

**Integração de Dados de Geospatial Crowdsourcing em IDE's
Utilizando o INSPIRE**

António José Silva

d2011090@isegi.unl.pt

JIIDE
2014

Sumário

- Motivação e Objectivos
- Utilizadores e “Produtilizadores”
- Desafios
- Arquitectura
- Caso de Estudo
- Conclusões

Motivação e Objectivos

- Motivação
 - Os utilizadores tornaram-se produtores de informação o que alterou completamente a forma como é criada Informação Geográfica;
 - Deixou de ser um processo conduzido pelas autoridades competentes envolvendo especialistas, para um processo informal que envolve um grande número de amadores, com interesses vários e habilidades;
 - Do lado da administração, os *workflows* actuais não consideram os dados obtidos pelos utilizadores;
- Objectivos
 - Com este trabalho pretende-se actualizar o conceito de IDE do INSPIRE, com o objectivo de incluir um novo workflow, capaz de tomar todas as vantagens dos dados recolhidos por Geospatial Crowdsourcing;
 - Pretende-se desenvolver um serviço aberto de endereços postais baseado na directiva INSPIRE, integrando dados abertos e gratuitos disponibilizados pela OpenStreetMap e por fontes oficiais;

Utilizadores e “Produtilizadores”

- Um endereço é uma colecção de informação, apresentada geralmente sob um formato fixo, utilizado para descrever uma localização, utilizando como referência limites administrativos e nomes de ruas;
- Todas as autoridades governamentais locais ou nacionais, devem possuir informação sobre endereços e simultaneamente ser capazes de explorar o seu potencial;
- Avanços tecnológicos recentes, aliados ao decréscimo no seu custo das unidades de GPS, smart phones redes de sensores, leitores RFID , ligações de alta velocidade de Internet, e nos computadores, ie, especialmente todas as tecnologias livremente denominadas de Web 2.0, fizeram com que um novo paradigma de comunicação fosse introduzido neste contexto, onde os utilizadores se tornam simultaneamente produtores de informação;
- Estes acontecimentos tornaram as tecnologias geoespaciais moveis e ubíquas e transformaram completamente a forma como os dados espaciais são obtidos, armazenados, analisados, disseminados, visualizados e utilizados, especialmente quando se assiste ao declínio da produção e actualização de IG pelas autoridades.

Utilizadores e “Produtilizadores” (2)

- Os dados recolhidos pelos utilizadores, podem auxiliar na correcção de erros, no preenchimento de falhas ou mesmo na predição das actividades das pessoas;
- Os dados gerados pelo geospatial crowdsourcing, estão a crescer a uma velocidade incrível, muito mais rápido do que o desenvolvimento de workflows capazes de tirar total partido;
- Tem sido investido muito esforço em IDE's na última década, um pouco por todo o mundo, no que diz respeito à interoperabilidade de serviços e criação de standards;
- Para integrar as IDE's internacionalmente, na Europa o maior acontecimento deu-se com a entrada da Directiva INSPIRE.

Desafios

- É de extrema importância ter uma percepção clara dos desafios a ultrapassar para atingir os objectivos propostos.
- Estes desafios podem ser de origens distintas:
 - Modelos de dados;
 - Inconsistências nos dados;
 - Semânticos;
 - Comparação de Dados;
 - Bidireccionalidade.

Desafios – Modelos de Dados

- Modelos de dados OSM
 - Modelo de dados bastante simples;
 - Devido à sua natureza é bastante dinâmico, interactivo, eficiente e flexível;
 - As Geometrias são representados por nós, que descrevem pontos georreferenciados;
 - Um conjunto de nós podem ser ainda combinados para criar “ways”, que descrevem linhas ou polígonos;
 - Para a representação de atributos de uma entidade OSM, é possível de atribuir um “Tag” à geometria com pares de valores-chave “key-value pairs”;
 - Embora, o modelo de dados do OSM estar melhor posicionado para a visualização e edição de dados, acaba por ser algo problemático a utilização para análise de dados devido a ser algo redundante;
- Modelo de dados INSPIRE
 - Abordagem top-down orientada politicamente com regras bem definidas a suportá-la;
 - Adaptado para representar um modelo hierárquico, muito bem definido, especificado e documentado para cada um dos temas de dados espaciais.

Desafios – Inconsistências Dados

- Inconsistências no dados OSM
 - Devido à sua natureza colaborativa, e à sua contínua actualização, este modelo é mais propício ao aparecimento de erros e inconsistências;
 - Estas inconsistências, provêm de diferenes práticas, das fontes e das limitações do próprio modelo.

Desafios – Semântica

- O crescimento da informação geográfica disponível online potencia a problemática da semântica;
- No conjunto de dados vectorial do OSM, os objectos do mapa são associados a propriedades que codificam o seu conteúdo semântico;
- As propriedades dos objectos são denominadas “tags”;
- O significado das “tags” pode ser alterado ao longo do tempo, num processo emergente de semântica, onde os conceitos emergem, se alteram e desaparecem numa complexa e evolutiva negociação;
- A heterogeneidade semântica num conjunto de dados, significa que termos distintos, possam referir-se a conceitos semelhantes, enquanto termos semelhantes podem significar conceitos distintos.

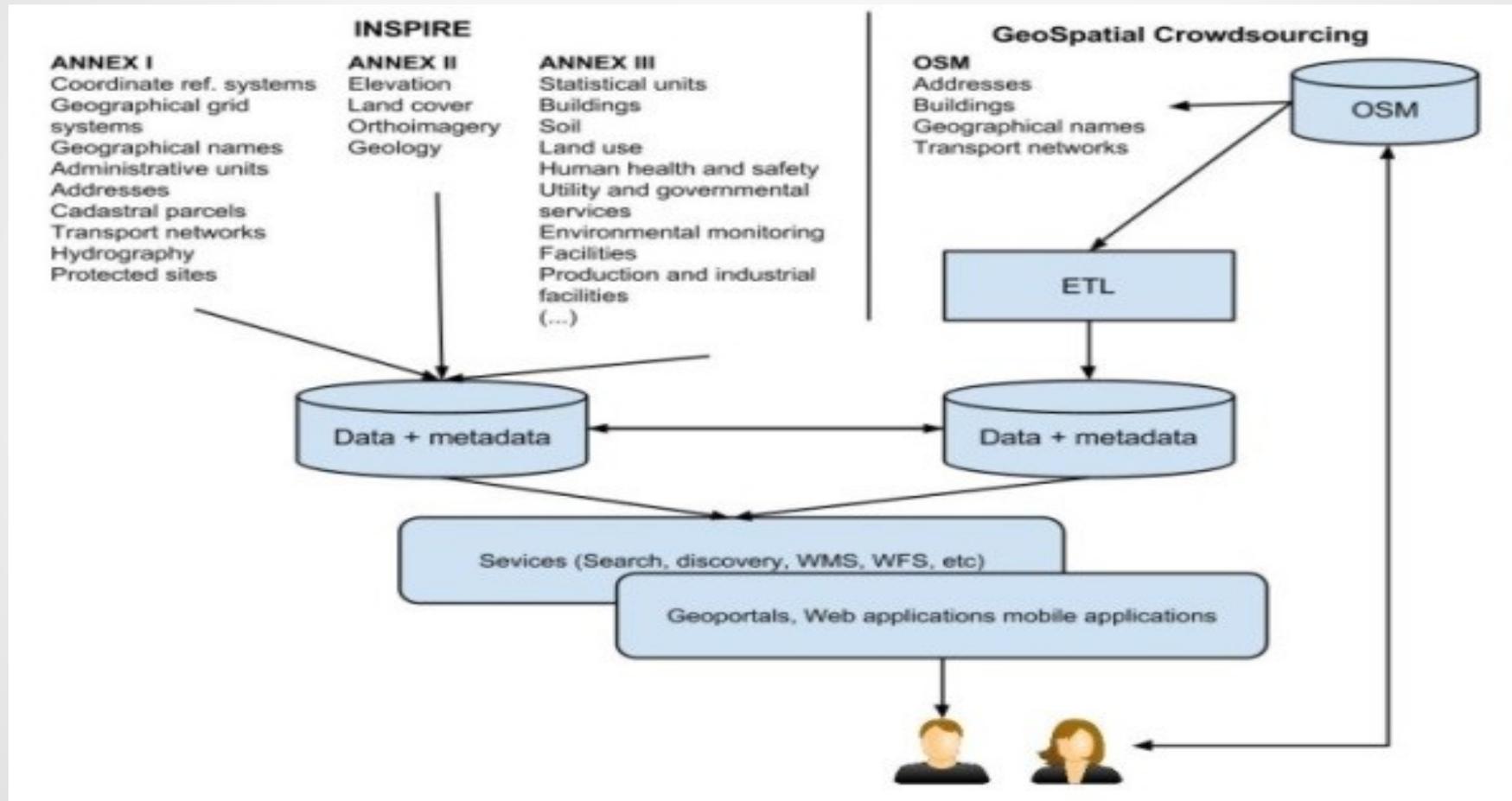
Desafios – Comparação de dados

- É muito difícil comparar conjuntos de dados do OSM, com conjuntos de dados oficiais referentes à mesma área;
- Simultaneamente, as diferenças topológicas e geográficas, aumentam as dificuldades de percepção.

Desafios – Bidireccionalidade

- Até agora, apenas foi discutido fluxo de dados numa direcção – OSM para INSPIRE.
- A bidireccionalidade refere-se à capacidade da total interoperabilidade entre OSM e INSPIRE, ie, a transformação de dados não ocorre apenas na direcção do OSM para INSPIRE, mas também no sentido contrário.
- É assim necessário, garantir a existência de fluxos do INSPIRE para o OSM, com o objectivo de adicionar valor aos utilizadores e ao OSM, sem no entanto, desvirtuar o espírito do projecto OSM, e em especial da sua comunidade.

Arquitectura

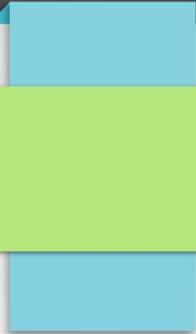


Caso de Estudo

- O workflow apresentado foi testado na transformação de dados OSM e oficiais no esquema de endereços do modelo de dados INSPIRE;
- Foram utilizados dados oficiais de um município, e os dados OSM da mesma área;
- Este teste foi levado a cabo apenas para observar os endereços do OSM e do município lado a lado. Foi também avaliada a dificuldade da transformação;
- Foi criado um conjunto de dados em formato INSPIRE com os Limites Administrativos (Administrative Units) e de códigos postais. Para cada um destes dois temas, foram criados identificadores únicos;
- Estes conjuntos de dados foram inseridos num sistema de gestão de base de dados espacial relacional;
- Foram levados a cabo várias tarefas com o objectivo de preparar os dados OSM para os processos subsequentes. Estas tarefas são compostas por uma série de processos ETL (Extract, Transform and Load) que genericamente convertem dados OSM em endereços formato Geography Markup Language (GML), de acordo com o modelo de dados INSPIRE;
- Com estas transformações, um simples endereço, que geralmente é escrito em três ou quatro linhas (Nome da rua, edifício e código postal), é agora descrito em mais de 200 linhas de GML, o que demonstra a complexidade do modelo de dados INSPIRE.;

Conclusões

- Com este trabalho acreditamos ter provado o interesse de um workflow para integrar geospatial crowdsourcing em INSPIRE, utilizando assim esta directiva, como denominador comum e alternativa para confrontar dados oficiais;
- Um melhor conhecimento do novo papel dos utilizadores em IDE's INSPIRE, torna os dados INSPIRE mais acessíveis ao público, uma vez que o INSPIRE foi criado antes do advento do Geospatial Crowdsourcing;
- O melhor do OSM é a sua base de dados. Integrar dados OSM em INSPIRE, torna possível tirar partido do melhor do OSM, os seus dados.
- O modelo de dados INSPIRE, é muito mais poderoso mas adiciona um elevado grau de complexidade. A médio prazo, será necessário desenvolver novas ferramentas que auxiliem na manutenção dos conjuntos de dados INSPIRE.
- Este é um trabalho que se encontra em desenvolvimento, pelo que para melhor ultrapassar os desafios identificados previamente, será necessário desenvolver ajustes ao workflow.
- Apesar da natureza local deste trabalho, o esforço aplicado pode e deve ser replicado por toda a Europa.
- Será muito importante num futuro próximo, observar de que forma o INSPIRE reflecte as necessidades locais, regionais e globais.



Obrigado

Questões?