

Métricas de qualidade para VGI para o projeto Pauliceia

Rodrigo M. Mariano¹, Karine R. Ferreira¹, Luis Ferla²

¹Instituto Nacional de Pesquisas Espaciais (INPE)
São José dos Campos – SP – Brasil

²Universidade Federal de São Paulo (UNIFESP)
Guarulhos – SP – Brasil

{rodrigo.mariano, karine.ferreira}@inpe.br, ferla@unifesp.br

Abstract. *Pauliceia is a project whose aim is to develop a computational platform for historical data manipulation collaboratively. Researchers will contribute with these data and assist in quality control. Apart from the insertion of historical data, there will be the opportunity for the volunteers to make the mapping of old maps and it is intended that the system be in charge of verifying a consensus in the geometries obtained. Quality questions and metrics were evaluated for data collection by VGI, in order to improve data quality. With it, a protocol for the Pauliceia's project was developed, because the lack of it can generate contributions with poor quality.*

Resumo. *Pauliceia é um projeto cujo propósito é desenvolver uma plataforma computacional para manipulação de dados históricos colaborativamente. Os pesquisadores contribuirão com esses dados e auxiliarão no controle de qualidade. Fora a inserção de dados históricos, haverá a oportunidade dos voluntários fazerem a vetorização de mapas antigos e pretende-se que o sistema se encarregue de verificar um consenso nas geometrias obtidas. Questões e métricas de qualidade foram avaliadas, para dados coletados por VGI, com o intuito de melhorar a qualidade dos dados. Com isso, desenvolvendo-se um protocolo para o projeto Pauliceia, porque a falta dele pode gerar contribuições com má qualidade.*

1. Introdução

O projeto Pauliceia tem como objetivo desenvolver uma plataforma computacional online para gerenciamento colaborativo de dados históricos, que contêm uma localização geográfica (i.e. dados espaço-temporais). Os pesquisadores poderão produzir mapas e visualizações de suas próprias pesquisas e ao mesmo tempo, contribuir para os dados dentro do sistema. Uma atividade online com participação da comunidade em um propósito, é conhecido como crowdsourcing e quando esses voluntários contribuem com informações geográficas, é conhecido como VGI. O projeto irá enriquecer a compreensão da história de São Paulo (SP) durante o período de 1870 a 1940, além de oferecer um modelo inovador de pesquisa para as Humanidades Digitais, que promova o trabalho colaborativo e o fluxo de conhecimento gratuito. O recorte histórico de 1870 a 1940 foi escolhido, pois foi uma época em que a cidade de SP cresceu muito, saindo de aproximadamente 30.000 habitantes para 1.300.000 em torno de 70 anos [Secretaria Municipal de Urbanismo e Licenciamento 2017]. Os principais motivos para

este acontecimento, se dá no aumento da produção de café, vinda de imigrantes, criação da ferrovia e desenvolvimento da indústria [Mota et al. 2007] [Carvalho 2017].

[Estellés-Arolas and González-Ladrón-de Guevara 2012] estudaram e produziram uma única definição de crowdsourcing: “*a type of participative online activity in which an individual, an institution, a non-profit organization, or company proposes to a group of individuals of varying knowledge, heterogeneity, and number, via a flexible open call, the voluntary undertaking of a task. The undertaking of the task, of variable complexity and modularity, and in which the crowd should participate bringing their work, money, knowledge and/or experience, always entails mutual benefit. The user will receive the satisfaction of a given type of need, be it economic, social recognition, self-esteem, or the development of individual skills, while the crowdsourcer will obtain and utilize to their advantage what the user has brought to the venture, whose form will depend on the type of activity undertaken*”. Enquanto a coleta de dados em projetos de ciência cidadã pode ser feitos, por exemplo, com formulários em papel, por definição o crowdsourcing tem natureza online. Isto torna o crowdsourcing mais restrito. Esta categoria de participação não precisa ser abertos a todos, podendo ser restringidos a certos grupos [See et al. 2016].

Originalmente o termo Volunteered Geographic Information (VGI) aparece por Goodchild [Goodchild 2007a] como “*the harnessing of tools to create, assemble, and disseminate geographic data provided voluntarily by individuals*” [See et al. 2016] e quando o termo é definido, ele diz: “*the widespread engagement of large numbers of private citizens, often with little in the way of formal qualifications, in the creation of geographic information. They are largely untrained and their actions are almost always voluntary, and the results may or may not be accurate. But collectively, they represent a dramatic innovation that will have profound impacts on geographic information systems (GIS) and on the discipline of geography and its relationship to the general public. VGI is a special case of the more general Web phenomenon of user generated content*”, em outro trabalho, ele com Li [Goodchild and Li 2012] descrevem VGI como sendo “*a version of crowd-sourcing 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*”. Em suma, é um fenômeno recente que oferece um mecanismo alternativo para a aquisição e compilação de informações geográficas de voluntários [Goodchild 2007a]. Essas contribuições contêm uma localização geográfica e uma descrição, com vários atributos, recorrentes dessa localização [Goodchild and Li 2012]. Os usuários são, frequentemente, não treinados, e apesar de seus conhecimentos e antecedentes, criam informações geográficas em plataformas web, como por exemplo: OpenStreetMap (OSM), Flickr ou Wikimapia [Senaratne et al. 2017]. A representação do dado pode ser feita por um ponto, uma linha ou polígono. Mesmo com alto potencial, por adquirir informações geográficas de maneira rápida, detalhada e com baixo custo, o VGI por padrão não oferece garantia de qualidade em seus dados [Goodchild and Li 2012].

O OpenStreetMap (OSM) é o mais conhecido sistema de VGI. Ele permite trabalhar com dados geográficos gratuitos, tendo uma licença de conteúdo aberto [OpenStreetMap 2017a]. Vários estudos medem a precisão dos dados de VGI com base no OSM, por exemplo: [Haklay 2010], [Girres and Touya 2010] ou [Ciepluch et al. 2010]. O trabalho de [Senaratne et al. 2017] faz uma revisão detalhada sobre os métodos para avaliar a qualidade dos dados de VGI. Todos os estudos fornecem

conhecimentos proveitosos sobre a precisão do VGI, ajudando a garantir e melhorar a qualidade.

A imprecisão dos dados de VGI é explicado pelo fato de que os seres humanos expressam as regiões geográficas e suas relações imprecisamente, através de conceitos vagos. A imprecisão na conceptualização humana da localização deve-se, não apenas ao fato de que as entidades geográficas são de natureza contínua, mas também por conta da qualidade e limitações do conhecimento espacial [Montello et al. 2003] [Hollenstein and Purves 2010]. Fornecer serviços confiáveis ou retirar informações úteis desses dados, requer das contribuições pelo menos um padrão de qualidade. Informações imprecisas, sejam elas maliciosas ou não, podem ser minimizados por indicadores de qualidade apropriados e medidas para essas várias contribuições VGI.

As Agências Nacionais de Mapeamento (NMAs) e as Empresas de Topografia Comercial (CSCs) utilizam protocolos robustos e padronizados que regem e orientam a coleta de dados geográficos, no entanto os projetos VGI muitas vezes carecem de padrões, ou apenas fornecem diretrizes e sugestões soltas, em vez de especificações rigorosas. Embora o VGI possa teoricamente atingir altos padrões de qualidade sem padrões rigorosos, sua ausência é muitas vezes uma fonte importante de erros nos dados, representando uma barreira à sua maior difusão e reutilização [Mooney et al. 2016].

A necessidade de estabelecer padrões e protocolos para projetos VGI não é uma novidade. Girres e Touya [Girres and Touya 2010] citam que a falta de especificação é um dos pontos chave que pode causar a má qualidade nos dados. Que a heterogeneidade das contribuições podem ser causadas pela falta de especificação na coleta dos dados. De acordo com Mooney et al. [Mooney et al. 2016] alguns pesquisadores alertaram sobre as ameaças para a comunidade e a sociedade, colocadas pela falta de protocolos e mencionaram a relevância deles para projetos de VGI, sugerindo a definição de protocolos para garantir alta qualidade de dados. Os protocolos são importantes para facilitar e ampliar a reutilização dos dados do VGI, para fins e aplicativos diferentes do que originalmente foram coletados.

Muitas obras desenvolveram métodos para avaliar a qualidade do VGI, mesmo sendo perceptível que não haja somente um único método que resolva todos os problemas, afinal as contribuições de VGI tem perfil heterogêneo [Senaratne et al. 2017]. A criação de um protocolo bem definido, pode diminuir as barreiras descritas, por conta da padronização ou formalização criada, provendo uma melhora na coleta de dados. Por conta disso, o presente trabalho fará uma revisão das técnicas de qualidade de VGI existentes e apresentará as que melhores se aplicam ao contexto do projeto Pauliceia, desenvolvendo um protocolo para ele.

1.1. Trabalho Relacionado

Nesta seção, será apresentado projetos que possuem características semelhantes ao projeto Pauliceia.

Wikimapia¹ é um sistema semelhante ao Wikipedia, permitindo aos usuários contribuírem com descrições sobre pontos de interesse, contendo coordenadas geográficas. Ele é um dicionário geográfico voluntário, produzido por pessoas que oferecem

¹<http://wikimapia.org>

explicações ricas sobre os lugares [Goodchild 2007b]. Para o controle de qualidade, o Wikimapia fornece de acordo com as contribuições feitas: prêmios, hierarquia de papéis e responsabilidades. Por exemplo, o papel de Usuário Avançado é dado a um voluntário com boa reputação, dando a ele ferramentas e responsabilidades extras. Isto gera usuários prolíficos (não garantindo a precisão), porém como normalmente estes usuários tem conhecimento local sobre um assunto, tendem a ter boa qualidade [Goodchild and Li 2012].

O Building Inspector² é um projeto dos Laboratórios da Biblioteca Pública de Nova York em colaboração com Lionel Pincus e Princess Firyal Map Division da Biblioteca Pública de Nova York, feito para gerenciar dados de mapas históricos. Os computadores são treinados para fazer o levantamento pesado e distribui-se as demais tarefas de controle de qualidade para os cidadãos, produzindo um diretório abrangente da antiga Nova York. Com essa informação é possível explorar o passado da cidade. O projeto permite vincular documentos históricos (arquivos, jornais antigos, fotografias e etc.) aos lugares, dando oportunidade de novas maneiras de pesquisar, aprender e descobrir o passado. No controle de qualidade é usado um algoritmo proposto por Budig et al. [Budig et al. 2016] para a extração da melhor representação poligonal, dado vários polígonos, de construções dos atlas do século XIX e início do século XX. Dado um conjunto de polígonos, que descrevem o mesmo objeto (e.g. um edifício), o algoritmo retorna o melhor consenso.

OpenHistoricalMap³ (OHM) é um projeto que usa a infraestrutura OSM para criação de um mapa universal e detalhado da história mundo, de forma colaborativa. Pode-se inserir dados de margens, fronteiras políticas, edifícios e caminhos, semelhantes ao OSM. O HistOSM⁴ é um aplicativo para exploração de objetos históricos do OSM. Pode-se rastrear as características históricas regionais ao ampliar e filtrar o mapa, mostrando os objetos de interesse. Se clicar nos objetos de maneira individual, será mostrado as informações detalhadas deles, como tags e links (e.g. imagens associadas ou sites), podendo também ir para o site do OSM para editar os objetos e se necessário, adicionar ou atualizar informações. Os dados OHM e o HistOSM se baseiam no OSM, portanto o controle de qualidade são feitos por ferramentas do OSM.

O OSM utiliza ferramentas para melhorar a qualidade dos seus dados. Estas ferramentas fornecem uma lista de erros nos dados, que os voluntários podem corrigir através das ferramentas de edição. Algumas dessas ferramentas se dividem em: reporte de bugs (e.g. Notes⁵); detecção de erros (e.g. Keep Right⁶); monitoramento (e.g. History Browser⁷); assistência nas etiquetas (e.g. Traffic Sign Tool⁸) e agregação de etiquetas (e.g. Taginfo⁹) [OpenStreetMap 2017c].

O projeto Pauliceia tem características parecidas com esses trabalhos, também sendo influenciados por alguns. Enquanto o OpenHistoricalMap e o HistOSM trabalham atualmente só com objetos históricos (e.g. monumentos), no Pauliceia também haverão

²<http://buildinginspector.nypl.org/>

³<http://www.openhistoricalmap.org>

⁴<http://histosm.org/>

⁵<http://wiki.openstreetmap.org/wiki/Notes>

⁶http://wiki.openstreetmap.org/wiki/Keep_Right

⁷<http://osm.virtuelle-loipe.de/history/>

⁸http://osmtools.de/traffic_signs/

⁹<http://wiki.openstreetmap.org/wiki/Taginfo>

eventos históricos (e.g. assaltos de uma época). Há um projeto chamado Historic Event¹⁰, cujo objetivo era adicionar eventos históricos no OSM, mas ele não foi continuado. Um dos principais focos do projeto Pauliceia é o crowdsourcing, pois quem alimentará a plataforma serão os usuários finais, que são pessoas com dados históricos interessadas em colaborar. O VGI será utilizado para motivar os cidadãos a participar na coleta de informações espaciais com qualidade, atuando como sensores. Os usuários providenciarão feedbacks (revisões, comentários, votações e etc.) para melhora dos dados.

1.2. 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 e como incorporar isto no serviço web para VGI que está sendo desenvolvido.

2. Informação Geográfica Voluntária

Com o VGI, se tornou possível as pessoas criarem sua própria informação geográfica digital, através de mapas on-line, sem custo. Isto se dá por conta que as coordenadas podem ser obtidas com o GPS ou utilizar imagens disponibilizadas por terceiros (e.g. Google Earth). Conhecimento em cartografia já não é necessário, pois há softwares open-source para construção de mapas com alta qualidade. Um usuário pode desenvolver mapas de sua área local, podendo ser mais eficaz, do que um especialista em mapeamento, devido seu próprio conhecimento local [Goodchild and Li 2012]. Por conta disso é interessante estudar a qualidade e melhoramento dos dados gerados por VGI, que será descrito nas próximas seções.

2.1. Medidas, indicadores e abordagens para a qualidade do VGI

A qualidade das contribuições do VGI podem ser descritas em essência por: medidas de qualidade e indicadores de qualidade. As medidas de qualidade utilizam os dados autoritários (gerados por agências confiáveis, e.g. NMAs ou CSCs), como um conjunto de dados de referência, para avaliar os dados gerados pelo VGI, comparando-os. Isto é feito, pois acredita-se que os dados autoritários têm sempre alta qualidade. Os indicadores de qualidade são utilizados quando os dados autoritários não estão disponíveis, o que pode ser frequente, dependendo da categoria dos dados. Além de que, Over et al. [Over et al. 2010] observam que a qualidade dos dados do OSM difere de dados autoritários. A natureza do voluntariado pode apresentar tendências nas contribuições. Isto ocorre por vários fatores, desde capacidade técnicas até diferenças culturais das pessoas. Conforme os dados de VGI ficam mais detalhados ao longo do tempo e em determinadas áreas, torna-se menos proveitoso a utilização de dados autoritários para avaliar a qualidade dos dados gerados por VGI. Isto ocorre, pois os dados de VGI são muitas vezes, mais completos e precisos, do que os conjuntos de dados autoritários existentes [Antoniou and Skopeliti 2015].

Em relação a medidas de qualidade para VGI, a Organização Internacional de Normalização (ISO) definiu a qualidade da informação geográfica como: *“a totalidade das características de um produto que tem sua capacidade de satisfazer as necessidades declaradas e implícitas”*¹¹. Existe um conjunto de padrões explicados pela

¹⁰http://wiki.openstreetmap.org/wiki/Proposed_features/historic_event

¹¹<https://www.iso.org/obp/ui/#iso:std:iso:19109:ed-1:v1:en>

ISO, que define as medidas de qualidade da informação geográfica, que são [TC 2009] [Jakobsson and Giversen 2007]:

- integridade: descreve a presença e ausência de dados, seus atributos e relacionamentos entre objetos. Para ser avaliado pode-se utilizar: comparação do número de características, comparação do comprimento ou área total, ou índice de integridade;
- consistência lógica: nível de coerência às regras lógicas das estrutura de dados (como conceitual, lógica ou física), atribuição e relacionamentos. Para ser avaliado pode-se utilizar: semelhança semântica entre as etiquetas, identificação de entidades com classificação inadequada ou sistema de recomendação de etiquetas (algoritmo que sugere etiquetas relevantes);
- precisão do posicionamento: precisão da posição dos recursos. É a proximidade entre uma medida de uma quantidade e o valor verdadeiro aceito dela. Para ser avaliado pode-se utilizar: a distância euclidiana dos atributos pontuais ou distâncias entre polígonos centroides;
- precisão temática: correção de classificação, correção de atributos não quantitativos e precisão de atributos quantitativos. Para ser avaliado pode-se utilizar: medição da porcentagem (%) da classificação correta, matriz de confusão ou análise padrão do índice kappa;
- precisão temporal: precisão de uma medida de tempo, consistência temporal e validade dos dados em relação ao tempo. Para ser avaliado pode-se estudar a evolução dos dados VGI.

No que diz respeito aos indicadores de qualidade para VGI, eles são qualitativos (expressam a qualidade dos dados), como o propósito, uso e linhagem. O propósito expõe o objetivo dos dados. O uso indica quais as funcionalidades dos dados. Linhagem se refere ao histórico dos dados, desde sua coleta, aquisição e formação para o seu uso final. Os seus indicadores podem ser descritos como [Senaratne et al. 2017]:

- confiabilidade: um julgamento baseado em características subjetivas, como confiança. Sendo adquiridas por boas classificações das contribuições ou maior frequência de uso delas [Flanagin and Metzger 2008];
- experiência: experiência do usuário na plataforma VGI. Pode ser capturado quando o voluntário se registra no portal, pela quantidade de contribuições feitas (seja adicionadas ou editadas) ou número de vezes que o usuário utilizou os fóruns online para discussão dos dados;
- credibilidade: definida como a credibilidade de uma fonte ou mensagem, na qual têm duas dimensões: confiabilidade e experiência. Para avaliá-lo, usa-se a fonte da informação como base, porém não é direto, pois como os dados de VGI não são autoritários, portanto a fonte talvez não seja disponível. É necessário considerar fatores de confiabilidade e experiência, para conseguir fazer esta categoria de avaliação. Os metadados sobre a origem do VGI podem fornecer uma base para isso;
- qualidade do conteúdo de texto: é a qualidade dos dados de texto baseada no uso das características do texto, como: comprimento dele, estrutura, legibilidade, histórico de revisões, uso de termos específicos e entre outros. Normalmente aplicável em VGI baseado em texto;

- imprecisão: ambiguidade na captura dos dados, como por exemplo: imprecisão por baixa resolução;
- conhecimento local: conhecimento dos usuários em relação ao ambiente geográfico que ele está mapeando;
- reconhecimento: dar recompensas ao contribuidor pela utilização da plataforma, como prêmios virtuais, conhecido como Gamificação, e oportunidade de revisão de suas contribuições por outros usuários;
- reputação: é a capacidade de avaliar, marcar, discutir e anotar as contribuições, afetando na reputação do usuário. Avaliação do histórico das interações do voluntário entre os outros colaboradores.

Segundo Goodchild e Li [Goodchild and Li 2012] há ainda três abordagens para assegurar a qualidade do VGI:

- crowdsourcing (geração de informação por várias pessoas): é o envolvimento de um grupo para validar e corrigir erros cometidos por um colaborador individual;
- abordagens sociais: são indivíduos confiáveis que tem uma boa reputação com suas contribuições para o VGI, podendo atuar como *gatekeepers* (porteiros) para manter e controlar a qualidade de outras contribuições de VGI;
- abordagens geográficas: é o uso de leis e conhecimento da geografia, como a primeira lei da geografia¹², para avaliar a qualidade.

Cechanowicz et al. [Cechanowicz et al. 2013] e Hamari et al. [Hamari et al. 2014] recomendam o uso de técnicas de Gamificação para incentivar a participação dos usuários, instigando-os na qualidade dos dados. Gamificação é o uso de elementos de jogo em contextos que não sejam jogos. É uma forma de melhorar o engajamento e a motivação dos usuários, aumenta a participação e os proporciona uma melhor experiência. Isto pode ser utilizado para obter mais de um contribuinte: mais dados, dados de maior qualidade, aumento da frequência e duração da participação [Cechanowicz et al. 2013]. É um processo de melhoramento dos serviços com recursos estimulantes, utilizando as experiências de jogos para obter resultados comportamentais adicionais. É usado para apoiar o empenho dos usuários e melhorar o uso de serviços. Por exemplo: aumentar a atividade do usuário, interação social, qualidade e produtividade de suas ações [Hamari et al. 2014].

3. Protocolo para projetos VGI

Mooney et al. [Mooney et al. 2016] propõem um protocolo genérico, para ser aplicado a diversos projetos de VGI. Este protocolo visa 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. Por conta disso, foi-se capturado ideias deste artigo para serem aplicadas ao projeto Pauliceia.

4. Protocolo VGI para o projeto Pauliceia

Como descrito na seção 2.1, as medidas de qualidade usam conjunto de dados autoritários para comparação. No contexto do projeto Pauliceia, até onde se tem conhecimento, não

¹²“Everything is related to everything else but near things are more related than distant things” [Tobler 2004]

existe dados de referência históricos que abordem a cidade de São Paulo no período do projeto. Por esta razão, essa abordagem não é aplicável, sendo utilizados, portanto, os indicadores e abordagens de qualidade. Isto será descrito nesta seção, aplicado ao protocolo do projeto Pauliceia.

4.1. Inicialização

Os pesquisadores de história, alunos ou quaisquer cidadãos entrarão no portal Pauliceia utilizando um navegador, pelo computador. Para fazer o gerenciamento dos dados, é necessário que eles façam um cadastro ou acessem com um login social. Antes de começar a contribuição, o usuário deverá aceitar um termo de Política de Uso do projeto. Este termo descreve, que o sistema não se responsabiliza pelas contribuições e que os dados da plataforma são públicos, portanto não se pode inserir dados com direitos autorais, logo o usuário se responsabiliza disso.

Os voluntários são encorajados a preencherem os campos opcionais em seu perfil, como instituição que faz parte, nível (graduação, mestrado ou etc), entre outros. Isto é feito, pois como o projeto é voltado para academia, é conveniente que se tenha esse tipo de informação para determinar a confiabilidade das contribuições de um usuário, baseado no seu perfil acadêmico. Por exemplo, supõe-se que os dados contribuídos por um usuário com pós-doutorado seja mais confiável, do que o de um aluno de graduação. O procedimento que o voluntário deve seguir antes de iniciar a coleta de dados é o seguinte:

- ler sobre o projeto Pauliceia, suas especificações e objetivos, que estarão disponíveis no site. Isto fará que o usuário se familiarize mais com o sistema, fazendo-o entender o porquê e como coletar os dados. Os usuários são encorajados a fazer comentários, tirar dúvidas ou dar sugestões da plataforma, enviando um e-mail para support_pauliceia@googlegroups.com.
- simular a coleta de dados em um ambiente “sandbox”, que é um ambiente que contém exemplos de como utilizar o sistema. Neste ambiente o usuário poderá sanar suas principais dúvidas, ter um treinamento prévio da ferramenta e se acostumar com os processos da coleta de dados, fazendo-o se sentir confiante para inserir os dados reais.
- ter certeza que os dados são de domínio público. Os dados não podem ter alguma licença privada.

4.2. Modelo de Dados

Em relação a abordagem de crowdsourcing para assegurar a qualidade do VGI, Goodchild e Li [Goodchild and Li 2012], descrevem que é importante dar a oportunidade de um grupo de validar e corrigir erros que existam no sistema. No projeto Pauliceia será utilizado o conceito de projeto e revisões para isto. Um projeto é um conjunto de contribuições feitas por um ou mais usuários, que indica uma camada temática no sistema. Um historiador pode criar quantos projetos quiser. Este conjunto de dados será relacionado a sua pesquisa (e.g. estudo de assaltos em 1920). O criador desse projeto pode adicionar outros voluntários nele, que terão a oportunidade de gerenciar as contribuições. Teoricamente o historiador só irá adicionar no projeto pessoas confiáveis, que tenham capacidade de adicionar e avaliar os dados corretamente, gerando um controle de qualidade local. Pode haver, por exemplo, voluntários fora desse projeto que descubram erros nos dados. Estes

usuários poderão solicitar uma revisão ou adicionar comentários, indicando os problemas que existam. A solicitação da revisão será feita na página “Revisões” do projeto. Assim um membro do projeto pode avaliar se aquela revisão é válida ou não, decidindo se corrige ou não aquele dado. É aconselhável que o número de contribuidores de um projeto não exceda 13, pois de acordo com Haklay et al. [Haklay et al. 2010] acima de 13 voluntários manipulando os mesmos dados, pode haver uma perda de precisão. A utilização de um conhecimento local dos usuários atrelados a um projeto, é um indicador de qualidade, como mostrado na seção 2.1.

O Sistema de Referência de Coordenadas (CRS) escolhido para a vetorização dos mapas e inserção dos dados é o WGS84 (EPSG:4326). Este CRS foi definido por cobrir o mundo inteiro [Spatial Reference 2017], deixando o sistema mais genérico e possivelmente internacionalizável, para caso o projeto ganhe repercussão em outros países.

Como o projeto Pauliceia é voltado para pesquisa histórica urbana da cidade de São Paulo de 1870 a 1940, os dados a serem coletados devem se restringir a esse espaço e período de tempo. Mais precisamente, os dados devem abranger a área central de São Paulo, que constitui a área piloto definida. Os tipos de dado que podem ser coletados e inseridos na plataforma serão:

- Dados de vetoriais: dados históricos com uma localização geográfica, por exemplo: ocorrência de um assalto em 1930, na Rua São Bento, n. 3 (um ponto); uma área de epidemia de uma doença ou edifício (um polígono); vetorização de ruas antigas (uma linha) e etc. Os dados históricos contêm atributos fixos (como nome, data início e fim, fontes), todavia o usuário pode inserir atributos dinamicamente. O atributo fonte descreve a origem dos dados, se ele é de um artigo, livro, dissertação ou outros, nele conterà a referência bibliográfica do documento original do dado. A referência pode ser descrita seguindo as normas da ABNT ou em bibtex (dica: no Scholar¹³ do Google, quando se pesquisa por algum trabalho, tem a opção “citar” onde têm as bibliografias prontas). Pode-se inserir mais de uma fonte.
- Contribuições textuais, que serão georreferenciadas usando o endereço fornecido. O usuário insere, por exemplo, a ocorrência de um roubo no ano de 1900, na rua São Pio X, n. 42; nisso o sistema se encarregará, dado o endereço, de geolocalizar no mapa o lugar. O dado será inserido como um ponto. Este tipo de dado será usado principalmente para importação de dados em grande volume, para evitar a geolocalização manual de grandes quantidades de dados.
- Dados de mídia (em anexo a um dado vetorial ou textual ou um dado de mídia georreferenciado) como fotos históricas, vídeos, depoimentos em áudio e etc. Esses arquivos devem ser armazenados em um repositório, cujo link será adicionado no dado. Por exemplo: os vídeos são colocados no Youtube, as imagens no Google Photos e documentos são adicionados no Google Drive (ou Dropbox); na contribuição será adicionada a URLs pública do anexo. É aconselhável ao usuário criar uma conta do projeto no repositório que for usar, para que seja evitado excluir acidentalmente, se estiver na conta pessoal. Em relação as fotos históricas, será possível fazer o upload no sistema, caso seja necessário.

¹³<https://scholar.google.com.br/>

- Dados de pesquisas, como artigos e outros trabalhos. Se o documento for o trabalho de origem do dado, é aconselhável que coloque a URL do site no atributo fonte. Caso contrário, o documento pode ser adicionado como um dado de mídia, como descrito acima.

Um atributo importante que os dados coletados devem possuir é uma data, que indica para qual período aquele dado é válido. A data de início representa quando a contribuição ocorreu ou começou a existir, e data final representa quando a contribuição terminou de acontecer ou deixou de existir. O formato da data pode ser com os três campos completos, com mês e ano ou apenas ano. Isto faz com que temos os seguintes casos:

- caso exista data início e final, o usuário deve colocá-los, pois representam a existência temporal dele neste período. Por exemplo: a contribuição tem data início igual à 13/05/1917 e final 13/10/1917, logo a contribuição existirá neste período.
- quando não houver certeza de início e fim do período, a data será considerada discreta, se encaixando em uma das seguintes categorias, onde é sabido a:
 - data início, mas não a final;
 - data final, mas não a de início;
 - data pontual.

Caso exista a data início, mas não a final, ou vice-versa, o usuário deve preencher somente a data que possuir certeza e a data em aberto (não preenchida) ficará disponível como uma data não certa, com uma abrangência máxima de dois anos. Por exemplo: a contribuição tem data início igual à 13/05/1917, mas a data final em aberto. Isto fará com que a contribuição “exista” entre 13/05/1917 e 13/05/1919, pois a data final está em aberto.

A data pontual é quando o usuário sabe somente uma aproximação da existência de uma contribuição, não sabendo uma data exata. Se isto ocorrer, o contribuinte deve informar o período que sabe da existência da dado. Por exemplo: o voluntário sabe que um teatro existiu em 1930, mas não sabe quando foi construído ou demolido, então ele deve colocar a data início e final iguais à 1930.

No caso de data discreta haverá um checkbox para o usuário indicar se há ou não este fato.

4.3. Métodos de Coleta de Dados

Certas recomendações são descritas para que seja este processo seja realizado com sucesso, como: (1) o usuário deve reservar um tempo suficiente para o sucesso dos processos do projeto; evitando distrações e se possível, concentrar o tempo em uma única quantidade, em vez de vários estágios curtos, porque melhora a qualidade dos dados; (2) ter acesso as especificações do projeto, para fazer consulta ou sanar dúvidas e (3) fazer a coleta de dados de acordo com as especificações do projeto. O contribuinte deve anotar qualquer problema técnico do sistema ou situação problemática que encontre, excluindo-se os por mal funcionamento do dispositivo.

A coleta de dados vetoriais pode acontecer em dois cenários diferentes: HistMapathons ou contribuições individuais. Um HistMapathon (Maratona de Mapeamento Histórico) será uma reunião organizada de pessoas com o objetivo de vetorizar mapas

antigos de São Paulo de 1870 a 1940, semelhantes aos realizados pelo Google Maps [Tech2 2017] e OpenStreetMap [OpenStreetMap 2017b]. Esses encontros serão promovidos pelos historiadores e seus alunos. O objetivo principal desses eventos é promover a contribuição em massa de dados vetoriais históricos para a plataforma Pauliceia. Um colaborador também pode vetorizar individualmente fora de um HistMapathon, se desejado.

A Figura 1 ilustra o processo de coleta de dados, onde um usuário poderá fazer a vetorização de mapas antigos ou inserir dados históricos. As contribuições de dados históricos poderão ser feitas manualmente ou por importação em massa, onde o próprio usuário poderá fazer uma auto-avaliação deles. Em relação a vetorização de mapas, vários usuários podem vetorizar o mesmo objeto (e.g. edifício), por conta disso, depois da coleta dos dados, um algoritmo achará o melhor consenso entre as geometrias inseridas. Todas as contribuições serão salvas em um banco de dados:

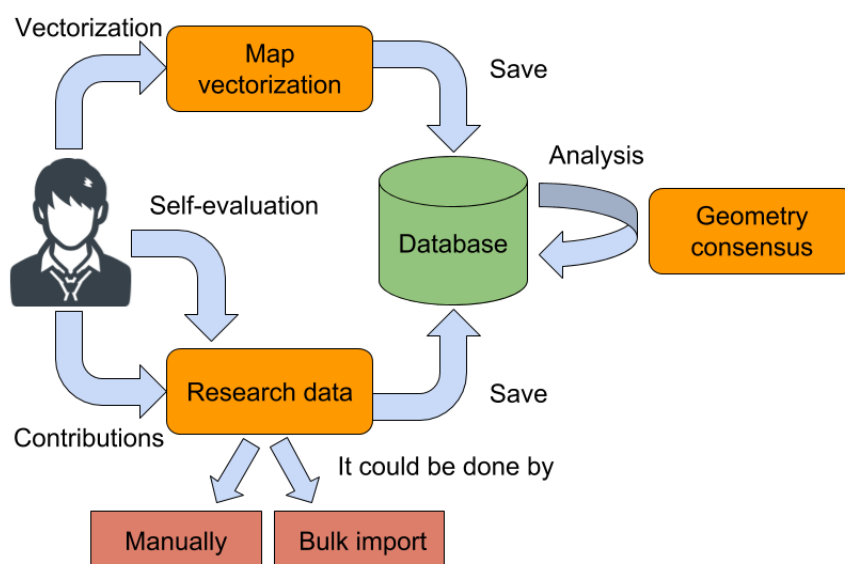


Figura 1. Processo de coleta de dados

Para que as contribuições feitas via vetorização manual obtenha o melhor resultado possível, são recomendadas algumas práticas durante o processo [Mooney et al. 2016]: (1) uso de um mouse ao invés de um touchpad, pois auxilia na precisão do posicionamento dos dados; (2) verificar o nível de detalhe (zoom) da imagem do mapa no navegador para a melhor coleta dos dados e (3) deve ser realizado de acordo com as especificações indicadas no modelo de dados e nas características dos dados.

A importação em massa se refere à integração de dados na plataforma, em grandes quantidades. Nesse projeto, a importação em massa será permitida para dados textuais, onde (1) os dados necessitam estar em formato CSV. Se eles não estiverem, é necessário convertê-los; (2) o usuário deve enviar um e-mail para import_pauliceia@googlegroups.com para discutir com os administradores sobre o plano dos dados. Enviar o nome do projeto em questão e uma amostra dos dados (pelo menos duzentos registros, se tiver) junto. Não será possível carregar os dados, sem antes os administradores

fazerem uma revisão do documento e dos dados e (3) se os administradores concordarem e for dada a permissão para importação em massa de dados, o contribuinte deve seguir com o plano de trabalho, fornecendo atualização frequente no mesmo e-mail.

4.4. Controle de qualidade

Serão utilizados indicadores de qualidade, em vez das medidas de qualidade (como explicado na seção 4). Isto será feito, pois para que segundo seja aplicável, necessita de um conjunto de dados de referência; o que não há no momento, para o contexto do projeto Pauliceia.

É necessário que os contribuintes leiam este protocolo, antes de começarem a coleta de dados. Isto colabora para a sanção de dúvidas, evitando que problemas possam acontecer, aumentando a qualidade dos dados inseridos.

Será utilizado o conceito de projeto e revisões nas contribuições feitas, como explicado na seção 4.2. Como somente usuários pertencentes a um projeto (pessoas confiáveis) podem manipular os dados, isto gerará uma qualidade local nas contribuições. Caso usuários externos encontrem dados de má qualidade, eles poderão adicionar comentários ou solicitar revisão na página de “Revisões”.

Os usuários devem fazer uma auto-avaliação dos dados que eles inseriram, uma revisão dos dados antes de serem enviados ao servidor. Esta revisão será principalmente em relação aos seus dados que acabou de contribuir. Eles revisarão se os dados estão coerentes, se tem qualidade adequada, verificando se a geometria e seus atributos estão corretos, de acordo com as especificações e baseados no seu conhecimento pessoal da área de pesquisa.

Esta análise é manual, o usuário verificará se a geometria e os atributos estão corretos. Este tipo de observação para o projeto Pauliceia é importante, pois público alvo do portal são os historiadores, que tem um conhecimento prévio do assunto, logo eles estão capacitados a corrigir eventuais problemas. Isto será feito aplicando a Lei de Linus¹⁴, que diz que, quanto mais voluntários contribuírem, maior será a qualidade. Goodchild e Li [Goodchild and Li 2012] indicam o uso dessa lei para fazer com que os dados gerados converjam para verdade. Haklay et al. [Haklay et al. 2010] estudaram a aplicação dela e concluíram que realmente, quanto mais usuários contribuindo, maior é a probabilidade dos dados terem qualidade. Além de que, foi concluído isto com base nos dados do OSM, onde geralmente os usuários são normais (comumente não tem estudo aprofundado do dado inserido, só conhecimento local). No caso do projeto Pauliceia, como dito, os voluntários geralmente terão conhecimento do assunto, por conta disso, possivelmente os dados terão ainda mais qualidade.

Após a coleta de dados e avaliação manual de seus dados, é aconselhável que o voluntário faça uma verificação nas contribuições de outros usuários. Isto é explicado em detalhes na seção 4.5, sendo uma abordagem para garantir a qualidade dos dados, conhecida como crowdsourcing, como descrito na seção 2.1.

Em relação aos dados vetoriais contribuídos nos HistMapathons, onde vários usuários poderão vetorizar os mesmos objetos, tem-se a intenção de utilizar um algoritmo (como o proposto por Budig et al. [Budig et al. 2016]) para a recuperação da me-

¹⁴“Given enough eyeballs, all bugs are shallow” [Raymond 2017]

lhor representação poligonal ou linear (melhor consenso), dado várias geometrias de um mesmo objeto.

Baseado nas recomendações de Cechanowicz et al. [Cechanowicz et al. 2013] e Hamari et al. [Hamari et al. 2014], no contexto do projeto Pauliceia, será utilizado algumas técnicas de Gamificação para instigar os usuários na qualidade dos dados. Uma das técnicas de Gamificação será é o de reputação, através de Ranking. Goodchild e Li [Goodchild and Li 2012] indicam como abordagem social o uso de um sistema de reputação para avaliar a confiabilidade dos usuários. No contexto do projeto, o voluntário começará com uma quantidade de “pontos” iguais à zero, podendo ganhar ou perder esses pontos ao decorrer da utilização do sistema. Isto instigará o usuário a contribuir mais e com mais qualidade, pois se contribuir com frequência, fazer comentários e receber boas votações em seus dados, seu Ranking aumentará. Se as contribuições forem de má qualidade, será recebido votações negativas, portanto seu Ranking diminuirá. A aquisição de pontos será feita através do esquema de votação nos dados e obtenção de emblemas. Um voluntário poderá votar nas contribuições de outros usuários. Haverá votação positiva ou negativa, onde o usuário votará dependendo da qualidade daquele dado, baseado em seu conhecimento. A quantidade de votações feitas pode interferir no seu Ranking. A reputação dos contribuintes será dada pelo seu Ranking em relação aos dos demais. Revelar os Rankings para a comunidade incentivará os usuários a terem mais cuidado (atenção) à qualidade de suas contribuições.

Por questões de segurança, as ações principais do usuário (como manipular dados históricos) serão salvos em históricos (arquivos de log). Isto poderá ajudar a rastrear, com mais facilidade, usuários mal intencionados que queiram danificar o sistema. Isto ajuda a definir a confiabilidade e reputação dos contribuintes [Davis Jr et al. 2013]. Para cada contribuição haverá um histórico relacionado também, podendo-se acompanhar a evolução dos dados e voltar uma correção no passado, caso necessário.

4.5. Última verificação da contribuição

A partir desse momento as contribuições estão disponíveis para a comunidade do projeto. O usuário deve fazer uma verificação final nos dados que acabou de enviar. Analisar a qualidade dos dados de acordo com a coerência do projeto e aos demais dados dos outros contribuidores. Caso os erros sejam detectados o usuário deve editá-los. Esta operação aplica-se tanto aos dados que o próprio usuário coletou quanto aos coletados por outros.

Este envolvimento das pessoas para validar e corrigir erros cometidos por outros usuários, é uma abordagem conhecida como crowdsourcing. É um processo para assegurar a qualidade do VGI de acordo com Goodchild e Li [Goodchild and Li 2012].

4.6. Feedback para a comunidade

Um projeto colaborativo melhora à medida que mais usuários contribuem para isso, portanto é interessante o usuário fornecer um feedback sobre sua experiência. O contribuinte pode expressar seus comentários, opiniões e observações nos canais disponíveis do projeto, como: lista de discussão¹⁵, grupo no Facebook¹⁶ e página no Facebook¹⁷.

¹⁵pauliceia@googlegroups.com

¹⁶<https://www.facebook.com/groups/pauliceia2.0/>

¹⁷<fb.me/pauliceia2.0>

O contribuinte é encorajado a descrever se o processo de coleta de dados foi fácil ou difícil, e o porquê. Escrevendo os problemas ou situações inesperadas que encontrou, sugerindo melhorias ou mudanças. Explicando precisamente o que aconteceu, para facilitar a compreensão dos administradores e aplicar uma correção adequada.

Os voluntários poderão receber notificações dos dados de projetos que faz parte, comentários e revisões de suas contribuições.

É aconselhável que o usuário divulgue o projeto para pessoas que conheça, com intenção de atrair novos contribuintes. Isto fará com que a plataforma cresça e melhore com frequência. Quanto mais houver participantes, mais o sistema se tornará rico em seus dados, conseqüentemente melhorando a qualidade deles.

5. Conclusão

VGI gera uma grande quantidade de dados com características heterogêneas, tornando a comparações tradicionais já não tão viáveis. Por conta disso, Senaratne et al [Senaratne et al. 2017] descrevem a utilização de indicadores de qualidade, quando não se há dados autoritários para comparação, como: confiabilidade, experiência, credibilidade, qualidade do conteúdo de texto, imprecisão, conhecimento local, reconhecimento, reputação. Goodchild e Li [Goodchild and Li 2012] propõem ainda três abordagens para garantir a qualidade do VGI: crowdsourced, sociais e geográficas.

Os indicadores e abordagens de qualidade serão utilizados para avaliar a qualidade dos dados do projeto Pauliceia. O primeiro deles é a definição de um protocolo, descrevendo os tópicos que o voluntário poderá fazer e como deve seguir. Isto fornece uma padronização para que se evite a heterogeneidade dos dados, acarretando uma melhora de qualidade.

A especificação de um protocolo bem estruturado para o projeto Pauliceia dá a oportunidade de ajudar a melhorar a qualidade dos dados que serão coletados na plataforma final. Sem contar que, contribui para a definição das funcionalidades do Serviço Web para VGI.

Na inicialização é descrito os pontos iniciais que o usuário deve fazer, como: (1) criar ou acessar o portal com um login, para se ter controle; (2) ler e aceitar a Política de Uso do projeto; (3) ler a especificação do projeto antes de começar a coleta de dados, para sanar as dúvidas; (4) entrar em contato com o suporte em caso de problemas; (5) antes do início da coleta de dados, deve-se fazer um treinamento em um ambiente de teste e (6) ter certeza que serão inseridos dados de domínio público.

No modelo de dados são explicadas algumas características dos dados, como: (1) criação de um projeto para controle local da área de pesquisa; (2) caso usuário fora do projeto encontre erro nos dados, pode adicionar comentários ou solicitar revisão; (3) uso do WGS84 como CRS dos dados e (4) descrição dos formatos das contribuições, evitando dados diferentes.

No métodos de coleta de dados descreve como a coleta de dados deve ser feita. No caso elas ocorrerão por HistMapathons ou contribuições individuais. Um HistMapathon é uma maratona para vetorização de mapas antigos e as contribuições individuais poderão ser feitas manualmente ou por importação em massa. Manualmente o colaborador inserirá

diretamente a informação no mapa e importação é quando um pesquisador faz o upload de um arquivo com inúmeros registros.

No controle de qualidade o voluntário fará uma auto-avaliação com base no conhecimento da área, aplicando-se a Lei de Linus. Em relação a vetorização dos mapas antigos, será utilizado um algoritmo para encontrar o melhor consenso dentre os mesmos objetos vetorizados. Técnicas de Gamificação serão usadas para instigar as pessoas, como um sistema de reputação baseado em Ranking e votação das contribuições. Todas manipulações de dados feitas pelos usuários e todas as contribuições gerarão históricos (arquivos de logs) por segurança.

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.

É visível que não existe um único método que possa resolver todos os problemas relacionados à qualidade das contribuições de VGI. Isto se dá essencialmente pela característica heterogênea que são esses dados. Por conta disso, temos que ter em mente que existem várias técnicas para resolver diversos tipos de problemas, havendo também suas limitações. Trabalhar em cima dessas limitações, pode, por exemplo, melhorar os métodos já existentes, aumentando a contribuição científica em estudos sobre VGI [Senaratne et al. 2017].

As questões principais discutidas neste trabalho estão sendo incorporadas no serviço web para VGI do projeto Pauliceia na forma de funções bem definidas. As funcionalidades já existentes são de autenticação de usuário e gerenciamento de geometrias. Fica-se então para trabalhos futuros: detalhar mais o protocolo do projeto, com o intuito de melhorá-lo; a implementação da Gamificação e construção do algoritmo para achar o melhor consenso, dado várias geometrias.

6. Agradecimento

Agradeço à FAPESP pela concessão da bolsa, referente ao processo nº 2017/03852-9, Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

Agradeço também ao Dr. Antônio M. V. Monteiro e ao Dr. Sergio Rosim pela oportunidade de fazer este trabalho na matéria e supervisão geral. Obrigado a Yasmin Wassef pela revisão do texto.

Referências

- Antoniou, V. and Skopeliti, A. (2015). Measures and indicators of vgi quality: An overview. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*.
- Budig, B., van Dijk, T. C., Feitsch, F., and Arteaga, M. G. (2016). Polygon consensus: smart crowdsourcing for extracting building footprints from historical maps. In *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, page 66. ACM.
- Carvalho, D. F. (2017). Café, ferrovias e crescimento populacional: o florescimento da região noroeste paulista. <http://www.historica.arquivoestado.sp>.

gov.br/materias/anteriores/edicao27/materia02/. Accessed on 04/04/2017.

- Cechanowicz, J., Gutwin, C., Brownell, B., and Goodfellow, L. (2013). Effects of gamification on participation and data quality in a real-world market research domain. In *Proceedings of the First International Conference on Gameful Design, Research, and Applications*, pages 58–65. ACM.
- Ciepluch, B., Jacob, R., Mooney, P., and Winstanley, A. C. (2010). Comparison of the accuracy of openstreetmap for ireland with google maps and bing maps. In *Proceedings of the Ninth International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences 20-23rd July 2010*, page 337. University of Leicester.
- Davis Jr, C. A., de Souza Vellozo, H., and Pinheiro, M. B. (2013). A framework for web and mobile volunteered geographic information applications. In *GeoInfo*, pages 147–157.
- Estellés-Arolas, E. and González-Ladrón-de Guevara, F. (2012). Towards an integrated crowdsourcing definition. *Journal of Information science*, 38(2):189–200.
- Flanagin, A. J. and Metzger, M. J. (2008). The credibility of volunteered geographic information. *GeoJournal*, 72(3-4):137–148.
- Girres, J.-F. and Touya, G. (2010). Quality assessment of the french openstreetmap dataset. *Transactions in GIS*, 14(4):435–459.
- Goodchild, M. F. (2007a). Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(4):211–221.
- Goodchild, M. F. (2007b). Citizens as sensors: web 2.0 and the volunteering of geographic information. *GeoFocus. Revista Internacional de Ciencia y Tecnología de la Información Geográfica*, (7):8–10.
- Goodchild, M. F. and Li, L. (2012). Assuring the quality of volunteered geographic information. *Spatial statistics*, 1:110–120.
- Haklay, M. (2010). How good is volunteered geographical information? a comparative study of openstreetmap and ordnance survey datasets. *Environment and planning B: Planning and design*, 37(4):682–703.
- Haklay, M., Basiouka, S., Antoniou, V., and Ather, A. (2010). How many volunteers does it take to map an area well? the validity of linus’ law to volunteered geographic information. *The Cartographic Journal*, 47(4):315–322.
- Hamari, J., Koivisto, J., and Sarsa, H. (2014). Does gamification work?—a literature review of empirical studies on gamification. In *System Sciences (HICSS), 2014 47th Hawaii International Conference on*, pages 3025–3034. IEEE.
- Hollenstein, L. and Purves, R. (2010). Exploring place through user-generated content: Using flickr tags to describe city cores. *Journal of Spatial Information Science*, 2010(1):21–48.
- Jakobsson, A. and Giversen, J. (2007). Guidelines for implementing the iso 19100 geographic information quality standards in national mapping and cadastral agencies. *Eurogeographics Expert Group on Quality*.

Montello, D. R., Goodchild, M. F., Gottsegen, J., and Fohl, P. (2003). Where's downtown?: Behavioral methods for determining referents of vague spatial queries. *Spatial Cognition & Computation*, 3(2-3):185–204.

Mooney, P., Minghini, M., Laakso, M., Antoniou, V., Olteanu-Raimond, A.-M., and Skopeliti, A. (2016). Towards a protocol for the collection of vgi vector data. *ISPRS International Journal of Geo-Information*, 5(11):217.

Mota, P. d. B. et al. (2007). A cidade de são paulo de 1870 a 1930: café, imigrantes, ferrovia, indústria.

OpenStreetMap (2017a). About openstreetmap. http://wiki.openstreetmap.org/wiki/About_OpenStreetMap. Accessed on 01/09/2017.

OpenStreetMap (2017b). Mapathon. <http://wiki.openstreetmap.org/wiki/Mapathon>. Accessed on 18/09/2017.

OpenStreetMap (2017c). Quality assurance. http://wiki.openstreetmap.org/wiki/Quality_assurance. Accessed on 25/09/2017.

Over, M., Schilling, A., Neubauer, S., and Zipf, A. (2010). Generating web-based 3d city models from openstreetmap: The current situation in germany. *Computers, Environment and Urban Systems*, 34(6):496–507.

Raymond, E. S. (2017). Release early, release often. <http://www.catb.org/~esr/writings/cathedral-bazaar/cathedral-bazaar/ar01s04.html>. Accessed on 01/09/2017.

Secretaria Municipal de Urbanismo e Licenciamento (2017). História demográfica do município de são paulo. http://smul.prefeitura.sp.gov.br/historico_demografico/tabelas/pop_brasil.php. Accessed on 04/04/2017.

See, L., Mooney, P., Foody, G., Bastin, L., Comber, A., Estima, J., Fritz, S., Kerle, N., Jiang, B., Laakso, M., et al. (2016). Crowdsourcing, citizen science or volunteered geographic information? the current state of crowdsourced geographic information. *ISPRS International Journal of Geo-Information*, 5(5):55.

Senaratne, H., Mobasher, A., Ali, A. L., Capineri, C., and Haklay, M. (2017). A review of volunteered geographic information quality assessment methods. *International Journal of Geographical Information Science*, 31(1):139–167.

Spatial Reference (2017). Epsg:4326. <http://spatialreference.org/ref/epsg/wgs-84/>. Accessed on 04/08/2017.

TC, I. (2009). Iso/tc 211 geographic information/geomatics.

Tech2 (2017). Why is google's mapathon in hot waters in india? all you need to know. <http://www.firstpost.com/tech/news-analysis/why-is-googles-mapathon-in-hot-waters-in-india-all-you-need-to-know->html. Accessed on 18/09/2017.

Tobler, W. (2004). On the first law of geography: A reply. *Annals of the Association of American Geographers*, 94(2):304–310.