

## PAPER REVIEW: SMART WEB SERVICES FOR BIG SPATIO-TEMPORAL DATA IN GEOGRAPHICAL INFORMATION SYSTEMS

JOURNAL	AUTORES	LINK - SCOPUS	COMPILADO POR
Fusion Engineering and Design	Matthias Frank, Stefan Zander	<a href="https://goo.gl/s8j3BN">https://goo.gl/s8j3BN</a>	Luiz Gustavo Dias - UFF

### RESUMO

A informação gerada em sistemas geográficos depende de precisão consistência e integridade dos dados coletados. Com isso é difícil manter uma base e manter requisitos como proveniência, veracidade, velocidade, e heterogeneidade estrutural e semântica. É apresentado no trabalho uma abordagem de integração de informações bem definida e colaborativa, por meio da web semântica, ontologias e APIs que enfatizam especificamente uma relação espaço-temporal e permitem uma nova geração de sistemas de informação geográfica. Foram empregados conceitos de serviços da web inteligente para trabalhos compostos dinamicamente para lidar com características dos fluxos de valores em big data e gerar dados mais elaborados.

### VISÃO GERAL

Sistemas de informação geográfica (SIGs) são ferramentas importantes para apoio a decisão com base em dados espaço-temporais, e são utilizadas em vários campos, como planejamento civil, gerenciamento de emergência, agricultura, proteção da natureza e etc. É apresentada no artigo uma nova abordagem de como SIGs podem explorar e integrar o Big Data baseado em tecnologia web semântica, ontologias e interfaces de programação de aplicativos (APIs). Além dos dados gerados por esses sistemas serem de tamanho grande, a confiabilidade desses dados pode ser incerta, além de dados espaço-temporais de diferentes fontes utilizarem diferentes esquemas para descrever locais, como endereços, relações espaciais relativas ou diferentes sistemas de coordenadas. Sendo assim, GIS estão enfrentando desafios em todas as quatro dimensões do big data:

- Volume: A prevalência e onipresença da tecnologia de sensores e onipresença das fontes de dados expõem desafios relativos aos volumes de dados a serem integrados;
- Variedade: dados não estruturados são novos tipos de dados para o GIS que requerem métodos inovadores de interpretação de dados para analisar, interpolar, prever e visualizar;
- Velocidade: A fim de integrar permanentemente os dados dos sensores, o processamento em lote desses sistemas tem que ser tecnicamente reorganizados conceitualmente, para permitir a análise em tempo real;
- Veracidade: A integração de informação geográfica voluntária (VGI) e outros conteúdos criados pelo usuário, bem como integração de sensoriamento remoto, podem estar incompletos, impedindo a suposição dos dados coletados, deixando em dúvida sua completude e correteude.

Na abordagem proposta pelos autores, foi utilizada a web semântica para:

1. Descrever as fontes de dados e serviços de transformação de dados para GIS
2. Desta permitiu que sistemas inteligentes: Compusessem um workflow que gera um conjunto de dados mais elaborados quando relacionados a precisão, consistência e integridade

Desta forma a proveniência foi alcançada. O esquema deste conjunto de resultados tem que ser flexível e corresponder aos requisitos de diferentes casos de uso. Isso exige uma estrutura para metadados que represente as relações dos dados que devem ser integrados. Ao descrever fontes de dados heterogêneos para GIS, bem como seus parâmetros de entrada e saída dos serviços de transformação de dados, o processo para processar esses dados pode ser composto dinamicamente para cumprir todos os requisitos específicos do domínio. Desta forma a abordagem apresentada na pesquisa, permite que especialistas de domínio selecionem uma combinação de fontes de dados e

definam estruturas de dados necessária para um caso de uso específico em relação a quantidades, unidades, granularidades, precisão das medidas, período e área de investigação.

Relacionado a proveniência e fontes de dados heterogêneas, foram levantadas as seguintes questões de pesquisa:

1. Que tipo de informação de proveniência é relevante para tornar o processamento de valores de fontes de resultados heterogêneos comparáveis dentro de um conjunto de resultados?
2. Como é uma abordagem colaborativa de descrição de fontes e serviços em escala de dados espaço-temporal para o trabalho de processamento de big data em GIS?

## ABORDAGEM

Foram abordados os diferentes requisitos em dados para análise, descrevendo fontes de dados espaço-temporais e APIs utilizando ontologias e web semântica. Para contribuir para a flexibilidade e usabilidade do GIS, foram focados os seguintes requisitos:

1. Colaboração: os usuários devem poder adicionar fontes de dados espaço-temporais e APIs relevantes ao GIS;
2. Semântica: todas as fontes de dados espaço-temporais e APIs relevantes precisam ser descritas de uma maneira interpretável pela máquina;
3. Eficiência: Fluxos de valores de observações espaço-temporais tem que ser transmitidos e processados com o menor número possível de despesas gerais a fim de lidar com dados de alto volume.

O primeiro passo constou em construir um sistema colaborativo baseado em SMW (semantic media wiki) para gerenciar metadados. Este sistema importa e reutiliza o vocabulário comumente usados no domínio de GIS como SSN, QUDT, GeoVocab que também cobre o Basic Geo Vocabulary para descrever semanticamente as fontes de dados e Hydra Vocabulary para APIs e serviços de transformação. Para um processamento eficiente dos dados espaço-temporais apenas os metadados desses fluxos são modelados com a flexibilidade de formatos de dados da web semântica como RDF, enquanto os valores observados são transmitidos e transformados com a menor quantidade possível de sobrecarga.

Os usuários criam páginas de descrição de fontes de dados ou APIs disponíveis no SMW, e o sistema fornece uma representação dessas informações no RDF usando a ontologia específica anotada com vocabulários comuns. Os autores forneceram uma API restful que permite que especialistas em domínio consultem as quantidades, unidades, granularidades, precisão de medições, período, e área sob investigação para um caso de uso específico.

## CASO DE USO

Foram analisadas as características das ilhas de calor suburbanas (SUHIs), ou seja, ilhas de calor que ocorrem dentro de cidades em dias quentes, desta forma foram necessários conjuntos de dados relacionados a temperatura de fontes heterogêneas. Assim, foram coletados dados de autoridades regionais do meio ambiente, do serviço de clima alemão, medições de um trem urbano, e dados de sensoriamento remoto de satélites.

Com os dados reunidos como o conjunto de dados de treinamento, foram planejadas previsões de SUHIs dentro de uma cidade da Alemanha, e foram avaliadas com o conjunto de dados de teste. Variando as fontes usadas como entrada para as previsões, foram avaliados os impactos dessas fontes no suporte de decisões.

Utilizando a abordagem colaborativa de integração de informações, possibilitou incluir fontes heterogêneas de temperatura utilizados no caso de uso.

## CONCLUSÃO

A exploração e integração de dados espaço-temporais em SIG dependem fortemente de um entendimento comum de conceitos entre dados heterogêneos. Os autores mostram na pesquisa que combinando a dinâmica de uma abordagem colaborativa com o poder expressivo de ontologias, esse requisito pode ser preenchido. Para a avaliação da abordagem proposta por eles, foram observados resultados experimentais no caso de uso relacionado a SUHI que mostram como descrições colaborativas através de descrições de fontes de dados podem ser utilizados para gerar uma ontologia interpretável por computadores e empregar o conceito de serviços da web inteligentes para compor dinamicamente o workflow. Acredita-se que os resultados preliminares indicarão que a abordagem permite que usuários mesmo sem histórico de engenharia web podem adicionar fontes e serviços para um GIS existente. Para uma avaliação significativa, é necessário criar mais exemplos de trabalho composto

dinamicamente a partir de um grande volume de dados espaço-temporais. Os valores transformados de grandes fontes podem ser unidos e investigados.

## ANÁLISE SEGUNDO O LEITOR

1. Qual tipo de proveniência abordada no trabalho?  
**A pesquisa não foca em um tipo de proveniência específica de forma explícita, entretanto, tratando-se de histórico de dados, assume-se que a proveniência abordada é do tipo retrospectiva.**
2. Qual tipo de ontologia utilizada no trabalho?  
**É percebido que a ontologia criada e as demais utilizadas no trabalho como vocabulário são do tipo de domínio, visto que são voltadas para dados espaço-temporais.**
3. Qual a principal vantagem em se utilizar ontologia no contexto da pesquisa?  
**Colaboração e compartilhamento de informações.**
4. Questões de granularidade são abordadas no artigo?  
**A granularidade não é abordada de maneira direta, entretanto percebe-se que dados relacionados ao domínio espaço-temporal podem necessitar de diferentes níveis de granularidade, uma vez que são utilizados nestes tipos de sistemas dados advindos de fontes heterogêneas, desta forma, utilizar ontologias pode proporcionar diferentes níveis de granularidades, uma vez que os componentes do domínio são formalizados e relacionados.**