

Interoperabilidade entre Objetos Geográficos Heterogêneos

Victor H. M. Azevedo¹, Margareth S. P. Meirelles^{1,2}, Rodrigo P. D. Ferraz²,
Antônio Ramalho Filho²

¹Programa de Pós-Graduação em Engenharia de Computação / Geomática –
Universidade do Estado do Rio de Janeiro (UERJ) – Rio de Janeiro, RJ – Brasil

²Empresa Brasileira de Pesquisa Agropecuária - Centro Nacional de Pesquisa de Solos
(Embrapa Solos) – Rio de Janeiro, RJ – Brasil

vhmeirelles@gmail.com, maggie@eng.uerj.br, rodrigo@cnps.embrapa.br,
ramalho@cnps.embrapa.br

Abstract. *The absence of interoperability is a problem found when it is necessary to use the information available in institutions that generate or detain the spatial information. This paper describes a methodology to obtain the interoperability between heterogeneous and distributed geographic objects based on the use of the Open Geospatial Consortium (OGC) Web Feature Service specification (WFS), as a mechanism to make data formats uniform, and the utilization of the domain ontology, through Ontology Web language (OWL), as a syntactic and semantic reference shared schema to integrate these objects. This approach allows for a higher interoperability and is being applied to obtain soil information under the agro-ecological zoning for dendê cultivation in deforested areas at Amazônia, coordinated by Embrapa Solos.*

Resumo. *A falta de interoperabilidade é um problema encontrado quando se faz necessário utilizar informações disponíveis em diversas instituições geradoras ou detentoras de informações espaciais. Este artigo descreve uma metodologia para se obter a interoperabilidade entre objetos geográficos heterogêneos e distribuídos baseada na utilização da especificação do serviço Web Feature Service (WFS) do Open Geospatial Consortium (OGC), como mecanismo para tornar os formatos de dados uniformes, e de ontologias do domínio, através da Ontology Web Language (OWL), como esquema sintático e semântico de referência compartilhado para a integração destes objetos. Esta abordagem permite um alto grau de interoperabilidade e está sendo aplicada na obtenção de informação de solos no projeto de zoneamento agroecológico da cultura do dendê em áreas desmatadas da Amazônia, coordenado pela Embrapa Solos.*

1. Introdução

A integração de objetos geográficos armazenados em fontes de dados distintas e com estrutura sintática e semântica heterogênea, tem sido alvo dos pesquisadores que trabalham com sistemas computacionais em ambiente distribuído de geoprocessamento

nos últimos anos. Este fato ocorre devido a crescente necessidade de troca de informação processadas pelas instituições geradoras de dados geográficos.

Muitas iniciativas têm sido realizadas para alcançar a interoperabilidade entre instituições que desejam trocar informações entre si. Segundo [Fonseca e Egenhofer 1999], as primeiras tentativas de se obter interoperabilidade em Sistemas de Informação Geográficos (SIG) foram através da tradução direta de formatos de dados entre fabricantes de software. Atualmente, profissionais e instituições da área de Geomática formam o Consórcio OpenGeospatial (OGC) [OGC 2006], tendo como objetivo definir um conjunto de especificações padrão para interoperabilidade em SIG. A iniciativa GeoBR [Lima et al 2002] propõe um esquema de dados único, com elementos pré-definidos, que inclui modelo de dados, metadados, projeções, geometrias e atributos que possa ser acessado por uma única interface de programação.

As cooperativas de dados geográficos, como descrito em [Câmara 1996], são uma solução emergente para tratar do gerenciamento de informações geográficas em grande volume, de forma a permitir o seu uso cooperativo por agências governamentais ou privadas.

Segundo [Hartman 1998], em projetos envolvendo um ambiente heterogêneo, a aquisição dos dados representa o maior custo, representando de 60% a 80% do valor total de implementação dos SIGs. Sendo assim, o alto custo da coleta e produção de dados geográficos é um fator de incentivo à interoperabilidade das informações espaciais já produzidas por diversas instituições. Neste cenário, a automatização desta interoperabilidade de objetos geográficos em um ambiente distribuído torna-se uma poderosa ferramenta cujo principal objetivo é viabilizar a cooperação entre as instituições produtoras de informações espaciais.

Alcançar a completa interoperabilidade de objetos geográficos de tal maneira que os mesmos sejam interpretados segundo uma mesma ótica não é tarefa simples [Lima et al 2002]. Segundo Casanova 2005, a integração completatanto de estrutura e forma dos dados, quanto de significado e interpretação deve resolver a incompatibilidade em três níveis: formato e estrutura; sintático; e semântico.

Muitos são os esforços no sentido de fornecer mecanismos computacionais de padronização de formatos e uniformização sintática e semântica de objetos geográficos. Segundo [Lima et al 2002], atualmente é inquestionável o uso da *Extensible Markup Language (XML)* (www.w3.org/xml) como padrão de troca de dados. O OGC fornece um conjunto de especificações para padronizar o processo de interoperabilidade entre diferentes formatos de dados baseado na tecnologia *XML*. A *Geographic Markup Language (GML)* (www.opengeospatial.org/standard/gml) pode ser considerada a principal delas, por ser utilizada em muitas outras especificações. Ela foi concebida com o objetivo de representar as informações geográficas, incluindo tanto as informações espaciais quanto as não espaciais. Segundo [Davis Jr. 2005], o objetivo da GML é oferecer um conjunto de regras com as quais o usuário pode definir sua própria linguagem para descrever os objetos geográficos que pretende manipular.

O OGC fornece um *framework* arquitetural que define por meio de especificações formais o escopo, objetivos e comportamento de uma série de serviços *web*, denominado *OpenGIS Services Framework*. O objetivo deste *framework*, é fornecer um

mecanismo capaz de garantir a interoperabilidade entre as instituições, utilizando *internet*.

As especificações do OGC não incluem a preocupação com os aspectos semânticos da interoperabilidade na sua abordagem, nem tão pouco seguem as recomendações do consórcio *W3C* para a utilização de *Web* semântica, como afirma [Davis Jr. 2005].

O termo “ontologia” vem sendo utilizado nas ciências da informação para representar uma “especificação explícita formal de uma conceituação compartilhada” [Klien et al 2004]. A utilização de ontologias como estratégia de representação do conhecimento sobre um determinado domínio de interesse, fornece um esquema semântico que vem se mostrando eficaz, já que se pode especificar de maneira explícita e formal os termos do domínio bem como o relacionamento entre eles. A *OWL* (www.w3.org/TR/owl-ref) é uma linguagem que tem o objetivo de definir, publicar e compartilhar ontologias na *web*, podendo viabilizar a interoperabilidade semântica.

Considerando-se o cenário exposto, este artigo propõe uma abordagem metodológica capaz de automatizar a interoperabilidade entre objetos geográficos disponíveis em diversas instituições, visando reduzir os custos desta integração e tornar mais ágil o processo de tomada de decisão.

Esta metodologia foi aplicada no projeto de zoneamento agro-ecológico para a cultura do dendê visando a produção de biocombustível nas áreas desmatadas da Amazônia, atendendo a demanda do governo por aumento da produção de biocombustível a partir de culturas oleaginosas e palmáceas. Para a execução deste zoneamento, as informações de solos da Amazônia Legal disponíveis na Embrapa Solos, no SIPAM e no IBGE tiveram que ser integradas e a interoperabilidade dos bancos de dados destas instituições passou a ser uma demanda.

2. Material e Métodos

As heterogeneidades sintática, semântica e de formato/estrutura dos dados espaciais devem ser consideradas para a plena realização da integração. Para solucionar este problema, [Klien et al 2004] propõem uma arquitetura baseada em serviços *web* geográficos e utilizam o sistema BUSTER (<http://www.informatik.uni-bremen.de/agki/www/buster/new>) para integração das informações sobre tempestades em floresta. O sistema BUSTER é um sistema baseado em ontologias utilizado para busca e integração de informações em ambiente distribuído heterogêneo [Klien et al 2004]. Alternativamente, a metodologia proposta neste trabalho sugere a utilização das especificações da GML e dos serviços *web* do OGC como mecanismo tecnológico para a integração sintática e de estrutura entre objetos geográficos heterogêneos. Além disto, as ontologias e a descrição do domínio na forma de linguagem *OWL*, podem definir os aspectos semânticos e sintáticos da integração fornecendo o braço tecnológico da representação do conhecimento a cerca do domínio.

A metodologia possui seis etapas (Figura 1):

- Definição da arquitetura orientada aos serviços do OGC, com a definição dos Servidores de Objetos Geográficos (SOG) e do(s) Servidor(es) de Integração (SI);

- Utilização do serviço de WFS para prover os objetos geográficos de cada fonte de dados através dos SOG;
- Elaboração da base de conhecimento com a criação, em linguagem OWL, da(s) ontologia(s) do(s) domínio(s) que se deseja trocar informações, através de um processo de engenharia do conhecimento;
- Publicação dos serviços de cada SOG no SI;
- Disponibilização das informações integradas em formato de serviços WFS e WMS;

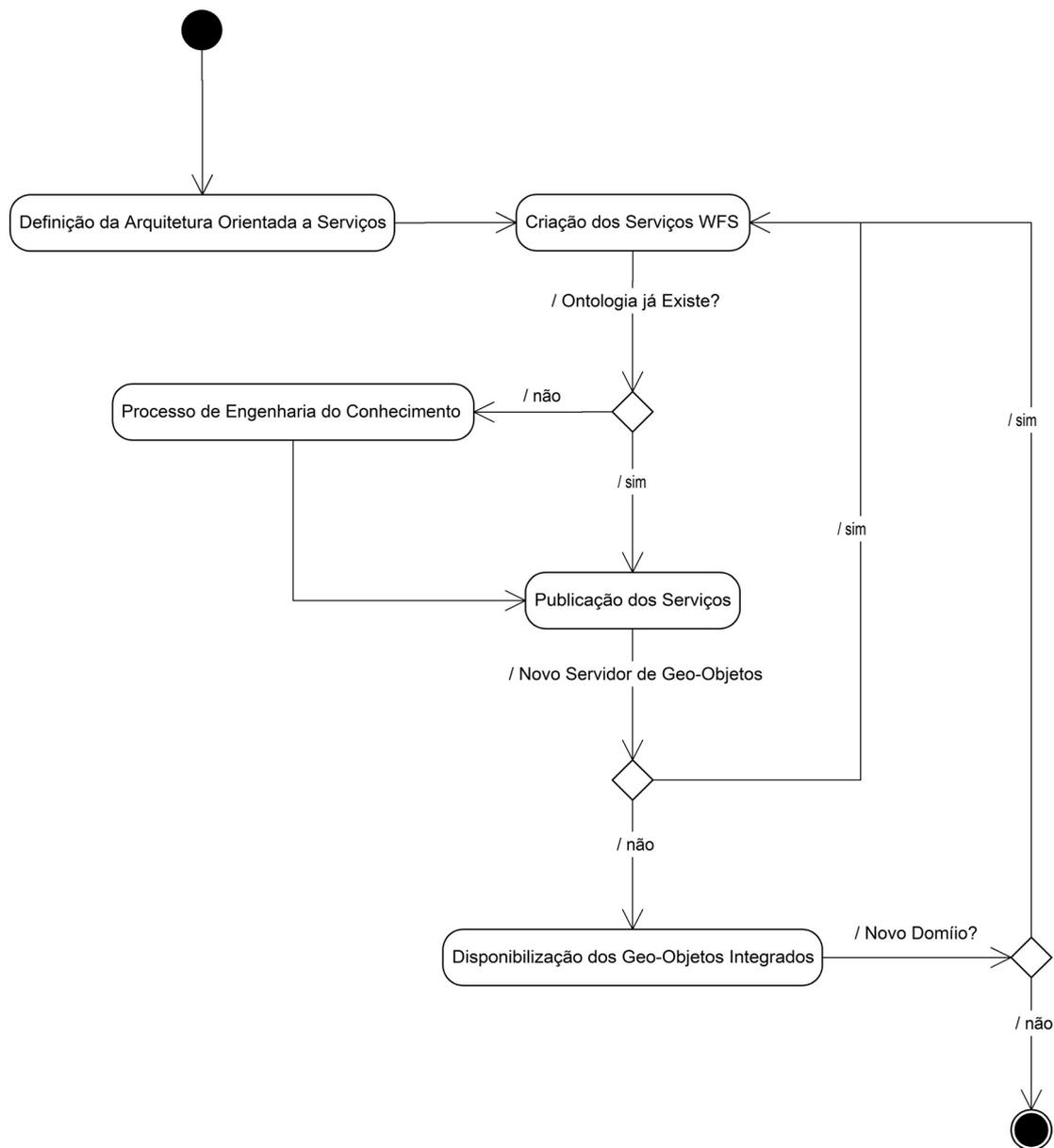


Figura 1. Diagrama de Atividades UML da Metodologia de Integração de Objetos Geográficos

2.1 Arquitetura Orientada a Serviços do OGC

Dois tipos diferentes de servidores devem ser implementados na arquitetura que será definida: Servidores de Objetos Geográficos (SOG) e Servidores de Integração (SI).

Os SOGs são responsáveis por fornecer as informações geográficas de cada instituição envolvida na integração, transformando os dados para o formato padrão GML. Estes dados em GML devem ser fornecidos através de serviços WFS.

Os SIs, por sua vez, são responsáveis por registrar os serviços WFS disponíveis nos SOGs e unificar a sintaxe e semântica de cada elemento destes serviços para que todos sejam interpretados da mesma maneira. Para isto, os SIs devem permitir o cadastramento das ontologias do domínio e fornecer um mecanismo de correlação entre os elementos disponíveis nos SOGs e a ontologia.

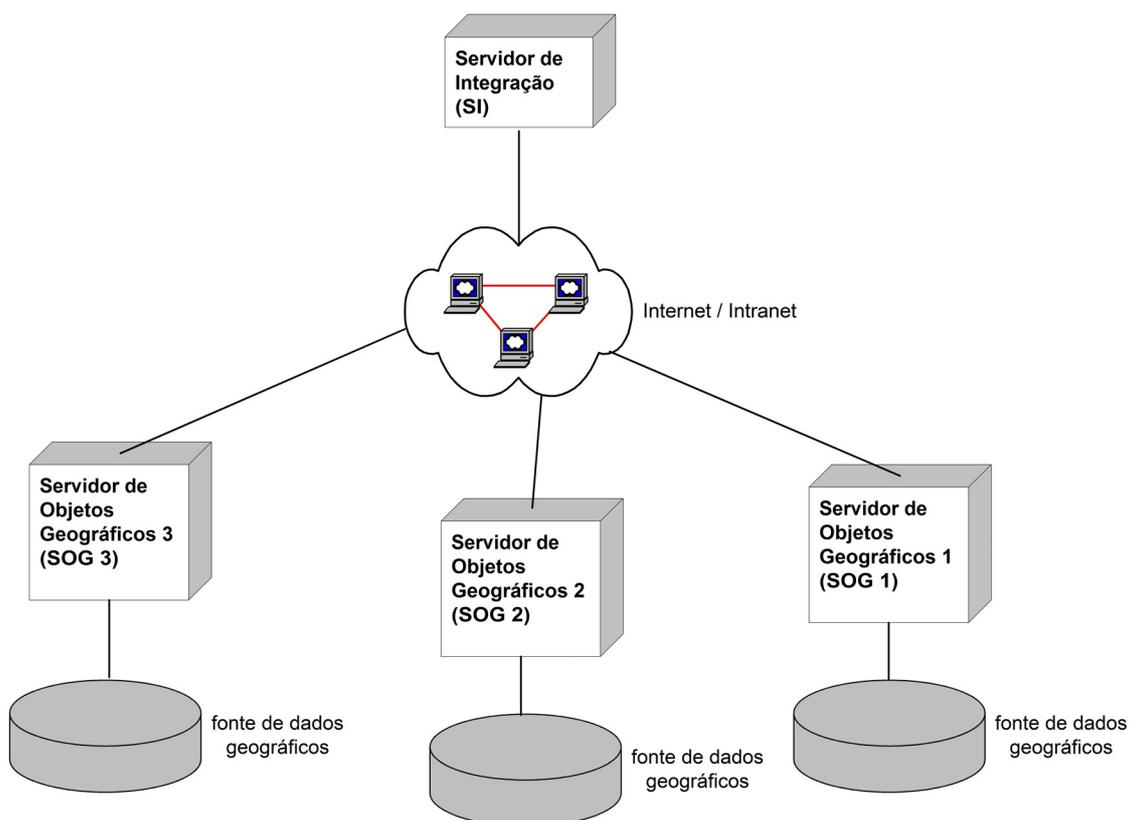


Figura 2. Exemplo de Arquitetura Orientada a Serviços OGC

2.2 Criação dos Serviços WFS

Depois de definida a arquitetura e conhecidos os servidores de objetos geográficos, é preciso automatizar o processo de busca e publicação destas informações armazenadas

em cada instituição. Para isto, são definidos em cada servidor de objetos geográficos os serviços de WFS necessários para tornar os objetos geográficos disponíveis no formato GML.

As diferentes formas de representação da informação espacial utilizada pelas instituições é inicialmente convertida para GML e depois disponibilizada ao requisitante do serviço neste formato. Alguns SIGs e servidores de mapas já possuem ferramentas que implementam automaticamente a especificação WFS do consórcio OGC e convertem os objetos geográficos para a sua respectiva representação em GML.

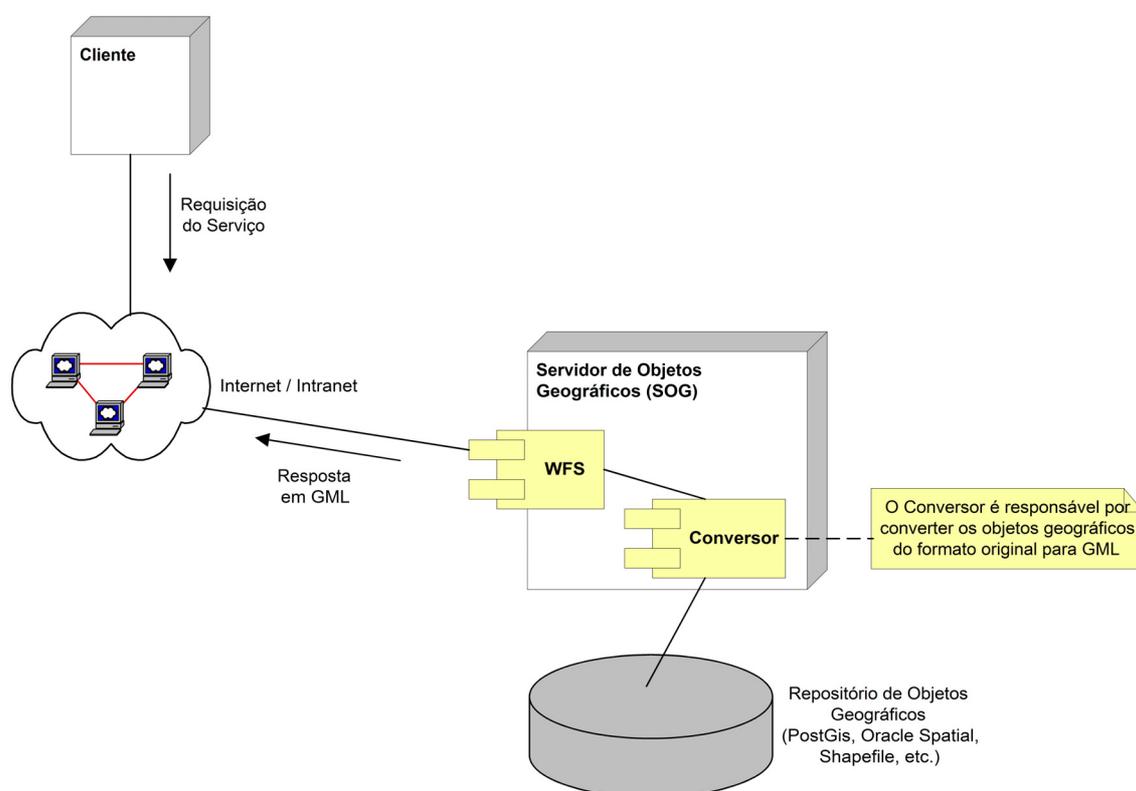


Figura 3. Funcionamento do Serviço WFS

2.3 Processo de Engenharia do Conhecimento

O objetivo desta etapa é gerar as ontologias dos domínios de integração com os seus termos e relações de maneira formal. É fornecido, desta maneira, um esquema conceitual de referência que possibilitará a realização da uniformização da sintaxe e da semântica das informações geográficas disponíveis em GML.

Os Engenheiros de Conhecimento e os Especialistas do Domínio devem interagir com o objetivo de formalizar uma linguagem padrão de representação do conhecimento dos termos do domínio (ontologia).

Ao gerar e compartilhar a ontologia em uma linguagem formal padrão, os usuários do domínio passam a possuir um referencial sintático-semântico sobre a área de conhecimento. A ontologia passa a ser um modelo de referência semântica para o processo de integração.

2.4 Publicação dos Serviços no Servidor de Integração

O objetivo desta etapa é criar um mecanismo automatizado de unificação dos dados geográficos disponíveis nas várias fontes de dados através do servidor de integração. A partir das informações cadastradas, o SI deverá fornecer o resultado da integração em formato GML seguindo a ontologia, ou seja, o modelo de referência definido pela base de conhecimento.

O servidor de integração deve fornecer as seguintes interfaces:

- Cadastramento dos domínios que formam a base de conhecimento com as suas respectivas ontologias em OWL;
- Publicação dos Serviços WFS com os objetos geográficos que deverão ser integrados, utilizando uma estrutura de metadados pré-determinada;
- Correlação sintática e semântica entre os esquemas conceituais utilizados nos serviços WFS e o esquema conceitual de referência definido pela ontologia em OWL.

Desta maneira, a configuração do servidor de integração é realizada através do cadastramento da localização dos serviços WFS disponíveis em cada fonte de dados na rede juntamente com as informações sobre os serviços em forma de metadados e a correlação sintático-semântica entre os esquemas conceituais. Com esta configuração realizada, o servidor de integração poderá de fato unificar os objetos geográficos disponíveis em cada uma das fontes de dados nele cadastradas.

2.5 Disponibilizar os objetos geográficos integrados

Os objetos geográficos unificados devem ser disponibilizados em formato GML por meio de serviços WFS. Além da publicação em formato GML, uma representação visual dos dados geográficos unificados deve ser fornecida utilizando o serviço WMS. O serviço WFS permitirá que os objetos geográficos unificados funcionem como uma nova fonte de dados, enquanto o serviço WMS permitirá a análise e manipulação do mapa gerado a partir dos objetos geográficos unificados.

3. Estudo de Caso

Com a finalidade de comprovar a eficiência da metodologia proposta, vem sendo elaborado um estudo de caso nos domínios de solos e aptidão agrícola das terras, no âmbito do projeto de zoneamento agro-ecológico para culturas oleaginosas com ênfase no cultivo do dendê (Zondendê), coordenado pela Embrapa Solos na região da Amazônia Legal. O projeto Zondendê utiliza como base, as informações obtidas pela Embrapa, projeto RadamBrasil e outras fontes de dados públicas e privadas. A base de dados do projeto conta ainda com a base de dados do Sistema de Proteção da Amazônia (SIPAM), além da criação de novos mapas de solos gerados pela Embrapa a partir dos mapas já existentes em escalas mais detalhadas e em áreas de maior interesse.

A avaliação de aptidão dos solos para uma determinada cultura é feita com base na interpretação dos dados das propriedades dos solos e das exigências da cultura avaliada. Sendo assim, existe a necessidade real de integração das diversas fontes de dados de solos disponíveis no projeto para se alcançar o objetivo esperado.

Neste estudo de caso, duas fontes de dados foram escolhidas para realizar a integração sintática, semântica e estrutural das informações geográficas de solos da região amazônica: base de recursos naturais do SIPAM e as informações georeferenciadas de recursos naturais do IBGE. A base de recursos naturais do SIPAM foi fornecida para o projeto em uma arquitetura *dual* com as informações geográficas em formato *shapefile* (www.esri.com) e as informações complementares em um banco de dados relacional *Oracle* (www.oracle.com). O banco de dados georeferenciado de recursos naturais do IBGE é um esforço realizado pela instituição, de sistematizar as informações absorvidas a partir do projeto Radambrasil. Estas informações são públicas e se encontram disponíveis na internet (www.ibge.gov.br).

Primeiramente, preparou-se um ambiente de simulação com a presença das duas fontes de dados mencionadas e posteriormente, foi aplicada a metodologia proposta em uma tentativa de verificar os resultados obtidos. A fonte de dados obtida junto ao SIPAM, foi convertida da arquitetura *dual* para um banco de dados geográfico implementado em Postgresql com extensão espacial Postgis.

No caso dos dados de solos do IBGE, como as informações geográficas e alfanuméricas estavam armazenadas, respectivamente, em arquivos em formato *shapefile* (separados por cartas ao milionésimo do IBGE) e em arquivos *dbf*, foi gerado um arquivo *shapefile* único contendo as cartas unificadas e os campos alfanuméricos relevantes.

As duas fontes de dados de solos utilizadas (SIPAM e IBGE) apresentam heterogeneidades em seus esquemas conceituais, tanto em termos de formato de dados (*postgresql/postgis* e *shapefile*) quanto a nível sintático e semântico. A Tabela 1 exemplifica estas diferenças:

Tabela 1 – Exemplo das heterogeneidades encontradas nas bases de dados de solos

Heterogeneidade	SIPAM	IBGE
Formato	Banco de Dados Espacial Postgis	Arquivo <i>ESRI Shapefile</i>
Sintática	Geo	SHAPE
Semântica	Polígono de unidade de mapeamento de solo na tabela espacial.	Polígono de unidade de mapeamento de solo no <i>shapefile</i> .
Sintática	cd_letra_simb	SIMB_UNID
Semântica	Símbolo da Unidade de Mapeamento.	Símbolo da Unidade de Mapeamento.
Sintática	cd_textura_1h1	GRUP_TEXT
Semântica	Textura do Horizonte Superficial	Grupamento de Textura
Sintática	cd_relevo_1	FS_RELEVO
Semântica	Relevo Predominante	Fase de Relevo
Sintática	cd_pre_casc_1_t1h1	FS_PEDREG
Semântica	Presença de cascalho no horizonte Superficial	Fase de Pedregosidade

Após a preparação do ambiente, a metodologia foi aplicada até a etapa de publicação dos serviços. Primeiramente, uma arquitetura simples foi definida, com dois servidores de objetos geográficos, o SOG-SIPAM e o SOG-IBGE, e um servidor de integração, o SI-SOLOS. Posteriormente, foram criados, para cada SOG, os serviços WFS utilizando a ferramenta GEOSERVER (<http://docs.codehaus.org/display/GEOS/Home>). A figura 4 exemplifica a fase de configuração do serviço para publicação de um mapa de solos do IBGE e a Figura 5 apresenta os objetos geográficos armazenados na base do IBGE convertidos para o formato GML.

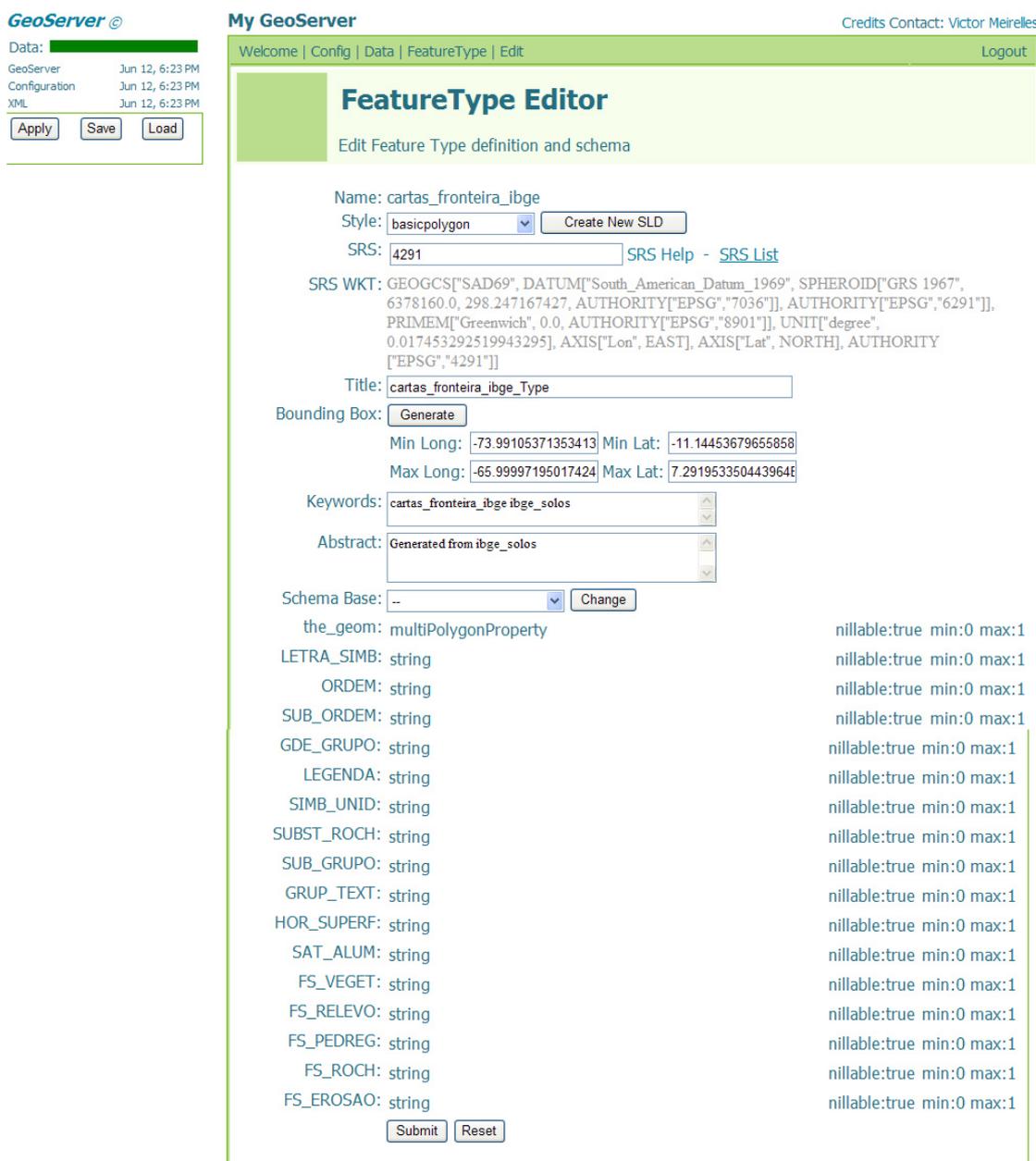


Figura 6. Publicação do Serviço no Geoserver

```

<?xml version="1.0" encoding="ISO-8859-1" ?>
- <wfs:FeatureCollection xmlns:wfs="http://www.opengis.net/wfs" xmlns:SOG_IBGE="http://geoserver/SOG-IBGE"
  xmlns:gml="http://www.opengis.net/gml" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://geoserver/SOG-IBGE http://geoserver:80/geoserver/wfs/DescribeFeatureType?
  typeName=SOG_IBGE:cartas_fronreira_ibge http://www.opengis.net/wfs
  http://geoserver:80/geoserver/schemas/wfs/1.0.0/WFS-basic.xsd">
- <gml:boundedBy>
- <gml:Box srsName="http://www.opengis.net/gml/srs/epsg.xml#4291">
  <gml:coordinates xmlns:gml="http://www.opengis.net/gml" decimal="," cs="," ts=">73.99105371,-11.14453953 -
  65.99997195,0.00000073</gml:coordinates>
  </gml:Box>
</gml:boundedBy>
- <gml:featureMember>
- <SOG_IBGE:cartas_fronreira_ibge fid="cartas_fronreira_ibge.1">
  - <SOG_IBGE:the_geom>
  - <gml:MultiPolygon srsName="http://www.opengis.net/gml/srs/epsg.xml#4291">
    - <gml:polygonMember>
      + <gml:Polygon>
    </gml:polygonMember>
  </gml:MultiPolygon>
  </SOG_IBGE:the_geom>
  <SOG_IBGE:LETRA_SIMB>LAd8</SOG_IBGE:LETRA_SIMB>
  <SOG_IBGE:ORDEM>LATOSSOLO</SOG_IBGE:ORDEM>
  <SOG_IBGE:SUB_ORDEM>AMARELO</SOG_IBGE:SUB_ORDEM>
  <SOG_IBGE:GDE_GRUPO>Distrófico</SOG_IBGE:GDE_GRUPO>
  <SOG_IBGE:LEGENDA>LATOSSOLO AMARELO</SOG_IBGE:LEGENDA>
  <SOG_IBGE:SIMB_UNID>SA19LAd8</SOG_IBGE:SIMB_UNID>
  <SOG_IBGE:SUBST_ROCH />
  <SOG_IBGE:SUB_GRUPO>típico</SOG_IBGE:SUB_GRUPO>
  <SOG_IBGE:GRUP_TEXT>argilosa sem cascalho/argilosa sem cascalho e média sem cascalho/média sem
  cascalho</SOG_IBGE:GRUP_TEXT>
  <SOG_IBGE:HOR_SUPERF>A moderado</SOG_IBGE:HOR_SUPERF>
  <SOG_IBGE:SAT_ALUM>állico</SOG_IBGE:SAT_ALUM>
  <SOG_IBGE:FS_VEGET>Floresta Ombrófila</SOG_IBGE:FS_VEGET>
  <SOG_IBGE:FS_RELEVO>plano</SOG_IBGE:FS_RELEVO>
  <SOG_IBGE:FS_PEDREG>não pedregosa</SOG_IBGE:FS_PEDREG>
  <SOG_IBGE:FS_ROCH>não rochosa</SOG_IBGE:FS_ROCH>
  <SOG_IBGE:FS_EROSAO>não erodida</SOG_IBGE:FS_EROSAO>
  </SOG_IBGE:cartas_fronreira_ibge>
</gml:featureMember>

```

Figura 7. Retorno em GML do Serviço WFS

4. Conclusões

Até o momento, os resultados obtidos na pesquisa se mostraram satisfatórios. Com um ambiente de simulação realmente heterogêneo e a utilização de informações e esquemas conceituais reais, foi alcançado o objetivo da unificação de formatos com a utilização dos serviços WFS. O formato GML, por ser baseado em XML, se mostrou adequado para troca de informações geográficas heterogêneas. A utilização de uma ferramenta de publicação de serviços WFS, com a conversão automática dos formatos originais das fontes de dados para GML, torna a tarefa de conversão dos objetos geográficos para um formato padrão mais simples.

Neste momento está sendo verificado uso da base de conhecimento e das ontologias geradas em OWL como esquema de referência para a integração sintática e semântica entre os objetos. Esta base de conhecimento além de servir de subsídio para outras aplicações do setor agrícola servirá também para validar a metodologia proposta.

5. Referências

- Fonseca, F. T. e Engenhofer, M. J. (1999) “Sistemas de Informação Geográficos Baseados em Ontologias”, National Center for Geographic Information and Analysis, Department of Spatial Information Science and Engineering, University of Maine, Orondo, ME 04469-5711, USA.
- Lima, P. ; Câmara, G. e Queiroz, G. (2002) “GeoBR: Intercâmbio Sintático e Semântico de Dados Espaciais”, INPE, São José dos Campos, SP, Brasil.
- Câmara, G.; Casanova, M. A.; Hemerly, A. S.; Magalhães, G. C. e Medeiros, C. M. B. (1996) “Centros de Dados Geográficos”, Anatomia de Sistemas de Informações Geográficas. Rio de Janeiro., p. 167- 173.
- Hartman, R. (1998) “GIS Data Conversion – Strategies, Techniques and Management”, New York: Onword Press.
- Casanova, M. A.et. al. (2005) “Integração e interoperabilidade entre fontes de dados geográficos”, Bancos de Dados Geográficos. Rio de Janeiro, p. 305- 340, Editora MundoGeo.
- Davis Jr., C. A. et al. (2005) “O Open Geospatial Consortium”, Bancos de Dados Geográficos, CASANOVA, M. et al., Rio de Janeiro, p. 367- 383, Editora MundoGeo.
- Klien, E., Lutz, M , Kuhn, W. (2004) “Ontology-Based Discovery of Geographic Information Services – An Application in Disaster Management”, 7th Conference on Geographic Information Science (AGILE 2004).