



PÓS-GRADUAÇÃO

COMPUTAÇÃO APLICADA

Métricas de qualidade para VGI para o projeto Pauliceia

Rodrigo Monteiro Mariano

rodrigo.mariano@inpe.br / rmmariano0@gmail.com

— Orientadores: Dra. Karine Reis Ferreira —
 Dr. Luis Antonio C. Ferla

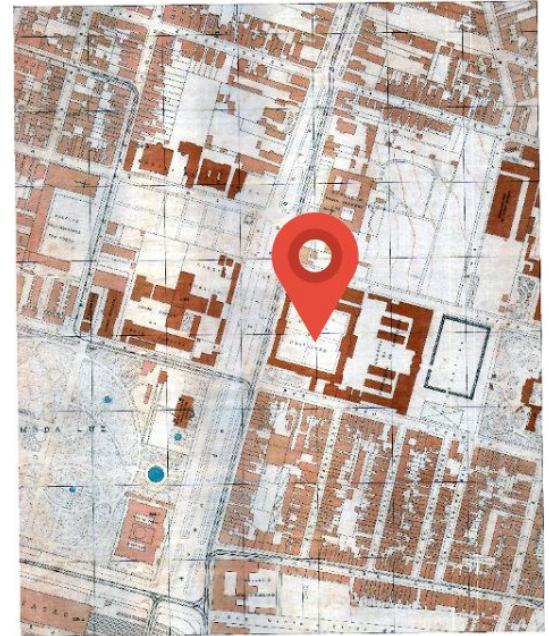


Objetivo

- Fazer uma revisão das métricas e estratégias utilizadas em VGI para avaliar a qualidade de dados colaborativos.
- Analisar quais métricas seriam viáveis para os dados históricos do projeto Pauliceia.

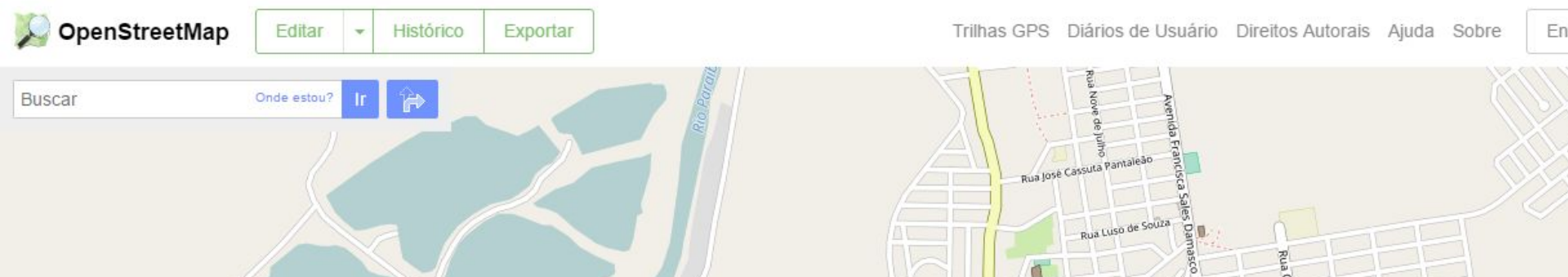
Projeto Pauliceia

- Coletar, selecionar, digitalizar e criar dados da história da cidade de São Paulo.
- Desenvolver uma **plataforma computacional** que permite pesquisadores **explorar** e **compartilhar** conjuntos de **dados históricos urbanos**.
 - Dados espaço-temporais.
- Dados geográficos coletados colaborativamente (**VGI**).



Volunteered Geographic Information (VGI)

- *“the version of crowdsourcing in which members of the general public create and contribute georeferenced facts about the Earth's surface and near-surface to websites where the facts are synthesized into databases” [Goodchild and Li 2012].*



Pauliceia e VGI

- Os historiadores podem **adicionar** e **compartilhar** conjuntos de dados geográficos resultantes de suas pesquisas de diferentes temas.
- Os **voluntários** podem nos ajudar a criar **dados vetoriais** (ruas e edifícios) a partir de mapas históricos (**vetorização**).
- Os cidadãos podem nos ajudar a colecionar outros tipos de informações históricas, como **fotos** associadas a **locais históricos**.



Como avaliar a qualidade dos dados de VGI?

- **Medidas de qualidade:** usa dados autoritários para fazer comparação (e.g. integridade);
- **Indicadores de qualidade:** expressam a qualidade dos dados (quando não há dados para comparação) (e.g. confiabilidade);
- **Abordagens de qualidade:** proposto pelo Goodchild e Li (e.g. crowdsourcing);
- **Protocolo:** padronização.

Medidas de qualidade para VGI

- Usa dados autoritários para fazer comparação.
 - **Integridade:** presença e ausência de dados.
 - **Consistência lógica:** coerência às regras lógicas das estrutura de dados.
 - **Precisão do posicionamento:** posição correta dos recursos.
 - **Precisão temática:** correção de classificação;
 - **Precisão temporal:** consistência temporal.
- Não se aplica ao projeto Pauliceia.

Indicadores de qualidade para VGI



- Expressam a qualidade dos dados.
- **confiabilidade:** boas classificações das contribuições ou maior frequência de uso delas.
- **experiência:** tempo de cadastro; quantidade de contribuições feitas ou número de vezes que o usuário utilizou os fóruns para discussão.
- **credibilidade:** credibilidade da fonte do dado (e.g. metadado que indica o artigo original do dado).

Indicadores de qualidade para VGI



- **conhecimento local:** conhecimento do ambiente geográfico.
 - Os historiadores têm conhecimento da área coletada.
- **reconhecimento:** dar recompensas ao contribuidor (e.g. prêmios virtuais, conhecido como Gamificação) e oportunidade de revisão de suas contribuições;
- **reputação:** avaliar o histórico das interações do voluntário entre os demais.
 - As ações (**adicionar, avaliar, ...**) do usuário afetam sua reputação.

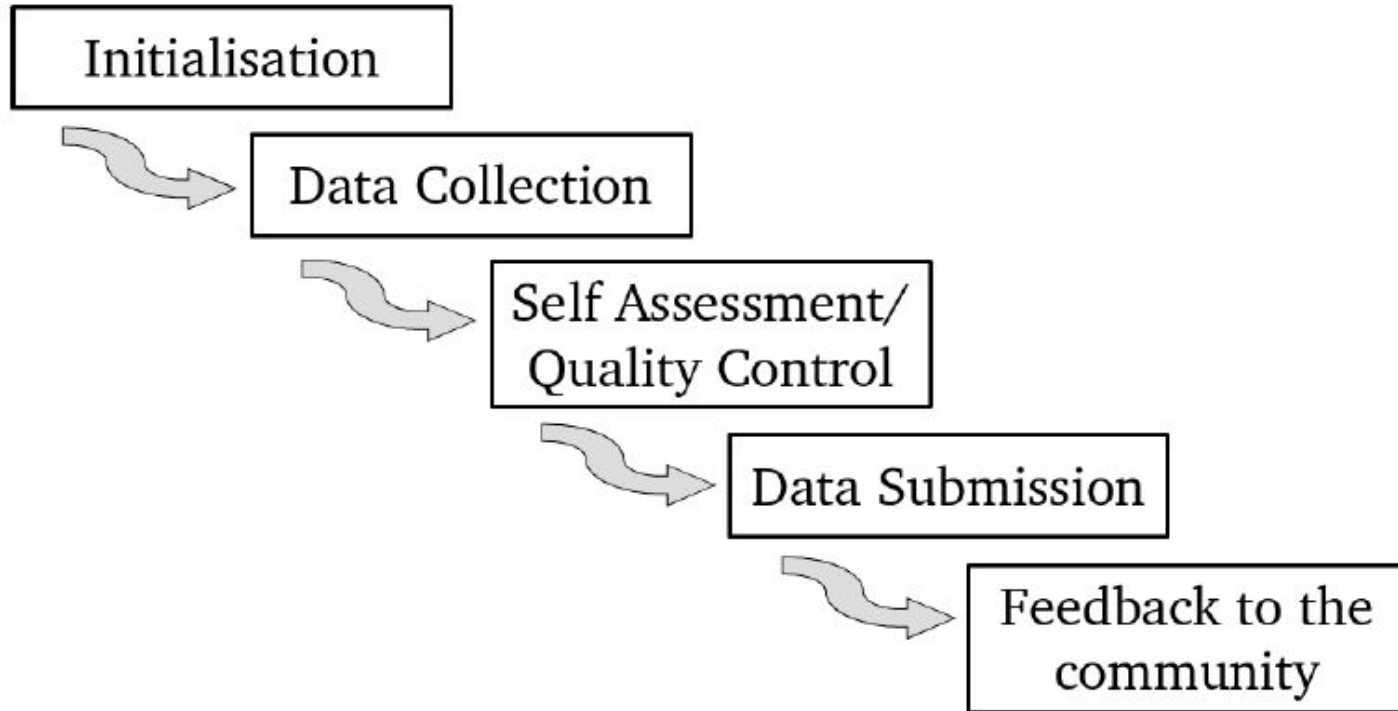
Abordagens de qualidade para VGI

- Proposto por Goodchild e Li (2012).
- **crowdsourcing**: envolvimento de um grupo para validar e corrigir erros cometidos;
- **abordagens sociais**: são indivíduos confiáveis (boa reputação) podem atuar como **porteiros** para manter e controlar a qualidade de outras contribuições.
- **abordagens geográficas**: é o uso de leis/regras para avaliar a qualidade.
 - Exemplo: alguém adicionar um hotel no meio de uma floresta (provavelmente é mentira).

Protocolo para projetos VGI

- Proposto por Mooney et al. (2016)
- Criar uma **padronização** para os projetos que trabalham com dados geográficos colaborativos.
- Esta padronização pode acarretar em uma **melhora na qualidade** das contribuições feitas pelos usuários.
- Falta de padronização pode gerar dados de má qualidade.

Protocolo para projetos VGI



Protocolo VGI para Pauliceia

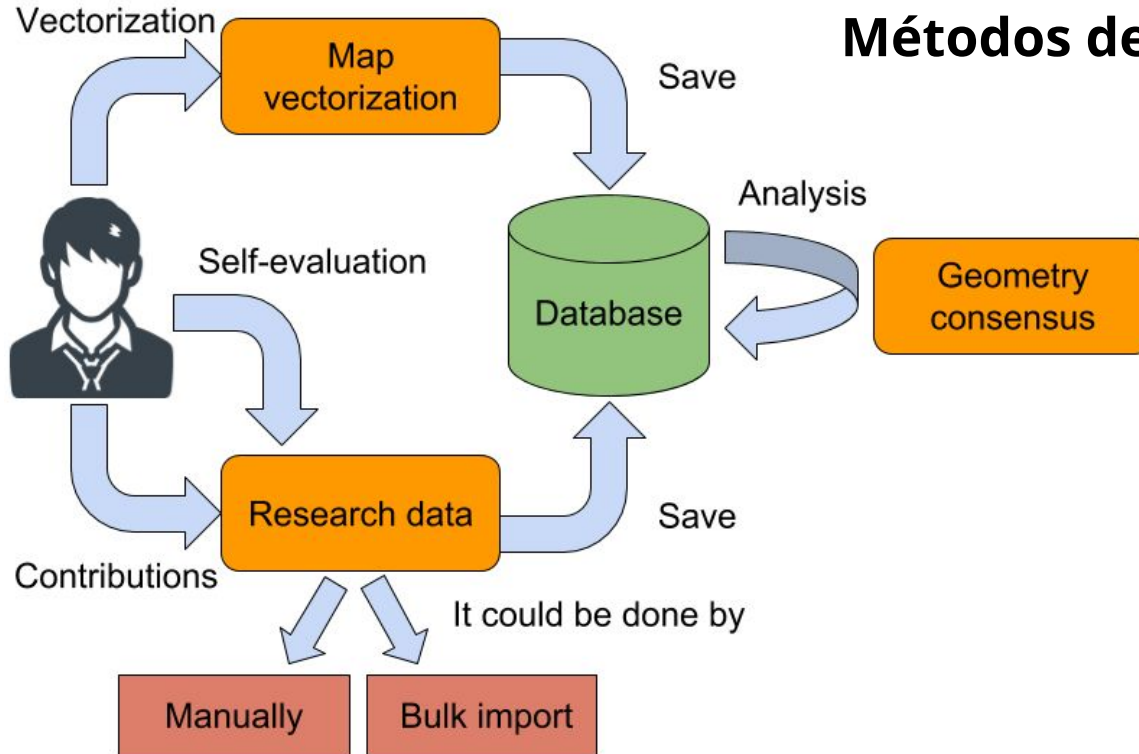
- **Inicialização:**

- Acessar o portal com um login;
- Deve-se fazer um treinamento em um **ambiente de teste;**



Protocolo VGI para Pauliceia

Métodos de coleta de dados



Protocolo VGI para Pauliceia

- **Controle de qualidade:**

- A avaliação será feita com base no conhecimento da área do usuário, aplicando-se a Lei de Linus. (“**Given enough eyeballs, all bugs are shallow**”).
- Vetorização dos mapas antigos: utilizará um **algoritmo** para encontrar o **melhor consenso** dentre os mesmos objetos vetorizados.

Protocolo VGI para Pauliceia

- **Controle de qualidade:**

- Técnicas de Gamificação: sistema de **reputação** baseado em **Ranking** e **votação** das contribuições.
- **Histórico** das contribuições.



Protocolo VGI para Pauliceia

- **Feedback:**

- Os usuários poderão **fornecer feedback** de sua experiência, expressando suas observações e opiniões nos canais disponíveis (e.g. lista de discussão).
- Os contribuintes poderão **ser notificados** sobre comentários, revisões e informações de projetos que faça parte.



Conclusões

- Não há dados autoritários para comparar: **indicadores de qualidade**.
- Das abordagens de qualidade, uso de pelo menos a de **crowdsourcing**.
- Definição de um protocolo (**padronização**) descrevendo os tópicos que o voluntário poderá fazer e como deve seguir.
- Não existe um único método que possa resolver todos os problemas (depende).
- As questões principais estão sendo incorporadas no serviço web para VGI do projeto Pauliceia na forma de **funções** bem definidas.

Obrigado!

Rodrigo M. Mariano

rodrigo.mariano@inpe.br / rmmariano0@gmail.com

<https://github.com/Pauliceia/vgiws>