

A Relação Entre os SIG's e as Estruturas de Disseminação de Informações

Jorge Henrique de Castro

UERJ – Mestrado em Engenharia de Computação - Concentração Geomática
Petrobras - Gás e Energia – Planejamento Estratégico
Av. Almirante Barroso, 81 33º andar
23030-003 Rio de Janeiro RJ – Brasil
jorgecastro@petrobras.com.br

Resumo – A informação está presente em todos os processos de uma empresa e desta forma em todos os níveis de tomada de decisão. Existem duas perspectivas a atender: a dos usuários, que trata da descoberta de conhecimento para o estabelecimento de estratégias de negócio; e a de tecnologia de informação que abrange os sistemas e processos de tratamento destas informações. Para tal é necessário desenvolver estruturas para armazenar, processar e recuperar dados. Este ambiente, intitulado “*Estrutura de Disseminação de Informações*”, constitui o alicerce da distribuição de informações na empresa. Nesta estrutura destaca-se a existência dos *DataWarehouse*'s que armazenam ordenadamente e temporalmente os dados, e os SIG's que permitem estabelecer padrões de comportamento a partir da utilização da informação espacial. Este artigo aborda a aplicação do SIG como ferramenta de manutenção de cadastros e as possíveis formas de associação deste à estrutura de disseminação de informações, seus benefícios e características.

Palavras- chave – Informação espacial, SIG, Conhecimento, Data Warehouse.

Abstract – The information is present in all enterprise process as well as in different decision levels. There are two views: the user's view, which is related to knowledge discovery to business strategies, and the information technology view, including both management information systems and process. It is necessary to develop structures to store, process and recover data. This structure, named “Corporate Information Factory” (CIF), is the basis of enterprise information distribution. The most important stage in this structure is the DataWarehouse (DW) that store data temporally. The GIS allows establishing pattern from use of spatial information. This paper presents the benefits and characteristics of GIS application as a tool of cadastre management and the possible ways of GIS association to CIF.

Keywords – Spatial information; GIS; knowledge; DataWarehouse.

1 - Introdução.

A geração de conhecimento a partir de análises preditivas coloca a empresa na terceira onda de inteligência, onde as perguntas são do tipo “O que vai acontecer”? Este estágio eleva as empresas a um patamar totalmente diferenciado e permite que elas antecipem tendências antes que estas se tornem conhecimento comum e se capitalizem com isto.

A necessidade de informações íntegras, integradas e históricas está presente em todos os processos do negócio e em todos os níveis de tomada de decisão (estratégico tático ou operacional). A incorporação efetiva da inteligência de negócios, em uma empresa, a induz a ter no contexto de inteligência a extração da informação, a partir dos dados, transformando-os em conhecimento, e mais importante que tudo, transformando este conhecimento em ação.

Nesta proposição devem ser vistas as perspectivas do usuário e da área de tecnologia de informação. Sob a perspectiva do usuário, o objetivo da inteligência de negócios foca o direcionamento desejado da organização, dentro do segmento de mercado que se encontra inserido. A necessidade de conseguir uma orientação estratégica voltada ao relacionamento com os clientes. Seu histórico e suas tendências. As modificações introduzidas no comportamento do cliente em virtude da evolução das relações e comportamentos da sociedade.

Sobre a perspectiva de TI os objetivos visam proporcionar a infra-estrutura necessária seja através da disponibilização de informações e conhecimentos, seja através da provisão de equipamento e soluções, para tratar, armazenar, transformar dados em conhecimento.

Captar e reter clientes é primordial para a sobrevivência da mesma em um ambiente de livre comércio e de livre concorrência. Neste sentido gerir o cadastro de clientes torna-se primordial no sentido de descobrir as necessidades dos mesmos. Saber onde o cliente se encontra, o que posso fornecer para ele naquele lugar onde ele se encontra (exemplo os serviços de empresas de utilidades) e em quanto tempo isto pode ser feito, de forma que não se crie dissonância cognitiva (quebra negativa da expectativa do cliente) (Castro, 2005).

2 - Gestão dos cadastros de Clientes e de endereços.

O processo de descoberta do conhecimento, ou melhor, de construção do conhecimento envolve a cadeia de transformação de dados em conhecimento propriamente dito. O tratamento necessário a este processo demanda o estabelecimento de estruturas para armazenar dados, com processamento e recuperação em alta velocidade; estas estruturas são denominadas sistemas de informações e se utilizam de bancos de dados como elementos de armazenamento de dados e informações (Castro, 2005). A partir destes bancos serão processadas as atividades necessárias à construção do conhecimento.

Porém é necessário entender como atender e, com o que atender os clientes. Neste ponto se introduz a informação espacial como elemento de descoberta de conhecimento; por exemplo, à capacidade de reprodução espacial da distribuição do consumo e dos clientes, aliada ao estudo destas interações, a partir de modelos de decisão e análises espaciais, ou o que hoje conhecemos como Geomarketing.

Analisando estes dois parágrafos anteriores pode-se relacionar quatro grupos de estruturas que atuam no processo de descoberta de conhecimento onde são considerados fatores críticos a existência e manutenção da base de clientes e de endereços (localização espacial):

- Base de endereços – de uma forma resumida, consiste no registro dos atributos em um banco de dados voltado para a criação e manutenção da base de dados de endereços. Por exemplo: bairro, código de logradouro, tipo do logradouro, título do logradouro, nome do logradouro, código de endereçamento postal (CEP).
- Base de clientes – de forma resumida, consiste em um modelo de dados para criação e manutenção da base de dados de clientes ou consumidores de uma determinada empresa. Neste cadastro serão armazenados os dados referentes ao endereço, informações pessoais, produtos consumidos, tempo de relacionamento, entre outros.
- MUB - Mapa urbano básico - Segundo Aguiar (Aguiar, 1996) constitui o conjunto de informações gráficas e alfanuméricas referentes à base cartográfica a ser utilizada, em geral relativa às plantas cadastrais, ou seja, os vários planos de informações (feições) que compõem o cadastro urbano e que permitem sua apresentação gráfica.
- Estrutura de disseminação da informação ou CIF – *Corporate Information Factory* – é um sistema com diferentes componentes onde cada um atende a uma especificidade, ao

mesmo tempo, trabalhando em conjunto com outros componentes, para garantir um ambiente de informações coeso e equilibrado (Cruz, 2006).

3 - A estrutura de disseminação de informações.

A estrutura de disseminação de informações constitui o alicerce para a disponibilização destas no ambiente da empresa e oferece suporte aos PTD's – Processos de tomada de decisão. Ao mesmo tempo em que é genérico apresenta particularidades para cada empresa em função do tipo de negócio e gestão associadas ao empreendimento. Segundo Cruz (Cruz, 2006) uma arquitetura funcional típica de disseminação de dados consta de seis grandes conjuntos de armazenamentos e acessos:

- OLPTS - *On Line Transaction Processing* – são os sistemas utilizados no ambiente de produção, ou seja, o conjunto de sistemas transacionais que suportam a operação da empresa.
- ETL aquisição – são os processos responsáveis pela aquisição de dados junto ao ambiente transacional.
- ODS – *Operational Data Store* – oferece uma visão integrada de dados para suportar os processos de tomada de decisão no nível tático. É uma base de dados orientada a assuntos, integrada, referente ao momento presente e de forma volátil.
- DW – *Data Warehouse* - oferece uma visão integrada de dados ao longo do tempo. É uma base de dados orientada a assuntos, integrada, temporal e não volátil.
- ETL disponibilização – são os processos responsáveis pela disponibilização dos dados para o ambiente gerencial.
- *Data Marts* – são bases de dados multidimensionais com o objetivo de atender as necessidades de um grupo específico de usuários. São compostos por informações oriundas do ODS ou do DW.
- Ferramentas OLAP – *On line Analytic Process* – ferramentas de análises de dados que operam sobre os *Data Marts* constituídos.

A figura 1 apresenta o esquemático de interação destes elementos e seus respectivos na arquitetura funcional de disseminação de informações.

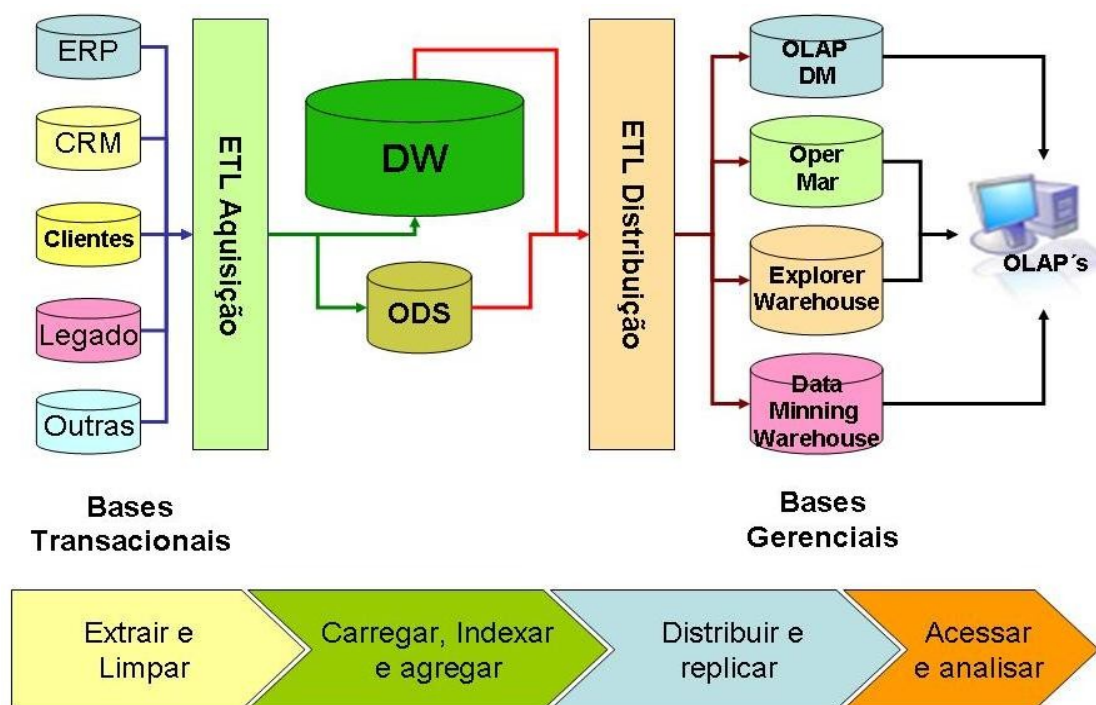


Figura 1 – Esquemático da estrutura de disseminação da informação ou CIF.

No estágio de extração e limpeza são tratados os assuntos relativos à aquisição de dados dos sistemas transacionais, a limpeza e o tratamento dos dados extraídos e a consolidação de arquivos de transferência para o posterior armazenamento no DW e no ODS.

No estágio de carga, indexação e agregação são executadas as operações de recebimento dos arquivos de transferência gerados pela ETL de extração e limpeza, sua posterior indexação aos meta-dados pertinentes e a agregação dos novos valores (*up-load*) as bases já existentes.

No estágio de distribuição e replicação são tratados os assuntos relativos à aquisição de dados dos sistemas transacionais, a limpeza e o tratamento dos dados extraídos e a consolidação de arquivos de transferência para o posterior armazenamento no DW e no ODS.

No ambiente de análise estão dispostos sistemas que realizam os processos analíticos e que utilizam as fontes armazenadas nos *data marts*.

4 - O papel do *DataWarehouse*.

O DW desempenha uma série de funções dentro da arquitetura funcional de disseminação de informações. Sobre o DW é estabelecido primeiramente o armazenamento dos dados dos sistemas transacionais.

Isto não se dá por uma simples cópia. Este processo é feito através de uma ação de *Data Warehousing*, ou seja, o processo de extrair, armazenar e transformar dados operacionais oriundo de diversas fontes de dados em dados adequados ao processo de tomada de decisão gerencial, fornecendo uma visão única de todo o negócio.

O DW possui uma perspectiva temporal o que implica que qualquer registro neste ambiente está relacionado a um determinado momento no tempo. Seu conjunto de dados não é volátil uma vez que normalmente não ocorrem atualizações. Quando ocorrem mudanças que devem ser armazenadas são geradas fotografias que refletem esta mudança ocorrida no tempo e desta forma constituem simples *up-grades* aos dados preexistentes. Ainda sobre o DW são estabelecidos os ETL que extraem informações para composição dos *Data Marts*. Estes têm visão dimensional e suportarão os processos de análises de dados.

Deste modo, o DW é integrado, centralizado e orientado a assuntos e desta forma atende ao suporte à decisão. Portanto é elemento fundamental na garantia da qualidade dos dados que serão replicados. Como o DW funciona como um centralizador da informação organizado por assuntos é imperativo que exista qualidade das fontes que alimentam o mesmo.

Para o cadastramento de um cliente é necessário que se conheça antecipadamente o seu endereço, ou seja, sem o endereço não é possível um completo e correto cadastramento. Em função desta pré-condição a integração das bases de endereços e do cadastro de clientes é quase que perfeita, pois a segunda depende da primeira para ser constituída. O que isto quer dizer? Quando estas bases são exportadas para um DW não deveriam apresentar problemas, pois são integradas no nascedouro. Então porque existem valores perdidos ou incoerentes após as migrações? A explicação é muito simples e ocorre por dois motivos:

- 1º - quando é realizada alguma atividade de limpeza e / ou padronização de dados nas ETL's surgem, ou são introduzidos, os valores fora de padrão;
- 2º - quando se utiliza uma referência de levantamento de campo, uma base de MUB e/ou uma base de endereços, os valores que são inválidos passarão a ser críticas de banco de dados. (não entendi)

Outra questão chave segundo Cruz (Cruz, 2006) é focar os processos de limpeza de dados, ou *data cleaning*, antes de serem carregados no DW evitando assim que estes tratamentos sejam realizados em todos os procedimentos posteriores. Desta forma o dado disponibilizado já se encontrará padronizado fisicamente, com o conjunto de valores de referência determinados, com valores convertidos e decodificados estabelecendo assim um padrão de dados.

5 – A entrada de clientes e o cadastro urbano.

Um logradouro quando criado esteja correto ou não, servirá a partir daquele momento como a referência básica de cadastramento para qualquer operação de manutenção sobre este cadastro. Uma vez que esta referência se estabelece, e esteja errada, qualquer agregação de informação que utilize esta referência

estará errada.

Quando um cliente omite ou adultera a informação de um bairro ou logradouro induz ao mantenedor do cadastro que uma nova referência passou a existir. Uma vez criada será replicada para todos os sistemas legados e desta forma tratada e analisada.

Independente do tipo de erro que possa surgir ou se propagar temos uma nova fronteira neste ponto: o custo da não qualidade. Para as empresas existe uma decisão entre se preparar para tratar as inconsistências oriundas dos diversos processos operacionais ou implantar ferramentas e estruturas de cadastro necessárias para garantir a qualidade.

Isto envolve prioritariamente decidir sobre construir uma base de endereços e mantê-la atualizada, o que demanda tempo e altos investimentos, ou criar *back offices* para tratamento de inconsistências de cadastro, o que pode se tornar uma operação sem fim visto que não se tratam as fontes de introdução de erros.

No caso do cadastro de endereços estar mal alicerçado as informações dos atuais clientes fazem parte de um passivo a ser tratado, isto explica a necessidade de existência de *back offices* para este tipo de ajuste. O dado do jeito que se encontra será carregado no DW e desta forma replicado a todo processo de KDD.

Na verdade isto depende se o cadastro de um cliente foi feito a partir de uma base certificada de endereços levantados em campo e armazenados em bancos de dados ou se este foi feito sem uma base certificadas de endereços, pois no primeiro caso se extinguirão as condições que levaram a adoção de limpeza de dados no que tange a parte do endereço do domicílio.

O DW por si só não tem condições de questionar e julgar as informações oriundas da integração da base de clientes e de endereços. Isto só poderá ser feito utilizando processos ou aplicações específicas. Nestas aplicações específicas serão implementadas as regras de negócios e/ ou os processos que permitam a utilização de uma base certificada de endereços levantados em campo e armazenados em bancos de dados com a utilização de uma base digital de mapas integrados, por exemplo, um SIG. Ou ainda serão implantados processos menos automáticos de controle, equipes de visitação a campo, que permitirão rastrear e tratar as não conformidades.

Desta forma, caso não se implante ferramentas específicas de certificação ou controle destes cadastros os *prospects* (*pessoas que ainda não são clientes, mas já foram prospectados*) e os *suspects* (*pessoas que ainda não são clientes, mas que tem potencial para tal*) passarão pelos atuais processos de cadastramento, e poderão transformar-se em passivo futuro para tratamento de dados, visto que todas as possibilidades de erros já citados tendem a se acumular. A figura 2 mostra uma representação esquemática desta situação.

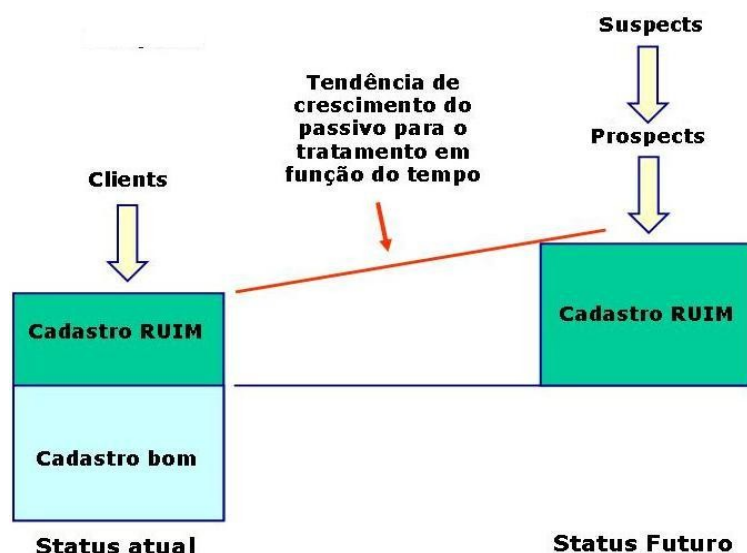


Figura 2 – Crescimento do passivo de cadastro.

Não atuar sobre os processos de entrada de clientes aumenta os custos com o tratamento das não conformidades. A existência de *back offices* não significa ter controle sobre o processo de manutenção de cadastro, pois esta equipe tratará o erro e não impedirá que novos aconteçam.

Quando uma empresa atuar em regiões urbanas de alta densidade populacional e desorganização estrutural, por exemplo, na periferia, estas equipes tendem a ficar gigantescas e com o custo muito alto, pois a natureza da desorganização urbana permeará a entrada de dados a partir do cadastramento dos clientes.

6 - Proposições de Ancoragem.

Os SIG's não são passíveis da classificação utilizada no item 3 visto que não são ferramentas puramente de cunho transacional ou gerencial. A estrutura básica de uma plataforma de SIG permite que o mesmo esteja em qualquer um dos estágios definidos, ou até em todos, visto que pode se apresentar como uma ferramenta de cadastro ou como uma ferramenta de análise.

Independente disto, uma vez que se adota uma plataforma SIG como ferramenta de manutenção de cadastros, seja isto de forma ativa ou re-ativa, será necessário realizar sua ancoragem a arquitetura funcional existente.

A forma como esta plataforma vai ser adicionada à arquitetura funcional e as regras de negócio que forem implantadas nas soluções de software é que irão definir a estratégia de qualidade que poderá ser obtida.

Por hora a referência de arquitetura funcional é a forma mais tradicional, ou seja, o cadastro alfanumérico e a plataforma digital (SIG ou CAD) associados ao DW (*Data Warehouse*) através de ETL (Estágios de Tratamento e Limpeza) e servindo como início do processo de disseminação de informação baseada na distribuição espacial, como se pode ver na figura 3.

Nesta arquitetura a plataforma de SIG é uma plataforma estanque se relacionando com o com o DW através de conexões de dados, por exemplo, as conexões ODBC – *Oracle Data Base Consult*.

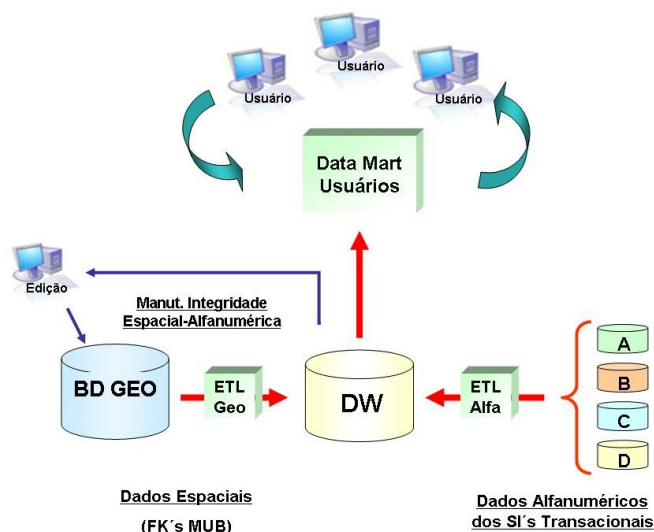


Figura 3 – Arquitetura funcional tradicional para ancoragem da plataforma SIG.

A partir desta arquitetura básica podem ser realizadas alterações nas formas de posicionamento e relacionamento entre a base SIG, o DW e a base de endereços de forma a melhorar o desempenho dos processos de cadastro, bem como sua qualidade.

A primeira variação proposta é a criação de uma interligação da base SIG com a base de endereços. Com esta interligação será disponibilizada para os postos de trabalho do ambiente transacional a consulta à base SIG. Com isto espera-se estabelecer uma operação constante de consulta à esta base, que passaria a servir como fonte de certificação para a base de endereços permitindo o controle dos desvios do cadastro e a certificação de endereços.

De certa forma isto não é o ideal por que não acontece sobre a forma de aplicação customizada que

obrigue a consulta à base SIG na operação de cadastro. Desta forma irá contar com a subjetividade do uso por parte dos operadores. Porém permite que sejam implantadas auditorias com o intuito de verificar se as não conformidades poderiam ter sido detectadas e desta forma tratar o problema com aplicação de conscientização, treinamento, ou até penalidades.

A relação com o DW acontecerá também com relação ao cadastro de clientes. Neste caso a plataforma de SIG irá receber o cadastro de clientes realizando sua espacialização, ou seja, a sua distribuição espacial ao longo dos mapas e retornará ao DW uma tabela com todos os códigos de endereçamento da base SIG. Como o DW trabalha sobre o conceito de atualizações após a certificação da relação endereço x cliente somente serão refeitas as espacializações dos novos clientes inseridos na base ou nas mudanças de endereço ocorridas. Na figura 4 está esquematizado o funcionamento da arquitetura funcional descrita.

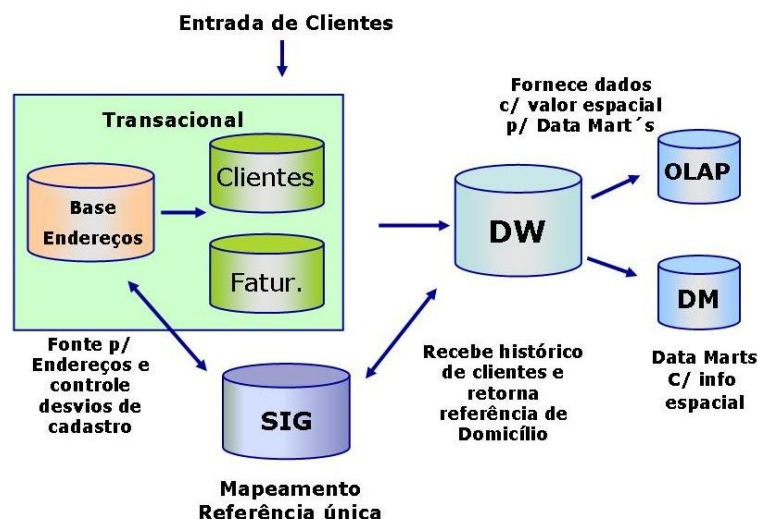


Figura 4 – Plataforma SIG – Interligação simultânea com o DW e base de endereços.

A segunda variação sugerida é implantar o cadastro de endereços dentro da plataforma SIG. Desta forma quando da consulta de um endereço ou cadastramento de um cliente novo os dados que serão retornados estão validados pela plataforma SIG, ou seja, tem correspondência com a realidade. O sistema realiza uma tentativa de espacialização baseada no seu acervo e nos dados fornecidos no cadastramento. Uma vez que esta espacialização foi satisfatória é retornado um valor válido de endereço para o referido domicílio do cliente.

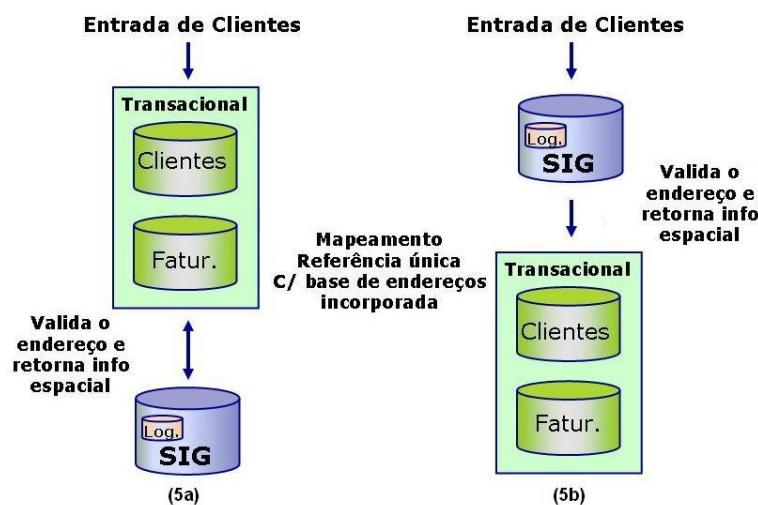


Figura 5 – Plataforma SIG – Base de endereços integrada.

Existem duas formas para isto ocorrer. Na primeira, a entrada de clientes é feita normalmente pelo transacional e quando forem solicitados os atributos para preenchimento do registro no meta-dados de cadastramento de endereços é enviado uma consulta à plataforma SIG que retorna os códigos referentes ao domicílio do cliente. Na segunda forma a entrada de clientes é feita pela plataforma SIG, desta forma não se cadastra um cliente sem que seu endereço exista na base. Estas duas opções são esquematicamente demonstradas na figura 5 – A e B.

A proposta 5b é com certeza a mais radical no que tange a cadastro de clientes, pois uma vez que não exista a correspondência o cliente não é cadastrado. Isto vai gerar um passivo de clientes para serem cadastrados, e compromete o atendimento ao cliente podendo inclusive significar perda de receita. Este passivo pode estar certo, caso o endereço informado no cadastramento esteja incorreto, ou pode ser errado visto que a base SIG pode conter desatualizações em função da dinâmica urbana. Esta estrutura demanda um forte apoio de *back office* para manter a plataforma SIG atualizada com os dados de campo.

Na proposta 5a isto é minimizado, pois permite o cadastramento e gera uma tabela de erros para posterior validação pela equipe de *back office*.

Em ambas as sugestões devem ser implantadas regras de negócio e aplicações que retornem os valores que apresentaram problemas. A partir disto será realizado o controle dos desvios do cadastro.

A terceira variação, apresentada na figura 6, trata da utilização da base SIG como componente do DW, neste caso nenhuma informação será armazenada no DW sem que passe pela plataforma SIG. Nesta arquitetura a referida plataforma serve como uma barreira garantindo que as informações constantes das bases gerenciais estejam certificadas. A espacialização dos endereços dos domicílios do cliente é realizada no DW e somente se permite o armazenamento dos dados para as espacializações realizadas. Uma vez que isto é feito cria-se uma tabela com o código do cliente para que nas atualizações de informações do DW somente os clientes não espacializados anteriormente passem novamente pelo processo de espacialização.

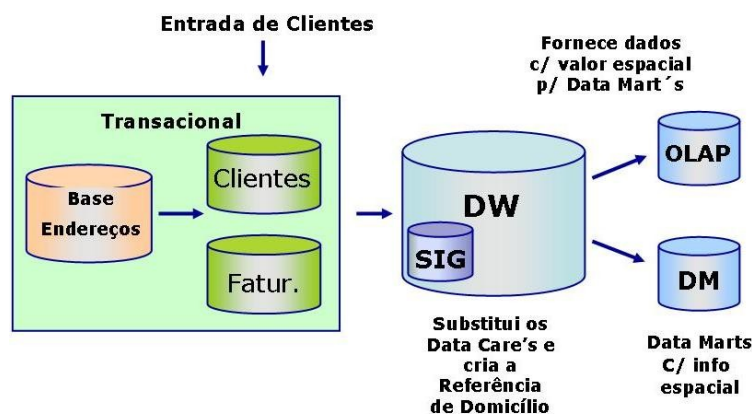


Figura 6 – Plataforma SIG incorporada ao DW.

Tal qual a proposta 5a será gerada uma tabela de erros para posterior validação ou pela equipe de *back office* ou pela equipe de cadastro.

Esta solução protege a análise final com relação à qualidade, pois as tuplas espúrias (*missing or noyse datas*) não serão replicadas para o processo de KDD – *Knowledge Discovery in Database*. Porém não resolve o problema do cadastramento básico, ou seja, as não conformidades podem até ser tratadas, mas não são inibidos novos acontecimentos e os custos com as não conformidades crescem. Como nas sugestões anteriores será necessário implantar uma solução para o controle dos desvios do cadastro.

7 - Conclusão

Os erros de cadastramento e sua propagação não são completamente eximidos com a adoção de uma plataforma de SIG, pois esta também apresenta dificuldade para manutenção e integração entre cadastros alfanuméricos e dados gráficos quando estes são feitos por cargas de dados externas. O divisor de águas deste assunto é o levantamento de campo. Uma vez que este exista se permitirá a certificação das informações e a correção dos valores dos atributos.

Como apresentado ao longo deste artigo o emprego de uma arquitetura funcional que implante na base dos processos de manutenção dos cadastros de clientes e endereços uma plataforma de SIG permitirá que a informação que segue adiante a partir do DW esteja organizada e neste caso não se tornam necessários estágios de limpeza pós DW.

A forma como a plataforma de SIG será implantada deverá estar em consonância com a estratégia de controle de qualidade que se pretende empregar, pois esta ancoragem influenciará no tipo, na forma e na estratégia com que os processos de controle poderão ser implementados.

O nível de qualidade esperado somente será alcançado se a base de dados cadastrais constante no SIG for certificada e se a arquitetura funcional empregada permitir que a estratégia de qualidade pertinente seja empregada.

Esta estratégia de qualidade irá se traduzir em processos e soluções customizadas que atendam a atividade de gestão dos cadastros de clientes e de endereços, bem como a manutenção do MUB que servirá como referência.

8 – Bibliografia

- [1] **Castro, J. H.** *A Utilização das Geotecnologias na Atividade de Planejamento Estratégico de uma Empresa*. XXIII Congresso Brasileiro de Cartografia, RJ, Macaé, 2005.
- [2] **Aguiar, C. D.; Medeiros, C. B.** Construção de um Modelo básico unificado a partir de sistemas Stand-Alone. DCC - UNICAMP - In. Anais Gis Brasil 1996.
- [3] **Cruz, C. F.** *Modelagem de dados para Business Intelligence*. Notas de treinamento, IIR Institute for International Research do Brasil – IIR Training desenvolvimento e Performance Empresarial, SP, São Paulo, 2006.