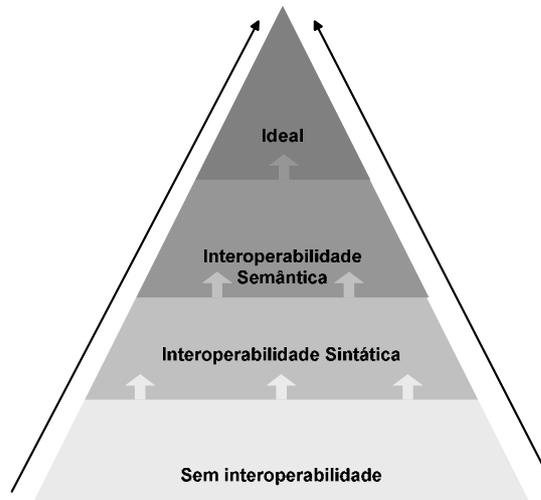


Figura 13: Hierarquia e interoperabilidade de dados espaciais.
Fonte: MOHAMMADI; RAJABIFARD; WILLIAMSON (2010).

2.2.1 Sistemas de integração de dados

Os sistemas de integração de dados são uma extensão dos SIs, com a característica de compreender como os dados de fontes distintas são mapeados para um esquema mediador usado por ele (DONG; HALEVY e YU, 2007). As fontes de dados desses sistemas podem ser independentes e dinâmicas, pois, as fontes de origem dos dados permanecem em atividade, alterando seus esquemas e atualizando seus dados com as transações diárias.

Um sistema de integração de dados, também precisa lidar com a heterogeneidade, apresentada no início da Seção 2.2, das fontes de dados. Dessa forma, esses sistemas precisam tratar a heterogeneidade de forma que uma visão integrada dos dados de todas as fontes de dados possa ser disponibilizada para o usuário. Porém, o sistema de integração de dados pode oferecer diferentes visões integradas, definidas conforme as necessidades dos usuários.

A comunidade científica tem trabalhado para desenvolver sistemas que possam fazer a integração de dados de forma automatizada, acelerando a criação de mapeamento de esquemas, reduzindo o esforço humano nessa tarefa.

Um dos principais obstáculos na criação de um sistema de integração de dados é o esforço necessário para criar as descrições de origem, e, mais especificamente, definir os mapeamentos semânticos entre as fontes e o esquema “intermediário”. Escrever tais mapeamentos (e sua manutenção) exige conhecimento de banco de dados (para expressá-los

em uma linguagem formal) e conhecimento do negócio (para entender o significado do que está sendo mapeado nos esquemas) (HALEVY; RAJARAMAN; ORDILLE, 2006).

A fim de alcançar uma boa descrição de origem dos dados, um sistema de integração de dados deve resolver vários problemas (BARLOW, 2000), como os três descritos a seguir.

Primeiro, ele deve saber onde os dados estão localizados e como podem ser acessados. Esse problema é normalmente resolvido por meio da utilização de *wrappers*. Um *wrapper* é uma camada de software que traduz a informação de uma fonte de dados específica em alguma forma padrão, entendido pelo sistema de integração de dados. Esse recurso pode ser definido através de algum modelo de objeto, descrito através da *Extensible Markup Language* (XML – Linguagem de Marcação Extensível) ou de qualquer outro modelo padrão. Dessa forma, o programador define os métodos utilizados para realmente obter as informações. Se a fonte de dados muda seu modelo interno, apenas o código no *wrapper* precisa ser modificado e o resto do sistema de integração de dados e todas as aplicações escritas com base nele permanecem sem alteração. Se uma nova fonte de informações for adicionada, tudo o que precisa ser feito é a implementação de um novo *wrapper*.

Segundo, deve fornecer alguns meios para corrigir as diferenças lógicas contidas até mesmo em dados muito semelhantes de diferentes fontes. Por exemplo, um sistema pode coletar dados sobre os postos de saúde de um município a partir de fichas eletrônicas. Apesar das informações das fichas serem muito semelhantes, ou seja, nome do paciente, o endereço, data de consultas, nome do médico, etc., pode ser muito difícil combinar essas informações em um único valor lógico. Mesmo essas questões simples como o nome do atributo de um registro pode causar problemas. Um posto de saúde pode usar um atributo "Nome" para o nome do paciente, enquanto outro, usa um atributo com o mesmo rótulo para descrever o nome do médico.

Tais inconsistências precisam ser eliminadas. Estes problemas são abordados pelo modelo interno padronizado do sistema de integração de dados, o *wrapper* possui uma referência dos dados para cada origem de dados. Assim, a partir dessas referências é possível identificar o nome correto de um atributo. Se o sistema utilizado tem os atributos "nomePaciente" e "nomeMedico" a aplicação pode ser programada com esses valores e o *wrapper* transforma os rótulos conforme necessário.

Finalmente, o sistema de integração de dados deve fornecer alguns meios eficazes para definir consultas sobre todas as fontes participantes. Esta é a camada que realmente representa o sistema de integração de dados com o seu poder e flexibilidade. A realização de uma consulta é alcançada por meio da definição de uma linguagem de consulta genérica projetada

para funcionar em conjunto com o modelo de informação interna, utilizada pelo sistema de integração de dados. As consultas são então definidas usando a notação do modelo interno do sistema de integração de dados e dependem do sistema para converter as subconsultas em uma forma compreendida por cada fonte de dados.

Como já descrito, um sistema de integração de dados fornece acesso integrado a fontes de dados independentes e heterogêneas. O objetivo é combinar dados residentes em diferentes fontes, proporcionando ao usuário uma visão integrada destes dados. Em um esquema de banco de dados, a visão integrada, representa os elementos obtidos a partir da integração de elementos de um ou mais esquemas de fontes de dados. Esta visão unificada é representada pelo esquema global, que oferece uma visão integrada de todos os dados, que podem ser consultados pelo usuário (LENZERINI, 2002).

Os principais componentes de um sistema de integração de dados são: o esquema global, as fontes de origem de dados, e o mapeamento. Os esquemas de origem descrevem a estrutura das fontes, onde os dados são reais, enquanto que o esquema global fornece uma visão integrada. As descrições feitas no mapeamento estabelece a ligação entre os elementos do esquema global e os dos esquemas de origem (LENZERINE, 2002). Estas descrições, os metadados, são importantes para o funcionamento do sistema de integração de dados, devem incluir: a identidade da fonte de dados; descrições do conteúdo e informações sobre o esquema.

Dentre as arquiteturas para sistemas de integração, existem esquemas globais (LENZERINE, 2002), sistemas federados (SHETH; LARSON, 1990), sistemas de mediadores (WIEDERHOLD, 2012) e de *DW* (INMON, 2005).

Os esquemas globais são visões que refletem os objetos dos bancos de dados utilizados, criando um esquema a partir da integração dos esquemas das fontes de dados. Geram um esquema lógico integrado onde os usuários acessam sem saber da localização física do SGBD. As diversas fontes de dados distribuídas e heterogêneas podem ser acessadas de maneira uniforme e transparente através desse esquema global. Como o esquema global é estático e oferece uma visão única dos dados integrados, esta abordagem não é muito adequada quando as fontes de dados a serem integradas são dinâmicas ou quando os usuários da informação integrada têm diferentes requisitos (BATINI; LENZERINI; NAVATHE, 1986). A construção de esquemas globais pode não ser uma tarefa fácil, pois requer conhecimento sobre os esquemas dos SGBDs e por dificilmente tratar o crescimento do próprio esquema.

Outra arquitetura para integração de dados é a abordagem federada (SHETH; LARSON, 1990). Um sistema de banco de dados federado é uma coleção de sistemas de bancos de dados cooperantes e autônomos que participam da federação para permitir um compartilhamento parcial e controlado de seus dados. Esta abordagem permite a cooperação entre sistemas independentes, oferecendo múltiplos esquemas integrados, de acordo com os requisitos das aplicações. Estes esquemas integrados são estáticos e definidos antecipadamente, não sendo apropriados para os casos em que as fontes de dados são dinâmicas e precisam ser adicionadas ou removidas, ou quando os requisitos das aplicações precisam ser atualizados.

Na arquitetura de mediadores são oferecidos serviços de integração e abstração dos sistemas de informação para suportar aplicações utilizadas pelos gestores, onde a escala, variedade e complexidade de dados e recursos de informação sejam tais que os sistemas comuns não apoiariam (WIEDERHOLD, 2012). Nessa abordagem é oferecida ao usuário uma interseção de várias fontes de dados por meio de visualizações integrada dos dados.

Contrário à arquitetura de mediadores, a arquitetura de *DW* trata da abordagem materializada onde os dados são recuperados, integrados e armazenados em um repositório de dados, de forma que as consultas podem ser realizadas diretamente (INMON, 2005).

Os sistemas de mediadores possuem arquitetura em camadas, na qual o nível mais baixo representa as fontes de dados, o nível intermediário resolve o processo de integração das fontes de dados da camada inferior, e o mais alto tem os sistemas para acesso ao sistema de integração. Na abordagem materializada do *DW*, a informação integrada de várias fontes está disponível em um único repositório.

2.2.2 Integração de dados espaciais

Em muitos Sistemas de Tomada de Decisão, a integração de conjuntos de dados de várias fontes é um dos últimos objetivos, e este geralmente não é alcançado com sucesso. Por isso, as pesquisas sobre a interoperabilidade de dados espaciais de diversas fontes e a integração técnica tem despertado o interesse científico. O desafio da integração de dados espaciais não está apenas nas questões técnicas, os requisitos não técnicos, que incluem os institucionais, sociais, jurídicos e políticos, também devem ser levados em consideração (MOHAMMADI et al., 2006).

Os problemas técnicos estão relacionados a padrões, interoperabilidade, topologia vertical, semântica, sistema de referência, modelo de dados, metadados, formato e qualidade de dados.

Os problemas não técnicos estão ligados a fase que fornece os conceitos básicos e os requisitos do domínio para as atividades da integração técnica, são eles: o uso de um único padrão para o modelo de dados, a elaboração de padrões intermediários, o desenvolvimento de soluções que facilitam a interação entre diferentes sistemas, o aumento da conscientização da existência de dados espaciais por meio da compreensão do catálogo de dados globais, o desenvolvimento de mecanismos de busca desses dados e a definição de modelos de colaboração.

Do ponto de vista técnico, a heterogeneidade pode ser considerada como sintática, estrutural e semântica, são do nível institucional, como visto no início da Seção 2.2. Para os dados espaciais a heterogeneidade sintática significa que vários sistemas utilizam diferentes linguagens de consulta (*Structure Query Language - SQL*, *Object Query Language - OQL*, etc.). A estrutural significa que diferentes sistemas espaciais guardam seus dados em diferentes estruturas. A semântica significa que os sistemas usam diferentes interpretações para os mesmos dados (MOHAMMADI, 2007).

A integração de bancos de dados espaciais de múltiplas fontes é um dos objetivos das Infraestruturas de Dados Espaciais (*Spatial Data Infrastructures - SDI*). Muitos trabalhos destacam a heterogeneidade e inconsistência encontradas nessas iniciativas e atividades, e tentam resolver estes problemas ao documentar as inconsistências técnicas. Todavia, as inconsistências técnicas, como apresentado anteriormente, tendem a surgir a partir de aspectos não técnicos e da fragmentação de acordos sociais, institucionais, jurídicos e políticos que afetam os responsáveis pelos dados, os individuais e as organizações.

O Quadro 1 apresenta uma descrição das questões técnicas e não técnicas que influenciam na integração de dados espaciais, segundo RAJABIFARD (2010) e WILLIAMSON et al. (2010).

A partir do exposto, para criar um ambiente em que diferentes conjuntos de dados podem ser integrados, uma infraestrutura deve fornecer um conjunto de ferramentas e orientações, incluindo padrões, políticas e requisitos de colaboração.

Assim, a padronização tecnológica busca facilitar o acesso aos dados, no qual diferentes formatos limitam os usuários, e não permitem a migração de um sistema para outro. Na busca pela padronização, instituições internacionais têm trabalhado pela sua definição, entre elas estão o *Open GIS Consortium (OGC)*¹² e a *Global Spatial Data Infrastructure*¹³ (*GSDI*).

¹² <http://www.opengeospatial.org/>

¹³ <http://www.gsdi.org/>

Quadro 1: Questões de integração de dados espaciais.

Questões de Integração			
Questões Técnicas	Questões não técnicas		
Questões Institucionais	Questões Políticas	Questões Legais	Questões Sociais
Heterogeneidade computacional (padrões e interoperabilidade)	Existência de legislação de apoio	Definições de direitos, restrições e responsabilidades	Questões culturais
Manutenção da topologia vertical	Consistência no direcionamento e prioridades políticas (desenvolvimento sustentável)	Consistência na abordagem dos direitos autorais e dos direitos de propriedade intelectual	Fragilidade na atividade de capacitação
Heterogeneidade semântica	Preços/tarifação	Diferentes acessos aos dados e políticas de privacidade	Pessoas interessadas de diferentes partes
Sistema de referência e consistência na escala			
Consistência na qualidade dos dados			
Existência e consistência dos metadados			
Consistência nos formatos dos dados			
Consistências nos modelos de dados			
Heterogeneidade atribuída			
Utilização de modelos de colaboração consistentes			
Diferenças no modelo de financiamento			
Conscientização da integração de dados.			

Fonte: adaptado de RAJABIFARD (2010); WILLIAMSON et al. (2010).

O *OGC* orienta o desenvolvimento de tecnologias que facilitam a interoperabilidade entre sistemas envolvendo informações espaciais e de localização, por meio de padrões. O *GSDI* promove cooperação internacional com a finalidade de apoiar o desenvolvimento de infraestruturas de dados espaciais em nível local, nacional e internacional, que permita as nações abordarem melhor suas prioridades de ordem social, econômica e ambiental.

2.2.3 Integração de dados no Brasil

No Brasil existem algumas ações sendo desenvolvidas com o objetivo de integrar dados e permitir a interoperabilidade entre SIs. Entre essas ações estão o Modelo Global de Dados (MGD) e a Infraestrutura Nacional de Dados Espaciais (INDE).

2.2.3.1 Modelo Global de Dados

O MGD é um modelo de dados e processos em alto nível, que tem como objetivo mapear as informações armazenadas nos sistemas do Governo Federal, considerando informações que permitam o entendimento do negócio, sem considerar detalhes operacionais e tecnológicos. Este modelo foi adotado pelo *e-Ping*¹⁴ e definido como arquitetura de interoperabilidade, possibilitando o compartilhamento e o reuso de informações nas soluções atuais e em desenvolvimento na Administração Pública Federal. Entende-se que “alto nível” é o mapeamento das principais informações que permite compreender o negócio sobre o ponto de vista dos dados utilizados (SERPRO, 2010b).

Apesar de ser um modelo adotado na esfera federal, serve de exemplo de iniciativa para que outras esferas do poder público comecem a pensar na integração entre sistemas. Atualmente, cada esfera do poder público trabalha isoladamente. Se os sistemas fossem integrados, a gestão administrativa do governo seria mais eficiente, ampliando a visão de contexto e possibilitando um melhor controle das informações geradas e tratadas por esses sistemas.

O MGD realiza o mapeamento dos dados levantados, registrando as informações relevantes da organização a partir da estrutura, da semântica e os dos processos que tratam desses dados.

Essa metodologia aborda as seguintes visões: de dados, apresenta um mapa com as principais entidades de negócio, onde estão guardadas as estruturas e as descrições; de negócio, mapeia os processos de negócio verificando quando cada informação é criada ou consumida; e de integração de processos, verifica-se onde um processo impacta no seguinte e quais processos podem ser integrados.

¹⁴ *e-Ping*: padrões de interoperabilidade de governo eletrônico do Brasil, formam um conjunto de premissas, políticas e especificações técnicas que regulamentam a utilização da Tecnologia da Informação e Comunicação no Governo Federal, estabelecendo as condições de interação com os demais poderes e esferas de governo com a sociedade em geral.

SERPRO (2011) detalha a metodologia e os padrões que permitem a realização de cada etapa para a definição do MGD, bem como quais são os artefatos gerados e como estes artefatos se integram.

Esse modelo é composto por sete produtos: Diagrama de Entidade Relacionamento (DER) em alto nível, apresentando conceitualmente as principais entidades do macroprocesso; Diagrama de Contexto; Matriz Entidade x Área de Negócio, Dicionário de Dados, Diagrama e documento de descrição de processos, Matriz de entidade x processo e Matriz de entidade x Gestor da Informação. O DER é o principal produto do MGD e fundamental na definição de requisitos das novas demandas. Os demais produtos são ferramentas auxiliares à análise de dados pelos usuários do modelo (SERPRO, 2010b; SERPRO, 2010a).

Há de se ressaltar que o Dicionário de Dados deve ser único, permitindo a centralização das informações diversificadas sobre um mesmo objeto e, com isso, estabelecendo uma semântica precisa para as informações do macroprocesso (SERPRO, 2011). O uso do Dicionário de Dados evita a ocorrência de ambiguidade: falta de clareza, falta de precisão, incerteza, dúvida, dualidade de interpretação, formação de conceitos equivocados etc. Trata-se de uma ferramenta essencialmente textual que define o significado de todas as entidades presentes no DER (SERPRO, 2009).

O MGD, sucessivamente refinado, serve como ferramenta auxiliar no processo geral de melhoria da qualidade da gestão pública, como elemento que permite uma visibilidade global e integrada da base de dados e informações ao gestor público. Também permite aos vários atores envolvidos num macroprocesso, tais como clientes, gestores públicos, analistas de negócios, analistas de sistemas, administradores de dados e administradores de banco de dados uma visão do todo. Com isso o conhecimento da organização é ampliado e pode ser disseminado entre as pessoas (SERPRO, 2009).

A gestão eficaz desse modelo permite a visibilidade e seu uso, pelas diversas áreas diretamente envolvidas, permitindo buscar a melhoria contínua do modelo, propiciando um ambiente colaborativo, compartilhado e integrado (SERPRO, 2009). É uma busca incessante em integrar dados sem duplicidades e com registro de procedência válido.

2.2.3.2 Infraestrutura Nacional de Dados Espaciais

A INDE, de abrangência nacional, foi estabelecida pelo Decreto Lei nº 6.666 de 27 de novembro de 2008 (CONCAR, 2009), que a define como um “conjunto integrado de tecnologias; políticas; mecanismos e procedimentos de coordenação e monitoramento; padrões e acordos, necessário para facilitar e ordenar a geração, o armazenamento, o acesso, o

compartilhamento, a disseminação e o uso dos dados geoespaciais de origem federal, estadual, distrital e municipal.” Coube à Fundação Instituto Brasileiro de Geografia e Estatística (IBGE) o início da implantação da INDE em 2010.

A INDE surgiu baseada em iniciativas internacionais, que buscaram a integração de dados geoespaciais, por meio de instituições que têm trabalhado no desenvolvimento de padrões e na implantação de Infraestruturas de Dados Espaciais. Como exemplo, as iniciativas *INSPIRE (Infrastructure for SPatial InfoRmation in Europe)*, *NSDI (National Spatial Data Infrastructure - Estados Unidos da América)*, *SNIG (Sistema Nacional de Informação Geográfica - Portugal)* e *IDEMex (Infraestrutura de Dados Espaciais do México)* (CONCAR, 2009).

A Infraestrutura de Dados Espaciais é um conjunto de dados espaciais fundamentais, com normas que lhes permitam ser integrados a rede de distribuição, com políticas e princípios administrativos que garantam a compatibilidade entre as jurisdições, agências e as pessoas, começando no nível local e passando para o estadual, nacional, regional até o nível global (WILLIAMSON et al., 2010).

Para a criação de uma Infraestrutura de Dados Espaciais deve-se considerar as características da região abrangida. Como pode ser visto na Figura 14, as estruturas organizacionais para gestão territorial precisam ter uma configuração que considere as constantes mudanças da cultura local, fatos jurídicos e acordos institucionais para apoiar a prática de políticas agrárias e de gestão, facilitando as funções da administração territorial. Dentro de cada país, as atividades de gerenciamento territorial, necessárias para apoiar o desenvolvimento sustentável, podem ser descritas por três componentes: política fundiária; infraestrutura de informação territorial; e as funções da administração territorial (WILLIAMSON et al., 2010).

A estrutura administrativa, para o compartilhamento dos dados, possui uma hierarquia composta de Infraestruturas de Dados Espaciais interligadas, em vários níveis, como ilustrado na Figura 15. Em geral, os vários níveis são uma função de escala. A administração local e estadual, da Infraestrutura de Dados Espaciais, gerenciam dados de média e grande escala, deixando a infraestrutura nacional para gerenciar dados de média e de pequena escala, com a infraestrutura regional e global adotando uma escala pequena para suas atividades (WILLIAMSON et al., 2010).

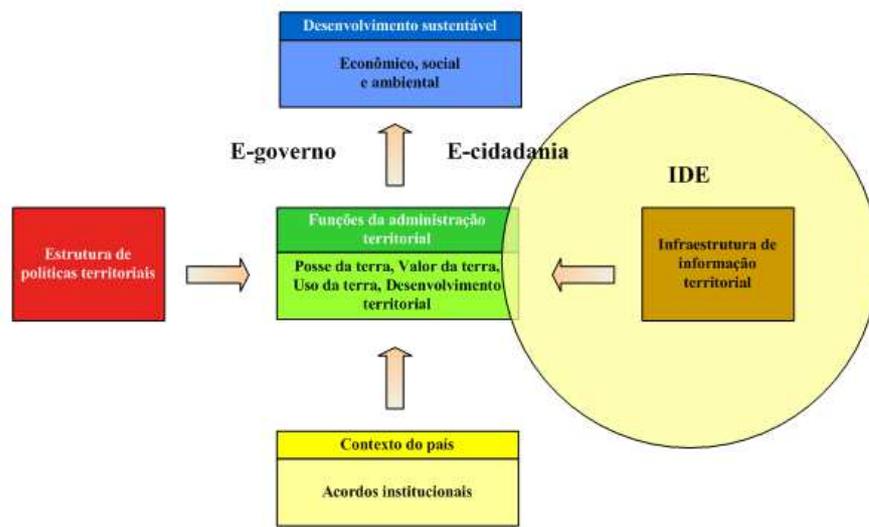


Figura 14: Estruturas organizacionais para gestão territorial.
Fonte: Adaptado de WILLIAMSON et al. (2010).

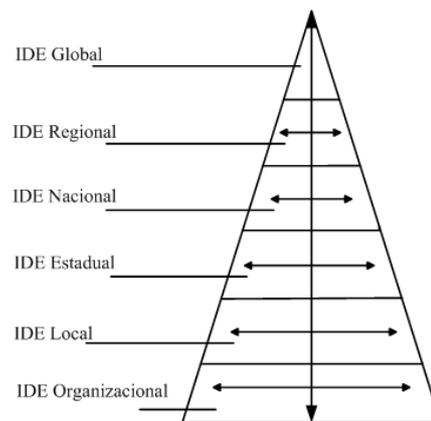


Figura 15: A hierarquia da Infraestrutura de Dados Espaciais.
Fonte: Adaptado de WILLIAMSON et al. (2010).

Groot (1997) identifica dois propósitos da Infraestrutura de Dados Espaciais: poupar tempo, esforço e dinheiro no acesso aos dados espaciais e utilizá-lo com responsabilidade; e, evitar a duplicação desnecessária dos dados espaciais pelo compartilhamento dos dados disponíveis.

Uma Infraestrutura de Dados Espaciais é formada por vários componentes fortemente relacionados. São eles, como mostra a Figura 16 (WARNEST, 2005; CINDE, 2010):

- dados – são os componentes centrais da Infraestrutura de Dados Espaciais e podem ser encontrados em diversos tipos de fontes, podem ser de três categorias: de referência, temático e de valor agregado;

- pessoas – são os responsáveis por coletar, manter e distribuir os dados espaciais. São responsáveis pelo treinamento e pesquisa em Infraestrutura de Dados Espaciais, e também pela determinação de quais dados são necessários e como serão acessados;
- instituições – são responsáveis pela legislação, coordenação e criação de políticas para a Infraestrutura de Dados Espaciais;
- tecnologia – descreve as necessidades físicas e de base da Infraestrutura de Dados Espaciais que permite buscar, consultar, encontrar, acessar, prover e usar os dados geoespaciais, ajudando a manter, processar, disseminar e acessar os dados espaciais;
- normas e padrões – permite a descoberta, troca, integração e usabilidade da informação espacial. São compostos por sistemas de referência, modelos de dados, dicionários de dados, qualidade dos dados, transferência de dados e metadados;

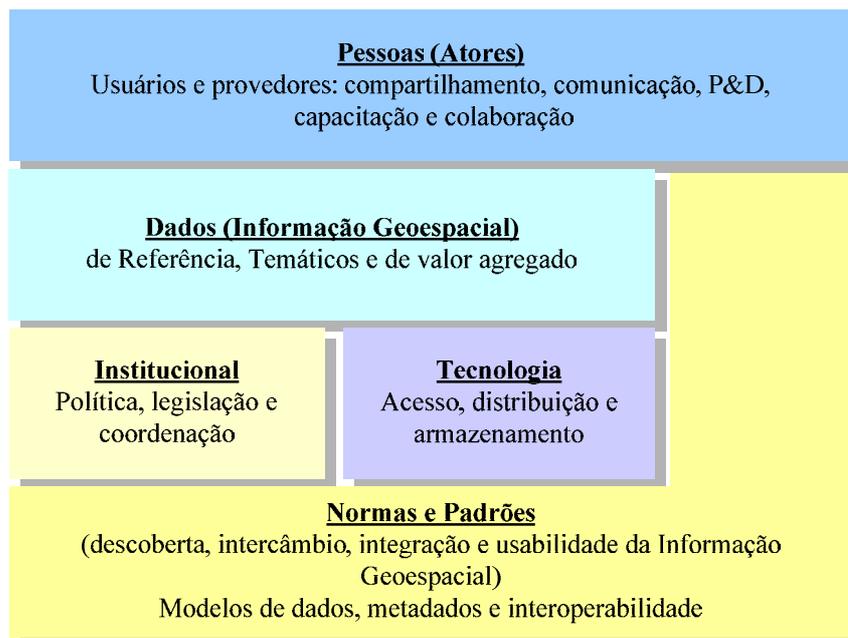


Figura 16: Componentes de uma Infraestrutura de Dados Espaciais.
Fonte: CINDE (2010).

O componente “Pessoas”, fundamental para o sucesso da Infraestrutura de Dados Espaciais, é dividido por responsabilidades: setor público e o setor privado são responsáveis pela aquisição, produção, manutenção e oferta de dados espaciais; o setor acadêmico é responsável pela educação, capacitação, treinamento e pesquisa em Infraestrutura de Dados Espaciais; e o usuário determina quais dados espaciais são requeridos e como devem ser acessados (CINDE, 2010). Ressalta-se essa distribuição, por identificar a interação necessária

para a implantação com sucesso de uma Infraestrutura de Dados Espaciais, o trabalho colaborativo entre as partes é fundamental.

2.3 Banco de Dados

Os dados representam fatos do mundo real que podem ser armazenados e possuem um significado próprio. Um banco de dados pode ser definido como “uma coleção de dados relacionados” (ELMASRI; NAVATHE, 2005), “uma coleção de dados persistentes utilizados pelos sistemas de aplicação” (DATE, 2000) ou como “uma estrutura computacional compartilhada e integrada que armazena um conjunto de: dados do usuário final [...] e metadados [...]” (ROB; CORONEL, 2011).

Os bancos de dados tradicionais armazenam dados alfanuméricos, são comumente chamados de banco de dados convencionais (exemplo: inteiro, data, reais e texto), quando que outros tipos de dados são armazenados, como dados geográficos e dados temporais, os bancos de dados chamados não convencionais. Para o armazenamento desses diferentes tipos de dados, espaciais e temporais, novas tecnologias para banco de dados foram desenvolvidas nos últimos anos, levando em consideração as suas peculiaridades.

O elemento fundamental dos bancos de dados são os dados, como já descrito, podem ser convencionais ou não convencionais. Os dados convencionais já fazem parte do dia-a-dia das organizações e são bem conhecidos, mas, os dados não convencionais ainda são objetos de muitas pesquisas. A seguir a descrição dos dois tipos de dados não convencionais envolvidos neste trabalho, dados temporais e dados espaciais, necessárias para compreender como são manipulados pelos SGBDs.

2.3.1 Dados temporais

Os dados convencionais representam os eventos do mundo real, logo os dados temporais representam os eventos do mundo real marcados com o tempo. Para definir quais eventos do mundo real serão registrados com o tempo de ocorrência é preciso fazer algumas considerações sobre o tempo. Essas considerações partem da definição de várias características temporais: dimensão temporal, ordem no tempo, tempo absoluto e tempo relativo, granularidade no tempo, variação temporal, instante no tempo, intervalo de tempo, elemento temporal e duração temporal. Esses conceitos básicos, sobre o dado temporal, foram reunidos no *Glossary of Temporal Database* (Glossário de Banco de Dados Temporal) (JENSEN et al., 1998), que representa a intenção de unificação dos conceitos ligados aos BDTs. Estas características serão descritas a seguir.

Os modelos de dados temporais possuem uma dimensão (dimensão temporal) a mais que os modelos de dados convencionais. Os modelos convencionais possuem duas dimensões, as linhas (instância de entidade) e as colunas (atributos) de uma tabela. Cada atributo, em um modelo de dados convencionais, de uma instância apresenta um único valor, se o valor é alterado o anterior é perdido.

A terceira dimensão¹⁵ associa alguma informação temporal a cada valor. Assim, quando um valor é alterado o valor anterior não é removido do banco de dados. Um tempo de validade é associada ao valor, quando o valor for alterado um novo tempo de validade é criado, desse modo, torna-se possível acessar toda a história dos atributos.

A Figura 17 mostra um exemplo da terceira dimensão, apresentando o histórico da escolaridade e da renda mensal (em salários mínimos) ligado às parcelas, com dois registros (id=212 e id=213), e cada um mostrando o histórico, com três registros temporais.

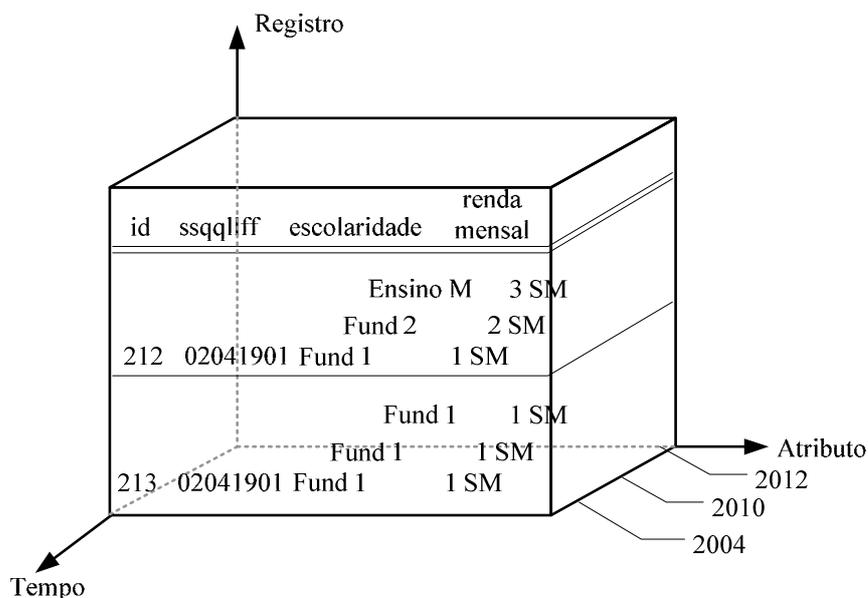


Figura 17: Relação temporal em 3-dimensões.

Outra característica é a **ordem** a ser seguida no tempo, no eixo temporal, os valores definidos para o tempo de início e o tempo de término devem estar em ordem cronológica, existem três tipos de ordem (EDELWEISS, 1998) (Figura 18):

- ordem linear: é o tipo mais comum, onde um ponto no tempo terá somente um antecessor e um predecessor;
- ordem ramificada: esse tipo permite que um ponto no tempo tenha mais de um antecessor e mais de um predecessor;

¹⁵ Terceira dimensão: a terceira dimensão nos modelos de dados temporais é diferente da terceira dimensão (3D) do Cadastro, esta se refere à representação gráfica tridimensional da parcela cadastral.

- ordem circular: modela processos recorrentes, como datas comemorativas, mas não é o tempo que se repete e sim os eventos.

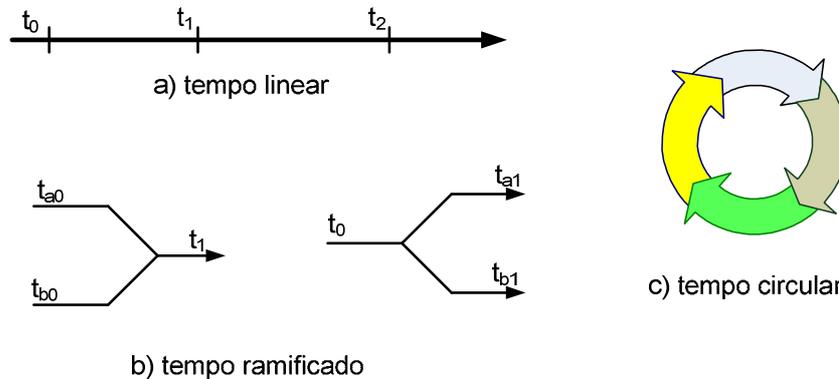


Figura 18: Tipos de ordem temporal.

Para o modelo de dados temporais é preciso diferenciar o **tempo absoluto** do **relativo**. O tempo absoluto indica que um tempo válido considera a granularidade associada com o fato, por exemplo, Antônio comprou o imóvel no dia 13/04/2000. O tempo relativo indica que o tempo válido de um fato está relacionado há algum outro tempo, por exemplo, o imposto subiu no mês passado (JENSEN et al., 1998).

A característica **granularidade** para o tempo de transação pode ser única para todo o banco de dados, sendo definida no momento da criação do banco de dados. Para o tempo de validade, a granularidade pode ser diferente para os objetos criados, permitindo uma melhor representação do mundo real.

Outra característica é a **variação temporal**. Existem duas formas de variação temporal, tempo contínuo e tempo discreto. O tempo é contínuo por natureza, entre dois instantes consecutivos temporais pode existir outro instante temporal, o modelo contínuo é semelhante aos números reais. No caso do tempo discreto, uma representação simplificada, baseando-se em uma linha de tempo composta de uma sequência de intervalos temporais consecutivos que não podem ser decompostos, de mesma duração. Os intervalos são definidos como *chronon*, sua duração não precisa ser fixada no modelo de dados, pode ser definida na implementação (EDELWEISS, 1998).

Date (2000) usa o termo *quanta* no lugar do *chronon*. Para ele um *quanta* é a menor medida de tempo que o sistema pode representar, pode ser de diferentes medidas de tempo, que chegam até frações minúsculas de um segundo. Entretanto, as unidades de tempo úteis em um sistema podem ser anos, meses, dias ou semanas, essas unidades são chamadas de ponto de tempo, e também são indivisíveis. Um ponto no tempo seria, então, uma seção da linha de

tempo, ou seja, um conjunto de *quanta* de tempo que vai de um *quantum* de tempo limite até o seguinte (exemplo, das oito horas de um dia até às oito horas do dia seguinte).

Em um modelo de tempo discreto a granularidade pode ser do tamanho de um *chronon*, especifica a unidade de medida. Todavia, para algumas aplicações podem ser necessário considerar diferentes granularidades (minutos, dias, anos) para permitir uma representação melhor da realidade.

A característica **instante temporal** é representada por um *chronon* no eixo temporal do modelo, para o tempo discreto. Na variação discreta, nenhum outro ponto existe entre dois pontos consecutivos. Existe um ponto denominado instante atual (*now*), na ordenação linear, que se movimenta constantemente ao longo do tempo, todo ponto anterior a ele é considerado passado e todo ponto posterior é considerado futuro.

Outra característica é o **intervalo de tempo** que possui tempo de início (*s*) e tempo de término (*e*), isso define o conjunto de todos os tempos (*t*), tal que $s \leq t \leq e$ (DATE, 2000). A linha de tempo, na qual o intervalo existe, consiste em uma sequência finita de *quanta* (ou *chronon*) de tempos discretos e indivisíveis. Sendo assim, um intervalo de tempo envolve um número finito de *quanta*.

O intervalo tem um ponto inicial e um final, portanto, deve ser tratado como pares de valores, e não separadamente. Para a definição dos intervalos são usados colchetes, [], ou parênteses, (), que representam se o ponto inicial ou final estão contidos no intervalo. Os colchetes representam intervalos fechados, os parênteses representam intervalos aberto. Por exemplo, um intervalo que vai do dia 12 ao dia 21, se for representado como [12,21] quer dizer que o intervalo começa no dia 12, incluindo o dia 12 e vai até o dia 21, inclusive. Outro exemplo, [12,21), quer dizer que o intervalo começa no dia 12, incluindo o dia 12 e vai até o dia 21, excluindo o dia 21 (DATE, 2000).

A característica **elemento temporal** representa a união finita de intervalos de tempo. Os casos especiais de elementos temporais incluem elementos de tempo de validade, elementos de tempo de transação, e elementos bitemporais. Eles são uniões finitas de intervalos de tempo válido, intervalos de tempo de transação, e intervalos bitemporais, respectivamente (JENSEN et al., 1998).

No elemento temporal as operações de união, intersecção e complemento da teoria dos conjuntos produzem um novo elemento temporal. O intervalo e o instante temporal são elementos temporais (EDELWEISS, 1998).

A característica **duração temporal** depende do contexto, pode ser fixa ou variável. A duração fixa independente do contexto, por exemplo, um minuto tem 60 segundos. Porém, a

duração variável depende do contexto, como, um mês pode ter 28, 29, 30 ou 31 dias (EDELWEISS, 1998).

Date (2000) faz um relato de todas as dificuldades que podem ser encontradas no projeto de um BDT, ressalta como o tratamento dos intervalos de tempo pode dificultar a definição de restrições e consultas. Discute também vários operadores usados para os dados de intervalos.

2.3.2 Dados espaciais

Outro tipo de dados não convencionais, citado anteriormente, é o dado espacial. O espaço representa a superfície terrestre denominado de espaço geográfico, conceituado como o meio físico onde as entidades geográficas coexistem. Essas entidades geográficas são quaisquer entidades identificáveis do mundo real, possuindo características espaciais e relacionamentos espaciais com outras entidades geográficas (GATRELL, 1991 apud SCHULTZ, 2007). O espaço geográfico é composto por dados geográficos e informação geoespacial que se diferenciam de outros dados pela componente espacial, ou seja, associam a cada entidade ou fenômeno uma localização na Terra em um dado instante ou período de tempo.

Muitos termos são usados para nomear os dados geográficos, segundo o CINDE (2010) **dados espaciais** são quaisquer tipos de dados que descrevem fenômenos aos quais estejam associada alguma dimensão espacial, e **dados geográficos, geoespaciais** ou **georreferenciados** são dados espaciais em que a dimensão espacial refere-se ao seu posicionamento na Terra e no espaço próximo, num determinado instante ou período de tempo.

Esses dados geográficos podem ser categorizados em dados de raster (Figura 19) e dados vetoriais (Figura 20). Dados de raster consistem em mapas de bits ou mapas de pixels em duas ou mais dimensões. Dados vetoriais são construídos a partir de objetos geométricos básicos, como pontos, segmentos de linha, triângulos e outros polígonos bidimensionais (SILBERSCHATZ; KORTH; SUDARSHAN, 2006).

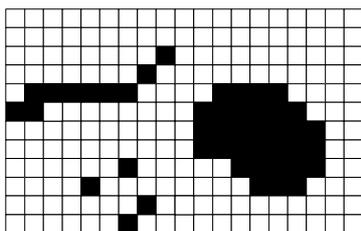


Figura 19: Representações de raster em duas dimensões.

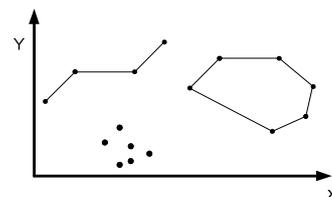


Figura 20: Representações vetoriais em duas dimensões.

Para representar os dados geográficos no computador é preciso diferenciar o espaço absoluto e o espaço relativo. “Espaço absoluto, também chamado cartesiano, é um container de coisas e eventos, uma estrutura para localizar pontos, trajetórias e objetos. Espaço relativo, ou leibnitziano, é o espaço constituído pelas relações espaciais entre coisas” (COUCLELIS, 1997¹⁶, apud CÂMARA, 2005).

Segundo Câmara (2005), quando a modelagem dos fenômenos geográficos é feita, precisa-se definir qual espaço será utilizado, o absoluto ou o relativo. Para o espaço absoluto, mapa à esquerda na Figura 21, existem dois modelos formais para entidades geográficas, o geo campos e o geo-objetos. O modelo de geo campos define o espaço geográfico como uma superfície contínua, sobre a qual variam os fenômenos a serem observados, como um mapa de cobertura com pontos identificando a cobertura vegetal. O modelo de geo-objetos representa o espaço geográfico como uma coleção de entidades distintas e identificáveis, onde cada entidade é definida por uma fronteira fechada, um lote do CTM, por exemplo. Para o espaço relativo, mapa à direita na Figura 21, existe o modelo de rede que define o espaço geográfico como um conjunto de pontos no espaço, conectados por linhas e ambos possuem atributos. Na Figura 21 à esquerda, distritos de São Paulo com suas fronteiras, à direita, grafo mostrando a rede de conectividade entre os distritos (espaço relativo) (CÂMARA, 2005).

Em SIG, descrito na Subseção 2.1.1, os dados geográficos são armazenados em Banco de Dados Geográficos (BDG), seus objetos primitivos são do tipo ponto, linha, polígono e superfície, dados vetoriais. Esses tipos de dados são usados para representar diferentes entidades do mundo real (LISBOA FILHO, 2001). Os mesmos elementos utilizados no CTM para a representação cartográfica.

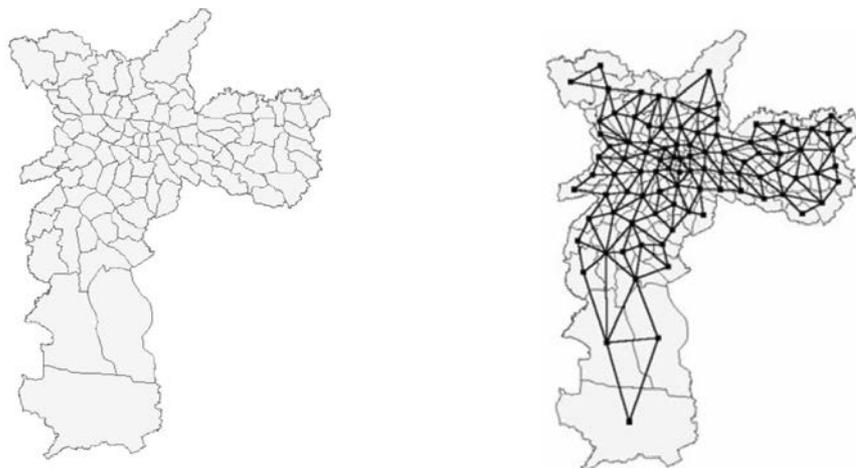


Figura 21: Representação do espaço absoluto e relativo.
Fonte: CÂMARA (2005).

¹⁶ COUCLELIS, H. From cellular automata to urban models: news principles for model development and implementation. **Environment and Planning B: Planning and Design**, v. 24, p. 165-174, 1997.

Entre esses objetos primitivos, os do tipo **ponto** são usados para representar entidades que não possuem dimensões significativas, de acordo com a escala em uso. Exemplos dessas entidades são: postes de iluminação pública, hidrantes e pontos de ônibus, para mapas de grande escala (1:5.000); escolas, hospitais, postos de saúde e rodoviária, para mapas de escala menores (1:20.000); cidades, para mapas em pequenas escalas (1:1.000.000).

O segundo desses objetos, são os do tipo **linha** representam entidades que possuem uma distribuição espacial linear, como: ruas, rodovias, cabos telefônicos e rios. Juntando linha com ponto cria-se uma estrutura de representação em rede, usadas para aplicações de malha viária, instalações de energia elétrica e telefone.

O último tipo dos objetos primitivos, os do tipo **polígono** permitem a representação bidimensional, é uma região limitada por uma ou mais linhas poligonais conectadas de tal forma que o último ponto de uma linha seja idêntico ao primeiro da próxima. Ele divide o plano em duas regiões: interna e externa. São usados para representar unidades espaciais individuais como, municípios, quadras, lotes e setores censitários.

2.3.3 Dados espaço-temporais

Unindo esses dois tipos de dados não convencionais, forma-se o conceito de dados espaço-temporais. Muitas aplicações de tecnologias de geoinformação utilizam representações estáticas de fenômenos espaciais, não permitindo visualizar as transformações ocorridas com o passar do tempo. Dessa forma, modelos espaço-temporais procuram representar adequadamente fenômenos que variam tanto no espaço como no tempo. Os modelos espaço-temporais possuem dois aspectos distintos: (1) os conceitos de espaço e tempo; e (2) a construção de representações computacionais apropriadas correspondentes a esses conceitos (DIAS; CÂMARA; DAVIS Jr, 2005).

Worboys (1995) apresenta um modelo semântico para o tratamento de mudanças espaço-temporais, com os mesmos elementos que compõem o SIT, descritos na subseção 2.1.1. Nesse modelo a informação geográfica possui três componentes: atributo, espaço e tempo (Figura 22), que possibilita responder as seguintes questões: O que? Onde? e Quando? Cada um desses componentes determina uma categoria de dimensão ao longo da qual os valores são medidos.

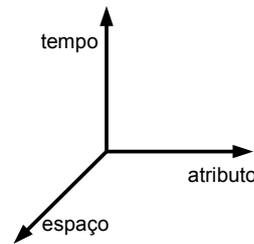


Figura 22: Categorias de dimensões da informação geográfica.
Fonte: LISBOA FILHO (2001).

O **componente atributo** apresenta características qualitativas e quantitativas descritas de forma textual e/ou numérica. O **componente espaço** define a localização geográfica, a forma geométrica do fenômeno descrito pela informação geográfica e o relacionamento com outros fenômenos geográficos. O **componente tempo** está associado a um instante ou intervalo de tempo em que ocorre ou é observado um evento (LISBOA FILHO, 2001).

Para analisar dados espaço-temporais é preciso identificar o momento em que o evento ocorreu, conforme a representação adotada (tempo de validade), e o momento em que essa ocorrência foi registrada no banco de dados (tempo de transação), que indica a partir de quando a informação correspondente ao evento se tornou disponível para o usuário (DIAS; CÂMARA; DAVIS Jr, 2005).

2.3.4 Banco de dados espaço-temporais

A necessidade de representar a posição geográfica dos dados, sua evolução temporal e a constatação de que muitas vezes é preciso guardar o histórico desta evolução levou à criação do conceito de Banco de Dados Espaço-temporais (BDET). Eles permitem armazenar todos os estados de uma aplicação (presentes, passados e futuros) registrando sua evolução com o passar do tempo. Informações temporais são associadas aos dados, geográficos e não geográficos, e armazenadas com tempo de transação e/ou tempo de validade para identificá-los ao longo do tempo.

Assim, um BDET incorpora dados espaciais, temporais e espaço-temporais, podendo capturar simultaneamente os dados espaciais e temporais. Todos os conceitos espaciais e temporais devem ser considerados, no entanto, devem ser observadas as questões de interseção entre as duas classes de conceitos, denominada classe de conceito espaço-temporal.

O BDET representa a junção dos conceitos do BDG com o BDT. Um sistema que utiliza o BDET pode descobrir se existe algum padrão temporal, quais tendências são visíveis

e qual é o processo responsável pelas mudanças. Os tipos de consultas espaço-temporais de acordo com o domínio de busca de consulta são (CASTRO, 2007):

1. Consulta temporal simples: “Qual é o estado do objeto A no tempo T?” (Figura 23 a);
2. Consulta temporal de intervalo: “O que aconteceu com o objeto A durante um intervalo T1 a T2?” (Figura 23 b);
3. Consulta espaço-temporal simples: “Qual é o estado da região R no tempo T?” (Figura 23 c);
4. Consulta espaço-temporal de intervalo: “O que aconteceu com a região R no intervalo T1 a T2?” (Figura 23 d).

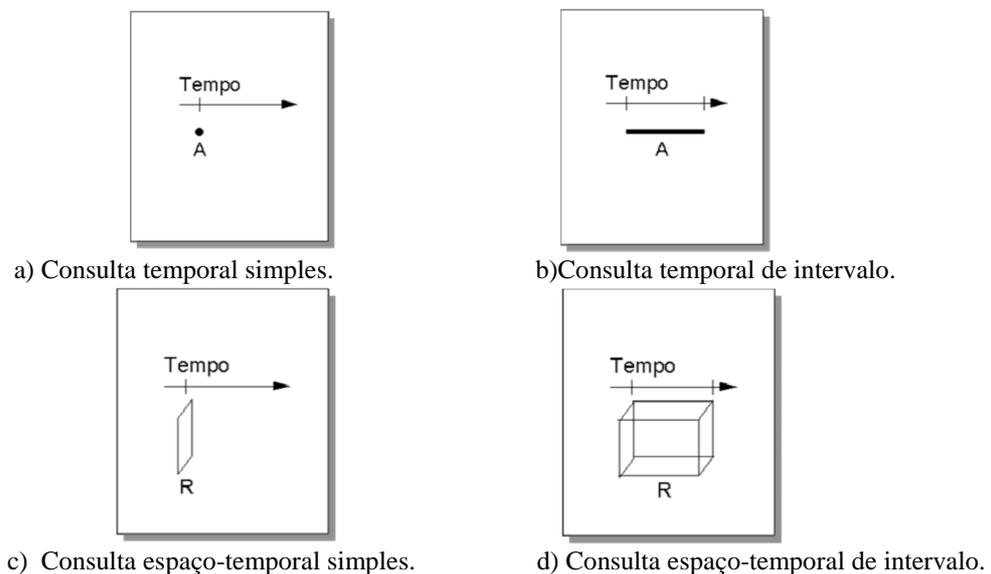


Figura 23: Tipos de consultas espaço-temporais.

Fonte: Castro (2007).

Para Medeiros e Jomier (1993) um SIG acrescido do componente tempo permite: prever o futuro baseado em registros do passado; analisar a evolução temporal dos dados; analisar comparativamente cenários simulados; e analisar de forma comparativa os dados atuais e cenários simulados.

2.4 Data Warehouse

“O DW é um tipo especial de banco de dados.” Surgiu para atender a necessidade de fornecer uma origem de dados única, limpa e consistente para apoio à decisão e de fazê-lo sem impactar os sistemas transacionais (DATE, 2000). São otimizados para recuperar dados e não processar transações, conseqüentemente, são diferentes dos bancos de dados convencionais (operacionais) em estrutura, funcionalidade, desempenho e propósito (ELMASRI; NAVATHE, 2005) como mostra o Quadro 2.

Quadro 2: Diferenças entre os Bancos de Dados Transacionais e os *DWs*.

Aspecto	Banco de Dados Operacionais	<i>DWs</i>
Usuário	Projetista de sistemas, administrador de sistemas, usuários de entrada de dados.	Responsável por tomar decisões, analistas, executivos.
Funções	Operações diárias, Processamento de transações (<i>on-line</i>).	Orientado a assunto.
Projeto do banco de dados	Orientado a aplicação.	Orientado a assunto.
Dados	Corrente, atualização atômica, Relacional (normalizado), isolado.	Histórico, resumido, multidimensional, integrado.
Uso	Repetitivo, rotina.	<i>Ad hoc</i> ¹⁷ .
Acesso	Simples transação de leitura/gravação.	Leitura mais complexa de consultas.
Requisitos de sistema	Processamento de transação, consistência de dados.	Processamento de consultas, precisão de dados.

Fonte: WU; BUCHMANN (1997).

Inmon (2005) define o *DW* como “um conjunto de dados baseado em assuntos, integrado, não volátil e variável em relação ao tempo para apoio às decisões gerenciais”.

O *DW* é orientado por assunto por tratar de assuntos de interesses específicos, organizados e resumidos por temas, como vendas, *marketing*, finanças, distribuição e transporte. Os projetistas preocupam-se com os dados e não com os processos que os modificam. Por exemplo, ao invés de armazenar uma fatura, o *DW* armazena os componentes de “venda por produto” e “vendas por cliente”, pois nas tomadas de decisões precisa-se recuperar resumos de vendas por produto ou cliente (ROB; CORONEL, 2011; MACHADO, 2000).

Os gestores buscam a visão da organização como um todo, apesar de todo o esforço para a obtenção desta visão, tratar essas informações ao mesmo tempo é uma atividade muito complexa. Assim, para uma melhor tomada de decisão, as informações devem estar organizadas por assuntos, durante o desenvolvimento busca-se a identificação desses assuntos como forma de facilitar o uso do sistema no futuro.

Outra característica do *DW* é ser integrado, pois integra dados recebidos de toda a organização e de várias fontes, com formatos diferentes. A integração significa que todos os processos e dados da organização estão descritos do mesmo modo, assumindo um formato comum e aceito pela organização.

Ele também é não volátil, ou seja, uma vez inserido um dado no *DW*, esse não será removido, são sempre inseridos, pois representam o histórico da organização. Com isso, o volume de dados está sempre crescendo, assim o SGBD dever ser capaz de armazenar um grande volume de dados. Como última característica, o *DW* é variável no tempo. Os dados

¹⁷ *ad hoc*: locução latina que significa "para isso", neste caso relatórios feito para um determinado propósito, não previsto anteriormente.

operacionais focam nas transações correntes, ao contrário, no *DW* os dados representam o fluxo através do tempo.

Além do termo *DW*, na literatura é comum encontrar o termo *Data Warehousing*, que é o processo de obtenção de dados de sistemas legados¹⁸ e banco de dados transacionais para transformá-los em informação organizada e em um formato amigável para incentivar a análise de dados e suporte à tomada de decisão de negócios baseada em fatos (KIMBALL; CASERTA, 2004).

O *Data Warehousing* engloba arquiteturas, algoritmos e ferramentas para reunir os dados selecionados a partir de múltiplos bancos de dados ou outras fontes de informação em um único repositório, chamado de *DW*, adequados para a consulta direta ou análise (WIDOW, 1995). Representa uma coleção de tecnologias de apoio à decisão, destinado a permitir que o profissional do conhecimento (executivo, gerente, analista) tome decisões melhores e mais rápidas (CHAUDHURI; DAYAL, 1997).

No *Data Warehousing*, os dados são extraídos de várias fontes, transformados e integrados antes de serem carregados no *DW* para serem disponibilizados para o usuário, Figura 24. Assim, diferentes tipos de dados têm que ser extraídos, transformados e carregados, utilizando aplicações de *Extraction, Transformation and Load (ETL - Extração, Transformação e Carga)*. As aplicações *ETL* são ferramentas desenvolvidas pela organização usuária, por terceiros ou adquirida de um fornecedor, como o *Kettle Pentaho*¹⁹ e o *SQL Server 2008 Integration Services*²⁰. Esse processo consiste na primeira etapa de obtenção de dados dos sistemas para o ambiente de *Data Warehousing*.

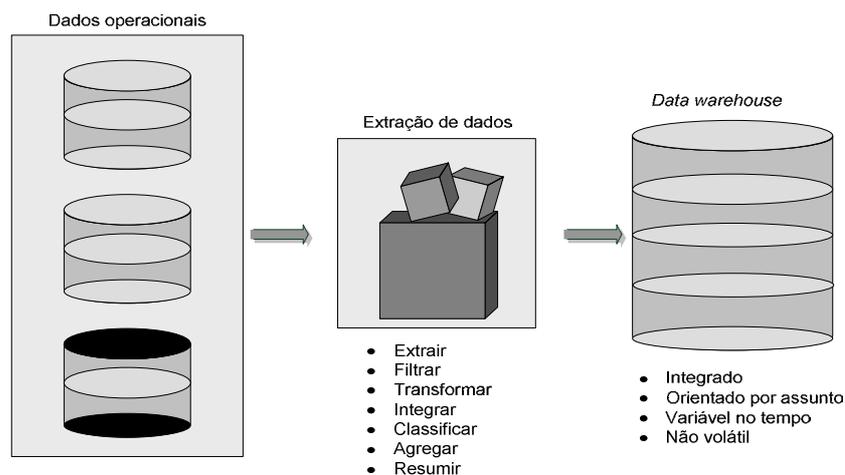


Figura 24: Fases de criação de um *DW*.
Fonte: ROB; CORONEL (2011).

¹⁸ **Sistema legado** é o termo utilizado para sistemas computacionais antigos que fornecem serviços essenciais a uma organização. Geralmente utilizam bancos de dados obsoletos.

¹⁹ <http://kettle.pentaho.com/>

²⁰ <http://technet.microsoft.com/pt-br/sqlserver/cc510302>

Inmon e Kelley²¹ (1994, apud ROB; CORONEL, 2011) criaram 12 regras que definem um *DW*:

1. os ambientes operacional e de *Data Warehousing* são separados;
2. os dados são integrados;
3. contém dados históricos por um longo tempo;
4. os dados constituem um retrato instantâneo tirado em determinado ponto do tempo;
5. os dados são orientados por assunto;
6. os dados são essencialmente para leitura, com atualizações periódicas em *batch*²² dos dados operacionais, não sendo permitidas atualizações *on-line*;
7. o ciclo de vida do desenvolvimento do *Data Warehousing* difere do desenvolvimento clássico de sistemas. O desenvolvimento do *Data Warehousing* é orientado para os dados, e a abordagem clássica para os processos;
8. contém dados com vários níveis de detalhes: dados atuais em detalhes, dados antigos em detalhes, dados levemente resumidos e dados altamente resumidos;
9. o ambiente de *Data Warehousing* é caracterizado por operações de leitura para conjuntos de dados muito grandes;
10. o ambiente de *Data Warehousing* possui um sistema que rastreia fontes, transformações e armazenamento;
11. o metadado é um componente fundamental desse ambiente. Ele identifica e define todos os elementos de dados, fornecendo a fonte, a transformação, a integração, o armazenamento, a utilização, o relacionamentos e o histórico de cada elemento dos dados;
12. contém um mecanismo de retorno da utilização de recursos que leva à aplicação ideal dos dados pelos usuários finais.

Os *DWs* mantêm a história da organização disponíveis e acessíveis para consultas e análises. Essas consultas são possíveis por meio de ferramentas de *On-Line Analytical Processing (OLAP- Processamento Analítico On-Line)* que permitem analisar os dados, descobrir informações e disponibilizar a visualização dos dados sob diversas perspectivas, com capacidade de navegar nos detalhes das informações (MACHADO, 2000).

²¹ INMOM, B; KELLEY, C. **The twelve rules of data warehouse for a client/server world.** Data Management Review, 4, 5, maio de 1994, p. 6-16

²² *Batch* ou arquivo de lote (.bat) é um arquivo de computador utilizado para automatizar tarefas. O *batch* neste caso, é que um conjunto de comandos que são rodados sequencialmente.

O termo *OLAP*, segundo Date (2000) foi definido em um *white paper* escrito para a Arbor Software Corp em 1993, como “o processo interativo de criar, administrar, analisar e gerar relatórios sobre dados”. Para Rob e Coronel (2011) as ferramentas *OLAP* permitem criar um ambiente avançado de análise de dados que dá apoio à tomada de decisões, modelagem comercial²³ e pesquisa operacional²⁴, com as seguintes características: utilizam técnicas de análise de dados multidimensionais; fornecem suporte avançado a banco de dados; têm interfaces fáceis de utilizar; e dão suporte à arquitetura cliente/servidor.

O *DW* ou *Data Warehousing* é um dos componentes do *BI*. Para Rob e Coronel (2011) o *BI* é um amplo, coeso e integrado conjunto de ferramentas e processos usados para coletar, integrar, armazenar e analisar dados para a geração e apresentação de informações que deem apoio à tomada de decisões em organizações. Ele não é simplesmente uma ferramenta, mas um modelo de conceitos, práticas, ferramentas e tecnologias que ajudam uma organização a compreender melhor seus recursos centrais, fornecem um retrato “instantâneo” da situação e identificam oportunidades.

2.4.1 Arquitetura de *Data Warehouse*

Abstratamente, arquitetura de *software* envolve a descrição de elementos a partir dos quais o sistema é construído, as interações entre esses elementos, padrões que guiam sua composição e as restrições desses padrões. Em geral, um determinado sistema é definido em termos de uma coleção de componentes e a interação entre estes componentes (SHAW; GARLAN, 1996; BASS; CLEMENTS; KAZMAN, 1998).

A partir desse conceito, a arquitetura de um *Data Warehousing* deve demonstrar seus componentes e a organização entre eles. A definição da arquitetura é uma tarefa importantíssima, devido à grande dependência existente entre a implementação dos componentes e sua organização. A seguir, estão descritas as duas arquiteturas mais comuns para *Data Warehousing*: *OLAP* multidimensional (*Multidimensional On-Line Analytical Processing – MOLAP*) e *OLAP* relacional (*Relational On-Line Analytical Processing – ROLAP*).

Os sistemas *MOLAP* utilizam Banco de Dados Multidimensionais, para armazenar o *DW* e as aplicações analíticas são construídas no *front end*²⁵, como mostra a Figura 25. O Banco de Dados Multidimensionais serve as camadas de armazenamento de dados e a camada

²³ Modelagem comercial: analisa a estrutura da organização, seus fluxos de trabalho e principais processos, o grau de eficiência dos mesmos e a existências de gargalos nos processos.

²⁴ Pesquisa operacional: faz uso de modelos matemáticos, estatísticos e de algoritmos no apoio à tomada de decisões.

²⁵ *Front end*: camada de apresentação dos dados.

de gerenciamento de dados. Os dados originais são adaptados para o modelo multidimensional, resumidos em todas as dimensões e pré-computados para melhorar o desempenho. O maior problema dessa arquitetura é a falta de escalabilidade, a vantagem de conseguir alto desempenho com a pré-geração de todos os cálculos no momento da criação dos cubos, faz com que os sistemas *MOLAP* sejam limitados a uma pouca quantidade de dados (WU; BUCHMANN, 1997; MALINOWSKI; ZIMÁNYI, 2008).

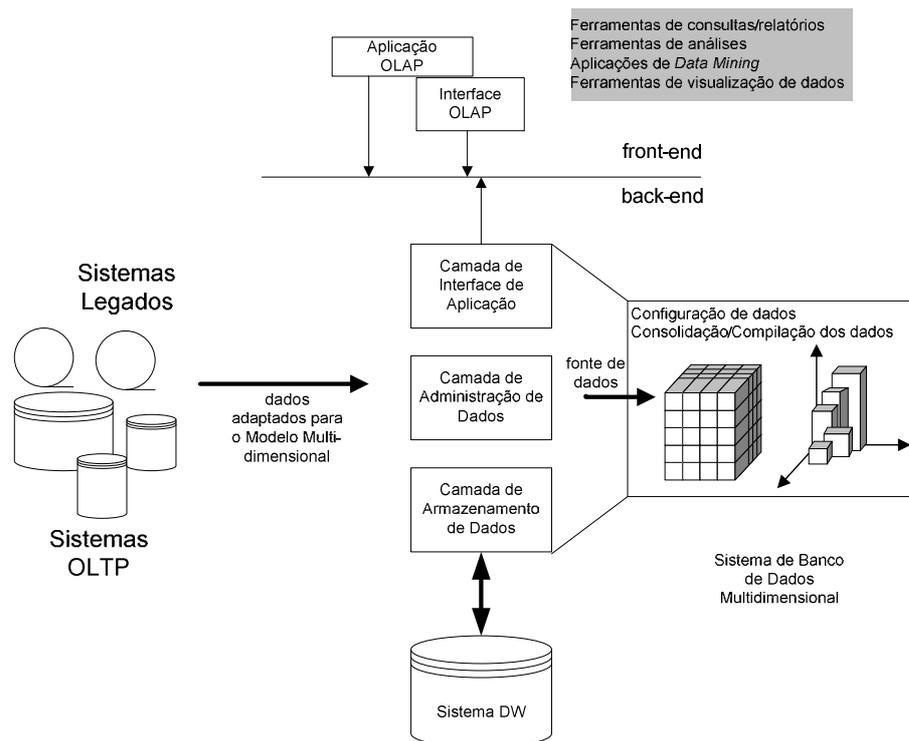


Figura 25: Arquitetura *Warehouse* para sistemas *MOLAP*.
Fonte: Adaptado de WU e BUCHMANN (1997).

Os sistemas *ROLAP* usam o Banco de Dados Relacional como *DW*, como mostra a Figura 26. As camadas de armazenamento de dados e gerenciamento de dados são relacionais, no entanto, são necessárias extensões do modelo relacional para garantir os requisitos da análise multidimensional da camada de interface de aplicação. A máquina *ROLAP*, na camada da interface de aplicação, visualiza a dimensionalidade dos dados, e fornece operações multidimensionais para seus usuários (WU; BUCHMANN, 1997; MALINOWSKI; ZIMÁNYI, 2008).

Em um ambiente de *Data Warehousing*, o padrão de acesso, operações, organização dos dados, critérios de desempenho, gerenciamento de metadados, gerenciamento de transações e processamento de consultas são diferentes dos bancos de dados operacionais (como apresentado no Quadro 2, Seção 2.4). Por consequência, se o SGBD Relacional é

voltado para o ambiente operacional, ele não pode ser diretamente transferido para o DW (WU; BUCHMANN, 1997).

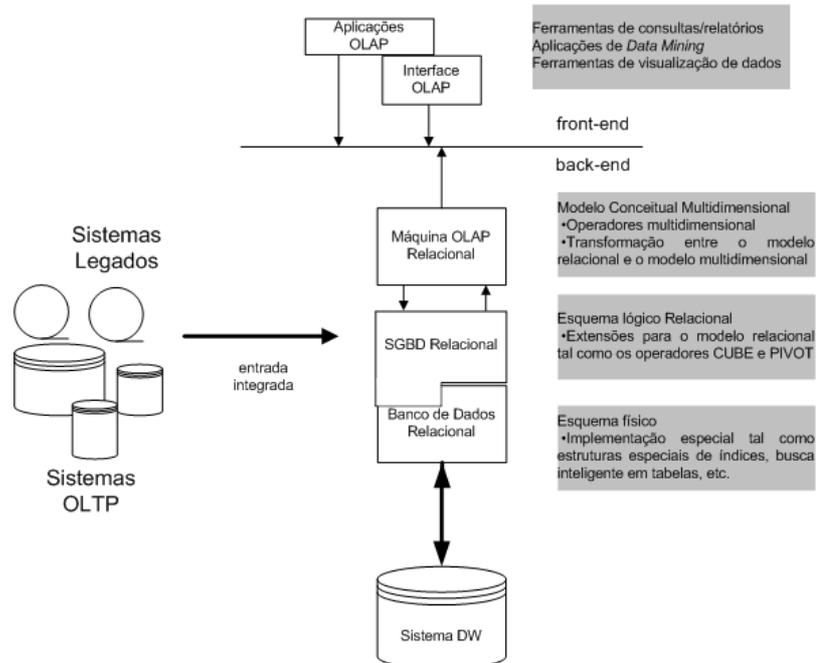


Figura 26: Arquitetura Warehouse para sistemas ROLAP .

Fonte: Adaptado de WU e BUCHMANN (1997).

Existem vantagens e desvantagens em ambas as arquiteturas, *ROLAP* e *MOLAP*, o Quadro 3 mostra algumas delas. A escolha entre uma arquitetura ou outra vai depender das necessidades dos usuários e do domínio da aplicação.

Alguns produtos combinam as abordagens *ROLAP* e *MOLAP*, resultando no *HOLAP* (*Hybrid OLAP*) (DATE, 2000; MALINOWSKI; ZIMÁNYI, 2008; SILVERS, 2008), não há na literatura um consenso de qual abordagem é melhor.

Quadro 3: *MOLAP* versus *ROLAP*.

Característica	MOLAP	ROLAP
Esquema	Utiliza cubos de dados. Dimensões adicionais exigem a recriação do cubo de dados.	Utiliza esquema estrela. É possível acrescentar dimensões adicionais dinamicamente.
Tamanho do Banco de Dados	Pequeno a médio.	Médio a grande.
Arquitetura	Cliente/servidor. Proprietária.	Cliente/servidor. Com base em padrões. Aberta.
Acesso	Limitado a dimensões predefinidas.	Suporte a solicitações <i>ad hoc</i> . Dimensões ilimitadas.
Recursos	Muito altos.	Altos.
Flexibilidade	Baixa.	Alta.
Escalabilidade	Baixa.	Alta.
Velocidade	Mais rápido com conjunto de dados pequenos ou médios; razoável para conjuntos médios.	Boa com pequenos conjuntos de dados; razoável para conjuntos médios ou grandes.

Fonte: ROB; CORONEL (2011).

2.4.2 Modelo Multidimensional

O *DW* e os sistemas *OLAP* são baseados no modelo multidimensional. Esse modelo permite o melhor entendimento dos dados para fins de análise e proporciona melhor desempenho para consultas analíticas complexas, pois, permite a visualização dos dados em um espaço de n -dimensões, chamado de cubo de dados ou de hipercubo (MALINOWSKI; ZIMÁNYI, 2008). Um exemplo de cubo de dados pode ser visto na Figura 27, com as Dimensões Parcelas, Tempo e Patologias.

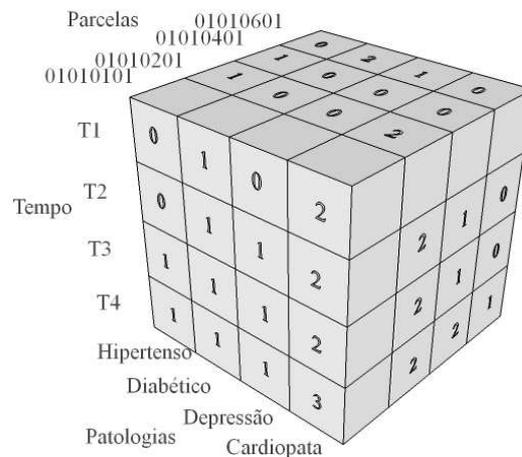


Figura 27: Cubo de três dimensões.

O cubo de dados é definido por Dimensões, Fatos e Medidas. As Dimensões são várias perspectivas que são usadas para analisar os dados, na Figura 27 são Parcelas, Tempo e Patologias, com atributos associados (Ano, Imóvel e Tipo de Patologia) que as descrevem. As células do cubo de dados, ou Fatos, têm associação com valores numéricos, chamadas Medidas, que permitem a evolução quantitativa em vários aspectos do problema analisado.

Um dado no cubo pode ser esparsos²⁶ ou denso dependendo se tem ou não a Medida associada com cada combinação dos valores das Dimensões. Em uma aplicação do mundo real é comum encontrar cubos esparsos. O esparsamento gerenciado adequadamente permite requerer espaço de armazenagem, para reduzir e melhorar o desempenho das consultas. Isso depende se todas as medidas sobre a quantidade de pessoas com patologias forem preenchidos em todo o período de tempo considerado, ressaltando que o valor zero corresponde que não há pessoas com a patologia e a medida sem um valor representa que o dado não foi coletado.

A Figura 28 exemplifica um cubo esparsos, neste caso foram feitos vários cortes no cubo, fatiando-o, para representar a ocorrência de patologias nas parcelas separadamente,

²⁶ Dados esparsos: ocorre quando alguns cruzamentos das dimensões não contêm dados.

supondo que alguns dados não foram coletados em uma atualização cadastral, o cubo fica sem valor na Medida. Nesse exemplo, na parcela 01010201, há quatro células vazias, para as patologias hipertensão (T3 e T4), diabético (T2) e depressão (T3). Na parcela 01010401, não foram coletados os dados de hipertensão (T2) e depressão (T3). Na parcela 01010601, não foram coletados os dados de hipertensão (T2 e T4) e depressão (T3).

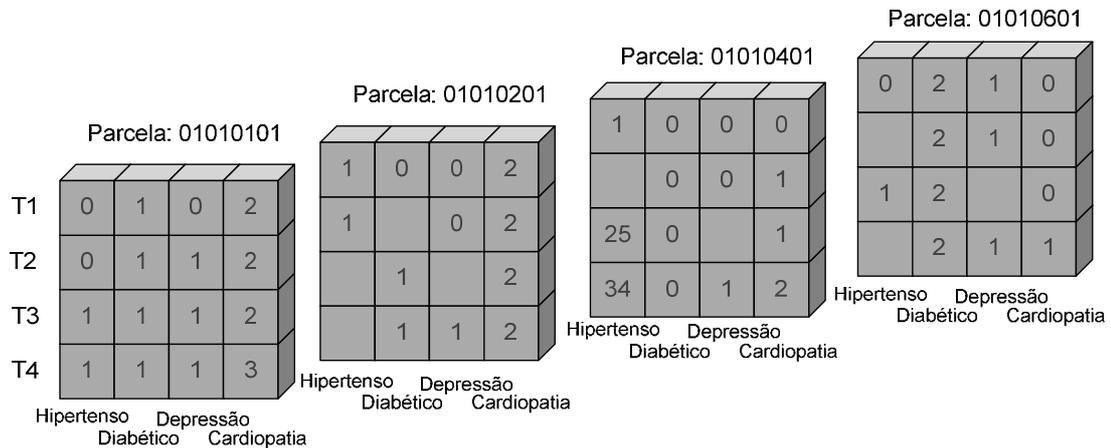


Figura 28: Exemplo de cubo esparso.

O nível de detalhe em que as Medidas são apresentadas é chamado de granularidade de dados. Essa é determinada pela combinação da granularidade de cada Dimensão do cubo. Quando é feita a extração de conhecimento do cubo é necessário, para as visões, várias granularidades. A hierarquia permite a granularidade, possibilitando a definição de uma sequência de mapeamentos relacionados em baixo nível, detalhando conceitos de alto nível, e conceitos mais gerais. Dado dois níveis relacionados na hierarquia, o nível mais baixo é chamado de filho e o nível mais alto é chamado de pai.

O modelo multidimensional, em sistemas *ROLAP*, é geralmente representado por tabelas relacionais organizados em estruturas especializadas chamadas de esquema estrela, esquema floco de neve, esquema floco de estrela ou esquema constelação. Nesses esquemas relacionais, há uma tabela Fato e várias tabelas Dimensões (MALINOWSKI; ZIMÁNYI, 2008). Normalmente, os *DWs* possuem muitas tabelas de Fatos, projetadas para responder questões específicas de suporte à decisão (ROB; CORONEL, 2011).

O **esquema estrela** é uma técnica de modelagem de dados, usada para mapear dados multidimensionais para um Banco de Dados Relacionais, criando uma estrutura multidimensional (ROB; CORONEL, 2011; SILVERS, 2008).

Esse esquema cria um modelo simples de implementar e preserva as estruturas relacionais do banco de dados. Possui quatro componentes: Fatos, Dimensões, Atributos (medidas) e Hierarquias de atributos (ROB; CORONEL, 2011). A Figura 29 representa um

esquema estrela para vendas com as Dimensões lote, localização e tempo, alguns exemplos de atributos estão ilustrados no Quadro 4.

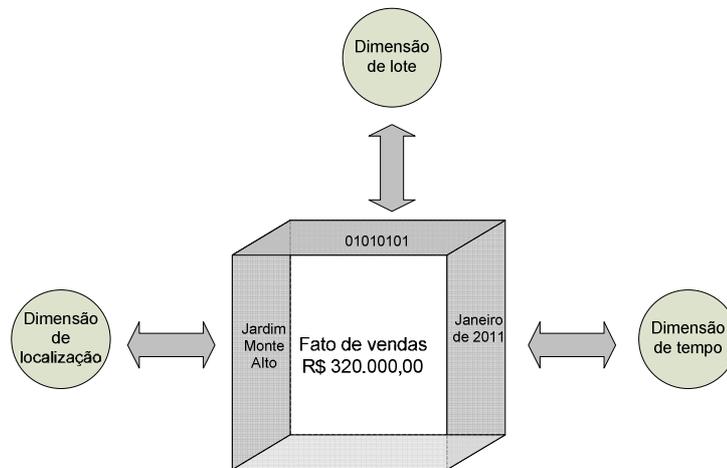


Figura 29: Esquema estrela.

Fonte: Adaptado de ROB e CORONEL (2011).

Quadro 4: Atributos para a dimensão.

Nome da dimensão	Descrição	Atributos possíveis
Localização	Qualquer dado que forneça descrição da localização. Por exemplo: Bairro, Região Sul, Centro, Periferia.	Região, bairro, lote.
Lote-Parcela	Qualquer dado que forneça descrição do lote vendido. Por exemplo: lote comercial, lote residencial, lote novo.	Tipo do lote, tamanho do lote, etc.
Tempo	Qualquer dado que forneça o período de tempo para a Fato vendas. Por exemplo: ano de 2010, mês de agosto.	Ano, trimestre, mês, semana, dia.

O esquema estrela, além das Dimensões, Fatos e Medidas, possui o componente Hierarquia de Atributos que oferece uma organização vertical utilizada para duas finalidades: agregação e análise de dados por *drill down* ou *roll up*, Figura 30. O *drill down* (decomposição) é o desmembramento de dados em dados de menor nível de agregação. O *roll up* (agrupamento) é o inverso ao *drill down*, ou seja, apresentar os dados em um nível mais elevado, resumido a partir de um nível mais detalhado.

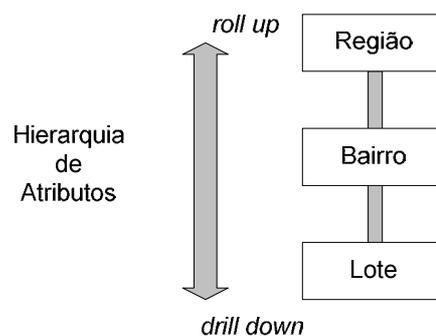


Figura 30: Hierarquia do atributo localização.

Fonte: Adaptado de ROB e CORONEL (2011).

A informação da Hierarquia é armazenada no dicionário de dados do SGBD e utilizada pela ferramenta *OLAP* para acessar o *DW*. A partir do acesso, as ferramentas de consulta devem estar integradas com os metadados do *DW* para dar suporte aos recursos analíticos.

No *DW*, as tabelas Fatos e Dimensões são representados por tabelas conceituais. A tabela Fato é relacionada com cada tabela de Dimensão em um relacionamento muito-para-um (*:1), isto é, várias linhas de Fato se relacionam a cada linha de Dimensão. A relação entre elas é feita por chaves estrangeiras, estão sujeitas as restrições comuns de chave primária/estrangeira. As chaves primárias das tabelas Dimensão são armazenadas como parte da chave primária da tabela de Fato.

O segundo esquema citado é o **floco de neve**, um tipo de esquema estrela como mostra o exemplo da Figura 31, com uma tabela Fato Vendas e com as tabelas Dimensão Lote, Bairro e Região, sendo as duas últimas normalizadas. Esse modelo evita a redundância do modelo estrela normalizando as tabelas de Dimensões. A normalização é feita para ter simplicidade semântica e facilitar a navegação do usuário final pelas Dimensões, contudo isso leva ao aumento da complexidade dos comandos para a execução de uma consulta (ROB; CORONEL, 2011; MALINOWSKI; ZIMÁNYI, 2008; SILVERS, 2008) e principalmente a degradação do desempenho no processamento de consultas OLAP, devido à necessidade de um maior número de operações de junção (CIFERRI, 2002).

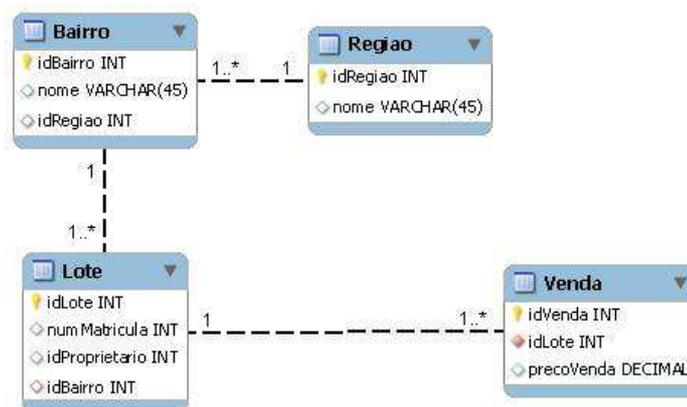


Figura 31: Esquema floco de neve.
Fonte: Adaptado de ROB e CORONEL (2011).

Além dos modelos descritos, Malinowski e Zimányi (2008) citam os modelos floco de estrela e constelação. O modelo floco de estrela é a combinação do modelo estrela e do modelo floco de neve onde algumas das Dimensões são normalizadas enquanto outras não são. O modelo constelação tem múltiplas tabelas Fato que compartilham as tabelas de Dimensões que podem estar ou não normalizadas.

2.4.3 Data Warehouse Espaço-Temporal

O *STDW* é um conjunto de dados baseado em assuntos, integrado, não volátil e variável em relação ao tempo, para apoio às decisões gerenciais, como definido por Inmon (2005) acrescido de dados espaciais e temporais. A literatura também apresenta o *DW* espacial/geográfico (*Spatial DW - SDW*) e o *DW* temporal (*Temporal DW - TDW*), descritos a seguir.

Estima-se que aproximadamente 80% dos dados armazenados nos bancos de dados têm um componente espacial ou de localização (RIVEST; B'EDARD; MARCHAND²⁷ apud MALINOWSKI; ZIMÁNYI, 2008). Por isso, a Dimensão localização tem sido amplamente usada em sistemas de *Data Warehousing* e *OLAP*. Entretanto, esta Dimensão é usualmente representada de maneira alfanumérica e não espacial, uma vez que estes sistemas não estão aptos a manipular dados espaciais. Apesar disso, entende-se que incluir o dado espacial no processo de análise, pode ajudar a revelar padrões que são difíceis de descobrir de outra maneira.

Para que a Dimensão localização seja baseada no dado espacial, alguns trabalhos têm sido desenvolvidos tratando do *SDW* e da *OLAP* espacial/geográfico (*Spatial OLAP - SOLAP*) (AHMED, 2008) (MATIAS; PIRES-MOURA, 2005) (CHAUDHURI; DAYAL, 1997). Fonseca et al. (2007) define um *SDW* como uma extensão da abordagem tradicional de *DW*, acrescentando um componente geográfico. Basicamente, inclui no modelo estrela propriedades geográficas (descritivas e geométricas), as quais podem ser definidas como Dimensões e ou Medidas do *SDW*. As Dimensões podem armazenar as geometrias e as descrições dos objetos geográficos, enquanto que as medidas espaciais só armazenam as geometrias. No entanto, o *SDW* tem que manter todas as características tradicionais de um *DW*, descritas na Seção 2.4, e oferecer suporte ao armazenamento, à indexação, à agregação e às análises, em mapas ou tabelas, de dados geográficos.

Apesar de existirem trabalhos relativos ao uso de *SDW*, ainda não existe uma padronização sobre o modelo de dados a ser especificado nas suas fases de projeto conceitual e lógico. Em alguns casos, propõe-se integrar ferramentas *OLAP* e SIG (SHEKHAR et al., 2001; KOUBA; MATOUSEK; MIKSOVSKÝ, 2000; FIDALGO; TIMES; SOUZA, 2001; SIQUERIA et al., 2010) e outros que visam o desenvolvimento de aplicações *SOLAP* (MANHÃES et al., 2008).

²⁷ RIVEST, S.; B'EDARD, Y.; MARCHAND, P. **Toward better support for spatial decision making: defining the characteristics of spatial on-line analytical processing (SOLAP)**. *Geomatica*, 55(4):539–555, 2001.

Por definição o *DW* é variável no tempo, no entanto os *DW* tradicionais e sistema *OLAP* não permitem a evolução de dados nas Dimensões (VAISMAN; ZIMÁNYI, 2009). Portanto para o dado temporal, as características dos BDTs são usadas para fornecerem estruturas e mecanismos para a representação e gerenciamento das informações que variam no tempo. Muitas pesquisas foram realizadas no campo dos BDTs nas últimas décadas. Por essa razão, a combinação das pesquisas em BDT e *DW* tem orientado para o conceito, de *TDW*.

No esquema tradicional, para *DW*, somente o tempo válido é modelado na Dimensão. Essa é uma solução simples, durante cada ciclo de atualização, um novo conjunto de dados é registrado, o que pode implicar em atualização dos registros de épocas anteriores. O momento em que os dados são registrados não está representado, assim valores anteriores deixam de existir (RIZZI; GOLFARELLI, 2006).

Em esquema bitemporais, *TDW*, ambos os tempos válidos e de transação são modelados como Dimensões. Em cada ciclo de atualização, podem ser adicionados novos eventos para os tempos válidos anteriores, e seu tempo de registro é rastreado, sem a substituição de eventos existentes, assim, nenhum dado é perdido (RIZZI; GOLFARELLI, 2006).

O *TDW* levantou muitas discussões, incluindo agregação consistente na presença do dado variável no tempo, consultas temporais, métodos de armazenamento e materialização de *view*²⁸ temporal. Contudo, há poucos modelos conceituais e lógicos para os *TDW*, ou para a análise, na qual exista o suporte temporal. A maioria dos trabalhos considera o dado temporal como características do *DW*, mas Malinowski e Zimányi (2008) tratam a informação temporal com mais detalhamento, levando a comparação aos BDTs.

Nesse contexto, o gerenciamento de dados espaço-temporais, por meio do *STDW*, precisa resultar da convergência das pesquisas em BDG e BDT, com o objetivo de fornecer abordagens de integração para o gerenciamento das informações espaciais e temporais. Um modelo espaço-temporal multidimensional é importante para o suporte dessas aplicações inovadoras, mas, a definição desse modelo tem que considerar questões específicas de cada domínio (MALINOWSKI; ZIMÁNYI, 2008).

O modelo MultiDim, publicado em 2008, desenvolvido por Malinowski e Zimányi (2008) apresenta um modelo para o desenvolvimento de *DW*, além de considerar o espaço e o tempo. Ele permite uma representação no nível conceitual de todos os elementos requeridos no *DW* e aplicações *OLAP*, por exemplo, Dimensões, Hierarquia e Fatos associados a

²⁸ *View*: Também chamadas de tabelas virtuais ou derivadas, apresentam dados derivados de outras tabelas (ELMASRI; NAVATHE, 2005).

Medidas, conceitos apresentados na seção 2.4, com algumas extensões descritas a seguir. São definidos os modelos de alto nível e do nível intermediário de acordo com a classificação de Inmon (2005). A notação gráfica do modelo MultiDim, apresentado na Figura 32, é similar ao modelo ER.

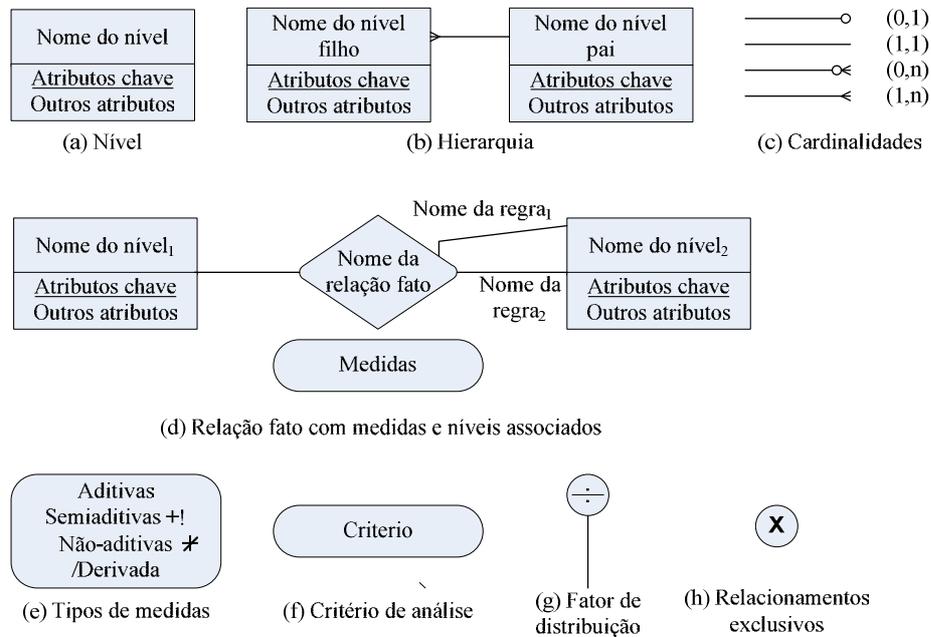


Figura 32: Notação do modelo MultiDim.
Fonte: Adaptado de MALINOWSKI e ZIMÁNYI (2008).

Nesse modelo, uma Dimensão é composta de um Nível ou por uma ou mais Hierarquias, nas quais estabelecem significativos caminhos para a agregação. Uma Hierarquia inclui vários níveis relacionados, como na Figura 32b.

Dados dois Níveis relacionados de uma Hierarquia, o Nível mais baixo é chamado de filho e o mais alto Nível é chamado de pai. Dessa forma, a composição das Hierarquias é chamada de relacionamentos pai-filho. Como estes relacionamentos são usados somente para passar de um nível para o outro, eles são simplesmente representados por uma linha para simplificar a notação.

As Medidas (Figura 32e) podem ser aditivas, semiaditivas e não aditivas, sendo que as duas últimas foram incluídas no modelo com os símbolos +! e ✖, respectivamente, próximo ao nome da Medida. Além disso, os Atributos de Nível e Medidas podem ser derivados, quando eles são calculados com base em outras Medidas ou Atributos no esquema. O símbolo / é usado para indicar um Atributo ou Medida derivada.

Malinowski e Zimányi (2008), levando-se em conta a crescente demanda por incorporar o dado espacial no processo de tomada de decisão, estenderam o modelo MultiDim para incluir o dado espacial, conforme apresenta o metamodelo da Figura 33.

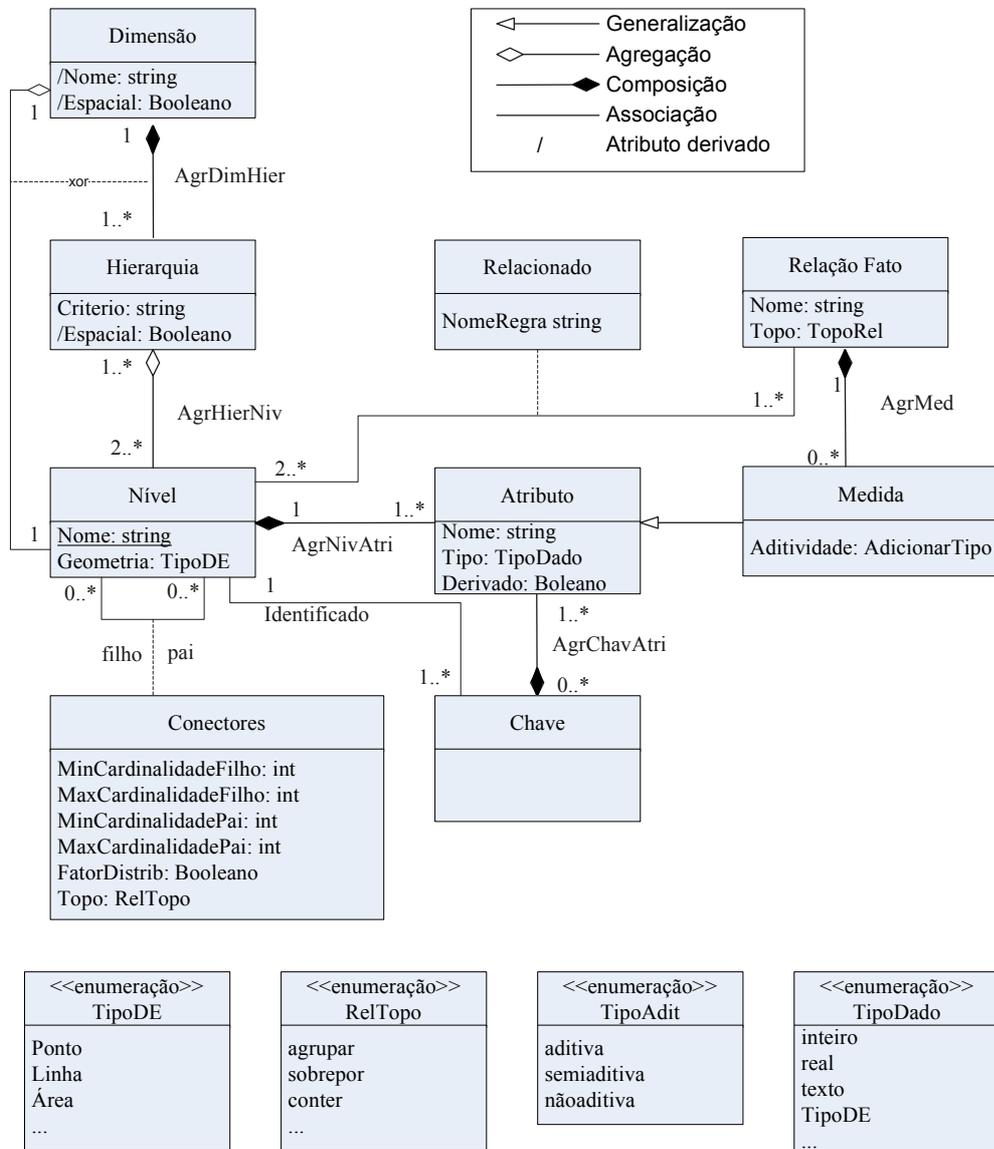


Figura 33: Metamodelo da extensão espacial do modelo MultiDim.
Fonte: Adaptado de Malinowski e Zimányi (2008).

As definições de Dimensão, Hierarquia e Nível são as mesmas do modelo MultiDim, apresentado anteriormente, acrescentando que um Nível pode ter uma geometria, que pode ser um tipo de dado espacial (TipoDE). Um Nível que tem uma geometria é chamado de Nível espacial. Uma Dimensão é espacial se tiver ao menos uma Hierarquia espacial, e a Hierarquia é espacial se ela tiver ao menos um Nível espacial. Isto é representado pelos Atributos espaciais na Dimensão e na Hierarquia.

Os Níveis incluem Atributos, alguns dos quais são chamados de Atributos chaves usado para fins de agregação, enquanto outros são Atributos descritivos. Os Atributos têm um nome e um tipo, que pode ser inteiro, real ou espacial (TipoDE).

Os Níveis formam hierarquias relacionadas por meio de conexões de associação das classes. Estas associações de classes são caracterizadas por cardinalidade máxima e mínima com papéis de pais ou filhos, um fator de distribuição opcional, e uma relação topológica opcional. A relação topológica está apresentada no MultiDim, Figura 33, como um tipo de dados RelTopo.

Uma Fato apresenta relacionamento n -ario entre os Níveis. Podem conter Atributos, chamados de Medidas, e estes podem ser aditivo, semiaditivo ou não aditivo (TipoAdit). Além disso, uma Fato pode ter um relacionamento topológico (RelTop). O relacionamento da tabela Fato possui uma restrição, um relacionamento topológico somente existirá se o relacionamento existir entre, pelo menos, dois Níveis espaciais.

Assim, os mesmos elementos que compõem o modelo MultiDim são tratados com características espaciais, agregando ao elemento a informação de localização. Como a informação espacial, a informação temporal também surge como elemento importante para análises e o acompanhamento da evolução dos objetos do mundo real.

Atualmente os modelos de *DW* e *OLAP* incluem uma Dimensão Tempo que, como nas outras Dimensões, é usado para fins de agrupamento (usando operações *roll-up*) ou na decomposição (usando a operação *slice-end-dice*). A Dimensão Tempo também indica a estrutura do tempo para as Medidas. No entanto, a Dimensão Tempo não pode ser utilizada para rastrear as mudanças em outras Dimensões. Consequentemente, as características de não volatilidade e a variação no tempo estão incluídas na definição do *DW*, aplicado somente para as Medidas, e esta situação deixa para as aplicações a responsabilidade de representar as mudanças nas Dimensões.

Kimball e Ross²⁹ (apud MALINOWSKI; ZIMÁNYI, 2008) propõem várias soluções para este problema em Banco de Dados Relacional, mas, nenhuma é satisfatória. Todavia, não consideram pesquisas realizadas na área de BDT.

A Figura 34 apresenta a extensão do modelo MultiDim para representar dados temporais. A definição de Dimensão, Hierarquia, Nível e Atributo são as mesmas do modelo MultiDim, apresentado anteriormente na Figura 32. O suporte temporal para os Níveis e Atributos é capturado pelo Atributo multivalorado SupTemp do tipo Temp. O Atributo, como

²⁹ KIMBALL, R.; ROSS, M. **The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling.** Wiley, second edition, 2002

mostra a classe Temp da Figura 34, é composto por três componentes: o primeiro, Tipo, representa os vários tipos de suporte temporal, ou seja duração, tempo de validade, tempo de transação e tempo de carregamento; o segundo componente, TipoDado, contém vários tipos de dados temporais definido no modelo, por exemplo, instante, intervalo, grupo de instantes etc.; o terceiro componente, Granularidade, contém várias granularidades, por exemplo, segundo, minutos horas etc.

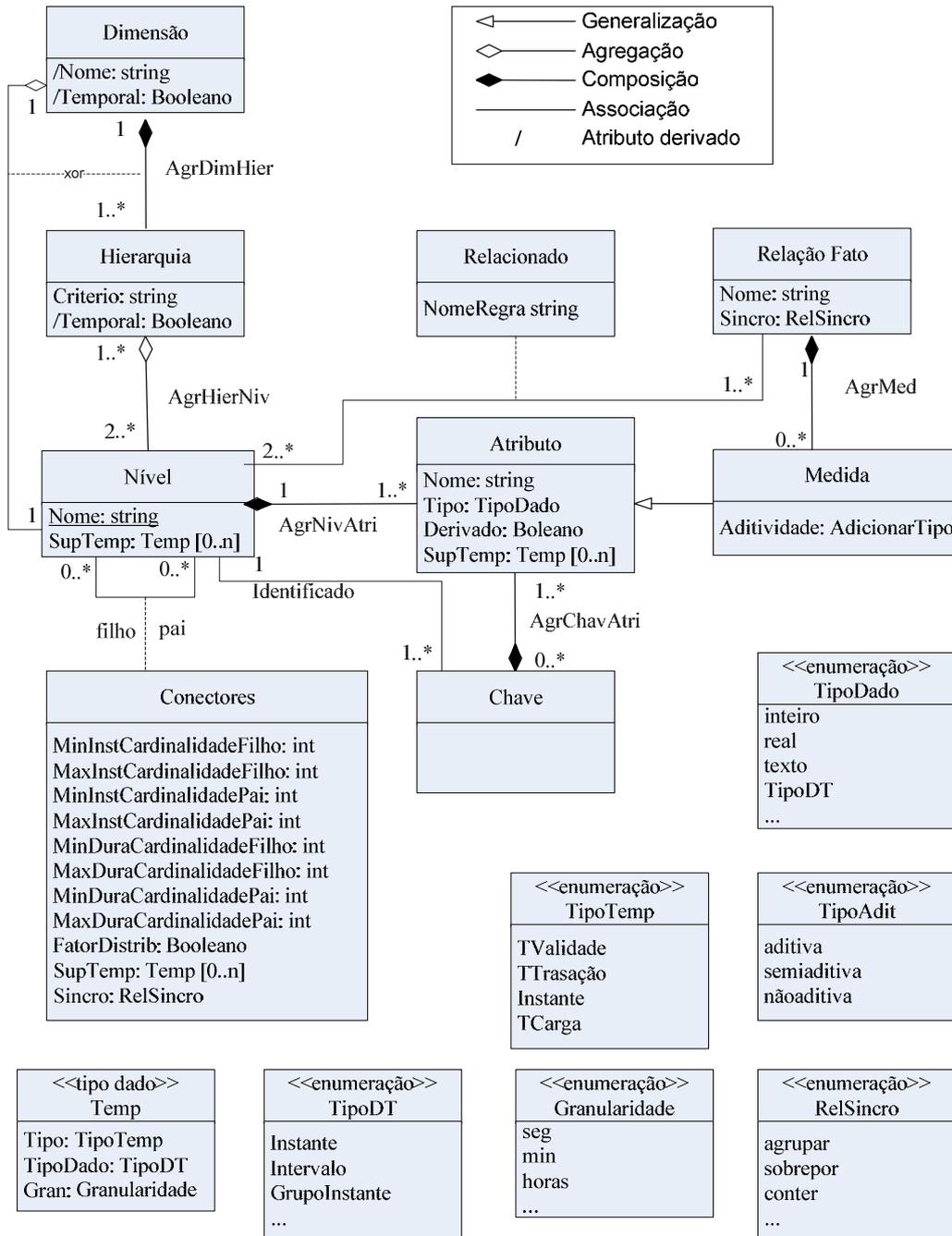


Figura 34: Metamodelo da temporalidade estendido do modelo MultiDim.
 Fonte: Adaptado de Malinowski e Zimányi (2008).

Os relacionamentos pais-filhos entre dois níveis são apresentados por Conectores na associação da classe. Este relacionamento pode ser temporal, independentemente se os Níveis têm suporte temporal ou não. Isto é indicado pelo Atributo SupTemp de Conectores de associação da classe. Além disso, os relacionamentos pai-filho são caracterizados por cardinalidade duração e instante. Ambos os tipos de cardinalidade incluem o valor mínimo e máximo expressado entre os papéis pai e filho.

Um relacionamento Fato representa uma associação n -ária entre os Níveis Folhas, que podem executar vários papéis nestas associações, o nome do papel é incluído na associação da classe Relacionado. Um relacionamento Fato pode conter Medidas temporais aditivas, semiaditivas ou não aditivas (TipoAdit).

Uma Dimensão é temporal somente se tem ao menos uma Hierarquia temporal. Uma Hierarquia é temporal se ela tem ao menos um Nível com suporte temporal ou um relacionamento pai-filho temporal. Isto é representado pelo Atributo derivado Temporal na Dimensão e na Hierarquia (MALINOWSKI; ZIMÁNYI, 2008).

As extensões do modelo MultiDim baseia-se no modelo ER (CHEN, 1976), com as extensões espaciais e temporais do modelo *MADS (Modeling of Application Data with Spatiotemporal features)* (PARENT; SPACCAPIETRA; ZIMÁNYI, 1999), e uma abordagem de projeto baseado em cascata. A próxima Subseção apresenta um resumo do modelo *MADS*.

2.4.3.1 *MADS*

O *MADS* é um modelo conceitual objeto-relacional, apresentado em 1995, estendido do modelo ER (CHEN, 1976). Possui recursos de representação dos objetos do mundo real com a adição de ícones ao diagrama, podem-se representar características espaciais e temporais das entidades, atributos e relacionamentos. Aceita propriedades espaço-temporais necessárias para as aplicações envolvidas com, principalmente, à gestão de territorial ou serviços públicos. Esse modelo foi concebido em colaboração com projetistas de SIG, e sua usabilidade foi verificada por meio de vários estudos de caso (PARENT; SPACCAPIETRA; ZIMÁNYI, 1999).

O *MADS* inclui características de objetos, atributos (mono/multivariado, simples/complexo, derivado), métodos, restrições de integridade, relacionamentos n -ários, relacionamentos simples, e relacionamentos de agregação. Possui uma solução conceitual para apoiar características de espaço e tempo por meio de tipos de dados abstratos (*abstract data types – ADT*) adequados.

O *ADTs* Espaciais (*Spatial ADTs – SADT*) prevêem a forma e as informações de localização, que incluem a representação de pontos, linhas e áreas simples. Os *ADTs* Temporal (*Temporal ADT - TADT*) apoiam dados *timestamp* (data e hora), ou seja, associa um período a um fato, que inclui a representação de instantes, intervalos, e elementos temporais. Esses *ADTs* consideram a granularidade espacial (ou seja, a escala) e a granularidade temporal (como, segundo, minuto, hora e dia).

Os *ADTs* Espaciais e Temporais podem ser implementados como hierarquias de generalização. No modelo da Figura 35, há a possibilidade de especificar conjuntos de *ADTs* (conjuntos de pontos, conjuntos de linhas, áreas complexas), tipos genéricos (por exemplo, geometria simples, geometria complexa, geometria), e definidos pelo usuário, estas definições, também, podem ser aplicadas aos tipos temporais, com o uso do ícone “⌚” para característica temporal.

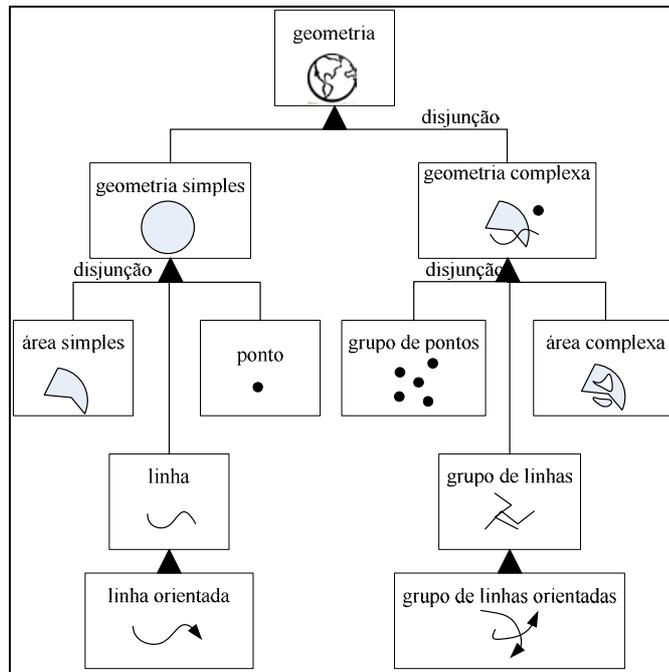


Figura 35: Hierarquia básica do MADS para tipos de dados abstratos espaciais.
Fonte: Adaptado de Parent, Spaccapietra e Zimányi (1999).

A informação espacial e/ou temporal pode estar associada a objetos, independentemente das características de seus atributos. Conseqüentemente, um tipo de objeto pode ser simples (nem espacial, nem temporal), espacial, temporal ou espaço-temporal.

Os relacionamentos são uma parte essencial do projeto conceitual. O *MADS* permite vários tipos de relacionamento, semelhantes aos objetos, podem ser localizado no espaço e no tempo, por meio da geometria e do estado do atributo. Neste caso, eles são chamados de relacionamento espaciais e/ou temporais, dos seguintes tipos: relacionamentos tradicionais

(do tipo ER), relacionamentos de agregação, relacionamento com restrições e relacionamento dinâmicos.

As características espaciais de *MADS* apoiam tanto as visualizações contínuas, como as discretas do espaço, a variação da informação espacial está na geometria do objeto. As características temporais do modelo ressalta que nenhuma restrição poder afetar os períodos relacionados aos fatos do mundo real.

2.5 Trabalhos Correlatos

Os conceitos que fazem parte desse trabalho são foco de muitas pesquisas atualmente. Dados temporais, espaciais e espaço-temporais têm despertado interesses por fornecer informações importantes para a gestão em diversas áreas, inclusive no CTM. A integração do *DW* com o SIG é objeto de pesquisa de vários grupos, que estão desenvolvendo ferramentas, metamodelos e índices para criar um *SDW* que possa ser usado em aplicações SIG (FONSECA et al. 2007) (SIQUEIRA; CIFERRI; TIMES, 2008) (SIQUEIRA et al., 2009) (SIQUEIRA et al., 2010).

Em várias pesquisas foram desenvolvidas aplicações utilizando as variáveis, tempo e espaço, para o CTM e áreas próximas, como mostra o Quadro 5.

Quadro 5: Comparativo dos trabalhos relacionados.

	Dado temporal	Dado espacial	DW	SIG	Cadastro Territorial	Método específico
<i>STDW-CTM</i>	X	X	X	X	X	X
Castro (2006)		X	X	X		
Amorim et al. (2007)		X	X		X	
Cömert e Alkan (2004)	X	X		X	X	
Devilleers, Bédard e Jeansoulin (2005)		X	X	X		X
Renolen (2000)	X	X			X	
Francisco e Imai (2003)	X	X		X	X	
Liang (2008)	X	X		X	X	
Song e Yang (2013)	X	X		X	X	X
Wang, Fang e Xie (2005)	X	X		X	X	X

Entre os trabalhos apresentados no Quadro 5, três utilizaram os conceitos de *DW*, são eles:

- Castro (2006), por exemplo, apresenta um estudo que trata do cadastro de clientes e seus endereços, discute a utilização de *DW* e *SIG*, mostra algumas arquiteturas para implantação e controle da qualidade dos dados resultantes do processo de inserção de clientes. Entretanto, o referido autor não trata do *CTM* e suas especificidades, da gestão pública dos dados, não considera o tempo fora do contexto do *DW*, e não avalia a geometria das parcelas, pois não usa *SDW*.
- Amorim et al. (2007) fizeram um experimento com *DW* usando um Banco de Dados Cadastrais para identificar a ocorrência de problemas de saúde em um determinado período de tempo. Esse trabalho não se baseou nos conceitos do *STDW* e não definiu um método de análise.
- O trabalho de Devillers, Bédard e Jeansoulin (2005) tem o objetivo mais próximo do que se propõe neste trabalho, pois apresentou um modelo multidimensional com acesso *SOLAP* (*Spatial On-Line Analytical Processing*) para aplicações *SIG*, porém o enfoque foi gerar informações de qualidade para diferentes níveis de análise. Para a implantação do modelo multidimensional eles utilizaram ferramentas proprietárias, com dados da Base de Dados Topográficos Nacional do Canadá.

Alguns outros trabalhos apresentam a importância do dado temporal para o *SIG*. Esses trabalhos criaram o atributo data no banco de dados:

- Cömert e Alkan (2004) apresentam um modelo de *SIG* temporal para o Cadastro e para o Registro de Imóvel na Turquia. Eles também usaram o recurso de inserir o atributo data no banco de dados, neste caso foram inseridos tempo inicial de validade e o tempo final.
- Renolen (2000) apresentou um modelo conceitual para *SIG* espaço-temporal, criando um descritor temporal no banco de dados.
- Francisco e Imai (2003) fizeram uma representação espaço-temporal para o cadastro de imóveis, baseado no registro do banco de dados das ocorrências de compra, venda, desmembramento e unificação de imóveis com a data em que os eventos aconteceram.
- O trabalho de Liang (2008) abordou o Cadastro temporal, criando um modelo de cadastro temporal usando o PostGIS para uma cidade chinesa. Como em outros trabalhos, ele também criou o modelo temporal por meio da inserção do

intervalo temporal, representado pelos campos *tmim* (representa a data de criação do dado) e *tmax* (representa a data da alteração do dado).

- O Modelo de dados espaço-temporais baseado na parcela cadastral (WANG; FANG; XIE, 2005) foi baseado em SIG temporal, e se propôs analisar somente a geometria da parcela. O protótipo para avaliação do modelo foi desenvolvido em VB 6.0 com MapObeect 2.3 da ESRI e com o Oracle 9i.
- Song e Yang (2013) apresentaram um modelo de dados cadastrais espaço-temporais implementado em SIG, com dados temporais e não temporais. Para a implementação do modelo foram utilizados o SGBD Oracle 10g para o gerenciamento dos dados e componente *Geomedia Control* da Intergraph para gerenciar e analisar os dados espaciais no ambiente de desenvolvimento VB.net.

Assim como os trabalhos citados acima, existem outros que discutem propostas para o armazenamento do tempo e do espaço. Contudo, não foi encontrado nenhum com o mesmo enfoque deste trabalho, a utilização do *Data Warehousing* para oferecer um método de análise temporal para o CTM, com ferramentas livres.

3 MATERIAIS E MÉTODOS

A área de teste utilizada neste trabalho foi o município de Ribeirão dos Índios – SP. O banco de dados possui dados reais do Cadastro Urbano do município, que vem sendo aprimorado, pelo Grupo de Pesquisa em Aquisição e Representação de Dados Espaciais (GARDE), desde 2009. Por isso, existem quatro bancos de dados desenvolvidos em anos diferentes e com estruturas diferentes, pois entre a realização do Cadastro Urbano de 1996, 2004, 2010 e 2012 novos requisitos e tecnologias foram adicionados ao CTM (MUNIZ et al., 1996) (AMORIM; SOUZA; DALAQUA, 2004) (MALAMAN; AMORIM, 2010).

O município possui uma população de 2.187 habitantes³⁰, segundo o censo de 2010 e com uma área de 196,34 Km². Localiza-se próximo a Presidente Prudente – SP, Figura 36.

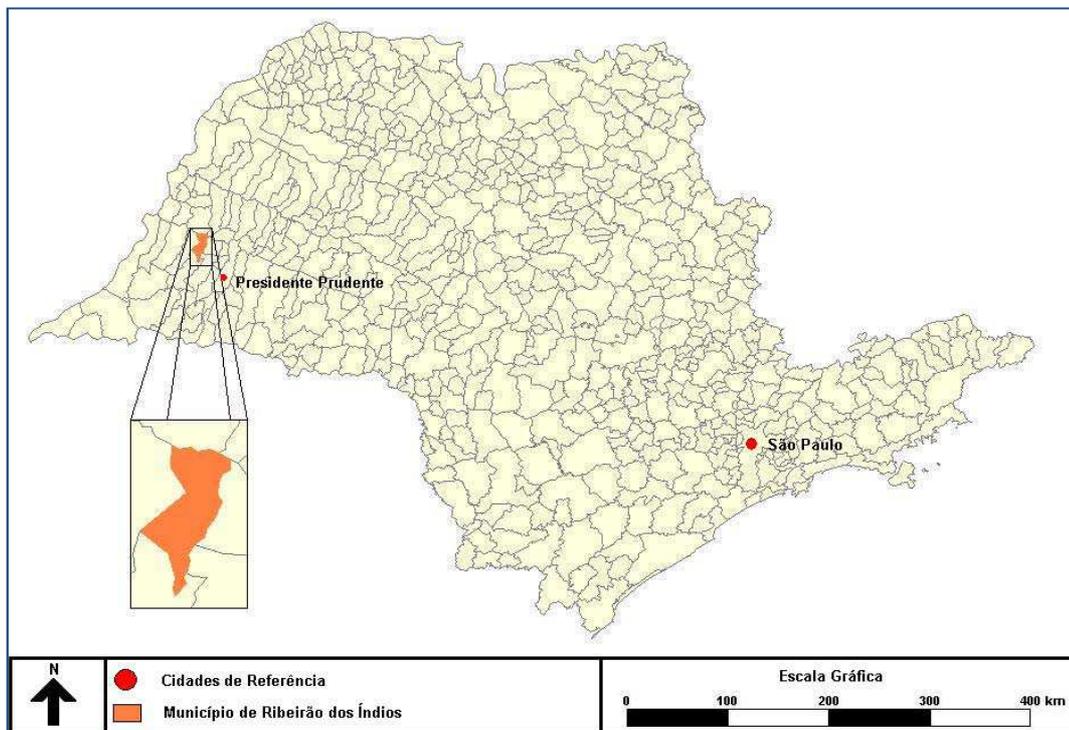


Figura 36: Mapa de localização do Município de Ribeirão dos Índios - SP.
Fonte: DINIZ (2004).

Considerando o exposto neste trabalho, a partir do processo de *BI*, o método de análise temporal para o CTM é composto pelas etapas apresentadas na Figura 37. Esse método permite a criação de um sistema de apoio à tomada de decisão considerando componentes espaço-temporais para o CTM.

Na primeira etapa o método foram definidas as fontes de dados utilizadas para a elaboração do *STDW*. Essas fontes de dados poderiam ser banco de dados convencionais,

³⁰ Disponível em: <http://www.ibge.gov.br/cidadesat/xtras/perfil.php?codmun=354323>. Acesso em: 30 jun 2012.

BDG, cópias de segurança de bancos de dados, arquivos texto e/ou arquivos de planilhas eletrônicas. No experimento realizado neste trabalho, os dados estavam armazenados em Paradox 4.0, MS ACCESS 2000, PostgreSQL 8.4 e PostgreSQL 9.1.

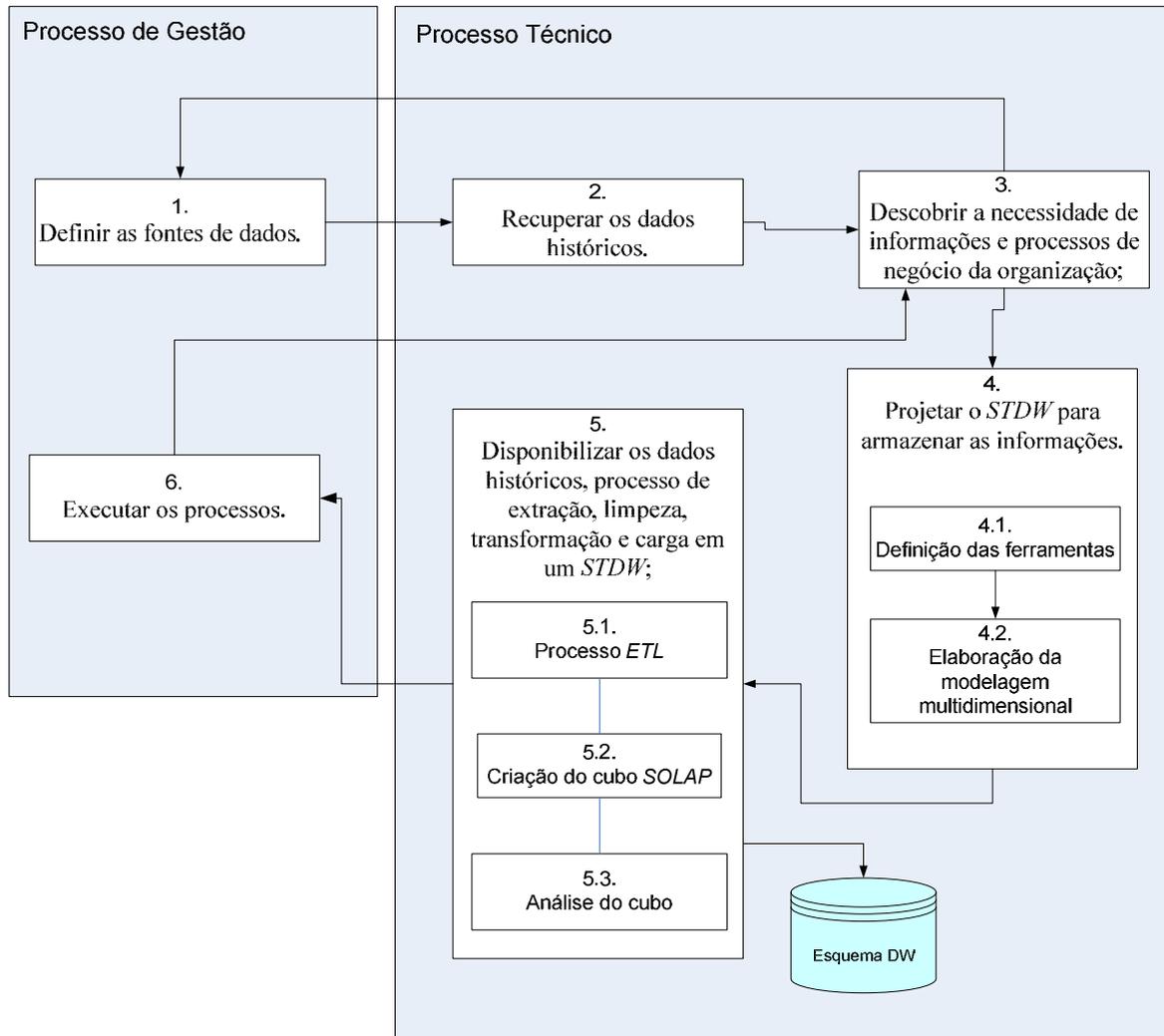


Figura 37: Etapas do Método de análise temporal para CTM.

Na segunda etapa, os dados foram recuperados. Inicialmente os dados foram restaurados em seus bancos de dados de origem para a verificação de quais dados existiam e como estavam organizados. A partir dessa atividade foi possível descrever e relacionar os dados existentes nos quatro levantamentos e definir os metadados do *STDW*.

A terceira etapa, com base nos dados recuperados, as necessidades de informações e indicadores de negócio da organização foram identificadas, necessários para as tomadas de decisões. Para apoiar a execução dessa etapa, uma descrição das fontes de dados foi elaborada, gerando os metadados individuais de cada CTM. Em algumas situações, no final dessa etapa, pôde ser identificada a necessidade de se trabalhar com outras fontes de dados. Se

isso ocorrer, volta o processo para a primeira etapa. No entanto, no experimento realizado, não foi necessário inserir novas fontes de dados.

Para a definição dos metadados foram considerados os conceitos de integração de dados, discutidos na Seção 2.2 conforme classificação apresentada por Sheth (1999), pois era necessário integrar os dados antigos aos novos no *STDW*.

A partir das informações resultantes da etapa três foram descritos os processos de negócio para o projeto do *STDW* (quarta etapa) responsável pelo armazenamento das informações. Segundo os conceitos de arquitetura apresentados na Subseção 2.4.1, o *STDW* foi projetado baseado na arquitetura *ROLAP* e no modelo MultiDim estendido (MALINOWSKI; ZIMÁNYI, 2008), descrito na Subseção 2.4.3. A quarta etapa é composta por duas subetapas: subetapa 4.1, na qual foram escolhidas as ferramentas utilizadas. Todas as ferramentas utilizadas nos experimentos são livres, PostgreSQL 9.1, Postgis 1.5, gvSIG 1.11 e a *suíte* Pentaho 4.8; subetapa 4.2, na qual foi elaborado o modelo multidimensional, diagramado em ArgoCaseGeo 3.0.

O PostgreSQL³¹ é um sistema gerenciador de banco de dados objeto-relacional, com código aberto e suporte ao padrão *SQL*. Possui uma extensão espacial, o PostGIS, que armazena os dados espaciais. Sua arquitetura é compatível com a especificação do *OGC*.

O gvSIG³² é um software de SIG livre, construído com a linguagem Java, que permite arquivos de diversos formatos, entre os quais estão tabelas de BDGs, como exemplo, o MySQL e PostGIS. Surgiu em 2004 e vem ganhando espaço entre os usuários de SIG, pois traz muitos recursos por meio de suas extensões. A instituição responsável pelo desenvolvimento do gvSIG é a *Generalitat Valenciana* (Espanha) por meio da *Conselleria de Infraestructuras y Transporte*.

O ArgoCaseGeo³³ é uma ferramenta *CASE* (*Computer Aided Software Engineering*), de código aberto que permite a modelagem de BDG e aspectos de modelagem temporal (LISBOA FILHO; RODRIGUES JUNIOR; DALCIO, 2004). Essa ferramenta é baseada na *UML* (*Unified Modeling Language* - Linguagem Unificada de Modelagem), utilizada para a modelagem orientada a objetos, no entanto, foi empregada para a modelagem relacional por oferecer os estereótipos de tempo e espaço para a diagramação.

³¹ <http://www.postgresql.org/>

³² <http://www.gvsig.gva.es/>

³³ <http://www.dpi.ufv.br/projetos/argocasegeo/>

A plataforma Pentaho³⁴ é uma *suite*³⁵ de aplicativos de código aberto para criação de soluções de *BI*. A versão livre é a *Community*. Entre os aplicativos dessa *suite* estão o *Pentaho BI Server*, o *ETL Geo Kettle*, o *Schema Workbench*, o *Mondrian OLAP System*, o *JPivot*, o *Saiku Charts Plus*, o *Metadata Editor* e o *Report Designer*. Eles permitem soluções que acompanham todo o ciclo de implementação de um sistema de apoio à decisão (PENTAHO, 2009) (PENTAHO, 2011).

Na quinta etapa os dados históricos, recuperados na etapa dois, foram disponibilizados para uso do *STDW*. Na subetapa 5.1 foi executado o processo de extração, limpeza, transformação e carga dos dados. Nessa subetapa foi utilizada a ferramenta *ETL Geo Kettle*. A subetapa 5.2, é responsável pela criação do cubo, elaborado no *Schema Workbench 3.5*. Com o *STDW* carregado, os processos de consultas foram disponibilizados para as análises, etapa 5.3, no qual diversos relatórios foram gerados com o *Report Designer 3.9.1-GA*. O *Metadata Editor 4.8* foi utilizado para geração do metamodelo que permite ao gestor gerar os próprios relatórios.

Na etapa seis o *Data Warehousing* foi disponibilizado para os testes no *BI Server* e no *gvSIG*. Com a utilização das informações geradas, novas necessidades provavelmente surgirão, assim, o processo volta para a etapa três, para verificar se novas fontes de dados são necessárias, caso não seja, o processo continua. O método se torna iterativo, para atender às necessidades dos usuários e para o refinamento do processo.

A partir do método apresentado na Figura 37, a arquitetura apresentada na Figura 38 foi projetada para o *STDW*.

Nessa arquitetura (Figura 38), a Camada 1 representa as possíveis fontes de dados que alimentam o *STDW*, são elas: os CTM Legado, o CTM baseado na INDE, PGV (Planta Genérica de Valores) e Registro de Imóveis, no experimento foram utilizados os dados de quatro levantamentos do CTM. A Camada 2 representa o processo de extração, transformação e limpeza dos dados para disponibilizá-los para a Camada 3, que representa o armazenamento do *STDW* no SGBD. A Camada 4 representa o processo de análise dos dados por ferramentas *SOLAP* ou *OLAP + SIG*. Na Camada 5 os dados são disponibilizados aos usuários: prefeitura, incorporadoras e imobiliárias, Organizações Não Governamentais, cartórios e cidadãos.

³⁴ <http://sourceforge.net/projects/pentaho/files/>

³⁵ *Suite*: (tradução do inglês) conjunto.

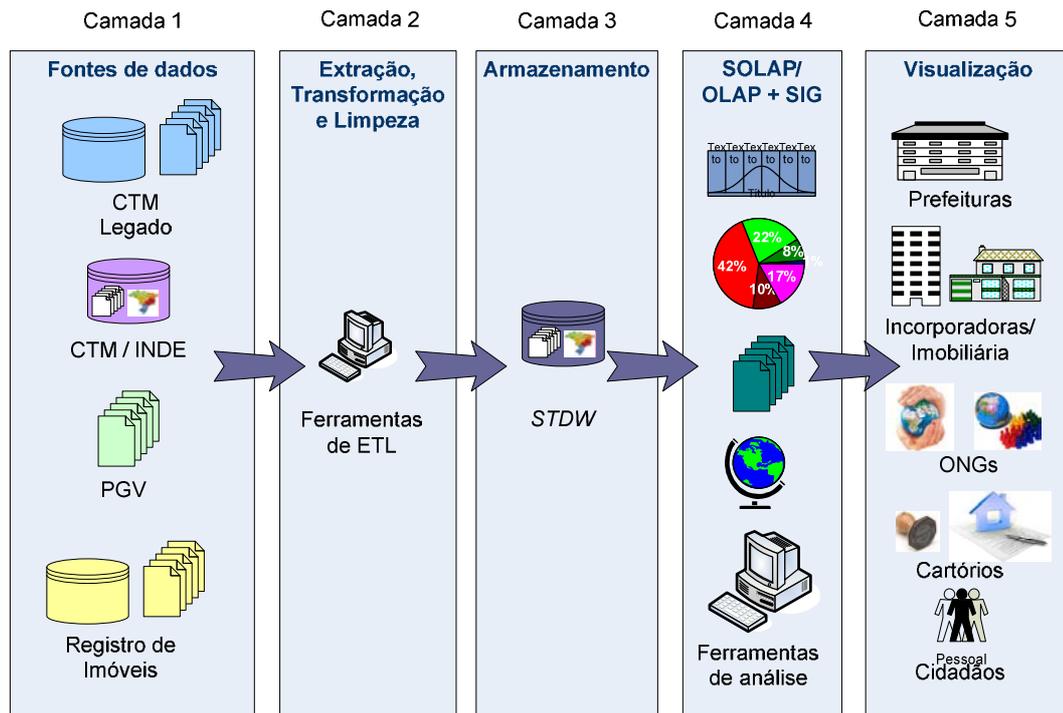


Figura 38: Arquitetura para o Método Temporal do CTM.

Para apoiar o experimento principal deste trabalho, e utilizar o *Data Warehousing* para criar um ambiente de apoio à tomada de decisão, o CTM do município de Ribeirão dos Índios – SP de 2012 foi desenvolvido baseado nos conceitos de BDET, seguindo as definições da modelagem espaço-temporal *MADS* (PARENT; SPACCAPIETRA; ZIMÁNYI, 1999). Nesse experimento, o SIT tem o objetivo de gerenciar o CTM, possibilitando a disponibilização dos dados alfanuméricos, geográficos e temporais para apoiar o ambiente transacional.

No CTM de 2012 os dados foram coletados em campo e registrados nos BICs, impressos em formulários para leitora óptica de marcas. Após a leitura dos BICs, um arquivo texto foi gerado e tratado para a importação dos dados para o PostgreSQL 9.1, conforme a especificação da modelagem. Além do CTM, alguns elementos do MGD (DER e Dicionário de Dados) foram especificados conforme as orientações do Guia Metodológico para Integração de Dados e Processos (SERPRO, 2009) para apoiar o experimento.

As características espaciais foram tratadas diretamente pela extensão PostGIS. No caso das características temporais, elas não são aceitas diretamente pelo PostgreSQL, portanto as funções de administração e gerenciamento das informações temporais ficam sob a responsabilidade da aplicação ou do projeto de desenvolvimento.

4 RESULTADOS

Neste Capítulo são apresentados os resultados dos experimentos realizados conforme a proposta metodológica apresentada no Capítulo 3. Os experimentos foram organizados segundo três elementos principais, sendo: a descrição do MGD, a modelagem do BDET, o projeto do *STDW* e as análises.

No decorrer deste trabalho, iniciado em 2010, um novo levantamento cadastral foi executado em 2012 na área de estudo. A partir das modelagens de dados dos CTMs anteriores e das novas demandas do grupo de pesquisa GARDE foi elaborada a modelagem conceitual do CTM de 2012, considerando os aspectos temporais e espaciais, como mostra parcialmente a Figura 39, a modelagem completa faz parte do Apêndice A. Essa modelagem foi baseada no *MADS* (PARENT; SPACCAPIETRA; ZIMÁNYI, 1999) e o esquema conceitual foi construído com a utilização da ferramenta *CASE* ArgoCaseGeo (LISBOA FILHO; RODRIGUES JUNIOR; DALTIO, 2004).

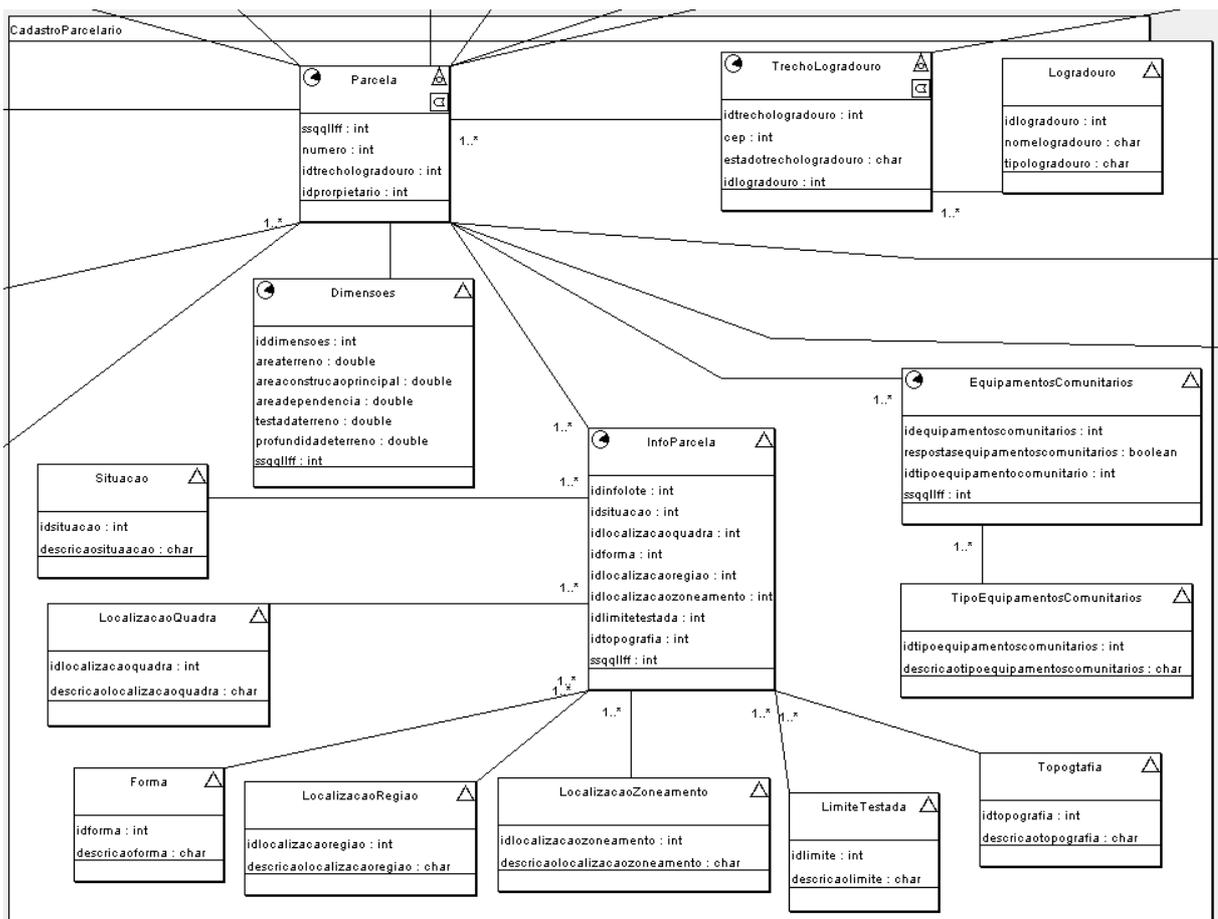


Figura 39: Modelagem parcial do CTM 2012 – Parcela Cadastral.

A ferramenta ArgoCaseGeo possui os estereótipos de tempo e espaço que facilitam a identificação de quais tabelas são comuns, temporais e/ou espaciais. Para a modelagem do BDET foram utilizados os seguintes estereótipos: Intervalo de tempo (🕒), ObjetoNãoGeográfico (△), ObjetoGeográfico (▲) e a representação espacial poligonal (📐).

Com a modelagem do BDET pronta, para a integração dos dados cadastrais no *STDW*, foram definidos os metadados do CTM da área de estudo. A partir dos modelos de dados dos Cadastros de 1996, 2004, 2010 e 2012, seguindo as orientações do MGD (descrito na Subseção 2.2.3), o DER (Figura 40) e o Dicionário de Dados (Quadro 6) foram elaborados. A descrição completa do Dicionário de Dados está no Apêndice B.

Na Figura 40, o DER representa as entidades definidas na modelagem do BDET. Segundo o MGD, as Entidades de Domínio (cor amarela) representam um conteúdo relativamente estável, geralmente composto de código e descrição, e as Entidades de Negócio, (cor cinza) representam um conteúdo mais dinâmico, trazendo para o modelo o dia-a-dia do negócio ou organização.

Quadro 6: Dicionário de Dados parcial.

Projeto: CTM Ribeirão dos Índios	Versão: 1.0
Tipo: entidades	Data: 01/07/2012
Nome	Descrição
parcela	Entidade que registra os dados básicos sobre a parcela territorial.
dimensao	Entidade que registra dados sobre as dimensões da parcela e das construções existentes.
infoparcela	Entidade que registra dados sobre as características físicas da parcela.
situacao	Entidade que registra os tipos de situações possíveis da(s) edificação(ões): não edificado; ruínas; paralisada; andamento; construído.
localizacaoquadra	Entidade que registra os tipos de localizações que a parcela pode ter: esquina; meio da face; mais faces.
forma	Entidade que registra os tipos de forma que a parcela pode ter: regular; irregular; alongada.

Esse Dicionário de Dados foi elaborado observando as características de integração, apresentadas na Seção 2.2, semântica e estrutural dos dados do CTM. A identificação dos dados para a integração foi feita a partir da verificação dos dados de 1996, 2004, 2010 e das novas demandas dos usuários. Para essa análise foi seguida a “Abordagem Orientada a Fonte/Análise” descrita por Malinowski e Zimanyi (2008) para o levantamento de requisito.

Depois de finalizada a modelagem do BDET e a construção do Dicionário de Dados, iniciou-se o projeto e a implementação do *Data Warehousing* para o CTM de Ribeirão dos Índios – SP. Conforme as etapas descritas na Figura 37 do Capítulo 3, os dados do CTM de

Ribeirão dos Índios - SP foram estudados (etapa 1 e 2) para a elaboração dos requisitos do *Data Warehousing*.

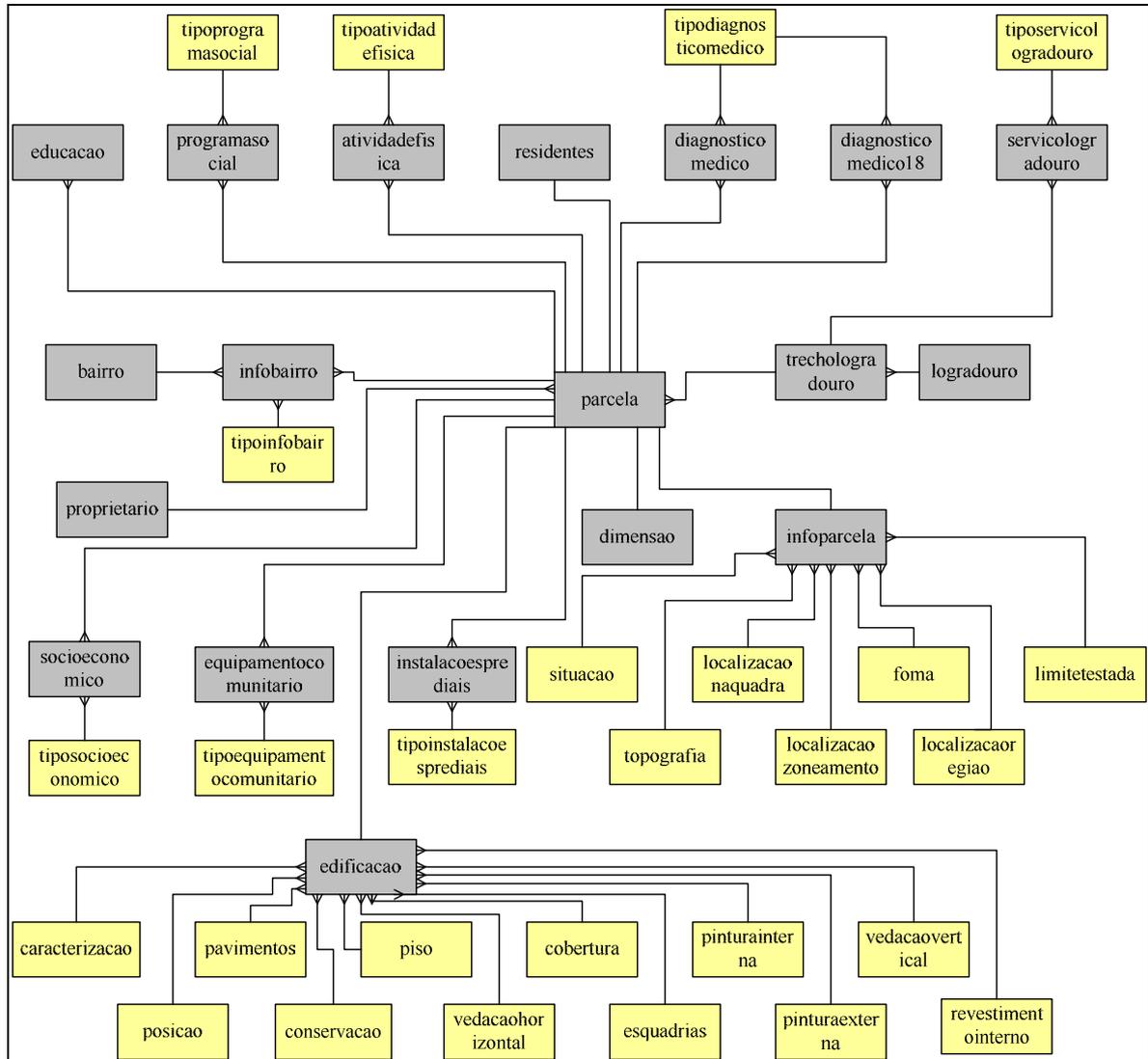


Figura 40: Diagrama de Entidade-Relacionamento de acordo com o MGD.

Na sequência, etapa 2, foram escolhidos os “processos de negócio” para o CTM. Conforme a conceituação do CTM, que reúne os dados do Cadastro e informações temáticas, esses elementos foram definidos como os assuntos do modelo de negócio, Quadro 7.

Após a definição dos “processos de negócios”, a etapa 4 foi realizada, compreendendo duas subetapas. Na primeira (4.1) definiu-se as ferramentas necessárias e utilizadas para a construção do *Data Warehousing*, apresentadas no capítulo anterior. Na segunda (4.2) foi elaborada a modelagem multidimensional para os processos de negócios. A modelagem do BDET (Apêndice A) foi utilizada como parâmetro para a definição da modelagem

multidimensional, apresentada no Apêndice C, e diagramada na ferramenta CASE ArgoCaseGeo. A Figura 41 apresenta a modelagem do assunto Parcela.

Quadro 7: “Processos de negócios” do Data Warehousing.

Assunto	Descrição
Parcela	Esse processo fornece as informações sobre o proprietário e as dimensões das parcelas e das edificações.
Edificação	Esse processo fornece as informações sobre as características das edificações, quando houver nas parcelas. Permite avaliar o padrão construtivo das edificações.
Residentes	Esse processo fornece as informações sobre a idade e o sexo das pessoas que residem nas parcelas.
Educação	Esse processo fornece as informações sobre o nível de escolaridade atual dos residentes.
Programas Sociais	Esse processo fornece as informações sobre os tipos de programas sociais que beneficiam os residentes.
Diagnóstico Médico	Esse processo fornece as informações sobre as patologias que acometem os residentes nas parcelas.
Socioeconomico	Esse processo fornece as informações sobre o tipo de trabalho dos residentes, a renda mensal geral e se possuem veículos com placa local ou de fora do município.
Instalações Prediais	Esse processo fornece as informações sobre os serviços de água, esgoto, energia elétrica telefonia e dados disponíveis nas parcelas.

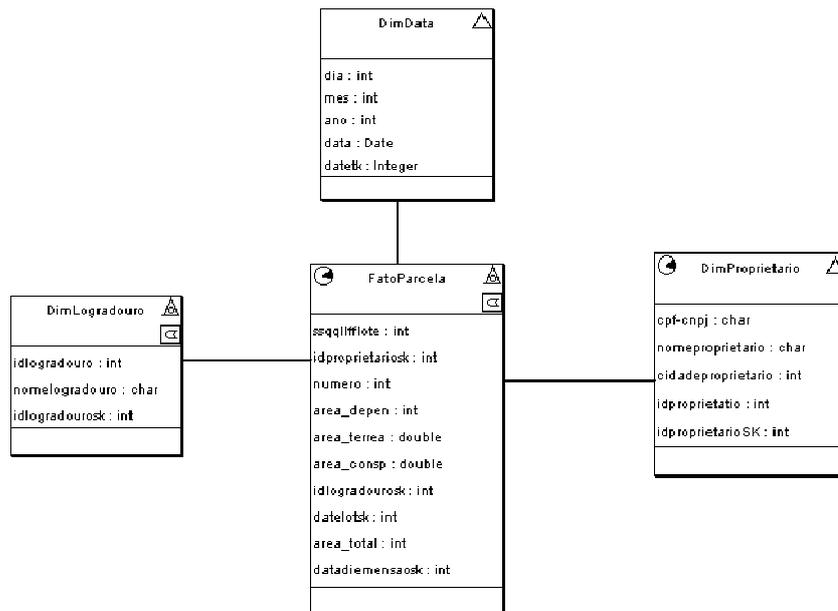


Figura 41: Processo de negócio, assunto Parcela.

Finalizada a modelagem multidimensional do *STDW*, a próxima etapa, a 5, foi executado o processo de *ETL*, com 3 subetapas, foram criados, carregados e disponibilizados os dados. Na subetapa 5.1 fez-se o processo de extração, limpeza e transformação dos dados na ferramenta *Geo Kettle*. A Figura 42 e a Figura 43 apresentam exemplos de transformações construídas na referida ferramenta. O *Geo Kettle* permitiu padronizar a entrada dos dados conforme o modelagem do *STDW*, gerar o *SQL* para a criação das tabelas Fatos e Dimensões

e carregar os dados de suas origens para o *STDW*, que foi implantando no PostgreSQL 9,1 com o PostGIS 1.5.

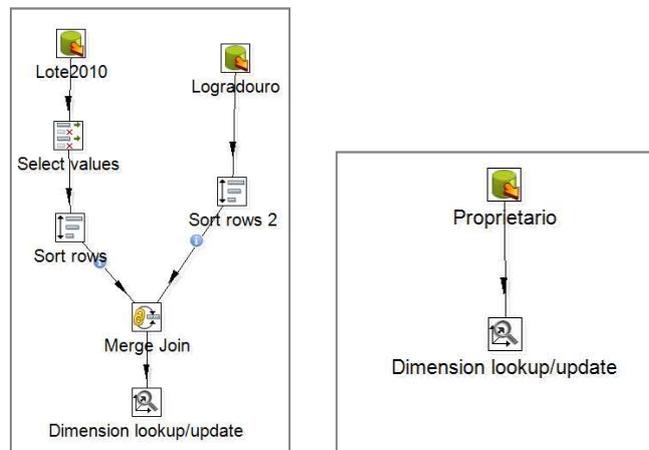


Figura 42: Transformação do processo de *ETL* para tabelas Dimensões.

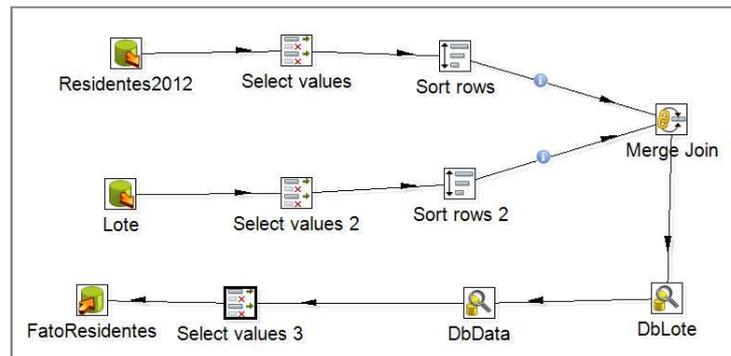


Figura 43: Transformação do processo de *ETL* para tabela Fato.

Com o *STDW* criado e carregado, na subetapa 5.2, foram elaborados os cubos com o *Schema Workbench*, para cada Fato foi criado um cubo, como mostra a Figura 44.

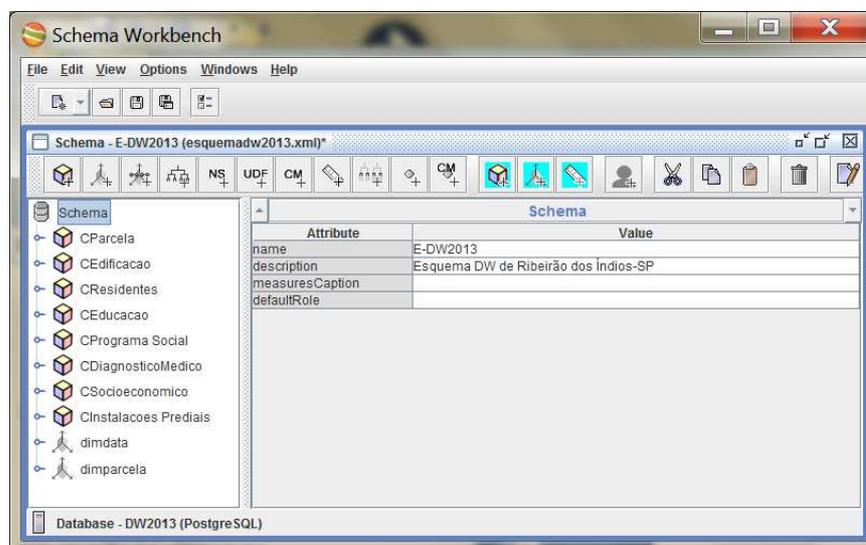


Figura 44: *Schema Workbench* apresentando os cubos.

A Figura 45 mostra os componentes do cubo Parcela, com a Fato Parcela, as Dimensões Logradouro, Data Lote, Data Dimensão e Proprietário, e as Medidas área do terreno, área construção principal, área de dependência e área total.

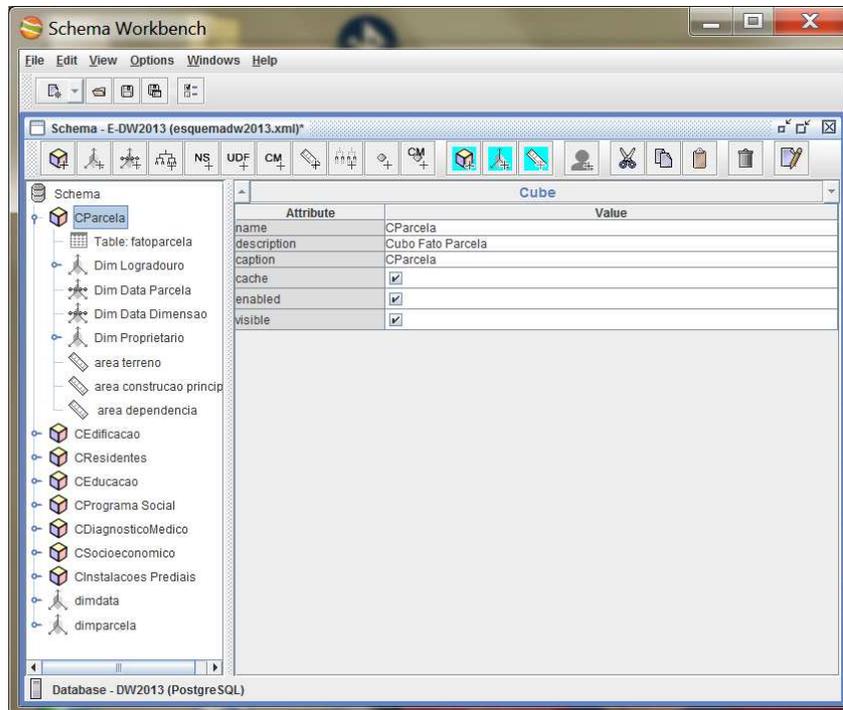


Figura 45: *Schema Workbench* com componentes do cubo Parcela.

Esse cubo, Parcela, pode ser visualizado no *BI Server* pelo pacote *jPivot*, visão analítica, que permite explorar os dados do cubo, como mostra a Figura 46. O *Sayku Analytics* também permite essa manipulação dos dados. Esses dois pacotes fornecem os recursos *OLAP* do *Data Warehousing*.

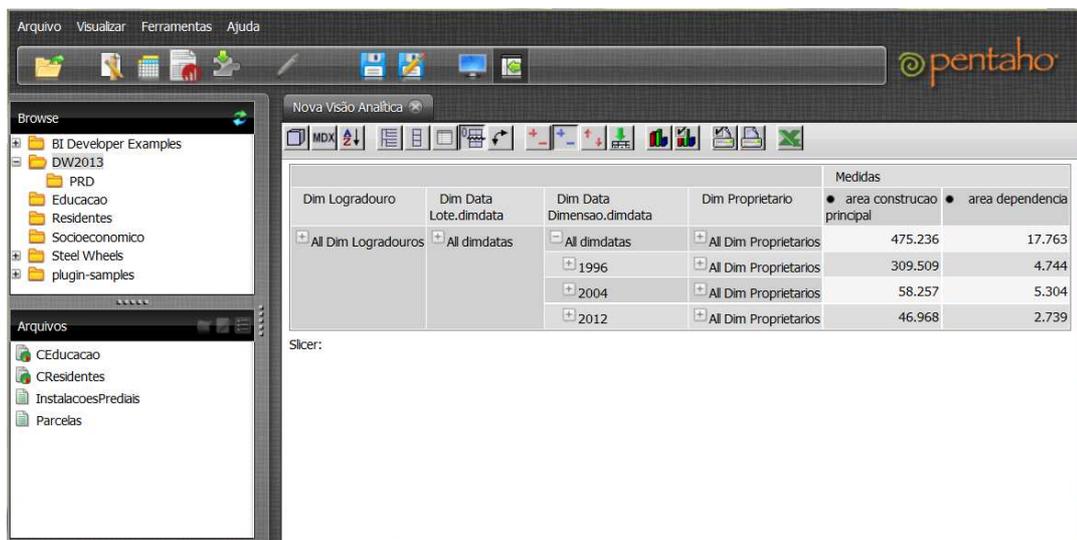


Figura 46: Interface do *jPivot* com o cubo Parcela.

Além da criação do cubo, nessa subetapa também foi utilizada a ferramenta *Metadata Editor* para a construção do metamodelo, um recurso do *BI Server* que permite ao usuário gerar seus próprios relatórios *ad hoc*. O metamodelo pode ser usado, também, como fonte de dados para o *Report Designer* para a construção de relatório *à priori*³⁶. A Figura 47 apresenta a interface do *Metadata Editor*, nessa figura aparece o metamodelo Fato Parcela.

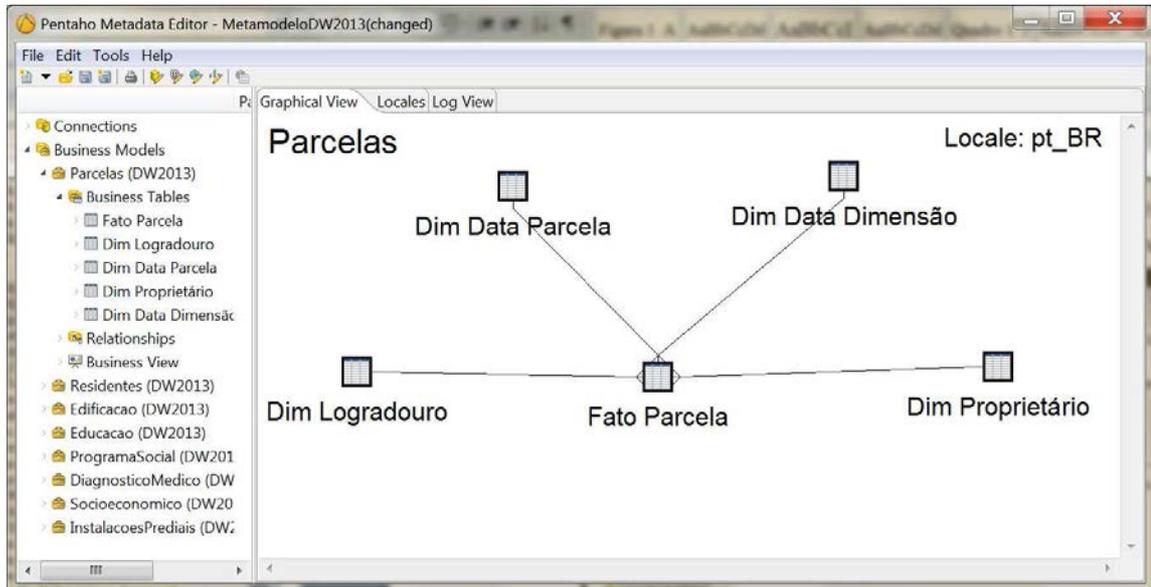


Figura 47: *Metadata Editor* com o metamodelo da Fato Parcela.

Esse metamodelo é disponibilizado pelo pacote *WAQR (Web Ad-Hoc Query & Reporting)* no *BI Server* como mostra a Figura 48, permitindo que o usuário crie seus próprios relatórios, oferecendo opções de formato, campos e filtros para a elaboração de relatórios *ad hoc*.

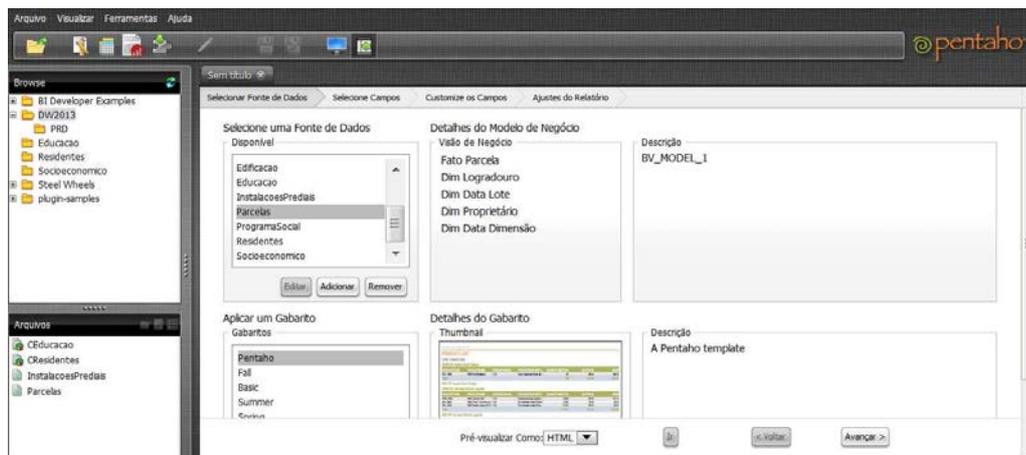


Figura 48: Interface *WAQR* com o metamodelo Fato Parcela.

³⁶ *à priori*: locução latina, com significado de "a partir do que é anterior", neste caso relatórios pré-formatados.

Com o *STDW* carregado, os processos de consultas foram disponibilizados para as análises, de acordo com a subetapa 5.3. Diversos relatórios foram gerados, *à priori*, com o *Report Designer*, para cada Fato (Parcela, Edificação, Residentes, Educação, Programa Social, Diagnóstico Médico, Socioeconômico e Instalações Prediais) foi elaborado um relatório geral com todos os atributos e com todos os anos de coleta de dados. As Figura 49 e Figura 50 mostram alguns desses relatórios, o primeiro sobre residentes, com as ocorrências referentes às parcelas, nos anos de 1996, 2004, 2010 e 2012 e o segundo sobre os programas sociais nos anos 2010 e 2012.

Residentes nas Parcelas												1 / 119
Anos 1996, 2004, 2010 e 2012												
SSQQLFF	Homens	Mulheres	Até 1 ano	De 1 a 3 anos	De 4 a 6 anos	De 7 a 9 anos	De 10 a 15 anos	De 16 a 21 anos	De 22 a 45 anos	De 46 a 60 anos	Mais de 60 anos	Ano
1020101	2	2							1	1		1996
	0	1	0	0	0	0	0	0	0	1		2004
	2	3	0	0	1	0	0	0	2	1	0	2010
1020102	3	1					2		1	1		1996
1020103	2	2			1		1		2			1996
1020104	4	1		1	1		1		2			1996
1020201	1	3				1	1		2			1996
	1	3	0	0	0	0	1	1	2	0		2004
	1	3	0	0	0	0	0	1	2	1	0	2010
	1	3	0	0	0	0	0	0	2	2	0	2012
1020301	1	1								2		1996
	0	1	0	0	0	0	0	0	0	1		2004
	1	1	0	0	0	0	0	0	1	1	0	2010
1020401	1	3	0	0	0	0	0	0	3	2	0	2012

Figura 49: Relatório de Residentes desenvolvido no *Report Desing*.

Programas Sociais						1 / 10
Anos 2010 e 2012						
SSQQLFF	Bolsa Familia	Renda Cidadã	Ação Jovem	Total de Benefícios	Ano	
1020301	0	1	0	1	2010	
1020701	1	0	0	1	2010	
	0	1	0	1	2012	
1021001	1	0	0	1	2010	
1040201	1	0	0	1	2012	
1040501	0	1	0	1	2012	
1041001	3	1	2	6	2010	
	0	0	1	1	2012	

Figura 50: Relatório de Programas Sociais desenvolvido no *Report Desing*.

A partir do *Data Warehousing*, muitos recursos foram disponibilizados para serem testados. A seguir, são apresentados alguns recursos que demonstram as informações espaço-temporais para apoio à tomada de decisão.

4.1 Consultas Espaciais, Temporais e Espaço-temporais

Nesta Seção são apresentadas algumas consultas escritas em *SQL*, demonstrando a aplicação dos conceitos espaciais, temporais e espaço-temporais, algumas dessas conforme a descrição apresentada por Castro (2007) na Subseção 2.3.1.

4.1.1 Consultas espaciais

Qual a localização da parcela '1010101'?

SQL:

```
select ssqllff, the_geom
from fatoparcela
where ssqllff = '1010101'
```

O resultado da consulta pode ser observado no mapa da Figura 51.



Figura 51: Resultado da consulta, parcela *ssqllff* = '1010101', no gvSIG.

Quais são os vizinhos da parcela '1051201'?

SQL:

```
select f1.ssqllff, f2.ssqllff as ssqllffvizinho,
f2.the_geom from fatoparcela f1, fatoparcela f2
where touches(f1.the_geom, f2.the_geom)
and (f2.ssqllff <> '1050401')
and (f1.ssqllff = '1050401')
```

O resultado da consulta pode ser observado no mapa da Figura 52.



Figura 52: Resultado da consulta, vizinhos da parcela '01050401', no gvSIG.

4.1.2 Consultas Temporais

Haviam aposentados e pensionistas no levantamento de 1996?

Consulta temporal simples (CASTRO, 2007).

SQL:

```
select      ssqllff,          datasocioeconomicoini,
apostados, pensionistas
from fatosocioeconomico
where
datasocioeconomicoini = '1996-01-01' and
apostados > '0' or
pensionistas > '0'
```

O resultado da consulta pode ser observado na Figura 53.

	ssqllff integer	datasocioeconomicoini date	apostados smallint	pensionistas smallint
1	102010	1996-01-01		1
2	102030	1996-01-01	2	
3	103010	1996-01-01	1	
4	103020	1996-01-01	1	
5	103050	1996-01-01		1
6	103060	1996-01-01	2	
7	103080	1996-01-01	1	

Unix Ln 6, Col 18, Ch 176 175 chars 277 rows.

Figura 53: Visualização da consulta *SQL* sobre aposentados e pensionistas, no PostgreSQL.

Haviam pensionistas nos levantamentos entre 1996 e 2010?

Consulta temporal de intervalo (CASTRO, 2007).

SQL:

```
select      ssqllff,          datasocioeconomicoini,
datasocioeconomicofim, pensionistas
from fatosocioeconomico
where
datasocioeconomicoini >= '1996-01-01' and
datasocioeconomicofim <= '2012-05-30' and
pensionistas > '0'
```

O resultado da consulta pode ser observado na Figura 54.

	ssqllff integer	datasocioeconomicoini date	datasocioeconomicofim date	pensionistas smallint
1	102010	1996-01-01	2003-12-31	1
2	103050	1996-01-01	2003-12-31	1
3	103130	1996-01-01	2003-12-31	1
4	105150	1996-01-01	2003-12-31	1
5	107020	1996-01-01	2003-12-31	1
6	107110	1996-01-01	2003-12-31	1
7	108090	1996-01-01	2003-12-31	1
8	201070	1996-01-01	2003-12-31	1
9	201110	1996-01-01	2003-12-31	1

OK. Unix Ln 5, Col 37, Ch 186 91 rows. 1

Figura 54: Visualização da consulta *SQL* sobre pensionistas, no PostgreSQL.

4.1.3 Consultas espaço-temporais

Em quais parcelas haviam aposentados e/ou pensionistas no levantamento de 1996?

Consulta espaço-temporal simples (CASTRO, 2007).

SQL:

```
select  ssqllff,  datasocioeconomicoini,  aposentados,
        pensionistas, the_geom
from fatosocioeconomico
where
datasocioeconomicoini = '1996-01-01' and
aposentados > '0' or
pensionistas > '0'
```

O resultado da consulta pode ser observado no mapa da Figura 55.

Em quais parcelas haviam aposentados nos levantamentos de 2004 e/ou 2010?

Consulta espaço-temporal de intervalo (CASTRO, 2007).

SQL:

```
select          ssqllff,          datasocioeconomicoini,
        datasocioeconomicofim, aposentados, the_geom
from fatosocioeconomico
where
datasocioeconomicoini >= '2004-01-01' and
datasocioeconomicofim <= '2012-05-30' and
aposentados > '0'
```

O resultado da consulta pode ser observado no mapa da Figura 56.

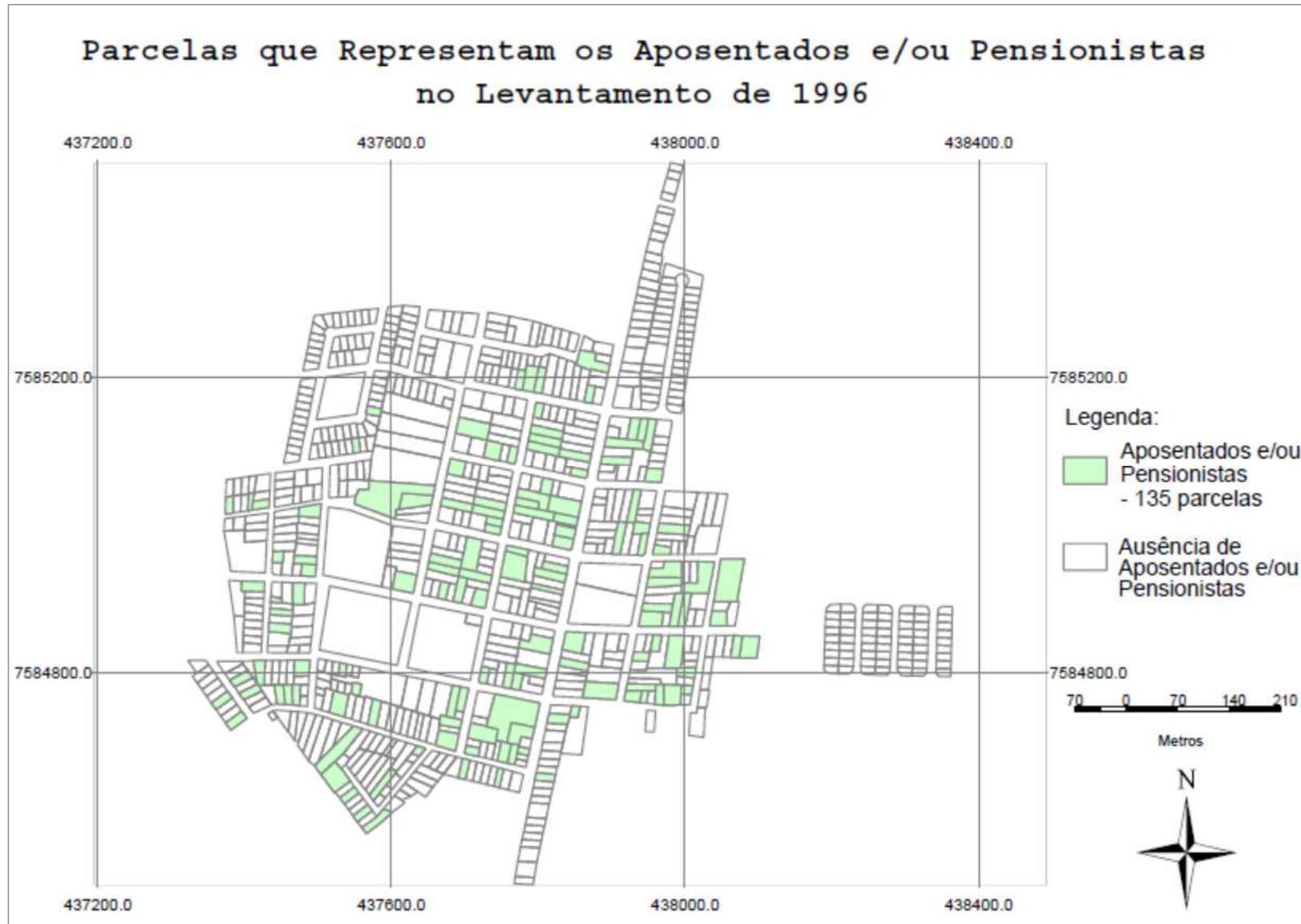


Figura 55: Mapa com a consulta *SQL* sobre aposentados e pensionistas, no gvSIG.

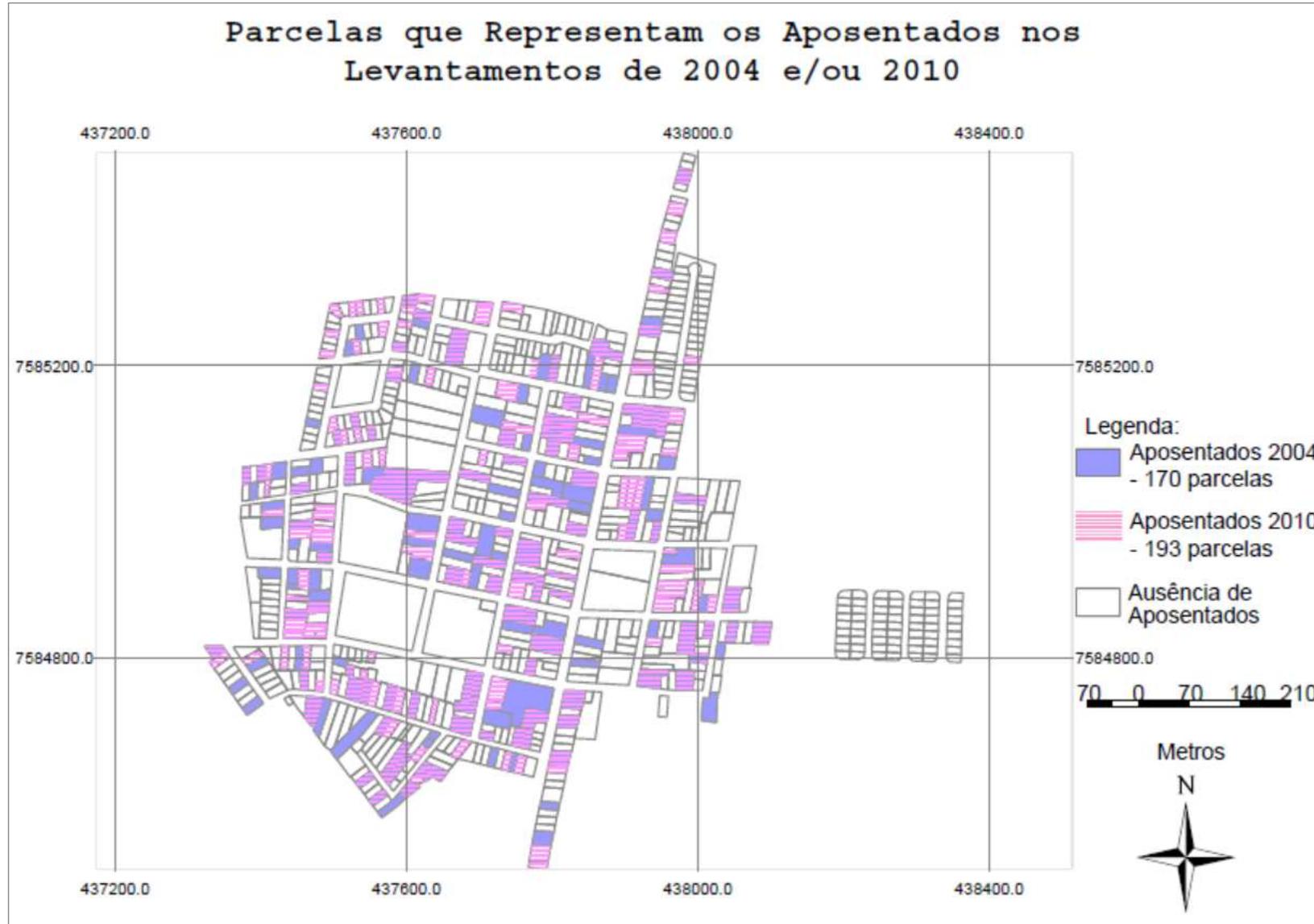


Figura 56: Mapa com a consulta *SQL* sobre aposentados, no gvSIG.

Em quais parcelas haviam aposentados no levantamento de 2004 e também em 2010?

Consulta espaço-temporal de intervalo (CASTRO, 2007).

SQL:

```
select  f1.ssqqllff,  f1.the_geom,  f1.aposentados  as
aposentados1996, f2.aposentados as aposentados2004
from fatosocioeconomico f1, fatosocioeconomico f2
where (intersects(f1.the_geom, f2.the_geom))
and f1.ssqqllff=f2.ssqqllff
and f1.datasocioeconomicoini = '2004-01-01'
and f2.datasocioeconomicoini = '2010-01-01'
and f1.aposentados>'0'
and f2.aposentados>'0'
```

O resultado da consulta pode ser observado no mapa da Figura 57.

Quais parcelas foram desmembradas entre os levantamentos de 2010 e 2012?

SQL:

```
select d1.ssqqllff, d1.the_geom
from dimparcela d1, dimparcela d2
where d1.ssqqllff=d2.ssqqllff
and area(d1.the_geom)< area(d2.the_geom)
and d1.date_from = '2010-01-01'
and d2.date_from = '2012-05-30'
```

O resultado da consulta, com dados simulados, pode ser observado na Figura 58.

Quais parcelas foram lembradas entre os levantamentos de 2010 e 2012?

SQL:

```
select d1.ssqqllff, d1.the_geom
from dimparcela d1, dimparcela d2
where d1.ssqqllff=d2.ssqqllff
and area(d1.the_geom)> area(d2.the_geom)
and d1.date_from = '2010-01-01'
and d2.date_from = '2012-05-30'
```

O resultado da consulta, com dados simulados, pode ser observado na Figura 59.

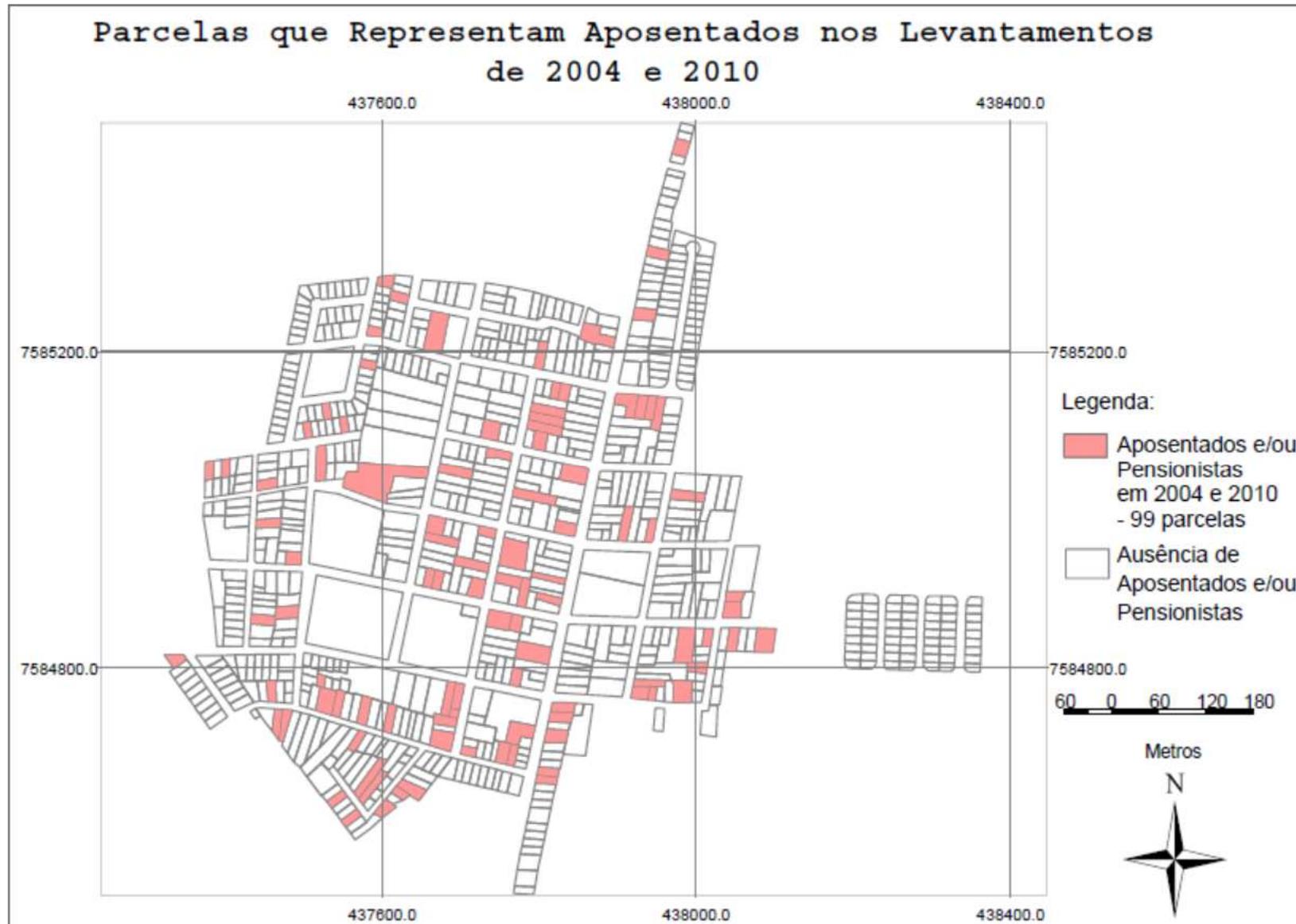


Figura 57: Mapa com a consulta *SQL* sobre aposentados em 2004 e 2010, no gvSIG.

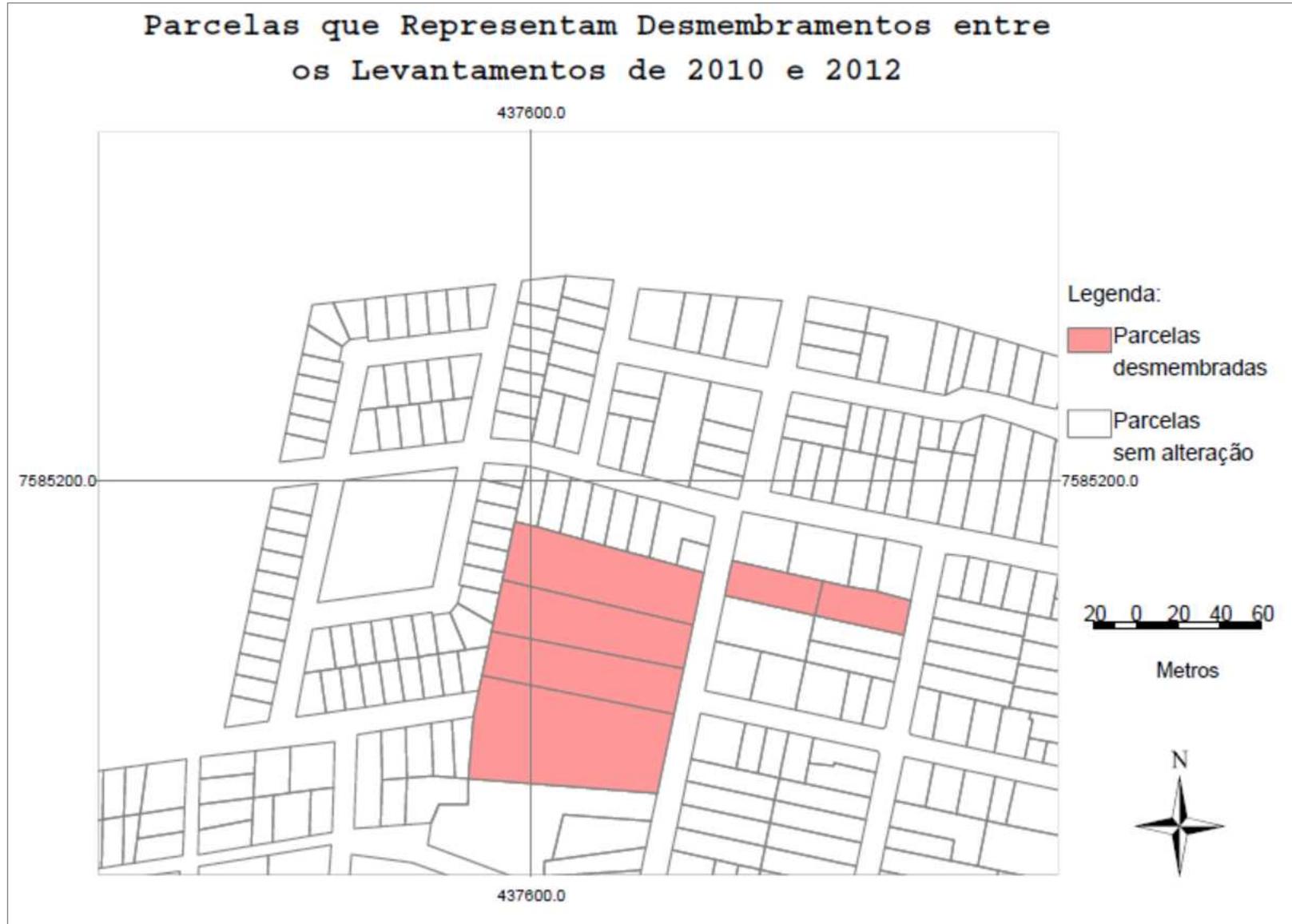


Figura 58: Mapa com a consulta *SQL* sobre desmembramento entre 2010 e 2012, no gvSIG.

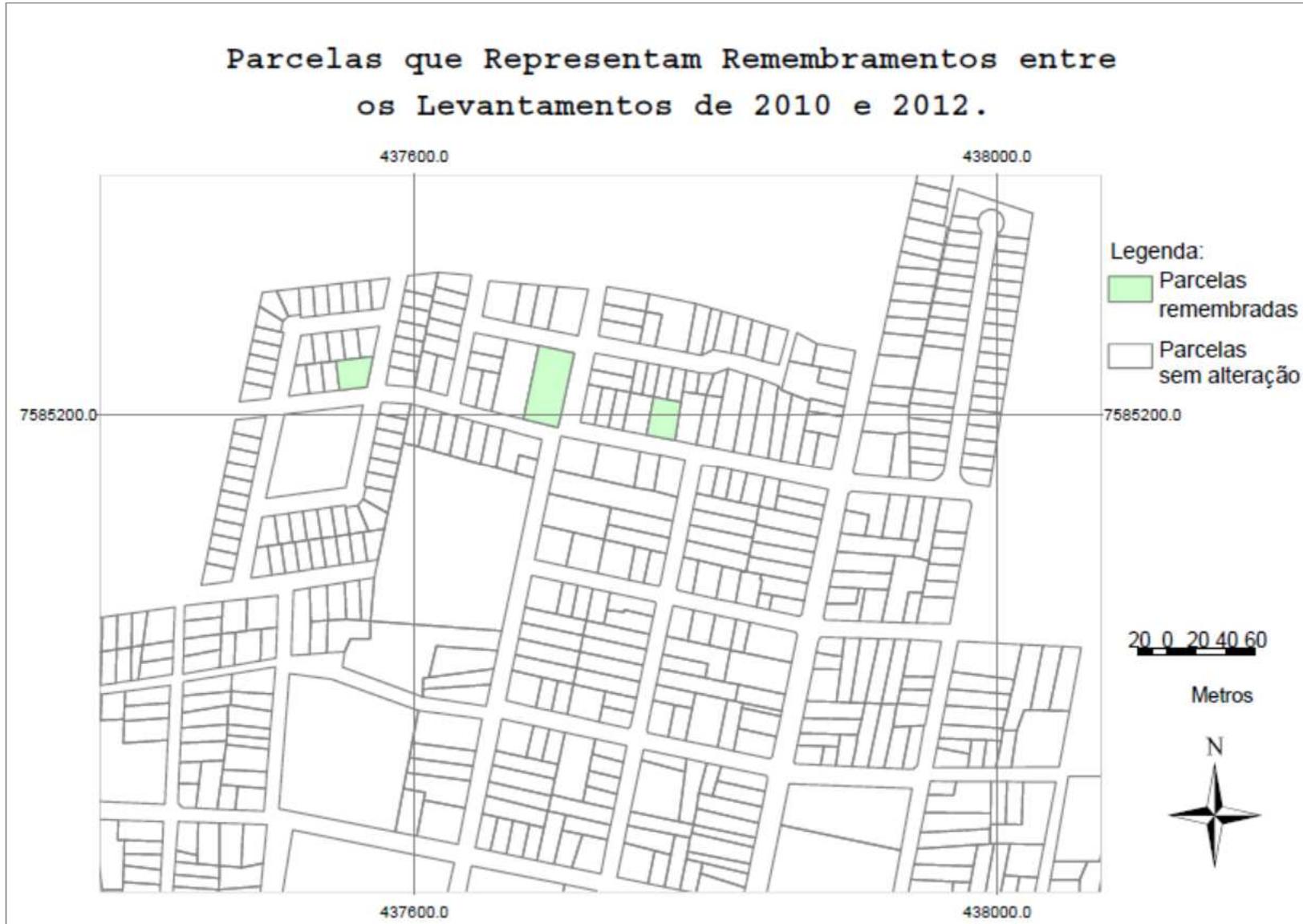


Figura 59: Mapa com a consulta *SQL* sobre remembramento entre 2010 e 2012, no gvSIG.

4.2 Análise espaço temporal: Área construída.

Nesta análise busca-se identificar a alteração de área construída, para isso as seguintes questões foram elaboradas:

- Qual é a diferença da área construída?
- Em quais parcelas ocorrem?
- Entre os levantamentos de 2004 - 2010 e 2010 - 2012

As Figura 60 e Figura 61 apresentam parte dos relatórios gerados pelo *Report Design*. Para este relatório uma consulta *SQL* foi elaborada, totalizando a área construída (área principal somada à área de dependências) entre os períodos (2004 - 2010) e (2010 - 2012). Com os totais por ano foi calculada a diferença entre esses dois valores, revelando se a área construída aumentou ou diminuiu no período selecionado.

Alteração da Área Construída nas Parcelas				1 / 18
Anos 2004 e 2010				
SSQQLFF	Área Total 2004	Área Total 2010	Diferença de Área	
1020101	135	118	-17	
1020201	111	103	-8	
1020301	80	85	5	
1020401	115	81	-34	
1020501	159	120	-39	
1020601	515	442	-73	
1020701	176	178	2	
1020801	133	0	-133	
1020901	431	140	-291	
1021001	133	133	0	
1021201	92	70	-22	

Figura 60: Relatório: Alteração da Área Construída nas Parcelas, 2004 e 2010.

Alteração da Área Construída nas Parcelas				1 / 12
Anos 2010 e 2012				
SSQQLFF	Área Total 2010	Área Total 2012	Diferença de Área	
1020201	103	117	14	
1020401	81	175	94	
1020701	178	15	-163	
1021201	70	94	24	
1021301	116	118	2	
1021401	60	66	6	
1030101	77	90	13	
1030201	117	124	7	
1030301	137	150	13	
1030401	151	94	-57	

Figura 61: Relatório: Alteração da Área Construída nas Parcelas, 2010 e 2012.

Após as consultas apresentadas nos relatórios das Figura 60 e Figura 61 foram elaborados os mapas no gvSIG, como mostram as Figura 63 e Figura 64. No mapa do período de 2004 a 2010 (Figura 63), seiscentos e cinquenta e duas parcelas possuíam dados sobre áreas nos dois anos consultados. Dessas parcelas, cento e sessenta e uma parcelas tiveram um decréscimo de área construída, trezentos e cinquenta e nove mantiveram a área e cento e trinta e duas parcelas aumentaram a área.

No mapa do período de 2010 a 2012 (Figura 64) quatrocentos e setenta e três parcelas possuíam dados sobre áreas nos dois anos consultados. Dessas parcelas, cento e nove parcelas tiveram um decréscimo de área construída, cento e quarenta e sete mantiveram a área e cento e oitenta e sete parcelas aumentaram a área.

4.3 Análise espaço temporal: Evolução das patologias cardiopatia, hipertensão arterial e diabetes.

Nesta análise o objetivo foi identificar em quais parcelas haviam pessoas com problemas de saúde, no caso com cardiopatia, hipertensão arterial e diabetes, para isso as seguintes questões foram elaboradas:

- Existem ocorrências de hipertensão, cardiopatia e diabetes?
- Em quais parcelas ocorrem?
- No período de 1996 a 2012.

A Figura 62 apresenta parte do relatório gerado pelo *Report Design*. Para esse relatório uma consulta *SQL* foi elaborada identificando em quais parcelas haviam residentes com alguma das patologias: cardiopatia, hipertensão arterial ou diabetes, entre o período de 1996 a 2012.

Diagnostico Médico				
Cardiopatia, Hipertensão Arterial e Diabetes				
Anos 1996, 2004, 2010 e 2012				
SSQQLFF	Cardiopatias	Hiper. Arterial	Diabetes	Ano
1020101	0	1	0	1996
1020201	0	1	0	2012
1020301	0	1	1	2004
1020501	0	1	1	1996
	0	1	1	2004
	1	1	1	2010

Figura 62: Relatório: Ocorrência de Cardiopatias, Hipertensão Arterial e Diabetes. Anos 1996, 2004, 2010 e 2012.

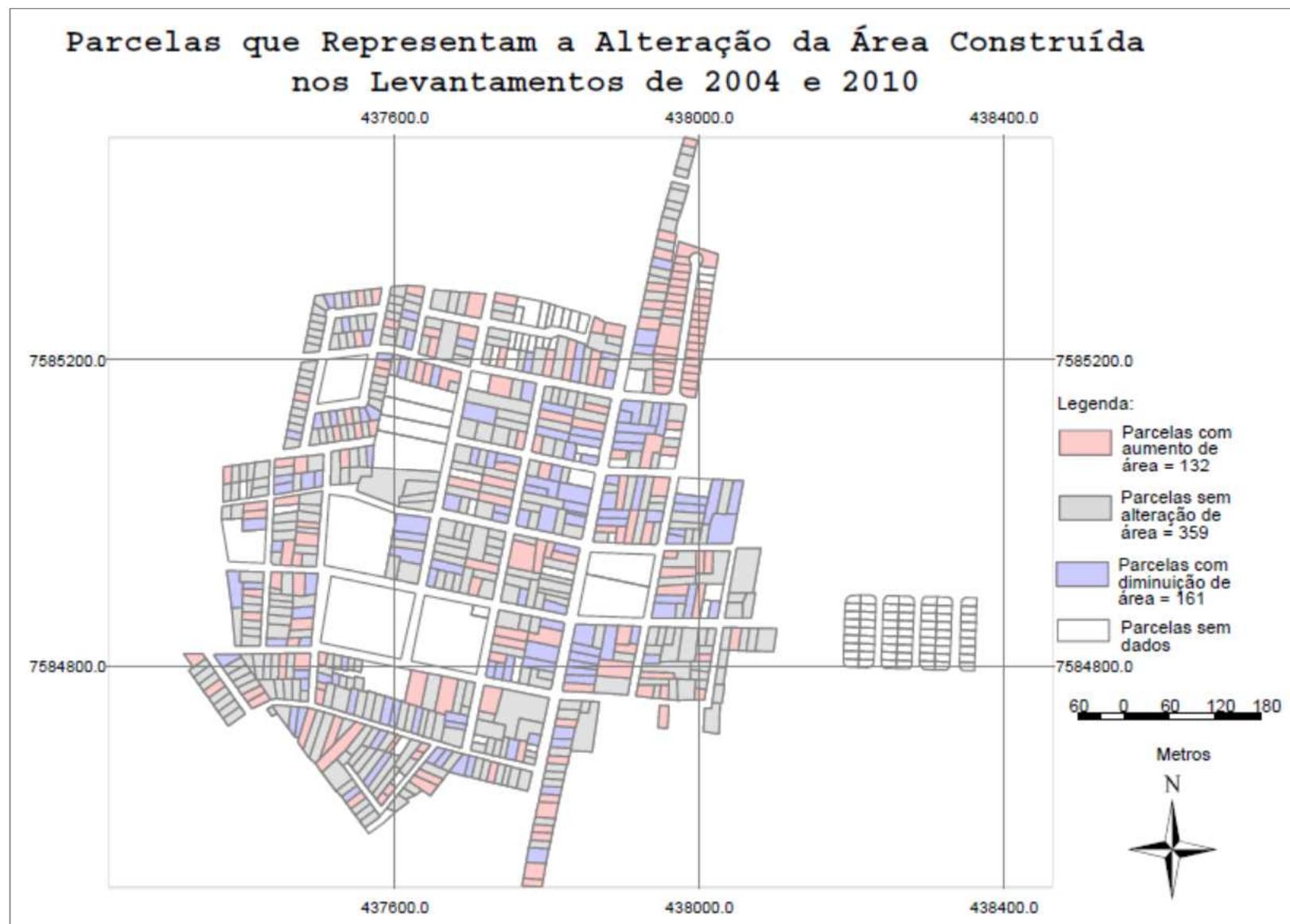


Figura 63: Mapa sobre a alteração de área construída entre 2004 e 2010.

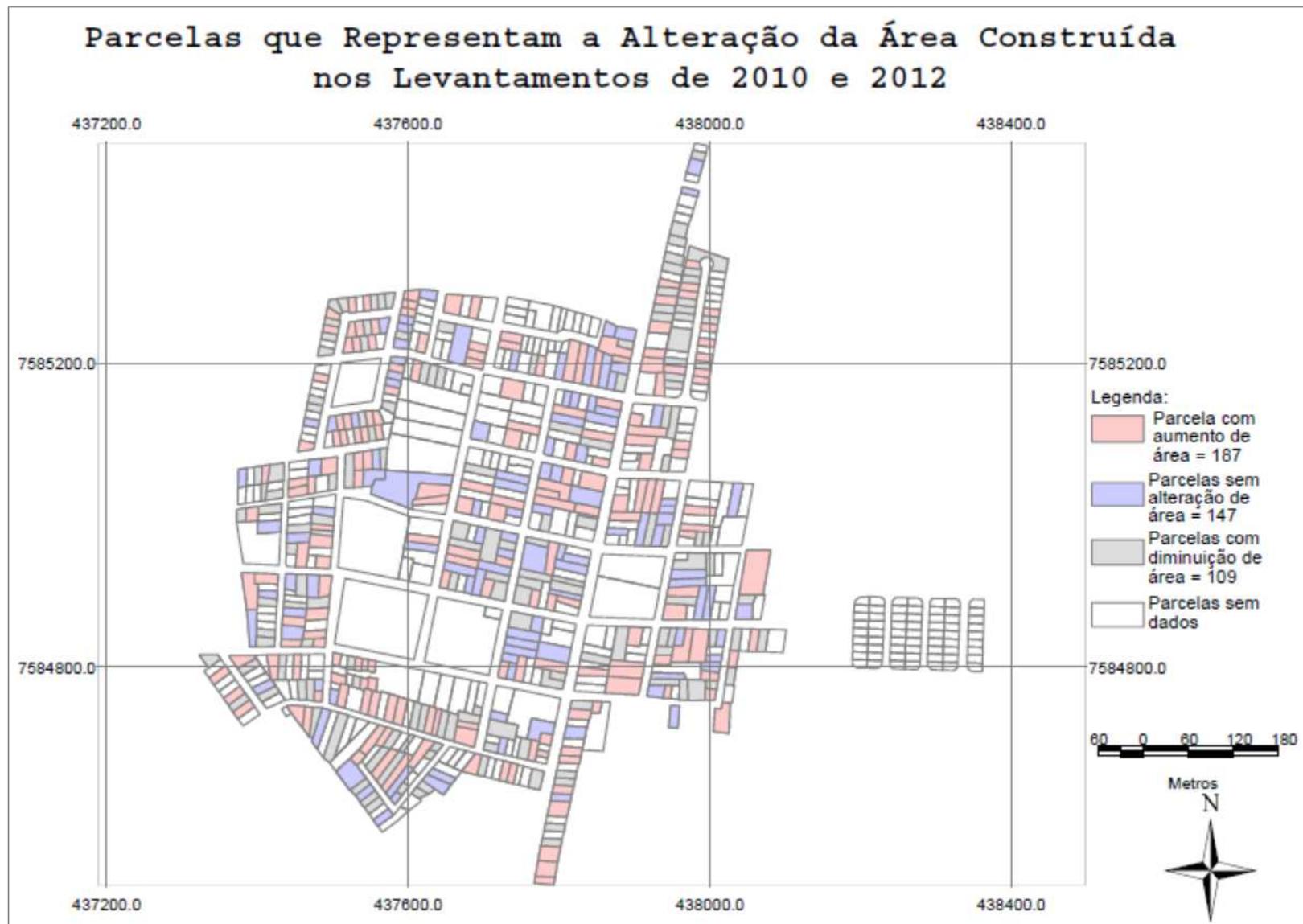


Figura 64: Mapa sobre a alteração de área construída entre 2010 e 2012.

A partir da consulta do relatório da Figura 62 foram elaborados os mapas no gvSIG, como mostram as Figura 66, Figura 67 e Figura 68 . No mapa apresentado na Figura 66, são apresentadas as parcelas que possuíam residentes com cardiopatia no período consultado. Em 1996, trinta e seis parcelas possuíam residentes com cardiopatia, em 2004, trinta e sete parcelas, em 2010 e 2012 quarenta e sete parcelas.

No mapa apresentado na Figura 67, são apresentadas as parcelas que possuíam residentes com hipertensão arterial no período consultado. Em 1996, cento e vinte e três parcelas possuíam residentes com hipertensão arterial, em 2004, duzentas e uma, em 2010, cento e setenta e seis, e em 2012, duzentas e uma.

No mapa apresentado na Figura 68, são apresentadas as parcelas que possuíam residentes com diabetes no período consultado. Em 1996, vinte e duas parcelas possuíam residentes com diabetes, em 2004, cinquenta e três, em 2010, setenta e uma, e em 2012, oitenta.

A Figura 65 apresenta um recorte do mapa da Figura 68, com essa aproximação fica mais fácil visualizar a sobreposição da informação de cada parcela.



Figura 65: Recorte do mapa de Ocorrência de Diabetes.

A Figura 69 apresenta a tela do *BI Server* com a ferramenta *Sayku Analytic*. Nessa figura é possível observar os dados de cardiopatia, hipertensão arterial e diabetes na sua totalização por ano. Nos mapas anteriores, visualizam-se as parcelas que possuem residentes com as patologias referidas, mas em uma parcela pode ocorrer mais de um residente com a mesma patologia.

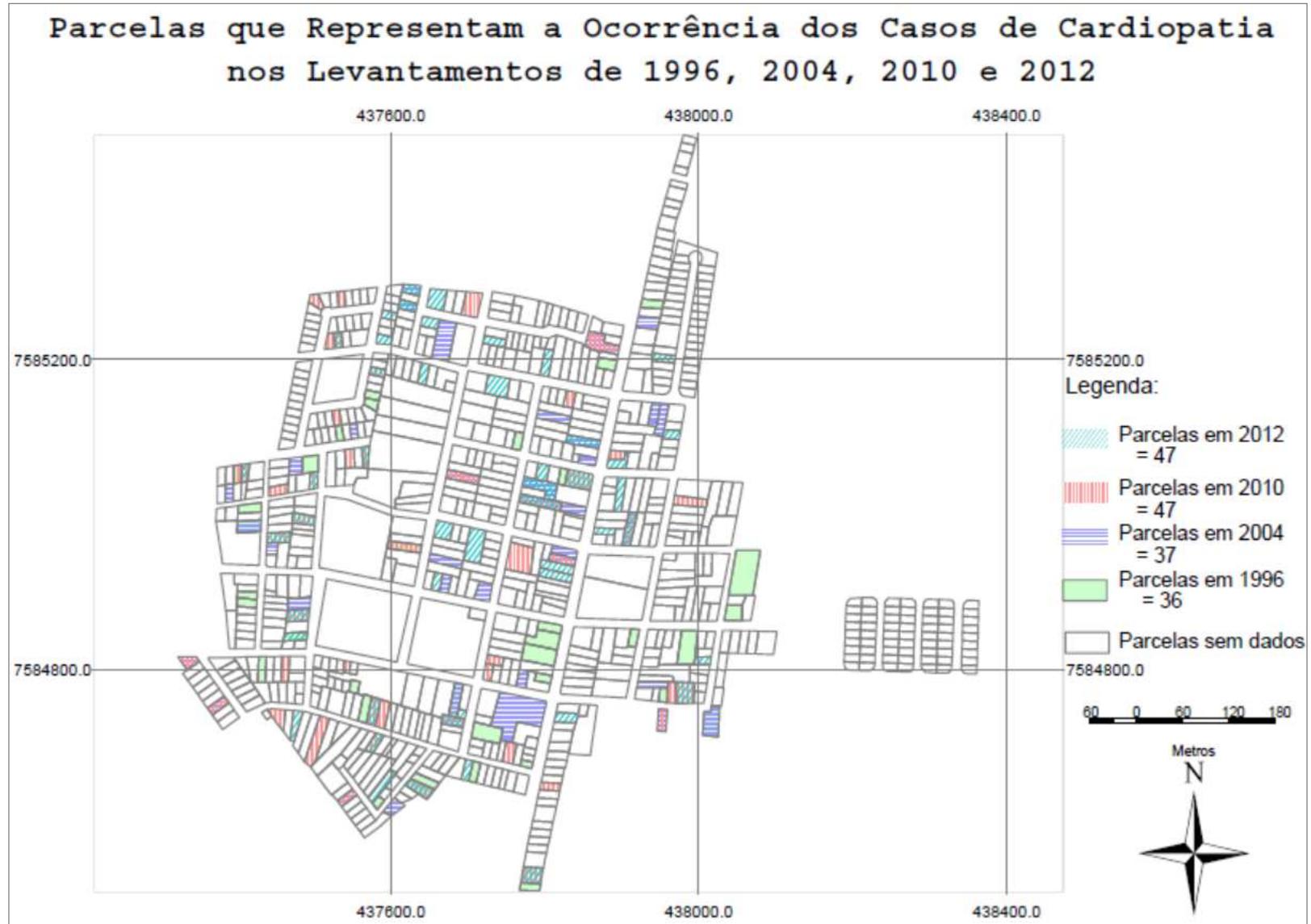


Figura 66: Mapa da ocorrência de Cardiopatia em 1996, 2004, 2010 e 2012.

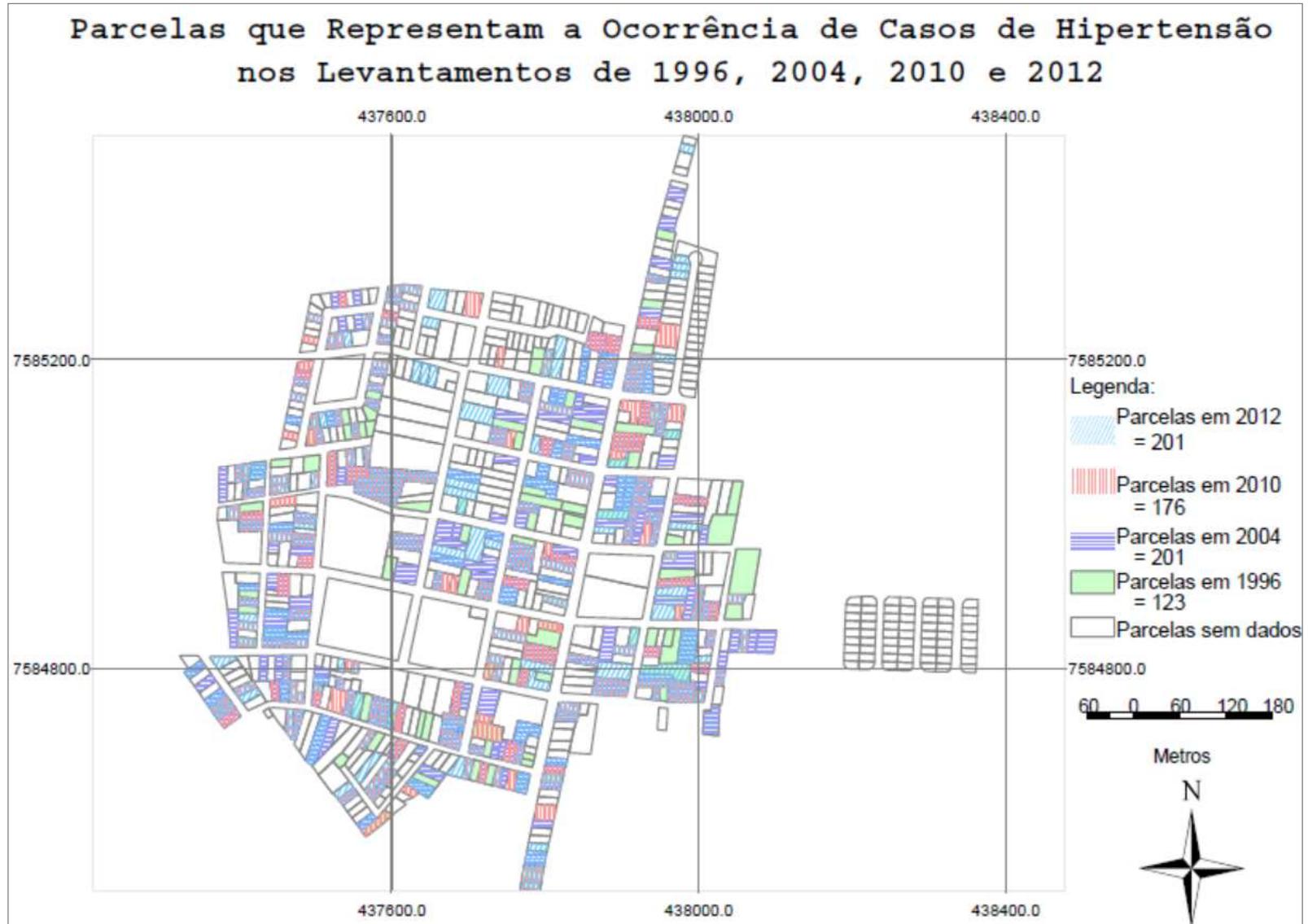


Figura 67: Mapa sobre a ocorrência de Hipertensão em 1996, 2004, 2010 e 2012.

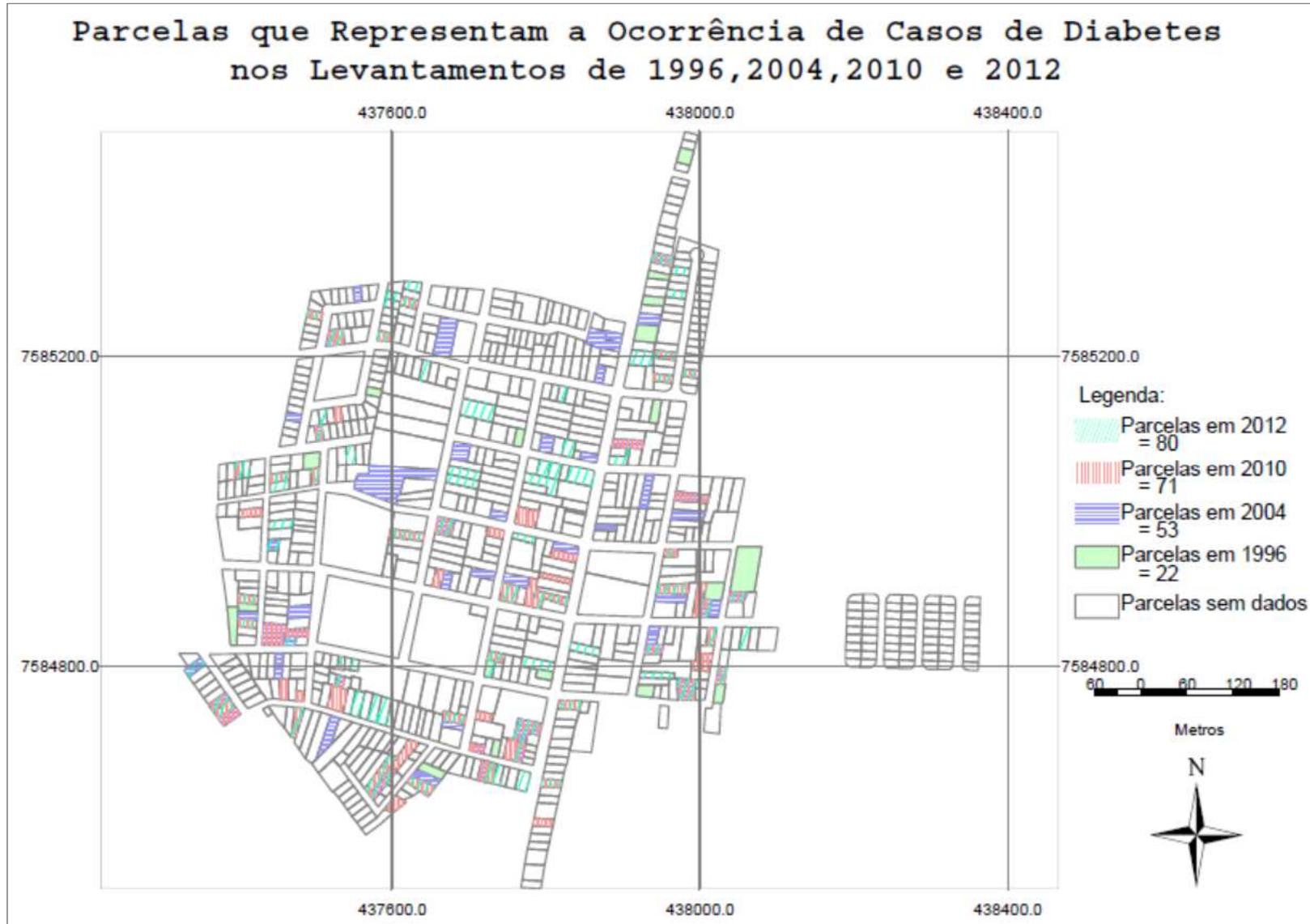


Figura 68: Mapa da ocorrência de Diabetes em 1996, 2004, 2010 e 2012.

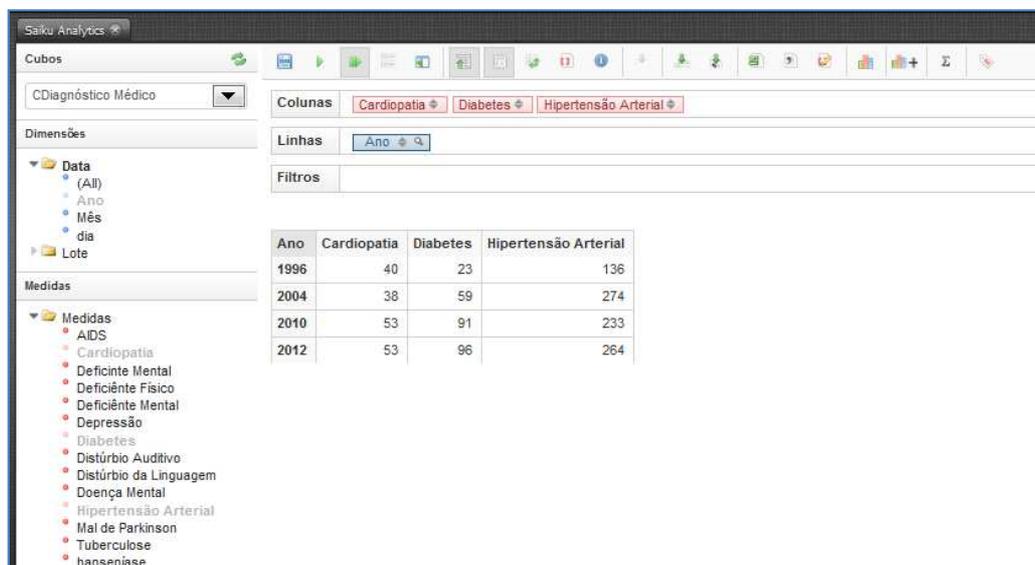


Figura 69: Visualização dos dados tabulados no *Saiku Analytics*.

A ferramenta *Saiku Analytic*, também permite visualizar os dados apresentados na Figura 69, a partir de gráficos, como mostra a Figura 70.

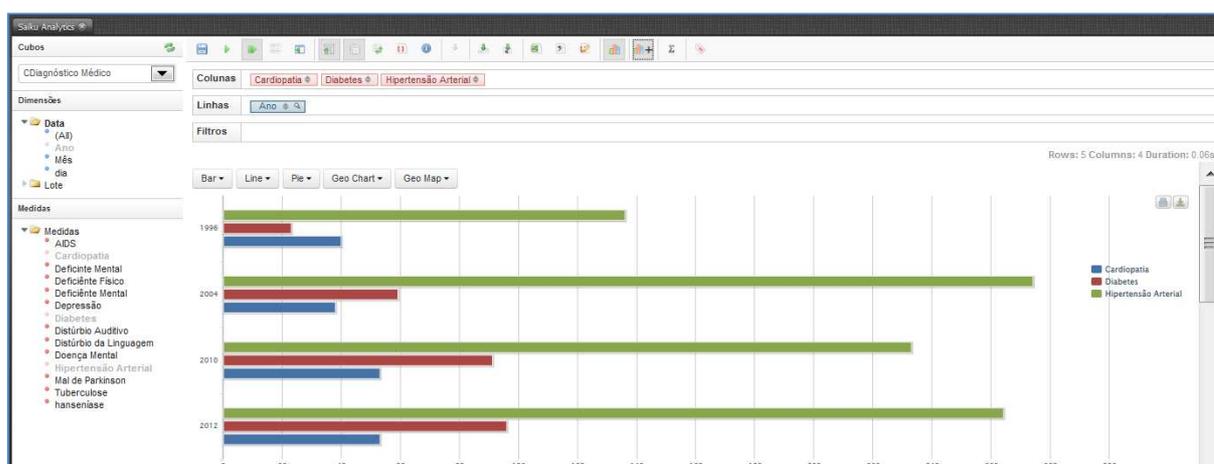


Figura 70: Visualização dos dados em gráfico no *Saiku Analytics*.

4.4 Análise espaço temporal: Ensino fundamental e creche.

Esta análise apresenta informações sobre alunos matriculados em creches e no ensino fundamental, para essa análise as seguintes questões foram elaboradas:

- Existem alunos que estudam no nível fundamental e usuários de creche?
- Em quais parcelas ocorrem?
- No período de 2004 – 2012.

A Figura 71 apresenta parte do relatório gerado pelo *Report Design*. Para esse relatório uma consulta *SQL* foi elaborada identificando em quais parcelas haviam residentes que eram

usuários de creche e/ou estudam no ensino fundamental, entre o período de 2004 - 2012, e o total de residentes que estudavam.

<u>Dados Educação - Alunos de Creche e Fundamental</u>					1 / 20
Anos 2004, 2010 e 2012					
SSQQLFF	Creche	Fund. Público	Fund. Particular	Total de Alunos	Ano
1020101	0	1	0	3	2010
1020201		1	0		2004
1020301	0	2	0	2	2010
1020501		1	0		2004
1020701	0	3	0	4	2010
	0	2	0	2	2012
1021001		1	0		2004
	0	2	0	2	2010
1021201					

Figura 71: Relatório: Alunos matriculados em creche e no ensino fundamental, nos anos de 2004, 2010 e 2012.

A partir da consulta apresentada no relatório da Figura 71 foram elaborados os mapas no gvSIG, como mostram as Figura 72, Figura 73, Figura 74 e Figura 75. No mapa apresentado na Figura 72, são apresentadas as parcelas que possuíam residentes usuários de creche ou alunos do ensino fundamental no ano de 2004.

O mapa da Figura 73 apresentada às parcelas que possuíam residentes usuários de creche e/ou alunos do ensino fundamental no ano de 2010.

No mapa apresentado na Figura 74, são apresentadas as parcelas que possuíam residentes usuários de creche e/ou alunos do ensino fundamental no ano de 2012.

No mapa apresentado na Figura 75, são apresentadas as parcelas que possuíam residentes usuários de creche e/ou alunos do ensino fundamental no período entre 2004 e 2012.

No mapa apresentado na Figura 76, são apresentadas as interseções entre as parcelas que possuíam residentes usuários de creche e/ou alunos do ensino fundamental no período entre 2004 - 2010 e 2010 - 2012.

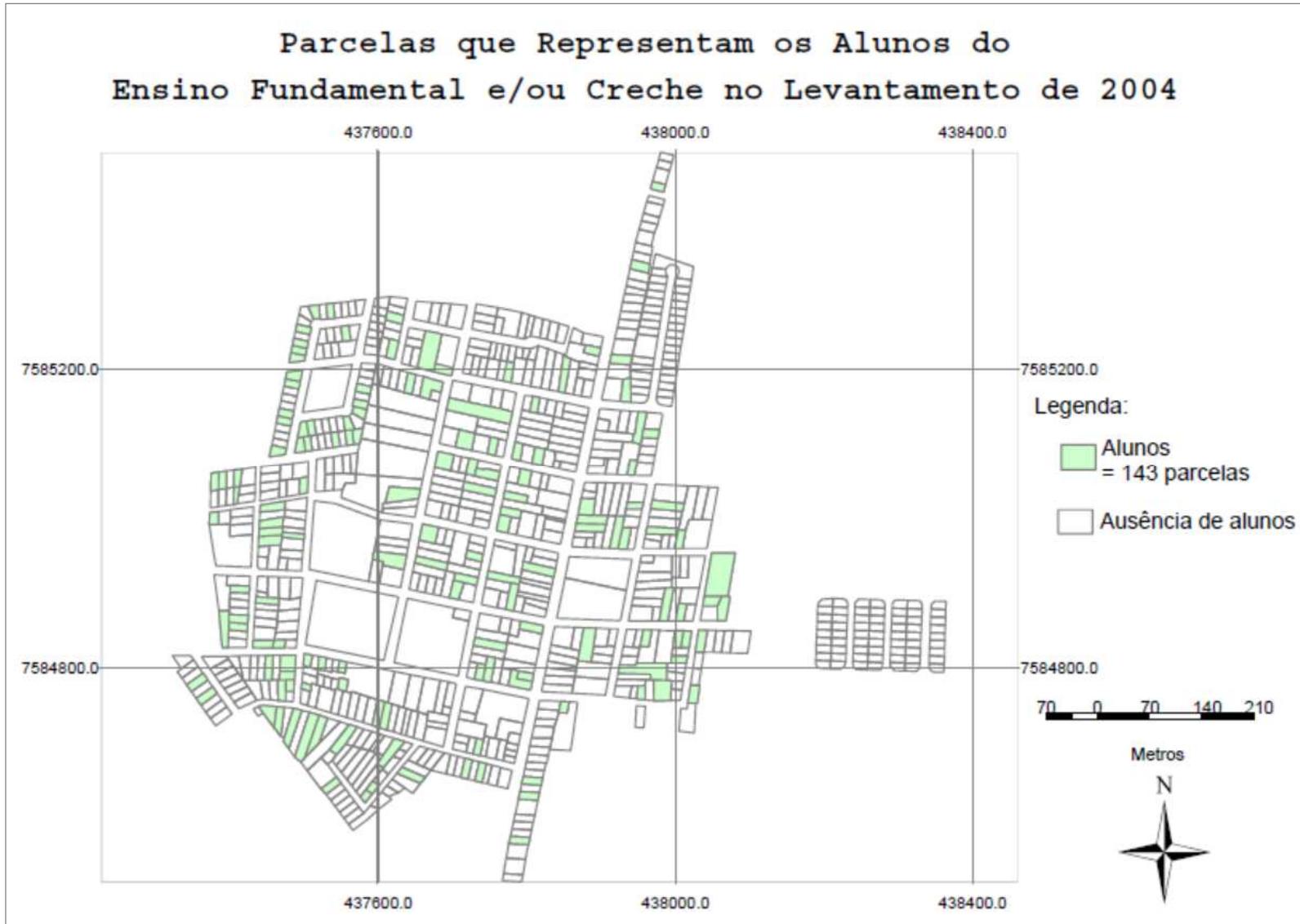


Figura 72: Representação gráfica das parcelas com alunos matriculados no ensino fundamental e creche em 2004.

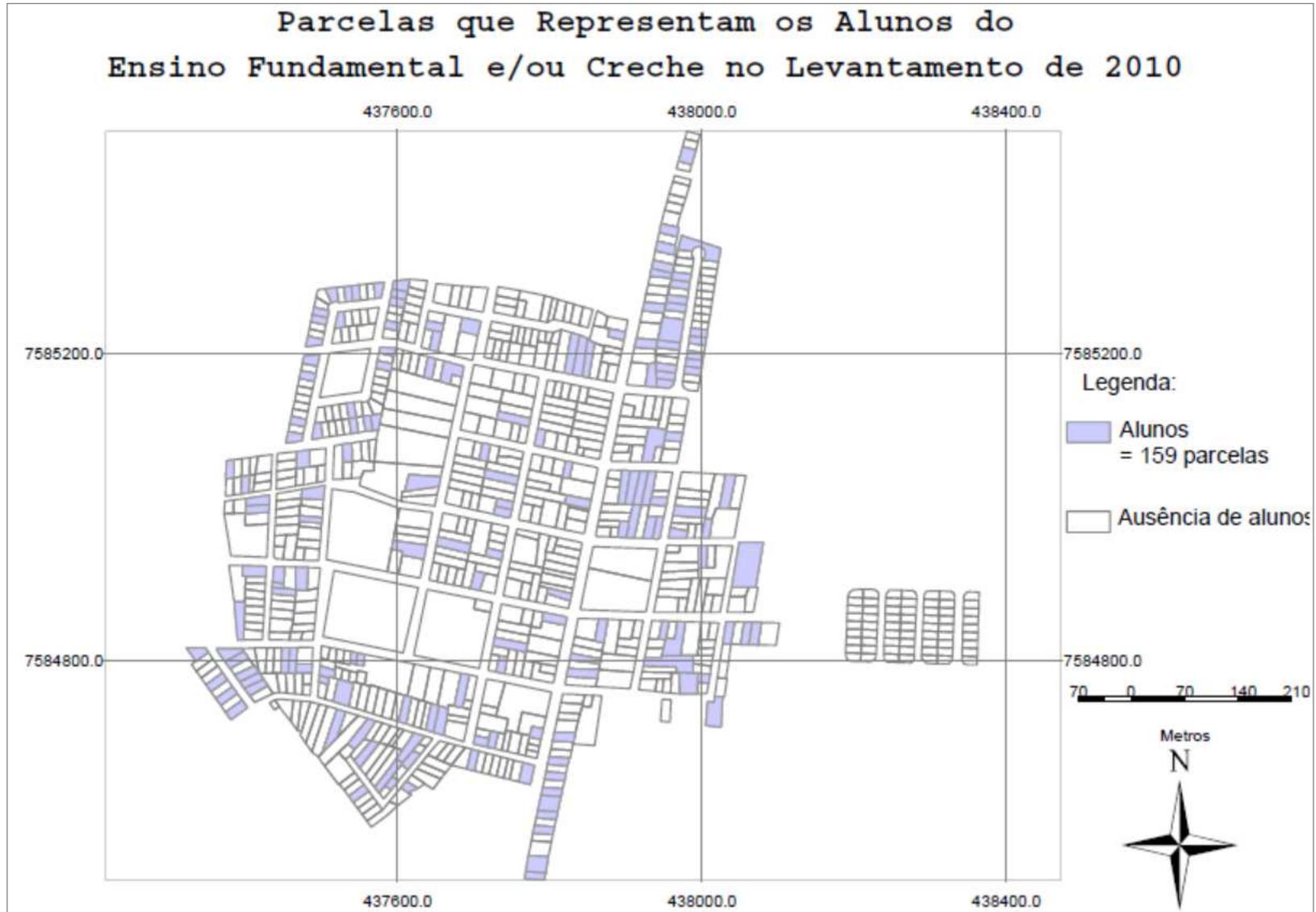


Figura 73: Representação gráfica das parcelas com alunos matriculados no ensino fundamental e creche em 2010.

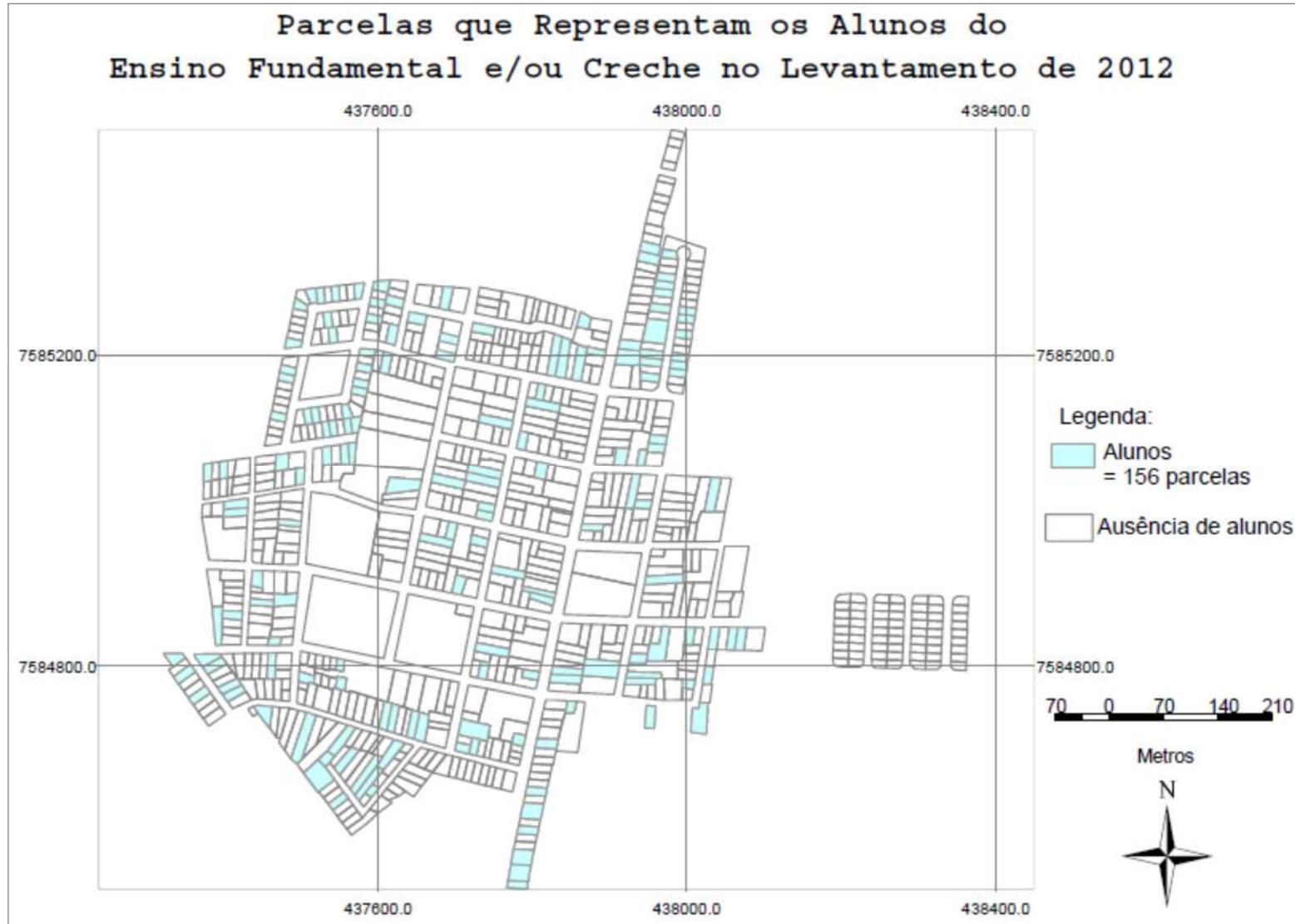


Figura 74: Mapa das parcelas com alunos matriculados no ensino fundamental e creche em 2012.

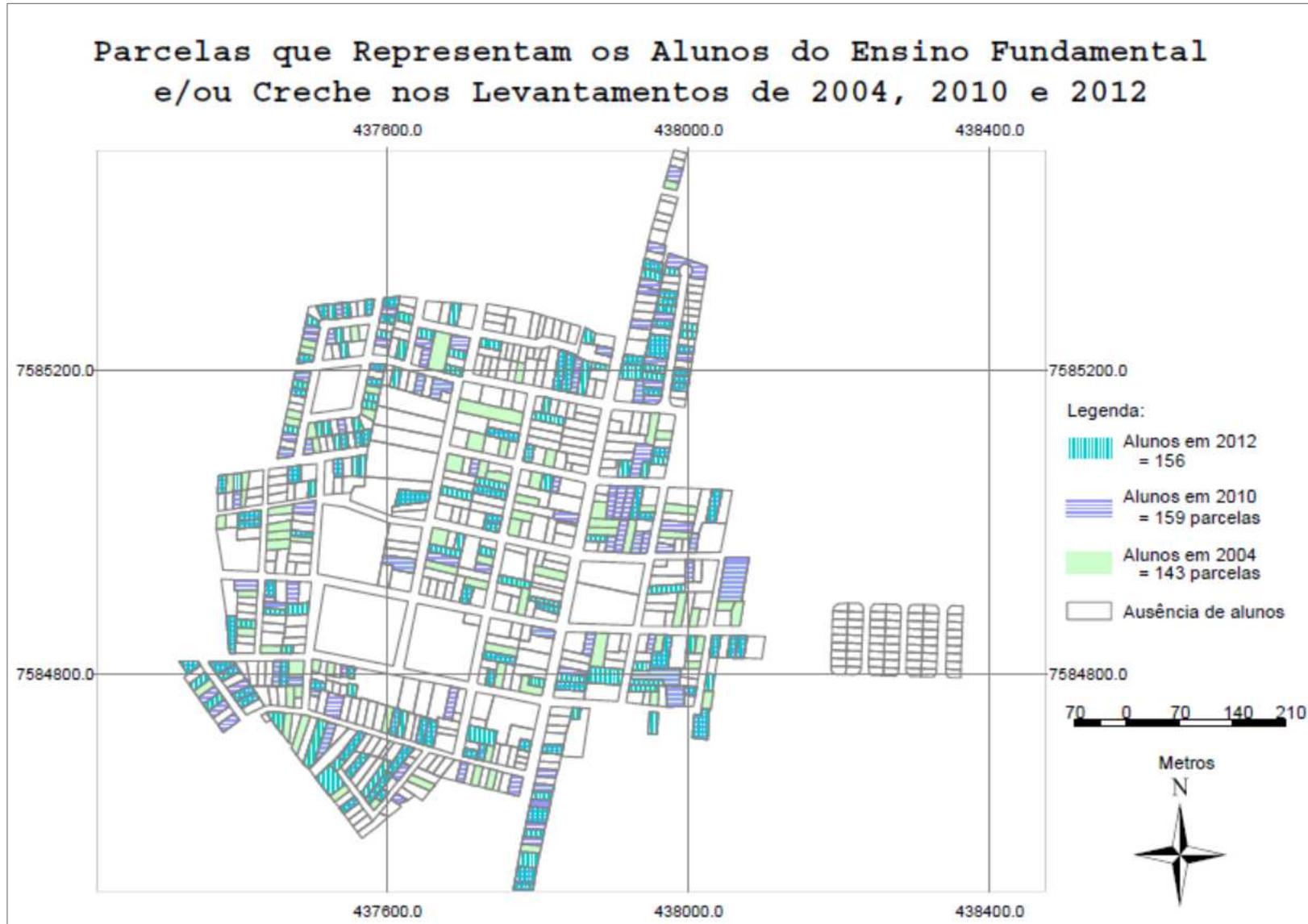


Figura 75: Mapa com as parcelas com alunos matriculados no ensino fundamental e creche em 2004, 2010 e 2012.

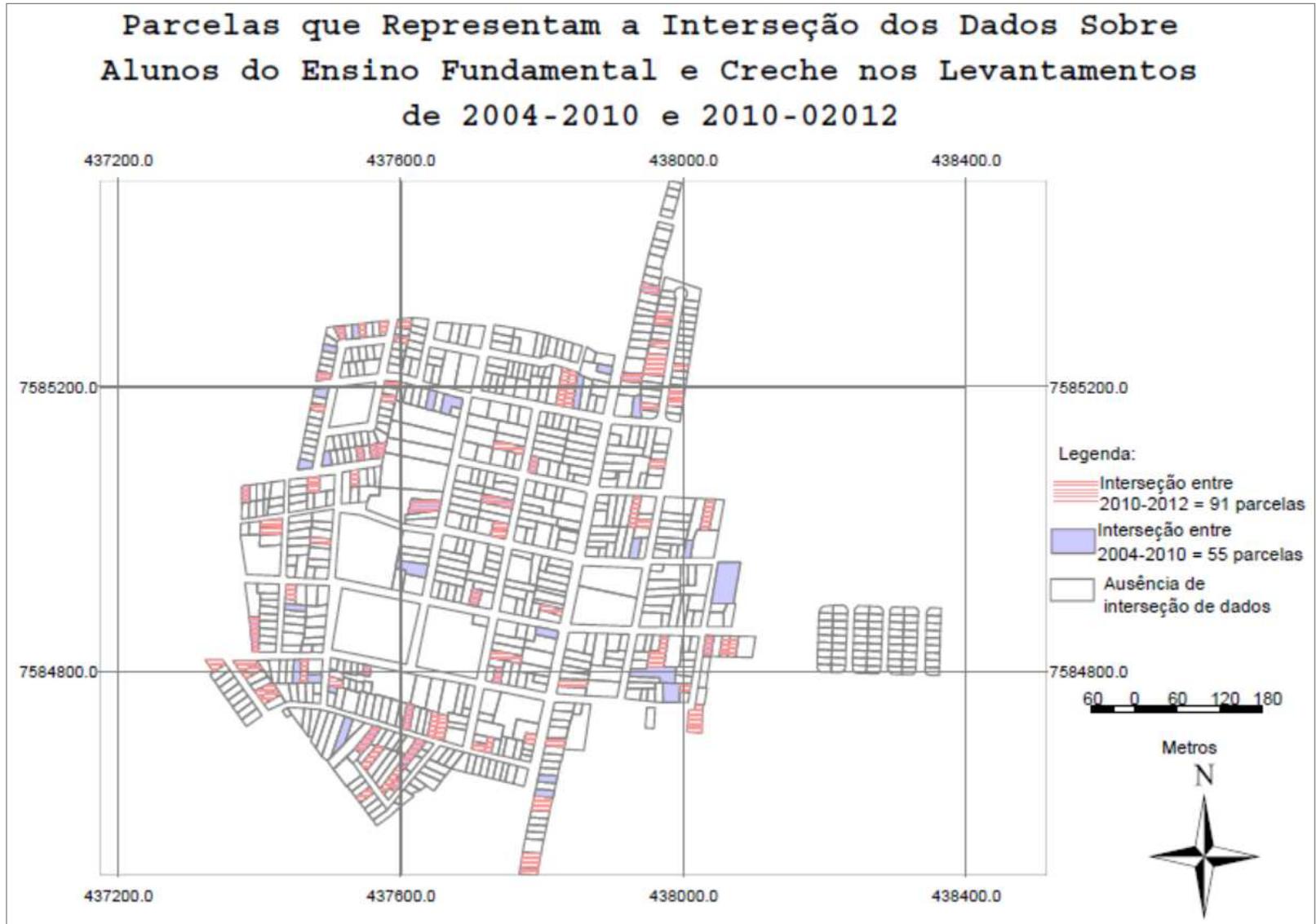


Figura 76: Mapa com interseção das parcelas: 2004 - 2010 e 2010 - 2012.

A Figura 77 apresenta a tela do *BI Server* com a ferramenta *Saiku Analytic*. Nessa figura é possível ver os dados de educação na sua totalização por ano. Nos mapas anteriores, visualizaram-se as parcelas que possuem residentes usuários de creche e/ou matriculados no ensino fundamental, porém em uma parcela pode ocorrer mais de um residente atendido pela educação básica.

The screenshot shows the Saiku Analytics interface with the following configuration:

- Cubos:** CEducacao
- Dimensões:** Dim Data, Lote, dimdata
- Medidas:** Analfabetos, Creche, Fundamental Particular, Fundamental Publico, Medio Particular, Medio Publico, Superior Particular, Superior Publico
- Colunas:** Analfabetos, Creche, Fundamental Particular, Fundamental Publico, Medio Particular, Medio Publico
- Linhas:** Superior Particular, Superior Publico
- Filtros:** Ano

The data table displayed is as follows:

Ano	Analfabetos	Fundamental Publico	Fundamental Particular	Medio Publico	Medio Particular	Superior Publico	Superior Particular	Creche
2004	171	203	3	89	2	4	16	
2010	116	214	22	52	6	8	40	29
2012	123	201	17	53	5	9	33	43

Figura 77: Visualização dos dados tabulados no *Saiku Analytic*.

A ferramenta *Saiku Analytic*, também permite visualizar os dados apresentados na Figura 77, a partir de gráficos, como mostra a Figura 78.

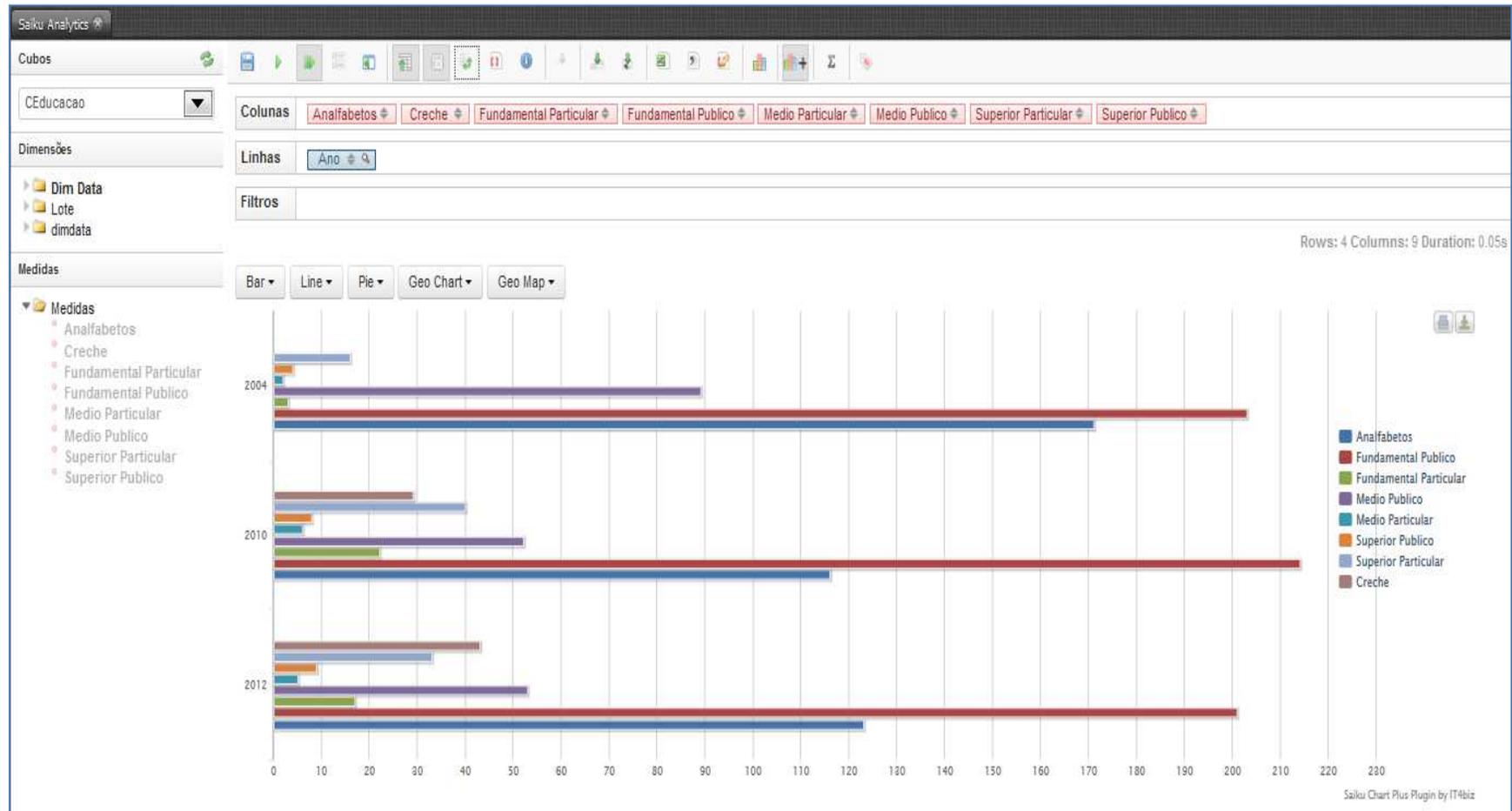


Figura 78: Visualização dos dados em gráfico no *Saiku Analytics*.

5 DISCUSSÃO

O CTM no Brasil é considerado novo, principalmente pela falta de legislação, entretanto as pesquisas acadêmicas vêm desenvolvendo vários trabalhos para mudar essa realidade. É uma área que precisa de muitos esforços e dedicação para se desenvolver e atingir os objetivos propostos pelo Cadastro 2014 (KAUFMANN; STEUDLER, 1998).

O CTM é um instrumento importante para a gestão territorial dos municípios e os sistemas cadastrais, em muitos municípios estão evoluindo dos tradicionais formulários e mapas convencionais, analógicos, para modernas plataformas e produtos digitais capazes de disponibilizar dados cadastrais rapidamente. Com isso, o SIG tem sido visto como uma eficiente ferramenta de análise e utilizado em grande parte dos Cadastros informatizados, gerando dados 2D para o CTM, em bancos de dados convencionais e geográficos, nos quais os dados são sobrescritos no momento da atualização, quando se perdem os históricos.

Mesmo com toda a evolução tecnológica nas últimas décadas, existe uma lacuna entre as possibilidades técnicas de gerenciamento de dados geográficos e a sua utilização por gestores e especialistas. O avanço das tecnologias tem disponibilizado uma série de ferramentas que podem tornar mais rápida a criação, a atualização e a manutenção dos Cadastros. A apresentação dos dados para os gestores tem que ser diferenciada da forma com que os usuários os utilizam em suas operações diárias no CTM. Normalmente, os gestores utilizam de dados resumidos, que demonstre uma situação que necessite ser alterada pela intervenção do poder público.

Nesse contexto, o CTM precisa de uma administração efetiva que abranja todas as suas características em benefício do próprio município e da população. Assim, considerando os problemas de integração, evolução histórica e disponibilização dos dados do CTM, neste trabalho o principal enfoque foi a etapa de execução do Cadastro: Trabalho de escritório (Figura 4). Um dos principais objetivos foi integrar a partir do processo de *BI*, utilizando *Data Warehousing*, dados cadastrais advindos de fontes distintas e fornecer ao administrador público informações para o apoio à tomada de decisão, a partir de um método de análise temporal com dados do CTM.

Para a definição do método, várias atividades foram executadas, relatadas no Capítulo 4, com a premissa de usar um ambiente de *BI* como recurso para desenvolver o método, o CTM 2012 e o MGD foram elaborados para apoiar esse ambiente. No início dos trabalhos de elaboração do BDET (CTM 2012), do DER e do Dicionário de Dados identificou-se que a arquitetura de dados do CTM 2010, não atendia mais às demandas atuais.

Frente a essa necessidade, um novo CTM foi projetado, baseado no conceito espaço-temporal. No desenvolvimento do CTM 2012 os dados dos CTMs anteriores, 1996, 2004 e 2010, foram estudados e analisados para a elaboração dos requisitos de dados, visando à integração desses dados no *STDW*.

A integração dos dados está ligada as questões técnicas apresentadas no Quadro 1, da Subseção 2.2.2, Essas questões apresentaram uma semelhança com os níveis de heterogeneidades discutidas por Widerhold (2000), enquanto as questões não técnicas representaram alguns dos problemas enfrentados pelo CTM.

Comparando às questões políticas, o CTM, em geral, tem problemas com a falta de legislação, falta de compromisso e prioridades dos gestores. Nas questões legais, ele possui dificuldade, em certos casos, em definir direitos, restrições e responsabilidade sobre a parcela por não estar diretamente ligado ao Registro de Imóveis. Por fim, nas questões sociais, o CTM, pela sua origem secular, traz arraigado o conceito de “uso para o estabelecimento de impostos”, as pessoas que trabalham com o CTM, em muitos casos, não possuem capacitação adequada para o trabalho e as pessoas interessadas nos dados cadastrais estão em diferentes organizações, sem acesso aos dados.

Para apoiar a integração o MGD, elaborado para o CTM 2012, foi fundamental, no caso dos experimentos, ele orientou na criação dos metadados do *STDW* e permitirá, no futuro, integrar os dados do *STDW* a outros dados externos ao CTM, como secretarias da própria prefeitura e com o registro de imóveis. O MGD também oferece os metadados necessários para a implantação dos dados temáticos da INDE local.

Após a elaboração do MGD, o SGBD foi implementado a partir dos conceitos do BDET para permitir, que mesmo no ambiente transacional, estejam disponíveis para os usuários recursos que permitam verificar e visualizar de forma detalhada as ocorrências “diárias” que afetam a parcela, e o histórico dessas ocorrências. A representação da informação geográfica sem a característica temporal é considerada “estática”. No CTM a representação estática não captura de forma adequada a ocorrência de fenômenos, sendo necessários os dados históricos para serem melhor interpretados. Essa visão é diferente do *STDW*, que apresenta uma visão resumida dos dados.

O projeto e implementação do *STDW* permitiu apresentar os dados de forma dinâmica, e o uso de ferramentas integradas que apresentam as informações de maneira diferente, permite ao gestor analisar os mesmos dados em diferentes formas. Como mostram as análises apresentadas nas Subsessões 4.1, 4.2, 4.3 e 4.4, é possível verificar as diferentes formas de apresentação dos dados.

Na Seção 4.1 foram descritos alguns exemplos de consultas espaciais, temporais e espaço-temporais. A Subseção 4.1.1 apresentou duas consultas espaciais, a primeira permitiu identificar uma parcela específica e mostrou a sua localização no mapa (Figura 51). A segunda consulta identificou quais eram os vizinhos de uma determinada parcela, como mostrou a Figura 52. Essas duas consultas foram executadas no *STDW*, contudo poderiam ser executadas no BDET que retornariam as mesmas informações.

As consultas das Subseções 4.1.2 e 4.1.3 seguem os quatro tipos de consultas descritas por Castro (2007) usando dados espaço-temporais. O primeiro tipo de consulta é a consulta temporal simples (Subseção 4.1.2), que localizou uma determinada informação em um ponto no tempo. Nesse caso a consulta apresentou todos os registros de aposentados e pensionistas no ano de 1996. Essa consulta também poderia ser realizada em um BDET, mas, nesse caso, os dados de 1996 não estavam em um BDET, o acesso a esses dados só foi possível por meio do *STDW*. Os dados de 1996 foram implantados no banco de dados Paradox 4.0, e por meio do processo de *ETL* foi transformado e inserido no *STDW* implantado no PostgreSQL.

O segundo tipo de consulta foi a temporal de intervalo (Subseção 4.1.2), nessa consulta foram listados todos os pensionistas registrados entre os anos de 1996 e 2010³⁷ (Figura 54). Nessa consulta ocorreu a mesma situação da consulta anterior, se houvesse um BDET ela poderia ser realizada sem problemas. No entanto, os dados de origem estavam em diferentes SGBDs, por isso só foi possível executá-la por meio do *STDW*.

A Subseção 4.1.3 traz os outros dois tipos de consultas descritas por Castro (2007). A primeira uma consulta espaço-temporal simples. Essa consulta identificou em quais parcelas haviam aposentados e pensionistas no ano de 1996 como mostra o mapa da Figura 55, que localiza onde ocorreu uma determinada informação em um determinado ponto no tempo.

O segundo tipo de consulta da Subseção 4.1.3 foi a consulta espaço-temporal de intervalo. Para esse tipo de consulta foram criadas quatro consultas, a primeira consulta identificou em quais parcelas haviam aposentados nos levantamentos de 2004 e/ou 2010, representados no mapa da Figura 56, as parcelas com cor lilás possuíam aposentados em 2004, as parcelas com hachurado rosa possuíam aposentados em 2010. O mapa (Figura 56) apresenta o histórico sobre os aposentados no município, sendo os dados de 2004 lilás, de 2010 hachurado rosa e a sobreposição do lilás/hachurado rosa aposentados nos dois levantamentos 2004 e 2010.

³⁷ O tempo de validade do levantamento cadastral de 2010 é de 01/01/2010 a 30/05/2012.

A segunda consulta da Subseção 4.1.3, identificou em quais parcelas haviam aposentados em 2004 e 2010 (Figura 57). Nessa consulta foi realizada uma interseção da geometria para identificar as parcelas que continham aposentados nos dois levantamentos. A terceira consulta mostrou em quais parcelas ocorreram desmembramentos, comparando as áreas das parcelas de 2010 e 2012 (Figura 58). A quarta consulta apresentou os remembramentos das parcelas, comparando as áreas das parcelas de 2010 e 2012 (Figura 59). Essas duas últimas consultas foram executadas com dados simulados, inseridos no *STDW*.

As consultas temporais só foram possíveis de executar porque os dados estão em um *STDW*. Isso ocorreu pelo fato dos dados de origem serem semânticos e sintaticamente diferentes e não poderem ser inseridos diretamente em um único BDET. Indicando que por meio do *STDW* é possível consultar outras fontes de dados externas e uni-las aos dados atuais, desde que seja possível fazer o mapeamento dos dados.

Na análise 4.2 sobre o tema Parcela, foram verificadas as formas de identificar as parcelas em que ocorreram alterações de área construída. Em um primeiro momento uma consulta englobando dados espaciais e temporais foi elaborada para calcular a diferença de área da mesma parcela entre duas épocas. Com esse relatório, além de identificar, principalmente, as parcelas onde ocorrera aumento substancial de área, que resulta em aumento da tributação da parcela, foi possível também identificar parcelas que tiveram diferenças exageradas, que podem representar erro da coleta de dados.

Os relatórios das Figura 60 e Figura 61 apresentados na análise 4.2 só foram possíveis de elaborar porque os dados temporais estavam em uma mesma base de dados. Se existissem dois SGBDs diferentes com esses dados, não seria possível executar a mesma consulta. Somente relatórios separados por ano, em cada SGBD, seriam possíveis de elaborar.

Se os dados não estivessem no mesmo SGBD, não seria possível comparar os dados coletados com os dados históricos. Outra vantagem de possuir os dados históricos em único SGBD é permitir identificar erros de coleta ou registro dos dados, para o controle de qualidade. Com isso, qualquer diferença expressiva nos dados pode ser identificada nos relatórios. Tornando-se foco de revisão da coleta, específica para aquela(s) parcela(s).

Na análise 4.3 sobre o tema Diagnóstico Médico, foram selecionadas para a consulta as patologias comumente encontradas entre a população, tais como, cardiopatia, hipertensão arterial e diabetes. Essas patologias, normalmente acometem pacientes que fazem uso de medicação de uso contínuo e são atendidos por políticas públicas de saúde.

Nessa análise, os mapas das Figura 66, Figura 67 e Figura 68 apresentam os dados coletados nos quatro anos, a sobreposição das informações prejudica um pouco a visualização

do mapa completo. No entanto, quando se usa diretamente a visualização no gvSIG e realizando uma aproximação (*zoom*) em uma determinada região do mapa (Figura 65), é possível identificar melhor a sobreposição das informações (hachuradas). Para complementar a análise, o uso de outras ferramentas disponíveis na *suíte* Pentaho, (Figura 69 e Figura 70) permite verificar os dados de diversas maneiras, por exemplo, a partir de relatórios *ad hoc*, detalhar determinadas consultas que serão importantes para à tomada de decisão.

O mapa da Figura 68 representa claramente a diferença em utilizar dados espaço-temporais e um SIG comum. Nesse mapa é possível visualizar a evolução da diabetes na população em uma única ferramenta, e por meio do *Data Warehousing* elaborar consultas *ad hoc* que detalham ou agrupam os dados, como mostrado na Figura 69 e na Figura 70. Essa situação também se aplica ao SIG temporal, pois os dados utilizados na análise são heterogêneos.

Continuando as análises dos resultados, na análise 4.4 sobre o tema Educação foram selecionados dados sobre creche e ensino fundamental. Pela Constituição Federal, Art. 211: “§ 2º Os Municípios atuarão prioritariamente no ensino fundamental e na educação infantil. § 3º Os Estados e o Distrito Federal atuarão prioritariamente no ensino fundamental e médio”. Nesse contexto é de interesse do Município conhecer esses dados, para gerenciar a oferta de vagas nas escolas municipais. Por Ribeirão dos Índios – SP ser um município pequeno verifica-se no relatório (Figura 71) que existem poucos alunos na faixa etária de 4 a 16 anos no ensino particular.

Nessa situação o governo municipal tem que atender toda a demanda da população. Os mapas das Figura 72, Figura 73 e Figura 74 mostram a distribuição espacial dos alunos por ano. No mapa da Figura 75, com a sobreposição dos dados nos anos de 2004, 2010 e 2012, fica evidente a mudança no nível escolar, muitas parcelas que possuíam alunos nesse nível no ano de 2004 já não o possuíam em 2010, pois entre um levantamento e outro se passaram seis anos. Continuando a análise desse assunto, o mapa da Figura 76 mostra que as parcelas que possuíam alunos no levantamento de 2004 e 2010 eram cinquenta e cinco, já entre os levantamentos de 2010 e 2012 foi possível verificar que há mais sobreposições, sendo noventa e uma parcelas. Essas informações só puderam ser elaboradas porque os dados eram espaço-temporais e estavam armazenados no *STDW*.

Outra característica interessante observada nos relatórios *ad hoc*, Figura 77, os alunos que frequentaram o ensino fundamental e médio público representam a maior parte dos alunos do ensino superior particular. Nesse caso, pode-se dizer que há uma tendência de que muitos dos alunos da escola pública (fundamental e médio) não acessam o ensino superior público.

Nesses experimentos pode-se perceber a importância do armazenamento temporal e espacial dos dados e a relevância em usar ferramentas de um *Data Warehousing* para gerar informações.

6 CONCLUSÃO E TRABALHOS FUTUROS

6.1 Conclusão

O objetivo dessa tese foi desenvolver um método para análise de dados temporais em Cadastro Territorial Multifinalitário Urbano a partir dos conceitos de *DW*, acessando fontes de dados heterogêneas. A construção do Cadastro não fez parte dos objetivos, mas sim, definir como usá-lo de maneira eficiente pelo gestor, por isso, foram investigados recursos tecnológicos para apoiar a definição do método.

Nesse contexto, identificou-se que o aspecto tempo não é um elemento novo na elaboração do CTM, no entanto, é um elemento importante para várias atividades, mas geralmente não implementado adequadamente. Sabe-se que os dados históricos são muito importantes para entender a evolução do cadastro das parcelas, mas atualmente, após cada aquisição e atualização, ocorre uma substituição desses dados, sendo eliminados os anteriores. Por isso, foi preciso organizar os dados antigos para integrá-los aos novos dados cadastrais.

Assim, a recuperação dos dados dos levantamentos anteriores permitiu elaborar um retrato da evolução de várias situações importantes para a administração pública e que indiretamente afeta a população. O CTM temporal visou representar as mudanças e seus efeitos ao longo do tempo. Ressalta-se que o “produto” é um CTM temporal e não um Cadastro 4D. O Cadastro não é 4D, pois não possui a terceira dimensão do Cadastro 3D.

A elaboração do MGD foi fundamental para o projeto do *STDW*, pois fornecerá os parâmetros para a integração dos dados do CTM com outros setores públicos. Almeja-se, também, que ele apoie uma futura integração dos dados do CTM com o Registro de Imóveis.

A partir dos esforços que vêm sendo despendidos pela padronização, definida pela *OGC*, e o esforço para a implantação da *INDE*, possivelmente, no futuro a integração de dados espaciais com diversos setores, público e privado, será uma realidade.

O projeto e execução do CTM 2012 com *BDET* permitiu aplicar novos conceitos tecnológicos ao CTM, que só foram possíveis de implementar por causa dos novos recursos oferecido pelas ferramentas. Mesmo em um “ambiente livre”, é possível explorar conceitos antes não aplicados ao CTM, que trazem grandes benefícios para o entendimento do ambiente urbano.

O método de análise de dados temporais, baseado no processo de *BI*, utilizando o *Data Warehousing*, se mostrou eficiente na elaboração de relatórios, mapas e disponibilização de recursos que permitem ao usuário criar relatórios *ad hoc*, com dados em forma textuais e gráficos. Dentre suas características, destaca-se a possibilidade de comparação entre dados

atuais e dados históricos, demonstrando a evolução dos objetos ou, em alguns casos, ser fonte de controle de qualidade dos dados coletados em campo.

O *STDW* representa um novo campo de investigação, não existindo até o momento uma abordagem metodológica consolidada e aceita para o seu desenvolvimento. No entanto, sabe-se que a inclusão de dados espaciais aumenta o poder do processo decisório do gestor, por expandir o escopo da análise e por fornecer um realce, facilitando a visualização. Semelhantemente, o suporte temporal possibilita que os elementos de um modelo multidimensional sejam variáveis no tempo.

As análises, realizadas com os dados do *STDW*, demonstrou que modelos espaço-temporais permitem um melhor entendimento das ocorrências registradas, por meio do acompanhamento da evolução dos dados. Tanto por meio de relatórios, como por meio gráfico, dos mapas, é possível acompanhar e entender onde e quanto os fenômenos acontecem. Garantindo aos gestores mais informações para apoiar suas decisões administrativas.

Este trabalho mostrou que é possível criar soluções de *BI* no setor público com baixo custo tecnológico, desde que existam recursos humanos capacitados. Espera-se que a partir das informações geradas em um ambiente como esse, no futuro, a cobrança de impostos possa ser mais equitativa e os recursos públicos possam ser mais bem aplicados no planejamento e desenvolvimento das cidades.

6.2 Trabalhos futuros

Nessa tese foi definido um método para apoiar a tomada de decisão com uma micro visão dos dados. Cita-se micro visão dos dados por tratar do CTM, que envolve dados do setor de Cadastro das prefeituras brasileiras. Todavia o Cadastro, quando agrega os dados temáticos conota, de certa forma, uma macro visão dos outros setores da prefeitura. Por isso, a partir da modelagem de dados do CTM, pode-se recomendar a elaboração de outros estudos que permitam integrar os outros dados distribuídos, macro visão, pelos setores da prefeitura.

Outro elemento importante é a integração dos dados do CTM ao Registro de Imóveis. Muitas pesquisas têm buscado de diversas maneiras essa integração, talvez, o uso do *Data Warehousing* possa garantir que os sistemas transacionais individuais das instituições responsáveis pelo Cadastro e o Registro permaneçam seguros e possam compartilhar informações que sejam de interesse dos gestores públicos e dos cidadãos.

Durante o processo de elaboração dos mapas, notou-se certa dificuldade em representar os dados temporais. As sobreposições dos dados nas parcelas em muitos casos

ficaram difíceis de interpretar. Por isso, sugere-se estudar a melhor forma de visualização cartográfica dos dados espaço-temporais para a elaboração dos mapas.

Outro ponto importante, alguns trabalhos têm discutido sobre a procedência dos dados, esse assunto é interessante quando se trata de integrar dados de fontes distintas. Dessa forma, para integrar dados oriundos de fontes diversas seria importante ter uma metodologia que armazenasse a procedência dos dados para sua validação.

REFERÊNCIAS

ÁGUILA, M.; ERBA, D. A. El Rol Del Catastro en el Registro Del Territorio. In: ERBA, D.A. (Ed.Org.). **Catastro Multifinalitario aplicado a la definición de políticas de suelo urbano**. Cambridge, MA : Lincoln Institute of Land Policy, 2007.

AHMED, O. T. **Spatial on-line analytical processing (SOLAP): overview and current trends**. In: International Conference on Advanced Computer Theory and Engineering - ICACTE '08. 2008. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4737127>. Acesso em: 01 de ago. 2011.

AMORIM, A. et al.. **A Modernização do cadastro técnico multifinalitário urbano e a influência da evolução tecnológica: uma reflexão sobre o futuro e a multidisciplinaridade do cadastro**. Info GPS/GNSS, Curitiba, 2007. p. 46 – 47.

AMORIM, A. SOUZA, G. H. B; DALAQUA, R. R. **Uma metodologia alternativa para otimização da entrada de dados em sistemas cadastrais**. Revista Brasileira de Cartografia. Rio de Janeiro, V.56, n. 1, 2004. p. 47-54.

AMORIM, A.; SOUZA, G. H. B.; YAMASHITA, M. C. **Cadastro técnico multifinalitário via internet: um importante instrumento de apoio ao planejamento municipal**. Revista Brasileira de Cartografia (*on-line*), v 60/2, 2008. p. 119-125. Disponível em: http://www.rbc.ufrj.br/_2008/60_2_02.htm. Acesso em: 27 set. 2011.

BARLOW, S. **Data Integration**. University of Passau. July 24, 2000.

BASS, L.; CLEMENTS, P.; KAZMAN, R. **Software architecture in practice**. Massachusetts: Addison Wesley, 1998.

BATINI, C.; LENZERINI, M.; NAVATHE, S. B. **A comparative analysis of methodologies for database schema integration**. ACM Computing Surveys, v. 18, n 4. 1986.

BRASIL – Ministério das Cidades. **Diretrizes para a criação, instituição e atualização do Cadastro Territorial Multifinalitário (CTM) nos municípios brasileiros**. Portaria nº 511, de 07 de dezembro de 2009.

CÂMARA G. Representação computacional de dados geográficos. In: LAENDER A. H. F. et al. **Banco de dados geográficos**. São José dos Campos, Rio de Janeiro, Belo Horizonte: Livro *on-line*, 2005. Disponível em: <http://www.dpi.inpe.br/gilberto/livro/bdados/>. Acesso em: 22 de set. 2009.

CÂMARA, G., et al. **Anatomia de Sistemas de Informação Geográfica**. Campinas: Instituto de Computação, UNICAMP, 1996. 197p.

CASTRO, A. F. **Sistemas computacionais espaço-temporais para tomada de decisão em questões ambientais relacionadas à indústria de petróleo e gás**. 2007. 178 f. Tese (Doutorado em Geodinâmica e Geofísica) Centro de Ciências Exatas e da Terra da Universidade Federal do Rio Grande do Norte. Natal – RN. Disponível em: ftp://ftp.ufrn.br/pub/biblioteca/ext/bdtd/AngelicaFC_tese.pdf Acesso em: 02 mai. 2011.

CASTRO, J. H. A relação entre os SIG's e as estruturas de disseminação de informações. In: **Congresso Brasileiro de Cadastro Técnico Multifinalitário**. 2006, Florianópolis. Florianópolis, 2006.

CHAUDHURI, S.; DAYAL, U. An Overview of data warehousing and OLAP Technology. In: **ACM SIGMOD Record**, v.26 n.1, 1997. p. 65-74.

CHEN, P. P. S. **The entity-relationship model**: toward a unified view of data. *ACM Trans. Database Syst.* 1, 1 (March 1976), 9-36. 1976.

CIFERRI, C. D. A. **Distribuição dos dados em ambientes de data warehousing**: o sistema WebD2W e algoritmos voltados à fragmentação horizontal dos dados. 2002. 263 f. Tese (Doutorado em Ciência da Computação) Centro de Informática da Universidade Federal de Pernambuco – UFPE. Recife – PE.

CINDE – Comitê de Planejamento da Infraestrutura Nacional de Dados Espaciais. **Plano de ação para a implantação da infraestrutura nacional de dados espaciais**. Rio de Janeiro: Ministério do Planejamento, Orçamento e Gestão, 2010. Disponível em: <http://www.concar.ibge.gov.br/arquivo/PlanoDeAcaoINDE.pdf>. Acesso em: 20 set. 2011.

CÖMERT, Ç.; ALKAN, M. **The design and development of a temporal GIS for cadastral and land title data of Turkey** In: XXXV International Society for Photogrammetry and Remote Sensing - ISPRS Congress, Istanbul, 2004. Proceedings of Commission IV. Disponível em: <http://www.isprs.org/proceedings/XXXV/congress/comm4/papers/309.pdf>. Acesso em: 03 de jan. 2012.

CONCAR. **Perfil de metadados geoespaciais do Brasil** (Perfil MGD). Comissão Nacional de Cartografia, Comitê de Estruturação de Metadados Geoespaciais CEMGCOMCAR. Conteúdo de Metadados Geoespaciais em conformidade com a norma ISO 19115:2003, Novembro, 2009.

DALE, P. F.; MCLAUGHLIN, J. D. **Land information management**: an introduction with special reference to cadastral problems in third world countries. Reprinted (with correction). Oxford. Oxford University Press, 1990.

DATE, C. J. **Introdução a sistemas de banco de dados**. Tradução: Vandenberg Dantas de Souza. 7ª ed. americana. Rio de Janeiro: Campus, 2000.

DEVILLERS, R.; BÉDARD, Y.; JEANSOULING, R. **Multidimensional management of geospatial data quality information for its dynamic use within GIS**. *Photogrammetric Engineering & Remote Sensing* Vol. 71, No. 2, February, 2005. p. 205–215.

DIAS, T. L.; CÂMARA G.; DAVIS Jr. C. A. Modelos espaço-temporais. In: LAENDER A. H. F. et al. **Banco de dados geográficos**. São José dos Campos, Rio de Janeiro, Belo Horizonte: Livro *on-line*, 2005. Disponível em: <http://www.dpi.inpe.br/gilberto/livro/bdados/>. Acesso em: 22 de set. 2009.

DINIZ, E. A. et AL. **Atualização do sistema cadastral da cidade de Ribeirão dos Índios – SP**. 2004, 124f. Trabalho de Graduação (Graduação em Engenharia Cartográfica). Faculdade de Ciência e Tecnologia, Universidade Estadual Paulista “Júlio de Mesquita Filho”,

Presidente Prudente, 2004.

DONER, F et al. **4D land administration solutions in the context of the spatial information infrastructure**. Integrating Generations FIG Working Week. Stockholm, Sweden. June 14-19, 2008. Disponível em: http://www.fig.net/pub/fig2008/papers/ts05e/ts05e_03_doner_et_al_3016.pdf Acessado: 21/09/2009.

DONG, X.; HALEVY, A. Y.; YU, C. Data integration with uncertainty. In: **Proceedings of the 33rd international conference on Very large data bases (VLDB '07)**. VLDB Endowment 687-698, 2007.

EDELWEISS, N. Banco de dados temporais: teoria e prática. In: Jornada de atualização em informática, 17^a. 1998, Belo Horizonte. **Anais do XVII Congresso Nacional da Sociedade Brasileira de Computação**. Belo Horizonte: SBC, 1998. p. 225-282.

ELMASRI, R.; NAVATHE, S. B. **Sistemas de banco de dados**. Revisor: Luis Ricardo de Figueiredo. 4^o ed. São Paulo: Pearson Addison Wesley, 2005.

FIDALGO, R. N.; TIMES, V. C.; SOUZA, F. F. **GOLAPA: Uma arquitetura aberta e extensível para integração entre SIG e OLAP**. In: III WorkShop Brasileiro de GeoInformática (GeoInfo), 2001, Rio de Janeiro. WorkShop Brasileiro de GeoInformática (GeoInfo2001), 2001. p. 111-118.

FIG – Fédération Internationale de Géomètres. **Statement on the cadastre**. International Federation of Surveyors. 1995. Disponível em: http://www.fig.net/commission7/reports/cadastre/statement_on_cadastre.html. Acesso em: 03 mai. 2010.

FONSECA, R. et al. Um metamodelo para especificação de data warehouse geográfico. In: **XXII Simpósio Brasileiro de Banco de Dados**, 2007. Disponível em: <http://www.lbd.dcc.ufmg.br/bdbcomp/servlet/Trabalho?id=6949>. Acesso em: 02 ago. 2011.

FRANCISCO, H. R.; IMAI, N. N. **Temporalidade em SIG: aplicação em cadastro de imóveis**. Boletim de Ciências Geodésicas, Curitiba, v. 9, n. 1, 2003. p. 89-104

GROOT, R. **Spatial data infrastructure (SDI) for sustainable land management**. ITC – Instrumental Transcommunication Journal, Enschede, Nr. ¾, 1997. p. 287-294.

HALEVY, A.; RAJARAMAN, A.; ORDILLE, J. Data integration: the teenage years. In Proceedings of the 32nd international conference on Very large databases (VLDB '06), Umeshwar Dayal, Khu-Yong Whang, David Lomet, Gustavo Alonso, Guy Lohman, Martin Kersten, Sang K. Cha, and Young-Kuk Kim (Eds.). VLDB Endowment, 2006. p. 9-16.

HULL, R. Managing semantic heterogeneity in databases: a theoretical perspective, In **Proceedings of the sixteenth ACM SIGACT-SIGMOD-SIGART symposium on Principles of Database Systems (PODS'97)**, ACM, New York, NY, USA, 1997. p. 51-61.

INMON, W. H. **Building the data warehouse**. 4^a ed. Indianapolis: Wiley Publishing, 2005.

JENSEN, C. et al. **The consensus glossary of temporal database concepts – February 1998 version**. Disponível em: <http://www.cs.aau.dk/~csj/Glossary/>. Acesso em: 15 jan. 2011.

KAUFMANN, J; STEUDLER, D. (Org). **Cadastró 2014: uma visão para um sistema cadastral futuro**. Grupo de Trabalho 1 da Comissão 7 da FIG. Julho 1998.

KIMBALL, R.; CASERTA, J. **The data warehouse ETL toolkit: practical techniques for extracting, cleaning, conforming, and delivering data**. Wiley Publishing, Inc: Indianapolis, 2004.

KOUBA, Z; MATOUSEK, K.; MIKSOVSKÝ, P. **On data warehouse and GIS integration**. In: Proceedings of the 11th International Conference on Database and Expert Systems Applications. Springer-Verlag, London, UK, 2000. p. 604–613.

KUDURRU. In: *Encyclopédia britannica*. Disponível em: <http://www.britannica.com/EBchecked/topic/324360/kudurru>. Acesso em: 21 out. 2011.

LANGRAN, G. **Time in geographic information systems**. Reprinted. London: Taylor & Francis, 1993.

LARSSON, G. **Land registration and cadastral systems**. Reprinted. England, UK, Longman Group, 1996.

LENZERINI, M. Data integration: a theoretical perspective. In: **Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems (PODS '02)**. ACM, New York, NY, USA, 233-246, 2002.

LIANG, Q. **User demands and access model for temporal cadastre in china**. 69 f. 2008. Thesis (Master of Science in Geo-information Science and Earth Observation). International Institute for Geo_Information Science and Earth Observation. Enschede, Netherlands.

LISBOA FILHO, J. **Projeto de banco de dados para sistemas de informação geográfica**. Revista Eletrônica de Iniciação Científica - REIC/SBC, v.1, n.2, 2001.

LISBOA FILHO, J.; RODRIGUES JUNIOR, M. F.; DALTIO, J. ArgoCASEGEO - Uma ferramenta CASE de código-aberto para o modelo UML-GeoFrame. In: **Workshop Iberoamericano de Ingeniería de Requisitos y Desarrollo de Ambientes de Software IDEAS, 7, 2004**, Arequipa-Perú. Memórias... Arequipa-Perú: Universidad Católica de Santa Maria, Universidad Católica San Pablo, 2004.

LOCH, C. A Realidade do cadastro técnico urbano no Brasil. In: **XIII Simpósio Brasileiro de Sensoriamento Remoto, 2007**, Florianópolis. SBSR 2007. Florianópolis: INPE, 2007. p. 5357-5364. Disponível em: <http://marte.dpi.inpe.br/col/dpi.inpe.br/sbsr@80/2006/11.14.18.04.51/doc/5357-5364.pdf>. Acesso em: 31 mar. 2010.

LOCH, C. Cadastro técnico multifinalitário instrumento de política fiscal e urbana. In: ERBA, D. A.; OLIVEIRA, F. L.; LIMA JÚNIOR, P. N. (Org.) **Cadastro multifinalitário como instrumento da política fiscal e urbana**. Rio de Janeiro, 2005. Disponível em: http://www.cidades.gov.br/index.php?option=com_content&view=article&id=547:cadastro-multifinalitario-como-instrumento-de-politica-fiscal-e-urbana&catid=48&Itemid=83. Acesso em: 01 out. 2011.

LOCH, C.; ERBA, D. A. **Cadastro técnico multifinalitário: rural e urbano**. Cambridge, MA: Lincoln Institute of Land Policy, 2007.

MACHADO, F. N. R. **Projeto de data warehouse: Uma Visão Multidimensional**. São Paulo: Érica, 2000.

MALAMAN, C. S.; AMORIM, A. **Utilização do software gvSIG no cadastro técnico multifinalitário do município de Ribeirão dos Índios - -SP**. In: Congresso Brasileiro de Cadastro Técnico Multifinalitário – COBRAC, 2010.

MALINOWSKI, E.; ZIMÁNYI, E. **Advanced data warehouse design: from conventional to spatial and temporal applications**. 2nd corrected printing. Springer-Verlag Berlin Heidelberg, 2008.

MANHÃES, R. S. et al. GeoOlap: An integrated approach for decision support . In: **Second IFIP TC8 International Conference on Research and Practical Issues of Enterprise Information Systems (CONFENIS 2008)**, Research and Practical Issues of Enterprise Information Systems. Springer-Verlag. Pequim, China. 2008. Disponível em: <http://www.uniriotec.br/~cgolap/doc/artigos/Geo-OLAP-CONFENIS-2007.pdf>. Acesso em: 01 de ago. 2011.

MATIAS, R.; PIRES-MOURA, J. Spatial On-Line Analytical Processing (SOLAP): A tool the to analyze the emission of pollutants in industrial installations . In: **12th Portuguese Conference on Artificial Intelligence, 5th International Workshop on Extraction of Knowledge from Databases & Warehouse (EKDB&W'05)** in Carlos Bento, Amílcar Cardoso, Gaél Dias (Eds.) UBI Proceedings, included in the IEEE Conference Publication Program with ISBN 0-7803-9365-1 and IEEE catalog number O5EX1157.

MEDEIROS, C. B.; JOMIER, G. Managing alternatives and data evolution in GIS. In: **ACM/ISCA Workshop on advances in geographic information systems**, Baltimore, USA, November, 1993. p. 36-39. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.54.6261>. Acesso em: 10 mai. 2011.

MICHAELIS. **Moderno dicionário da língua portuguesa**. Disponível em: <http://michaelis.uol.com.br/moderno/portugues/index.php>. Acesso em: 21 de out. 2011.

MOHAMMADI, H. et al. The Development of a Framework and Associated Tools for the Integration of Multi-Sourced Spatial Datasets. In: **Proceedings of 17th United Nations regional cartographic conference for Asia and the Pacific**, Bangkok, Thailand, 2006.

MOHAMMADI, H. **Spatial data integration: a necessity for spatially enabling government, towards spatially enabled society**. Edited by Abbas Rajabifard. The University of Melbourne, 2007. p. 333-343,

MOHAMMADI, H.; RAJABIFARD, A.; WILLIAMSON, I. P. Development of an interoperable tool to facilitate spatial data integration in the context of SDI. In: **International Journal of Geographical Information Science**, v. 24, n. 4, 2010. p. 487-505.

MUNIZ, D. P. et al. Implantação do Cadastro Técnico Multifinalitário em uma área teste. In:

II Congresso Brasileiro de Cadastro Técnico Multifinalitário. Florianópolis – SC, 1996. **Anais...** Florianópolis – SC, 1996.

O'BRIEN, J. A. **Sistemas de informação e as decisões gerenciais na era da internet.** Tradução: Célio Knipel Moreira e Cid Knip Moreira. 2ª edição. São Paulo: Editora Saraiva, 2004.

OLIVEIRA, F. H. Do cadastro territorial multifinalitário. In: CUNHA, M. P.; ERBA, D. A. (Org). **Manual de apoio – CTM:** Diretrizes para a criação, instituição e atualização do cadastro territorial multifinalitário nos municípios brasileiros. Brasília: Ministério das Cidades, 2010. Disponível em: <http://www.cidades.gov.br/images/stories/Arquivos/Capacitacao/Capacita%C3%A7%C3%A3o/livro%20diretrizes%20em%20alta.pdf>. Acesso em: 27 de out. 2011.

PARENT, C.; SPACCAPIETRA, S.; ZIMÁNYI, E. Spatio-temporal conceptual models: data structures + space + time. In **Proceedings of the 7th ACM international symposium on Advances in geographic information systems (GIS '99)**, Claudia Bauzer Medeiros (Ed.). ACM, New York, NY, USA, 1999. p. 26-33.

PHILIPS, J. Das disposições gerais. In: CUNHA, M. P.; ERBA, D. A. (Org). **Manual de apoio – CTM:** Diretrizes para a criação, instituição e atualização do cadastro territorial multifinalitário nos municípios brasileiros. Brasília: Ministério das Cidades, 2010. Disponível em: <http://www.cidades.gov.br/images/stories/ArquivosCapacitacao/Capacita%C3%A7%C3%A3o/livro%20diretrizes%20em%20alta.pdf>. Acesso em: 27 de out. 2011.

PIUMETTO, M.; ERBA, D. A. Sistema de información geográfica aplicadas al catastro urbano. In: ERBA, D.A. (Ed.Org.). **Catastro Multifinalitario aplicado a la definición de políticas de suelo urbano.** Cambridge, MA : Lincoln Institute of Land Policy, 2007. Cap 9, p. 242 – 266.

RAJABIFARD, A. Data integration and interoperability of systems and data, Under Review. **International journal of geographical information science**, v. 24, n. 9, 2010.

RAWAT, S. **Interoperable geospatial data model in the context of the Indian NSDI.** 2003. 97 f. Master's thesis, International Onstitute for Geo-Information Science and Earth Observation. Holland.

RENOLLEN, A. **Modelling the real world: conceptual modelling in spatiotemporal information system desing.** 2000. Transactions in GIS. Wiley on-line Library. v. 4, i.1, January 2000. p. 23-42.

RIZZI, S.; GOLFARELLI, M. What time is it in the Data Warehouse? In: **Proceedings 8th International Conference on Data Warehousing and Knowledge Discovery - DaWaK 2006**, Krakow, Poland, 2006. p. 134-144.

ROB, P.; CORONEL, C. **Sistemas de banco de dados:** projeto, implementação e gerenciamento. Tradução: All Tasks. Revisão Técnica: Ana Paula Appel. 8ª Edição norte-americana. São Paulo: Cengage Learning, 2011.

ROCHA, L. V.; EDELWEISS, N; IOCHPE, C.. GeoFrame-T: A temporal conceptual

framework for data modeling. In: **Proceedings of the 9th ACM international symposium on Advances in geographic information systems (GIS '01)**. ACM, New York, NY, USA, 2001. p. 124-129.

SASS, G. G; AMORIM, A. **Análise Temporal a Partir do Cadastro Territorial Multifinalitário**. RBC. Revista Brasileira de Cartografia (*on-line*), v. 65, 2013. p. 283-291.

SCHULTZ, R. E. O. **Inteligência computacional aplicada à predição espaço-temporal**. 2007. 117 f. Dissertação (Mestrado em Engenharia Elétrica e Informática [Industrial]) Universidade Tecnológica Federal do Paraná – UTFPR. Curitiba – PR.

SERPRO. **Modelo de governança e gestão do modelo global de dados – MGG**: modelo global de dados. Junho de 2010a. Disponível em: <http://modeloglobaldados.serpro.gov.br/modelo-global-de-dados/modelo-de-governanca-e-gestao/MGD%20-%20Modelo%20de%20Governanca%20e%20Gestao.pdf/view>. Acesso em: 02/10/2012.

SERPRO. **Modelo global de dados - MGD**: compendio integrado. Versão 1.1. Maio de 2011.

SERPRO. **Modelo global de dados - MGD**: guia metodológico para integração de dados e processos. Versão 1.0. Dezembro 2010b. Disponível em: <http://modeloglobaldados.serpro.gov.br/modelo-global-de-dados>. Acesso em: 02/10/2012.

SERPRO. **Projeto de integração do macroprocesso de planejamento, orçamento e finanças**: modelo conceitual e metodológico. Brasília, setembro de 2009.

SHAW, M.; GARLAN, D. **Software architecture - perspectives on an emerging discipline**. New Jersey: Prentice-Hall, 1996.

SHEKHAR, S. et al. **Map Cube: A visualization tool for spatial data warehouse**. In: MILLER, H. and HAN, J. (Eds). Geographic data mining and knowledge Discovery. 2001. Disponível em: http://www.spatial.cs.umn.edu/paper_ps/mapcube.pdf. Acesso em: 01 ago. 2011.

SHETH, A. Changing focus on interoperability in information systems: from system, syntax, structure to semantics. In: GOODCHILD, M.; EGENHOFER, M.; FEGEAS, R.; KOTTMAN, C. (Eds.) **Interoperating geographic information systems**. Kluwer Academic Publishers, Norwell, Ma. 1999.

SILBERSCHATZ, A.; KORTH, H. F.; SUDARSHAN, S. **Sistema de banco de dados**. 5ª ed. São Paulo: Makron Books, 2006.

SILVERS, F. **Building and maintaining a data warehouse**. Boca Raton: Taylor & Francis Group, 2008.

SIQUEIRA, T. L. L. et al. Benchmarking spatial data warehouses. In: **12th International Conference on Data Warehousing and Knowledge Discovery, 2010**, Bilbao. 2010. p. 40-51.

SIQUEIRA, T. L. L. et al. **The impact of spatial data redundancy on SOLAP query performance.** Journal of the Brazilian Computer Society, v. 15, 2009. p. 19-34.

SIQUEIRA, T. L.; CIFERRI, R. R.; TIMES, V. C. I-DWE: Uma estrutura de indexação para data warehouse espacial. In: **VII Workshop de Teses e Dissertações em Bancos de Dados**, 2008, Campinas. 2008. p. 79-84.

SONG, W.; YANG, X. A spatio-temporal cadastral data model based on space-time composite model. In: **Geoinformatics: IEEE**, p.1-4, 2013.

SOUZA, G. H. B. **Método de modelagem da parcela especial para o cadastro tridimensional.** 2011. 97 f. Tese (Doutorado em Ciências Cartográficas) Faculdade de Ciência e Tecnologia da Universidade Estadual Paulista - UNESP. Presidente Prudente – SP. Disponível em: http://www4.fct.unesp.br/pos/cartografia/docs/teses/t_souza_GHB.pdf. Acesso em: 10 jan. 2012.

STOTER, J. E.; VAN OOSTEROM, P. **3D cadastre in international context: legal, organizational, and technological aspects.** Boca Raton: CRC Press : Taylor & Francis, 2006.

VAISMAN, A.; ZIMÁNYI, E. What Is Spatio-Temporal Data Warehousing?. In: **Proceedings of the 11th International Conference on Data Warehousing and Knowledge Discovery (DaWaK '09)**, Torben Bach Pedersen, Mukesh K. Mohania, and A Min Tjoa (Eds.). Springer-Verlag, Berlin, Heidelberg, 2009. p. 9-23.

VAN OOSTEROM, P. et al. **Aspects of a 4D cadastre: a first exploration.** In: **Shaping the Change XXIII FIG Congress.** Munich, Germany. October 8-13, 2006. Disponível em: http://www.fig.net/pub/fig2006/papers/ts14/ts14_06_oosterom_etal_0576.pdf Acesso em: 21 set.2009.

WANG, Z; FANG, Y.; XIE, X. A spatio-temporal data model based on the parcel in cadastral. In: **Geoscience and Remote Sensing Symposium, 2005. IGARSS '05. Proceedings. 2005 IEEE International** , vol.2, no., pp.4 July 2005. P. 25-29.

WARNEST, M. **A collaboration model for national spatial data infrastructure in federated countries.** 2005. 261 f. Doctor of Philosophy. Department of Geomatics. The University of Melbourne. Melbourne, Australian.

WIEDERHOLD, G. Mediators, Concepts and practice. In: ÄYER, T.; KIANMEHR, K.; Mehmet TAN, M.; ZENNG, J. **Studies information reuse and integration in academia and industry**, Springer Verlag, Wien, 2012.

WIEDERHOLD, G. Precision in processing data from heterogeneous resources. In: LINGS, B.; JEFFREYS, K. (eds.): Advances in databases. In: **Proceedings 17th British National Conference on Databases**, Exeter, UK, July 2000, p. 1-18.

WILLIAMSON, I. et al. **Land administration for sustainable development.** First edition. Redlands – California, Esri Press, 2010.

WORBOYS, M.F. **GIS: A computing perspective.** London: Taylor and Francis, 1995.

WU, M.; BUCHMANN, A. P. **Research issues in data warehousing**. BTW'97, Ulm, Springer-VerlagMarço 1997. Disponível em: <http://www.dvs.tu-darmstadt.de/publications/pdf/btwllncs.pdf>. Acesso em: 01 de jun. 2011.

APÊNDICE A – Modelagem do Banco de Dados Espaço-temporais

Para apoiar o CTM 2012, foi desenvolvido a modelagem do SGBD com características temporais e espaciais para dar suporte a análise temporal.

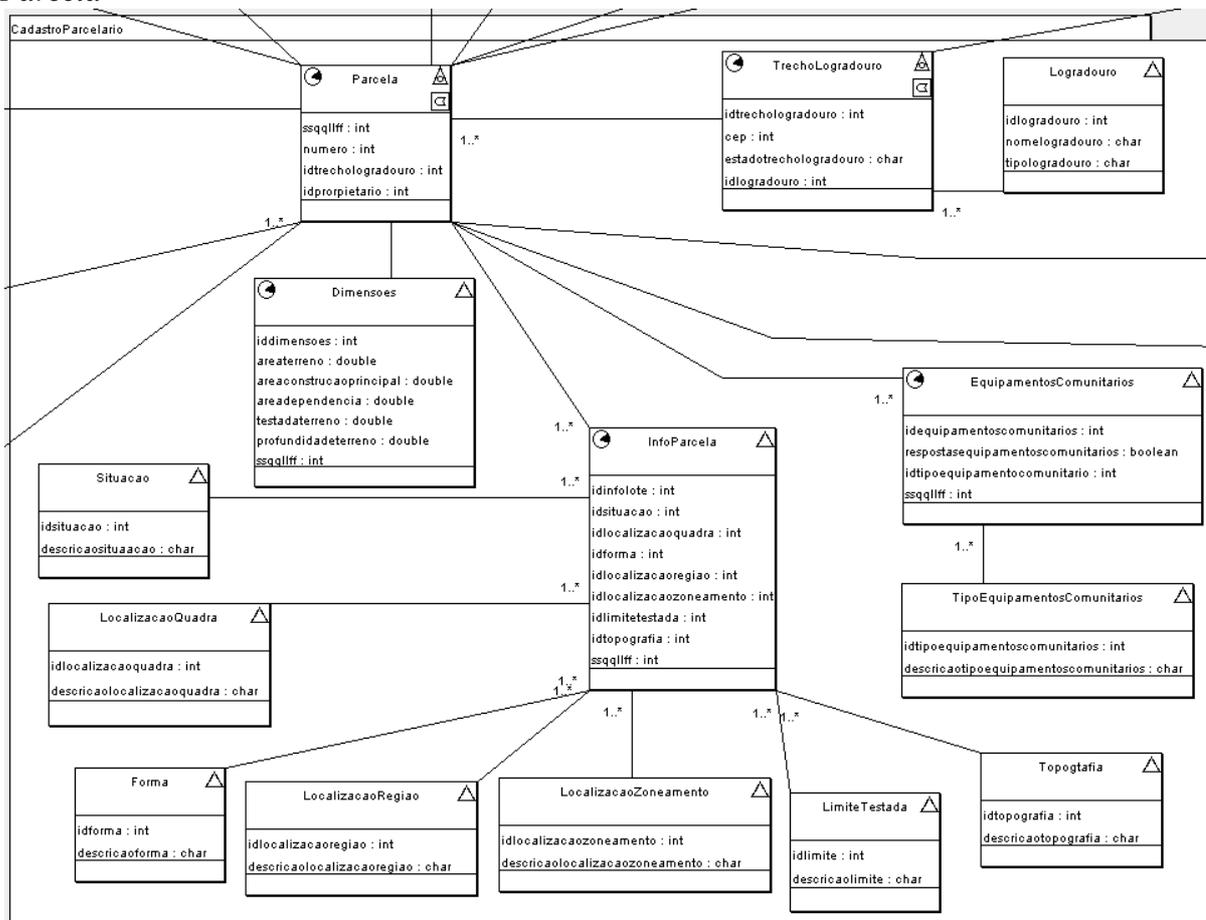
Requisitos do Sistema:

- Parcela: registra os dados básicos sobre a parcela territorial, incluído a geometria, informações sobre o logradouro, dados sobre as dimensões da parcela e das edificações existentes. Também registra dados sobre equipamentos comunitários (segurança, educação, saúde, cultura e lazer) próximos e dados que caracterizam a parcela: topografia, limite de testada, localização de zoneamento, localização na região, forma, localização na quadra e situação.
- Edificação: registra os dados sobre as edificações existentes nas parcelas. Várias características são coletadas: instalações prediais, pavimentos, posição, caracterização, pintura interna e externa, vedação vertical e horizontal, esquadrias, quartos, banheiros, cobertura, piso, estrutura, revestimento interno e conservação.
- Proprietário: registra os dados sobre os proprietários das parcelas.
- Bairro: registra os dados sobre os bairros. Informações básicas, como nome e geometria, e informações mais específicas, como se o bairro é violento ou tem estrutura para a prática de atividades físicas.
- Residentes: registra os dados sobre os residentes, se são homens ou mulheres e a qual faixa etária pertencem. Além de informações sobre atividades físicas praticadas pelos residentes: qualquer esporte, caminhada, vai trabalhar caminhando, vai trabalhar de bicicleta.
- Educação: registra os dados sobre qual o nível de escolaridade os residentes se encontram, incluindo, se existe algum analfabeto.
- Programa Social: registra os dados sobre os programas sociais que atendem os residentes, como: bolsa família, renda cidadã e ação jovem.
- Diagnóstico Médico: registra os dados sobre as patologias que acometem os residentes, entre elas: deficiência física, distúrbio de linguagem, deficiência visual, deficiência auditiva, deficiência mental, doença mental, depressão, mal de Parkinson, cardiopatia, hipertensão arterial, tuberculose, hanseníase, AIDS, diabetes. Além de algumas patologias que acometem residentes acima de 18 anos:

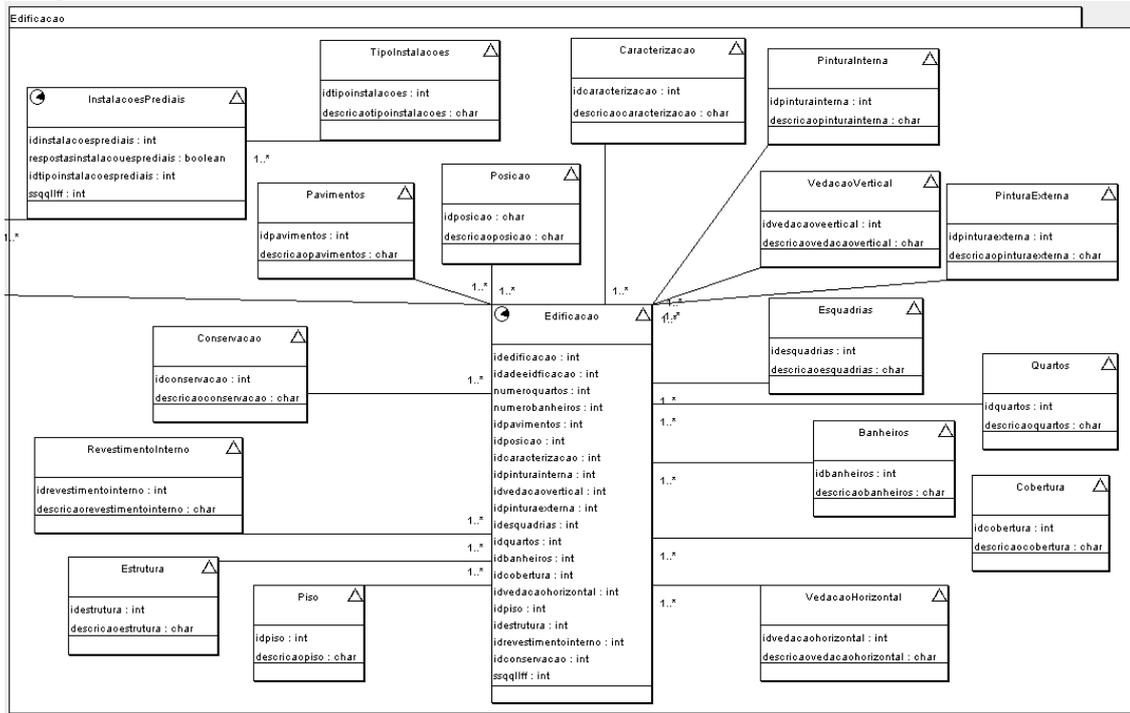
hipertensão com mais de 18 anos, diabético com mais de 18 anos e se faltou ao trabalho por motivo de doença com mais de 18 anos, no último ano.

- Serviço: registra os dados sobre os serviços disponíveis nos logradouros, como: água, esgoto, energia elétrica, iluminação pública, pavimentação, limpeza pública, galerias pluviais, fone, passeio, coleta seletiva.
- Socioeconômico: registra os dados sobre a atividade socioeconômica dos residentes (empregado da indústria, empregado do comércio, empregado rural, empregado de serviços, empregado público, empregado informal, pensionistas, aposentado, desempregados), renda mensal e se possuem veículos com placas local ou de fora do município.

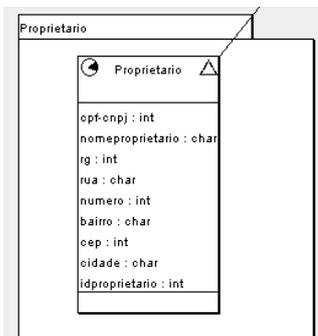
Parcela



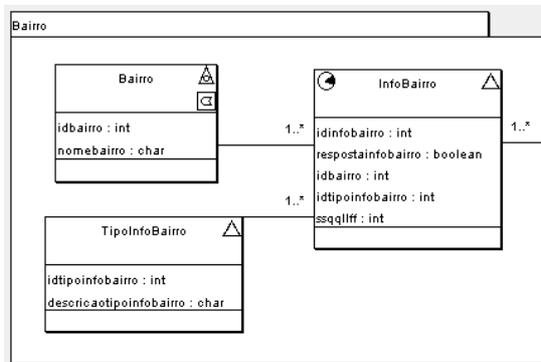
Edificação



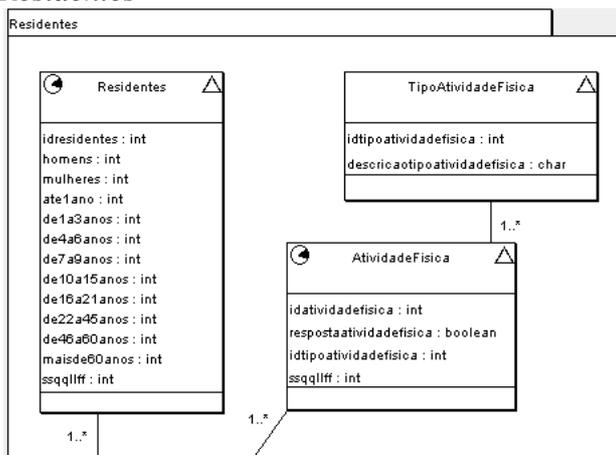
Proprietário



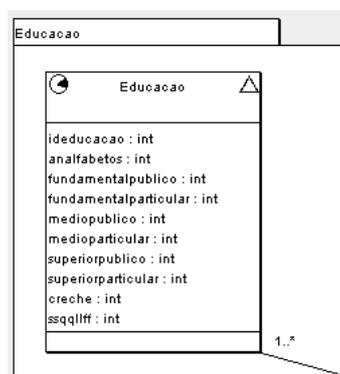
Bairro



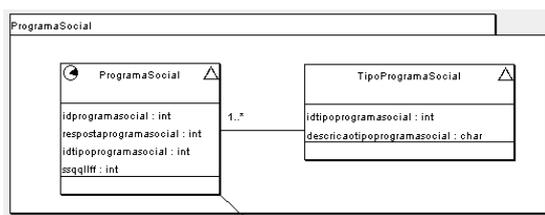
Residentes



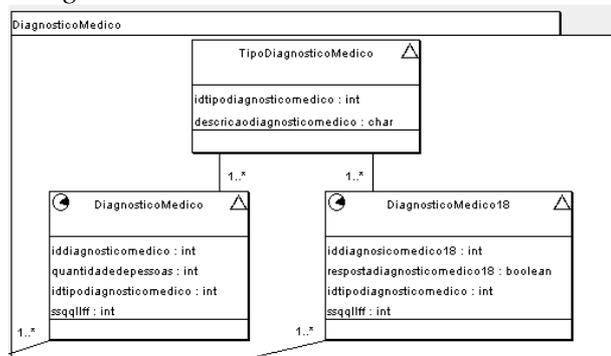
Educação



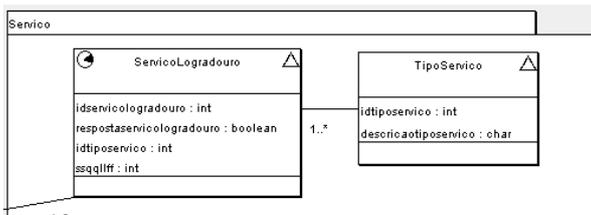
Programa Social



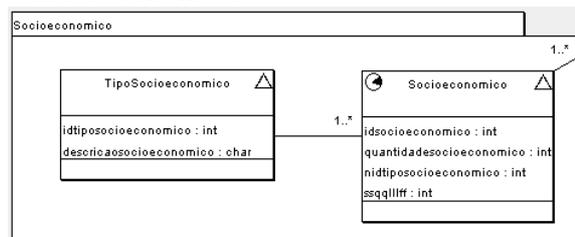
Diagnóstico Médico



Serviço



Socioeconômico



APÊNDICE B – Modelo Global de Dados

Dicionário de Dados.

Projeto: CTM Ribeirão dos Índios	Versão: 1.0
Tipo: entidades	Data: 01/07/2012
Nome	Descrição
parcela	Entidade que registra os dados básicos sobre a parcela territorial.
dimensao	Entidade que registra dados sobre as dimensões da parcela e das construções existentes.
infoparcela	Entidade que registra dados sobre as características físicas da parcela.
situacao	Entidade que registra os tipos de situações possíveis da(s) edificação(ões): não edificado; ruínas; paralisada; andamento; construído.
localizacaoquadra	Entidade que registra os tipos de localizações que a parcela pode ter: esquina; meio da face; mais faces.
forma	Entidade que registra os tipos de forma que a parcela pode ter: regular; irregular; alongada.
limitetestada	Entidade que registra os tipos de limite de testada: aberto; grade; cerca simples; alambrado; tapume; muros/placas.
topografia	Entidade que registra os tipos de topografia: plana; aclave; aclave acentuado; declive; declive acentuado; irregular.
localizacaozoneamento	Entidade que registra os tipos de localização de zoneamento: ZEIS (Zona Especial de Interesse Social); residencial; comercial; industrial; outros.
localizaoregiao	Entidade que registra os tipos de localização na região: centro principal; centralidade periférica; bairro.
logradouro	Entidade que registra os dados sobre os logradouros.
trechologradouro	Entidade que registra os dados sobre os trechos de logradouros.
bairro	Entidade que registra os dados sobre os bairros.
edificacao	Entidade que registra os dados sobre as edificações existentes nas parcelas.
caracterizacao	Entidade que registra os tipos de caracterização das edificações: residencial; comercial; industrial; equipamentos; uso misto.
posicao	Entidade que registra os tipos de posição das edificações: única; conjugada; geminada; superposta.
pavimentos	Entidade que registra os tipos de pavimentos das edificações: um; dois; três; quatro; mais que 4.
conservacao	Entidade que registra os tipos de conservação das edificações: ruim; regular; bom; ótimo.
estrutura	Entidade que registra os tipos de estrutura das edificações: convencional de CA; aço; madeira; alvenaria; outras.
piso	Entidade que registra os tipos de piso das edificações: contrapiso; concreto; todo em cerâmico; cerâmico e madeira; carpete ou equivalente; vinílico ou equivalente.
vedacaohorizontal	Entidade que registra os tipos de vedação horizontal das edificações: ausente; estuque; lambri madeira; forro PVC ou equivalente; laje; laje e forro.
cobertura	Entidade que registra os tipos de cobertura das edificações: ausente; zinco; amianto; cerâmica; aço; especial.
esquadrias	Entidade que registra os tipos de esquadrias das edificações: madeira rústica; aço pintado; aço galvanizado; alumínio; madeira de primeira; outras.
pinturaexterna	Entidade que registra os tipos de pintura externa das edificações: ausente; cal; látex; acrílica; texturizada; outras.
pinturainterna	Entidade que registra os tipos de pintura interna das edificações: ausente; cal; látex; acrílica; texturizada; outras.
vedacaovertical	Entidade que registra os tipos de vedação vertical das edificações:

Projeto: CTM Ribeirão dos Índios	Versão: 1.0
Tipo: entidades	Data: 01/07/2012
Nome	Descrição
	alvenaria de blocos de concreto; outras alvenarias; madeira; gesso acartonado.
revestimento interno	Entidade que registra os tipos de revestimento interno das edificações: ausente; alvenaria aparente; chapisco; reboco; massa corrida; cerâmico.
residentes	Entidade que registra os dados sobre os residentes.
educacao	Entidade que registra os dados sobre o nível de escolaridade.
programa social	Entidade que registra os dados sobre os programas sociais que atendem os residentes.
tipo programa social	Entidade que registra os tipos de programas sociais: bolsa família, renda cidadã e ação jovem.
diagnostico medico	Entidade que registra os dados sobre as patologias que acometem os residentes.
tipo diagnostico medico	Entidade que registra os tipos de patologias: deficiência física, distúrbio de linguagem, deficiência visual, deficiência auditiva, deficiência mental, doença mental, depressão, mal de Parkinson, cardiopatia, hipertensão arterial, tuberculose, hanseníase, AIDS, diabetes; hipertensão com mais de 18 anos; diabético com mais de 18 anos;
servico logradouro	Entidade que registra os dados sobre os serviços disponíveis nos logradouros.
tipo servico logradouro	Entidade que registra os tipos de serviços do logradouro: água, esgoto, energia elétrica, iluminação pública, pavimentação, limpeza pública, galerias pluviais, fone, passeio, coleta seletiva.
socioeconomico	Entidade que registra os dados sobre a atividade socioeconômica dos residentes, renda mensal e se possuem veículos.
tipo socioeconomico	Entidade que registra os tipos de dados socioeconômicos: empregado da indústria, empregado do comércio; empregado rural; empregado de serviço; empregado público; empregado informal; pensionista; aposentado; desempregado; renda mensal; se possui carro com placa local; se possui carro com placa de fora da cidade.
info bairro	Entidade que registra os dados sobre informações de segurança do bairro.
tipo info bairro	Entidade que registra os tipos de informações sobre o bairro: é violento, tem estrutura para a prática de atividades físicas.
equipamento comunitario	Entidade que registra os dados sobre equipamentos comunitários próximos a parcela.
tipo equipamento comunitario	Entidade que registra os tipos de equipamentos comunitários: segurança; educação; saúde; cultura; lazer.
instalacoes prediais	Entidade que registra os dados sobre as instalações prediais.
tipo instalacoes prediais	Entidade que registra os tipos de instalações prediais: água fria; água quente; esgotamento; águas pluviais; comunicações; dados; elétrica.
proprietario	Entidade que registra os dados sobre os proprietários das parcelas.
diagnostico medico 18	Entidade que registra os dados sobre patologias que acometem residentes acima de 18 anos.
atividade fisica	Entidade que registra os dados sobre a realização de atividades físicas pelos residentes.
tipo atividade fisica	Entidade que registra os tipos de atividades físicas: qualquer esporte, caminhada, vai trabalhar caminhando, vai trabalhar de bicicleta.

Projeto: CTM Ribeirão dos Índios		Versão: 1.0	
Tipo: atributos		Data: 01/07/2012	
Nome	Tipo	Entidade	Descrição
ssqllff	integer	parcela (PK), dimensao (FK), infoparcela(FK) edificacao (FK) residentes (FK) educacao (FK) programasocial (FK) diagnosticomedico (FK) servicologradouro (FK) socioeconômico (FK) infobairro (FK) equipamentocomunitario (FK) instalacoesprediais (FK) atividadefisica (FK) diagnosticomedico18 (FK)	Sigla de Setor (ss) Quadra (qq) Lote (ll) e Fração (ff), utilizada para identificar as parcelas.
the_geom	geometry	parcela trechologradouro logradouro bairro	Objeto geográfico com o polígono da parcela.
numero	integer	parcela	Número da parcela no logradouro.
iddimensao	integer	dimensao (PK)	Identificador da entidade dimensão.
areadoterreno	double	dimensao	Área do terreno que compreende a parcela, m ² .
areaprincipal	double	dimensao	Área da construção principal, m ² .
areadependencia	double	dimensao	Área da construção de dependência, m ² .
testada	double	dimensao	Medida da testada da parcela, metros.
profundidade	double	dimensao	Medida da profundidade da parcela, metros.
idinfoparcela	integer	infoparcela (PK)	Identificador da entidade infoparcela.
idosituacao	integer	situacao (PK), infoparcela (FK)	Identificador da entidade situacao.
descricaoosituacao	varchar	situação	Descrição dos tipos de situação.
idlocalizacaoquadra	integer	localizacaoquadra (PK), infoparcela (FK)	Identificador da entidade localizacaoquadra.
descricaolocalizacaoquadra	varchar	localizacaoquadra	Descrição dos tipos de localização na quadra.
idforma	integer	forma (PK), infoparcela (FK)	Identificador da entidade forma
descricaoforma	varchar	forma	Descrição dos tipos de forma.
idlimitetestada	integer	limitetestada (PK) infoparcela (FK)	Identificador da entidade limitetestada.
descricaoolimitetestada	varchar	limitetestada	Descrição dos tipos de limites da testada.
idtopografia	integer	topografia (PK) infoparcela (FK)	Identificador da entidade topografia.
descricaootopografia	varchar	topografia	Descrição dos tipos de topografias.

Projeto: CTM Ribeirão dos Índios		Versão: 1.0	
Tipo: atributos		Data: 01/07/2012	
Nome	Tipo	Entidade	Descrição
idlocalizacaozoneamento	integer	localizacaozoneamento (PK) infoparcela (FK)	Identificador da entidade localizacaozoneamento.
descricaolocalizacaozoneamento	varchar	localizacaozoneamento	Descrição dos tipos de localização de zoneamento.
idlocalizaoregiao	integer	localizaoregiao (PK) infoparcela (FK)	Identificador da entidade localizaoregiao.
descricaolocalizaoregiao	varchar	localizaoregiao	Descrição dos tipos de localização na região.
idtrechologradouro	integer	trechologradouro (PK) servicologradouro (FK) parcela (FK)	Identificador da entidade trechologradouro.
cep	integer	trechologradouro	Código de endereçamento postal.
estadotrechologradouro	varchar	trechologradouro	Estado de conservação do trecho do logradouro.
idlogradouro	integer	logradouro (PK), trechologradouro(FK)	Identificador da entidade logradouro.
nomelogradouro	varchar	logradouro	Nome do logradouro.
tipologradouro	varchar	logradouro	Tipo do logradouro.
idbairro	integer	bairro (PK) infobairro (FK)	Identificador da entidade bairro.
nomebairro	varchar	bairro	Nome do bairro.
idedificacao		edificacao(PK)	Identificador da entidade edificação.
idcaracterizacao	integer	caracterizacao (PK), edificacao (FK)	Identificador da entidade caracterização.
descricaocharacterizacao		caracterizacao	Descrição dos tipos de caracterização da edificação.
idposicao	integer	posicao (PK), edificacao (FK)	Identificador da entidade posição.
descricaooposicao		posicao	Descrição dos tipos de posição da edificação.
idpavimentos	integer	pavimentos (PK), edificacao (FK)	Identificador da entidade pavimentos.
descricaoopavimentos		pavimentos	Descrição dos tipos de pavimentos.
idconservacao	integer	conservacao (PK), edificacao (FK)	Identificador da entidade conservação.
descricaoconservacao		conservação	Descrição dos tipos de conservação.
idestrutura	integer	estrutura (PK), edificacao (FK)	Identificador da entidade estrutura.
descricaoestrutura		estrutura	Descrição dos tipos de estruturas.
idpiso	integer	piso (PK) edificacao (FK)	Identificador da entidade piso.
descricaoopiso		piso	Descrição dos tipos de pisos.
idvedacaohorizontal	integer	vedacaohorizontal (PK), edificacao (fk)	Identificador da entidade vedação horizontal.
descricaoovedacaohorizontal		vedacaohorizontal	Descrição dos tipos de vedação horizontal.
idcobertura	integer	cobertura (PK) edificacao (FK)	Identificador da entidade cobertura.

Projeto: CTM Ribeirão dos Índios		Versão: 1.0	
Tipo: atributos		Data: 01/07/2012	
Nome	Tipo	Entidade	Descrição
descricaocobertura		cobertura	Descrição dos tipos de cobertura.
idesquadrias	integer	esquadrias (PK) edificacao (FK)	Identificador da entidade esquadrias.
descricaoesquadrias		esquadrias	Descrição dos tipos de esquadrias.
idpinturaexterna	integer	pinturaexterna (PK) edificacao (FK)	Identificador da entidade pintura externa.
descricaopinturaexterna		pinturaexterna	Descrição dos tipos de pinturas externas.
idpinturainterna	integer	pinturainterna (PK) edificacao (FK)	Identificador da entidade pintura interna.
descricaopinturainterna		pinturainterna	Descrição dos tipos de pinturas internas.
idvedacaovertical	integer	vedacaovertical (PK) edificacao (FK)	Identificador da entidade vedação vertical.
descricaovedacaovertical		vedacaovertical	Descrição dos tipos de vedação vertical.
idadeedificacao	integer	edificacao	Idade da edificação.
idrevestimentointerno	integer	revestimentointerno (PK) edificacao (FK)	Identificador da entidade revestimento interno.
descricaorevestimentointerno		revestimentointerno	Descrição dos tipos de revestimentos internos.
idresidentes	integer	residentes (PK)	Identificador da entidade residentes.
homens	integer	residentes	Quantidade de homens residentes.
mulheres		residentes	Quantidade de mulheres residentes.
ate1ano		residentes	Quantidade até 1 ano.
de1a3anos	integer	residentes	Quantidade de 1 a 3 anos.
de4a6anos	integer	residentes	Quantidade de 4 a 6 anos.
de7a9anos	integer	residentes	Quantidade de 7 a 9 anos.
de10a15anos	integer	residentes	Quantidade de 10 a 15 anos.
de16a21anos	integer	residentes	Quantidade de 16 a 21 anos.
de22a45anos	integer	residentes	Quantidade de 22 a 45 anos.
de46a60anos	integer	residentes	Quantidade de 46 a 60 anos.
maisde60anos	integer	residentes	Quantidade com mais de 60 anos.
ideducacao	integer	educacao (PK)	Identificador da entidade educação.
analfabetos	integer	educacao	Quantidade de analfabetos.
fundpublico	integer	educacao	Quantidade de alunos no ensino fundamental público.
fundparticular	integer	educacao	Quantidade de alunos no ensino fundamental particular.
medpublico	integer	educacao	Quantidade de alunos no ensino médio público.
medparticular	integer	educacao	Quantidade de alunos no

Projeto: CTM Ribeirão dos Índios		Versão: 1.0	
Tipo: atributos		Data: 01/07/2012	
Nome	Tipo	Entidade	Descrição
			ensino médio particular.
suppublico	integer	educacao	Quantidade de alunos no ensino superior público.
supparticular	integer	educacao	Quantidade de alunos no ensino superior particular.
creche	integer	educacao	Quantidade de crianças na creche.
idprogramasocial		programasocial (PK)	Identificador da entidade programa social.
idtipoprogramasocial	integer	tipoprogramasocial (PK) programasocial (FK)	Identificador da entidade tipo de programa social.
respostaprogramasocial	booleano	programasocial	Resposta sobre o programa social, sim/não.
descricaotipoprogramasocial	varchar	tipoprogramasocial	Descrição do tipo de programa social.
iddiagnosticomedico	integer	diagnosticomedico (PK)	Identificador da entidade diagnóstico médico.
idtipodiagnosticomedico	integer	tipodiagnosticomedico (PK) diagnosticomedico (FK)	Identificador da entidade tipo de diagnóstico médico.
quantidadepessoas	integer	diagnosticomedico	Quantidade de residentes com a patologia.
descricaodiagnosticomedico	varchar	tipodiagnosticomedico	Descrição do tipo de patologia.
idservicologradouro	integer	servicologradouro (PK)	Identificador da entidade serviço logradouro.
idtiposervico	integer	tiposervicologradouro (PK) tiposervicologradouro (FK)	Identificador da entidade tipo de serviço.
respostaservicologradouro	booleano	servicologradouro	Respostas sobre o serviço, sim/não.
descricaotiposervico	varchar	tiposervicologradouro	Descrição do tipo de serviço.
idsocioeconomico	integer	socioeconomico (PK)	Identificador da entidade socioeconômico.
idtiposocioeconomico	integer	tiposocioeconomico (PK) socioeconomico (FK)	Identificador da entidade tipo de informação socioeconômica.
quantidadesocioeconomico	integer	socioeconomico	Quantidade de ocorrência da atividade socioeconômica.
descricaosocioeconomico	varchar	tiposocioeconomico	Descrição da atividade socioeconômica.
idinfobairro	integer	infobairro (PK)	Identificador da entidade infobairro.
idtipoinfobairro	integer	tipoinfobairro (PK) infobairro (FK)	Identificador da entidade tipo infobairro.
respostainfobairro	booleano	infobairro	Resposta das informações sobre o bairro.
descricaoinfobairro	varchar	tipoinfobairro	Descrição da informação relacionada ao bairro.
idequipamentocomunitario	integer	equipamentocomunitario (PK)	Identificador da entidade equipamento comunitario.
idtipoequipamentocomunitario	integer	tipoequipamentocomunitario (PK) equipamentocomunitario (FK)	Identificador da entidade tipo equipamentocomunitario.
respostaequipamentocomunitario	integer	equipamentocomunitario	Resposta sobre o

Projeto: CTM Ribeirão dos Índios		Versão: 1.0	
Tipo: atributos		Data: 01/07/2012	
Nome	Tipo	Entidade	Descrição
			equipamento comunitário.
descricaoequipamentocomunitario	varchar	tipoequipamentocomunitario	Descrição do equipamento comunitário.
idinstalacoesprediais	integer	instalacoesprediais (PK)	Identificador da entidade instalacoesprediais.
idtipoinstalacoesprediais		tipoinstalacoesprediais (PK) instalacoesprediais (FK)	Identificador da entidade tipoinstalacoesprediais.
respostainstalacoesprediais	booleano	instalacoesprediais	Respostas sobre as instalações prediais.
descricaoinstalacoesprediais	varchar	tipoinstalacoesprediais	Descrição do tipo de instalações prediais.
idatividadefisica	integer	atividadefisica (PK)	Identificador da entidade atividadefisica.
respostaatividadefisica	booleano	atividadefisica	Resposta sobre a atividade física.
idtipoatividadefisica	integer	tipoatividadefisica (PK) atividadefisica (FK)	Identificador da entidade tipo de atividadefisica.
descricaoatividadefisica	varchar	tipoatividadefisica	Descrição da informação sobre atividade física.
iddiagnosticomedico18	integer	diagnosticomedico18 (PK)	Identificador da entidade diagnosticomedico18.
idtipodiagnosticomedico	integer	tipodiagnosticomedico18 (PK) diagnosticomedico18 (FK)	Identificador da entidade tipodiagnosticomedico18.
respostadiagnosticomedico18	booleano	diagnosticomedico18	Resposta sobre o diagnóstico médico para residentes acima de 18 anos.
descricaoatividadefisica		tipodiagnosticomedico18	Descrição da informação sobre o diagnóstico médico para residentes acima de 18 anos.
idproprietario	integer	proprietario (PK) parcela (FK)	Identificador da entidade proprietario.
cpf-cnpj	integer	proprietario	CPF ou CNPJ do proprietário da parcela.
nomeproprietario	varchar	proprietario	Nome do proprietário da parcela.
rg	integer	proprietario	Registro Geral do proprietário da parcela.
rua	varchar	proprietario	Logradouro onde reside o proprietário.
numero	integer	proprietario	Número da parcela onde reside o proprietário.
bairro	varchar	proprietario	Bairro onde reside o proprietário.
cep	integer	proprietario	CEP onde reside o proprietário.
cidade	varchar	proprietario	Cidade onde reside o proprietário.

APÊNDICE C – Spatio-Temporal Data Warehouse

A modelagem do *STDW* é baseada no modelo MultiDim (MALINOWSKI; ZIMÁNYI, 2008). A ferramenta usada para a diagramação foi ArgoCASEGEO v3.0, escolhida por permitir a definição de estereótipos para a geometria e o tempo.

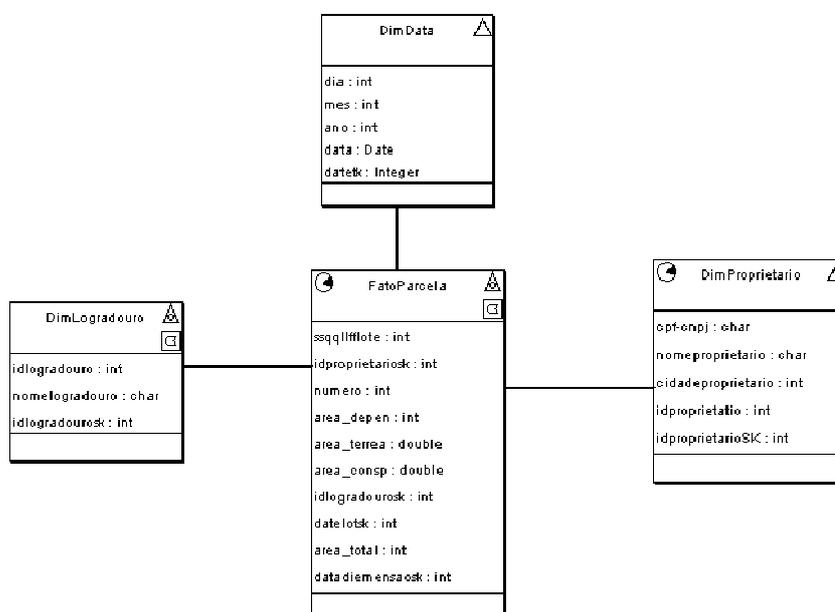
Parcela

Processo: Parcela

Granularidade: ano e parcela.

Dimensões: DimData, DimProprietario e DimLogradouro

Fato: FatoParcela



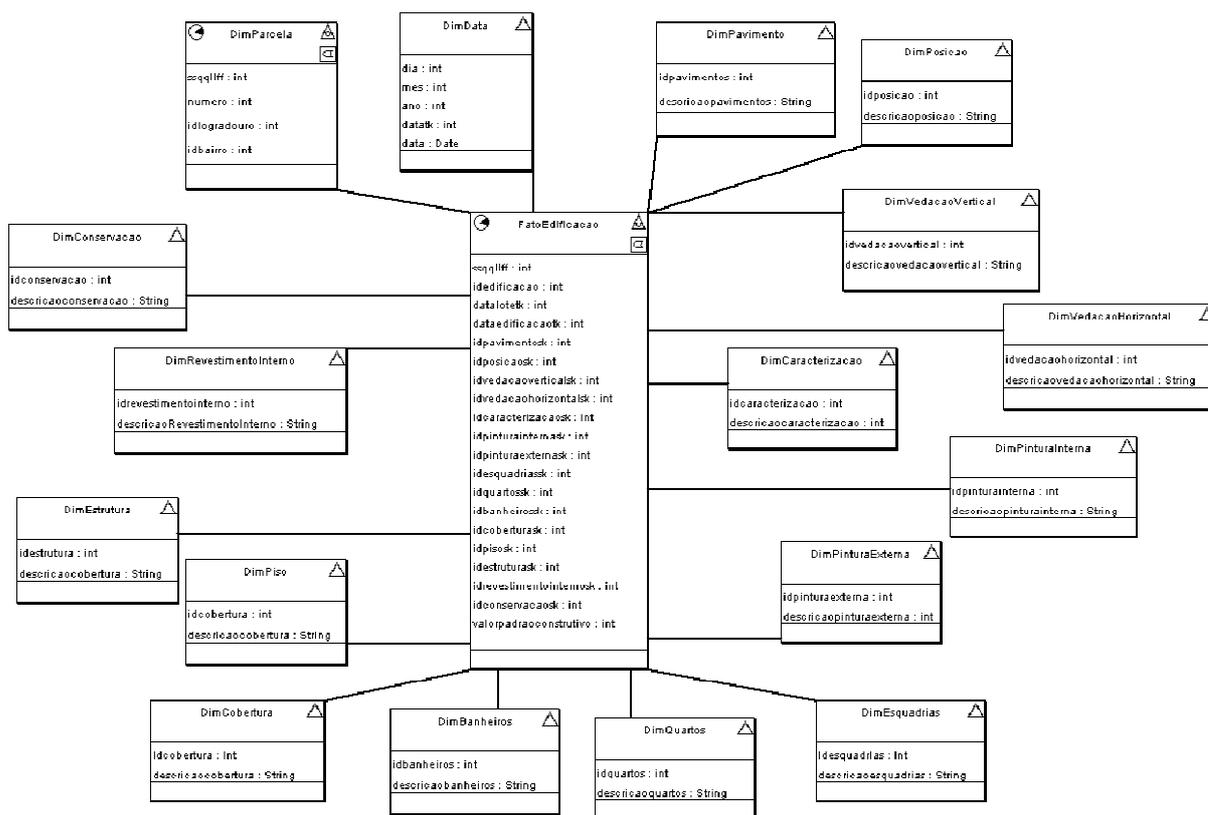
Edificação

Processo: Edificação.

Granularidade: ano e parcela.

Dimensões: DimData, DimParcela, DimPavimento, DimPosicao, DimVedacaoVertical, DimVedacaoHorizontal, DimCaracterizacao, DimPinturaInterna, DimPinturaExterna, DimEsquadrias, DimQuartos, DimBanheiros, DimCobertura, DimPiso, DimEstrutura, DimRevestimentoInterno, DimConservacao, DimInstalacoesPrediais.

Fato: FatoEdificacao.



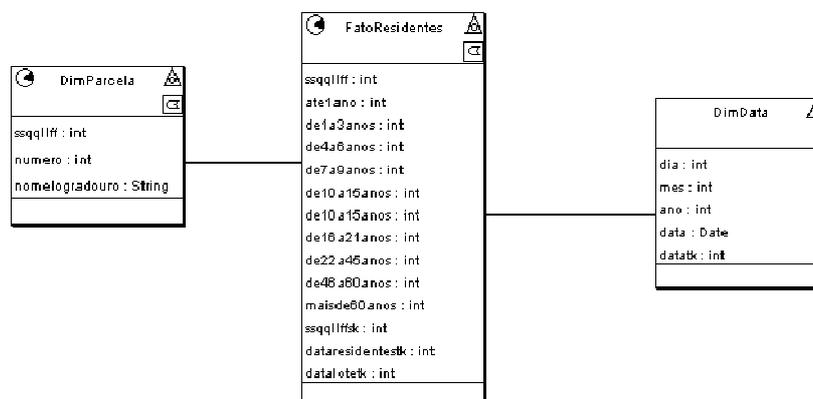
Residentes

Processo: Residentes.

Granularidade: ano e parcela.

Dimensões: DimParcela, DimData.

Fato: FatoResidentes.



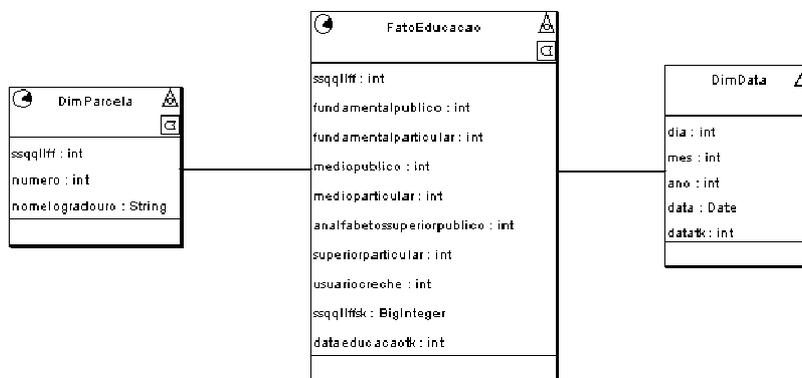
Educação

Processo: Educação

Granularidade: ano e parcela.

Dimensões: DimParcela, DimData.

Fato: FatoEducacao.



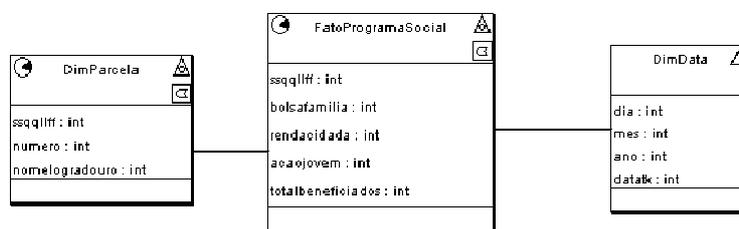
Programas Sociais

Processo: Programas Sociais.

Granularidade: ano e parcela.

Dimensões: DimParcela, DimData

Fato: FatoProgramaSocial.



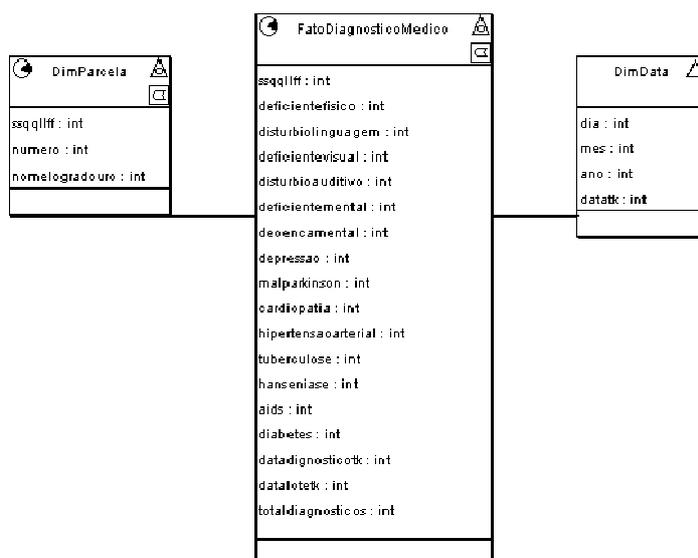
Diagnóstico Médico

Processo: Diagnóstico Médico.

Granularidade: ano e parcela.

Dimensões: DimParcela, DimData.

Fato: FatoDiagnosticoMedico.



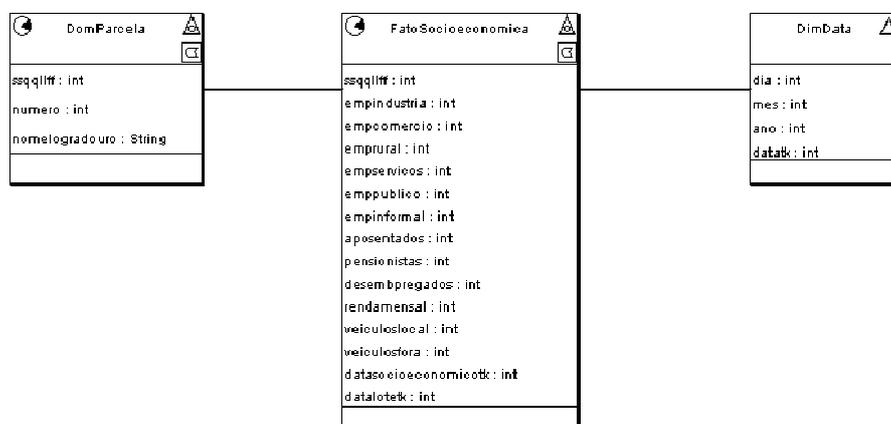
Socioeconômico

Processo: Socioeconomico.

Granularidade: ano e parcela.

Dimensões: DimParcela, DimData.

Fato: FatoSocioeconomico.



Instalações Prediais

Processo: Instalações Prediais.

Granularidade: ano e parcela.

Dimensões: DimParcela, DimData.

Fato: FatoDimensoesPrediais.

