

ÍNDICE DE DADOS ABERTOS PARA CIDADES

SÃO PAULO - SP



FGV DAPP



ÍNDICE DE DADOS ABERTOS PARA CIDADES

SÃO PAULO - SP

ABRIL | 2017





Instituição de caráter técnico-científico, educativo e filantrópico, criada em 20 de dezembro de 1944 como pessoa jurídica de direito privado, tem por finalidade atuar, de forma ampla, em todas as matérias de caráter científico, com ênfase no campo das ciências sociais: administração, direito e economia, contribuindo para o desenvolvimento econômico-social do país.

Escritório

Praia de Botafogo 190, Rio de Janeiro
RJ - CEP 222509000
Caixa Postal 62.591 CEP 22257-970
Tel (21) 3799-5498 | www.fgv.br

Presidente Fundador

Luiz Simões Lopes

Presidente

Carlos Ivan Simonsen Leal

Vice-Presidentes

Sergio Franklin Quintella,
Francisco Oswaldo Neves Dornelles e
Marcos Cintra Cavalcante de Albuquerque



Diretor

Marco Aurelio Ruediger

Diretor-Executivo

Ariel Kogan

DAPP

(21) 3799-4300
www.dapp.fgv.br | dapp@fgv.br

Equipe

Comunidade OKBr

EXPEDIENTE

EQUIPE DE EXECUÇÃO

Pesquisadores

Amaro Grassi
Andressa Falconieri
Bárbara Barbosa
Janaina Fernandes
Wagner Oliveira

Projeto gráfico

Arielle Asensi
Humberto Ferreira
Luís Gomes
Rebeca Liberatori Braga



SUMÁRIO

- 04 SOBRE A FGV/DAPP
- 04 SOBRE A OPEN KNOWLEDGE BRASIL
- 05 SOBRE A PARCERIA FGV/DAPP E OKBR
- 06 O ÍNDICE DE DADOS ABERTOS PARA CIDADES
- 08 RESULTADOS
- 14 SOBRE AS BASES DE DADOS
- 23 CONCLUSÃO
- 24 FONTES

SOBRE A FGV/DAPP

A Diretoria de Análise de Políticas Públicas da Fundação Getúlio Vargas (FGV/DAPP) é um centro de pesquisa social aplicada voltado à inovação para políticas públicas, produzindo análise de ponta com uso intensivo de redes sociais e conhecimento interdisciplinar. Tem como missão aprimorar a gestão pública brasileira e qualificar o debate público na sociedade em rede, por meio da transparência e do diálogo entre o Estado e a cidadania.

A FGV/DAPP desenvolve uma agenda de pesquisa a partir de metodologia própria de análise, aprimorada desde a sua criação, em 2012, e que reúne métodos quantitativos e qualitativos tradicionais com recursos inovadores de processamento e análise de bancos de dados públicos e de redes sociais. E disponibiliza à sociedade, afinal, um conjunto de ferramentas de visualização e análise de dados de fácil compreensão e acesso.

As análises de políticas públicas são produzidas por uma equipe interdisciplinar e diversificada, de formação em áreas como Sociologia, Ciência Política, Antropologia, Linguística, Economia, Administração Pública, Relações Internacionais, Estatística, Matemática e Comunicação Social, aliadas à Tecnologia da Informação e ao Design.

SOBRE A OPEN KNOWLEDGE BRASIL

A Open Knowledge Brasil (OKBr) é uma Organização da Sociedade Civil (OSC) sem fins lucrativos e apartidária, fundada em 2013. A OKBr representa no Brasil a Open Knowledge Internacional a partir de um acordo firmado também em 2013.

A OKBr utiliza e desenvolve ferramentas cívicas, faz análises de políticas públicas, trabalha com jornalismo de dados e promove o conhecimento livre para tornar a relação entre governo e sociedade mais transparente e para que haja uma participação política mais efetiva e aberta.

A OKBr acredita em um mundo onde as informações estão livres e disponíveis para permitir decisões e escolhas mais conscientes e inteligentes. Um mundo onde a informação e o conhecimento estão ao alcance de todos.

SOBRE A PARCERIA FGV/DAPP E OKBR

Um dos principais eixos de atuação da FGV/DAPP é a promoção da Transparência, entendida não apenas como a divulgação de dados públicos, mas como a capacidade de facilitar o entendimento das informações provenientes desses dados. A partir da integração do uso de tecnologia e design, a FGV/DAPP divulga informações de maneira simples e acessível sobre diversos temas, como o destino dos recursos públicos, a origem das doações de campanha eleitoral, a situação dos serviços públicos (educação, saúde, transporte, segurança), entre outros.

Tendo em vista seu papel de destaque na promoção dessa agenda, a FGV/DAPP firmou na segunda metade de 2016 uma parceria com a Open Knowledge Foundation no Brasil (OKBr) com o objetivo de contribuir para o levantamento anual das informações do Índice de Dados Abertos nacional e para detalhar e aprimorar a metodologia do índice aplicado em nível subnacional – cidades.

Trata-se de um projeto pioneiro na promoção da transparência nos municípios brasileiros, uma vez que o índice pode ser utilizado como ferramenta de avaliação e identificação de gargalos, de forma a orientar os municípios em relação ao aprimoramento de suas políticas de dados abertos. Em última instância, a parceria entre OKBr e FGV/DAPP visa a contribuir para que os municípios disponham de ferramentas capazes de aumentar a eficácia de suas políticas e sua capacidade de resposta e diálogo com os cidadãos.

O ÍNDICE DE DADOS ABERTOS PARA CIDADES

Dados são considerados abertos quando qualquer pessoa é livre para acessá-los, usá-los, modificá-los e compartilhá-los, sujeitos, no máximo, a medidas que preservem sua origem e sua publicidade¹. A Open Knowledge Foundation desenvolveu o Índice de Dados Abertos (ODI) global para comparar países em termos da sua capacidade de abrir dados aos cidadãos, à mídia e à sociedade civil. As informações que compõem o índice são levantadas por meio de uma plataforma crowd-sourced, ou seja, a partir das contribuições de agentes locais ligados à rede da OKFN em cada país, e depois revisadas por especialistas em dados abertos em cada país, gerando como resultado final um ranking de países.

O objetivo do índice é avaliar o estado da política de dados abertos de cada país, levando em consideração todas as suas características: o tipo de dado que é divulgado, os formatos, a facilidade de acesso e a transformação dos dados em informação, entre outras. O índice avalia diversas dimensões, tais como finanças públicas, dados socioeconômicos, legislativos e eleitorais, serviços públicos, informações geolocalizadas e indicadores ambientais.

O índice oferece, assim, um parâmetro de referência sobre a capacidade de fornecer dados abertos para todos os países, apresentando tal informação de forma clara e de fácil entendimento e usabilidade, algo que está em consonância com a missão de Transparência da FGV/DAPP.

1 Veja a “Open Definition” para mais detalhes em: <http://opendefinition.org/>

Este índice já existe para os anos de 2013, 2014 e 2015 e agora está sendo construído para o ano de 2016, usando uma metodologia já aprimorada pela experiência acumulada ao longo desse período. A FGV/DAPP, por meio da parceria firmada com a Open Knowledge do Brasil (OKBr), foi a entidade brasileira responsável pelo levantamento das informações que compõem o índice nacional nesta edição.

No entanto, além do índice global, que compara países, a Open Knowledge partiu para um esforço de replicação do ODI para a esfera subnacional, em especial, para cidades. Em sua ramificação brasileira, a OKBr contou com o apoio da FGV/DAPP para estruturação de uma metodologia pioneira de aplicação do índice compatível com a realidade das cidades brasileiras. Dentre as contribuições da FGV/DAPP, destaca-se a adequação das dimensões já existentes do índice para cidades e a inclusão de novas dimensões importantes para o estado das cidades brasileiras: transporte, criminalidade e educação.

Por se tratar de um projeto-piloto, o índice será apresentado inicialmente para três cidades brasileiras: Rio de Janeiro, São Paulo e Curitiba. Os resultados do índice podem ser consultados na página <http://br-cities.survey.okfn.org/place/sp>.

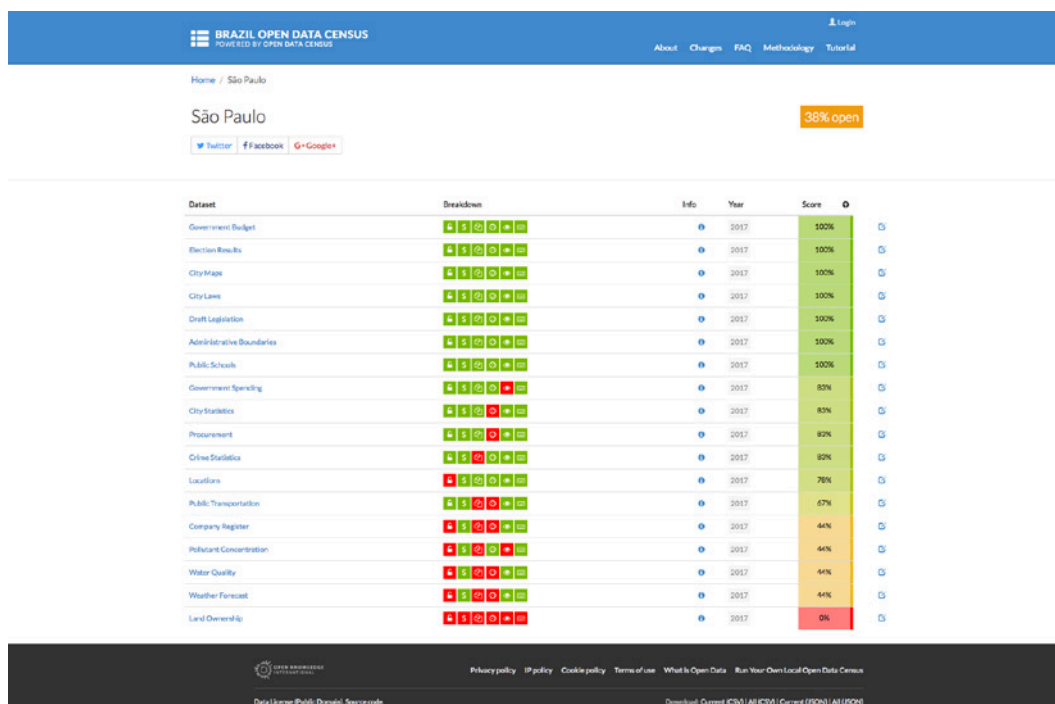
A próxima seção traz resultados baseados no levantamento realizado pela FGV/DAPP para a cidade de São Paulo. Tais insumos podem ser úteis para a elaboração de diretrizes para uma política de dados abertos no âmbito da cidade. Em seguida, cada dimensão do índice será detalhada de modo a ressaltar as fontes utilizadas, as boas práticas encontradas e os principais pontos de atenção.

RESULTADOS

Com base no levantamento realizado pela FGV/DAPP, é possível apresentar algumas considerações sobre a situação dos dados abertos no município de São Paulo. Em primeiro lugar, os resultados do próprio índice sinalizam que São Paulo apresenta um escore de 75%, o que pode ser considerado muito positivo. Ainda que não seja possível realizar uma comparação tecnicamente nivelada², se tomarmos por base o levantamento feito, não para outras cidades, mas para países, usando metodologia semelhante, apenas dois entre 94 países avaliados apresentam índice superior a este.

Dentro de suas dimensões de análise, como mostra a figura abaixo, sete dos 18 datasets avaliados obtiveram nota máxima, ou seja, 38% do total, percentual que aparece na figura abaixo do lado direito (box amarelo). Apenas uma dimensão obteve nota zero.

FIGURA 1 - AVALIAÇÃO DAS DIMENSÕES DE TRANSPARÊNCIA EM SÃO PAULO³



2 No momento de realização deste relatório o levantamento das outras cidades ainda estava sendo revisado, o que impede a comparação com os resultados no mesmo subnível.

3 Os nomes das dimensões do índice foram traduzidos para o português pelos autores deste documento, levando em consideração a capacidade explicativa que o nome tem em relação ao conteúdo da dimensão. Ou seja, não são traduções literais em grande parte dos casos. Seguem os nomes adotados em português na ordem apresentada na figura 1: Orçamento Público, Resultados Eleitorais, Mapas da Cidade, Leis em Vigor, Atividade Legislativa, Limites Administrativos, Escolas Públicas, Gastos Públicos, Estatísticas Socioeconômicas, Compras Públicas, Estatísticas do Crime, Localizações, Transporte Público, Registro de Empresas, Concentração de Poluentes, Qualidade da Água, Previsão do Tempo e Propriedade da Terra.

Para analisar o estado dos dados abertos para a cidade de São Paulo com ano-base em 2016, adotamos uma metodologia de contagem dos problemas encontrados em cada dimensão e categorizamos os problemas encontrados em dois tipos: usabilidade e processo. Na primeira categoria, estão listados problemas relacionados à dificuldade de uso do dado em si e à capacidade de transformação deste em informação. Como questões de processo, identificamos: dificuldades de se acessar o dado e omissão/inadequação da licença. A tabela abaixo apresenta a síntese de problemas em cada dimensão, ordenadas da mais para a menos urgente.

FIGURA 2: CATEGORIZAÇÃO DOS PROBLEMAS EM USABILIDADE E PROCESSO

USABILIDADE	PROCESSO
Dataset incompleto	Restrição de acesso
Dataset desatualizado	Dificuldade de localizar dados
Indisponibilidade de formato aberto	Download da base completa indisponível
Dificuldade de trabalhar dados	Licença não transparente

Em geral, encontramos proporcionalmente mais problemas de usabilidade do que de processo na prefeitura de São Paulo. A frequência de problemas de processo é maior nas dimensões de Registro de Empresas, Concentração de Poluentes, Previsão do Tempo e Estatísticas de Crime. A frequência de problemas de usabilidade é maior em Qualidade da Água, Previsão do Tempo e Registro de Empresas.

Dentre as 18 dimensões do índice, três não apresentaram qualquer gargalo, seja em relação à usabilidade, seja em relação ao seu processo de divulgação: Mapas da Cidade, Leis em Vigor, Limites Administrativos. Tais dimensões constituem exemplos de boas práticas que podem ser replicadas para as demais dimensões. Destacam-se como boas práticas destes datasets:

API (Application Programming Interface)⁴ para disponibilizar dados de leis em vigor (Câmara Municipal), com possibilidade de conectar à totalidade dos dados disponibilizados por meio de webservice;



Plataforma centralizada de informações geolocalizadas (mapas básicos da cidade e limites administrativos), com download disponível em formatos próprios (shapefiles);

Nestes casos obtemos bases completas, atualizadas, em formatos abertos, fáceis de encontrar e facilmente traduzidas em análises e com licenças de uso transparentes.

No extremo oposto estão as dimensões de Registro de Empresas, Qualidade da Água e Previsão do Tempo, que apresentaram problemas em quase todos os critérios da avaliação⁵. Além disso, o dataset de Propriedade da Terra não foi encontrado publicamente, o que impossibilitou a sua avaliação e o coloca como principal ponto crítico do estudo. A tabela abaixo apresenta a frequência de problemas identificados por dimensão do índice. São apresentadas também informações a respeito do ente administrativo que fornece a informação (poder e nível federativo).

4 Trata-se de um conjunto de rotinas acessíveis apenas por programação para que um usuário possa acessar as informações de um software de forma direta, sem necessariamente navegar através de sua interface tradicional.

5 São os critérios de avaliação de problemas: Dataset incompleto, Desatualizado, Indisponibilidade de formato aberto, Dificuldade de trabalhar os dados, Acesso restrito, Dificuldade de localizar os dados, Download da base completa indisponível e Licença não Transparente.

FIGURA 3 - RANKING DE DIMENSÕES DE DADOS ABERTOS PARA SÃO PAULO

RANKING	DIMENSÕES DO ÍNDICE	GARGALOS ENCONTRADOS	PODER RESPONSÁVEL	NÍVEL FEDERATIVO
1	Mapas da cidade	0	Executivo	Municipal
	Leis em vigor	0	Legislativo	Municipal
	Limites administrativos	0	Executivo	Municipal
2	Compras Públicas	1	Executivo	Municipal
	Atividade Legislativa	1	Legislativo	Municipal
	Orçamento Público	1	Executivo	Municipal
3	Escolas Públicas	2	Executivo	Federal
	Resultados Eleitorais	2	Judiciário	Federal
	Estatísticas Sócioeconômicas	2	Executivo	Federal
	Localizações	2	Executivo	Estadual
4	Gastos Públicos	3	Executivo	Municipal
5	Estatísticas do Crime	4	Executivo	Estadual
	Transporte Público	4	Executivo	Municipal
6	Concentração de Poluentes	5	Executivo	Estadual
7	Qualidade da Água	6	Executivo	Municipal
8	Previsão do Tempo	7	Executivo	Municipal
	Registro de Empresas	7	Executivo	Estadual
9	Propriedade da Terra	NA	Executivo	Municipal

O problema mais comum entre os bancos de dados é a dificuldade em trabalhar os dados, ou seja, na usabilidade dos mesmos para realizar análises e tirar informações. Esse critério se refere à condição geral de facilidade ou dificuldade de realização de análise a partir dos dados divulgados, o que deve depender de fatores como granularidade, organização do site e da base, entre outros. Em segundo lugar, vem a falta de transparência em relação ao tipo de licença dos dados divulgados, o que é um problema comum na realidade brasileira (veja a tabela abaixo para entender a frequência com que cada problema aparece). Essa informação é relevante porque permite ao cidadão e usuário do portal entender seus direitos em relação à forma com que pode manipular, utilizar e divulgar análises realizadas a partir de determinada base. Na outra ponta, está a restrição de acesso, encontrada em apenas dois datasets.

FIGURA 4 - FREQUÊNCIA DE PROBLEMAS ENCONTRADOS EM SÃO PAULO

RANKING	GARGALOS MAPEADOS	FREQUÊNCIA DE IDENTIFICAÇÃO DOS GARGALOS
1	Dificuldade de trabalhar dados	10
2	Licença não transparente	8
3	Download da base completa indisponível	6
	Desatualizado	6
	Dataset incompleto	6
	Dificuldade de localizar dados	6
4	Indisponibilidade de formato aberto	3
5	Acesso restrito	2

Em seis das 18 dimensões, identificou-se um problema considerado particularmente crítico na dimensão da usabilidade: a incompletude da base de dados (de acordo com os requisitos da metodologia). Na dimensão de processo, o problema caracterizado como muito crítico no estudo é a restrição do acesso (mediante necessidade de identificação), identificado apenas em relação aos datasets de Gastos Públicos e Concentração de Poluentes. Abaixo, seguem algumas prioridades de ação, com destaque para as dimensões sob competência do executivo municipal. O dataset de Propriedade da Terra é o que chama mais atenção, pois não foi encontrado para realizar a avaliação.

FIGURA 5 - PRIORIDADES DE AÇÃO

DATASETS	ACESSO RESTRITO	DATASET INCOMPLETO
Propriedade da Terra	NA	NA
Registro de Empresas		×
Qualidade da Água		×
Previsão do Tempo		×
Transporte Público		×
Gastos Públicos	×	
Concentração de Poluentes	×	×
Escolas Públicas		×

● Disponibilizado pela Prefeitura

● Disponibilizado por outros entes

Em síntese, essas informações podem ser úteis para subsidiar uma política de dados abertos no âmbito da cidade de São Paulo, pois fornecem os caminhos para replicação das boas práticas e correção dos pontos de atenção. Os benefícios de uma política de dados abertos são inúmeros e passam pela ampliação da eficiência da gestão, da criação de um instrumento de cobrança de resultados da administração pública, do fomento da accountability e do controle social, do engajamento da sociedade civil com a gestão pública e da melhora da imagem do ente público, com potencial de torná-lo referência internacional.

SOBRE AS BASES DE DADOS

Esta seção apresenta comentários a respeito dos conjuntos de dados analisados a fim de dar profundidade ao índice, explicando as questões particulares de cada dimensão analisada.



Orçamento Público (Nota: 100%)

Problemas encontrados - Processo

Falta de clareza sobre a licença pública de uso de dados

Diferentemente do dataset anterior, o orçamento público foi avaliado com base no Portal da Transparência do município de São Paulo, considerado a melhor fonte dessa informação. Todos os requisitos da metodologia foram cumpridos, mas dois detalhes merecem atenção: a licença não foi encontrada no site, mas esses dados são considerados, em geral, públicos pela Lei de Acesso à Informação. Neste caso, seria importante ter uma licença específica e transparente.



Resultados Eleitorais (Nota: 100%)

Problemas encontrados - Usabilidade

Dificuldades na usabilidade dos dados

Problemas encontrados - Processo

Falta de clareza sobre a licença pública de uso de dados

As informações a respeito dos resultados eleitorais no Brasil são consolidadas em nível federal pelo Tribunal Superior Eleitoral - TSE. Ainda que tribunais regionais levantem e disponibilizem tais informações em alguns casos, o repositório de dados eleitorais do TSE é uma fonte consolidada e confiável de informação. Neste caso, os dados do TSE foram avaliados a respeito das eleições municipais (prefeito e vereador) do município de São Paulo. No que diz respeito à metodologia do índice e ao preenchimento do survey, todos os pré-requisitos foram preenchidos. No entanto, dois problemas foram levantados: o primeiro refere-se à dificuldade de trabalhar a partir dos dados, uma vez que são codificados. A documentação facilita o estudo, mas é necessário tempo para se

familiarizar com a natureza da informação. O segundo problema se refere à licença. Não há, no website acessado, informações sobre licença ou termos de uso, mas os dados estão amparados pela legislação federal de acesso à informação, como informado pelos técnicos do TSE. Ou seja, ainda que os dados tenham licença pública, é importante que ela esteja disponibilizada de forma mais clara juntamente aos dados.



Mapas da Cidade (Nota: 100%)

Não foram encontrados problemas neste dataset

As informações de mapas da cidade estão disponíveis no portal GeoSampa hospedado pela SMUL - Secretaria Municipal de Urbanismo e Licenciamento da Prefeitura de São Paulo (antes denominada SMDU - Secretaria Municipal de Desenvolvimento Urbano). O mapa vem no formato aberto (shapefile e dbf) e pode ser plotado em escalas diferentes. No menu de opções é possível selecionar dados quanto à relevo, rotas de trânsito, bacias hidrográficas e limites administrativos. Neste caso obtemos bases completas, atualizadas, em formatos abertos, com clareza nos metadados e licenças de uso, fáceis de encontrar e transformar em análises. A consolidação das informações georreferenciadas em um portal único é considerada uma boa prática de dados abertos da cidade de São Paulo.



Leis em Vigor (Nota: 100%)

Não foram encontrados problemas neste dataset

As informações a respeito das leis em vigor estão disponíveis na Câmara Municipal de São Paulo. Há a possibilidade de pesquisar a legislação por tipo de norma, número, ano, autor ou palavras-chave e todas as informações existentes sobre a busca é exibida. Ressalta-se como boa prática a API com possibilidade de conectar à totalidade dos dados disponibilizados por meio de webservice.

Atividade Legislativa (Nota: 100%)



Problemas encontrados - Usabilidade

Dificuldades na usabilidade dos dados

As informações quanto a atividade legislativa estão disponíveis na Câmara Municipal de São Paulo. A dimensão atende a todos os requisitos do índice e, por isso, atingiu nota máxima. No entanto, vale ressaltar que no site da Câmara não é possível obter todas as informações quanto aos projetos de leis e as leis consolidadas em um conjunto de dados pronto para analisá-lo. O processo para obtenção de informações quanto a um projeto de lei ocorre de forma manual, deve-se procurar pelo número do projeto que contém informações quanto ao conteúdo, autor, status e votos. A partir destas informações é possível procurar as transcrições dos debates que ocorreram na sessão plenária no dia em que o projeto foi votado. Os dados também estão disponíveis via API (por isso o download da base completa é possível), mas para isso é necessário algum conhecimento básico em programação.

Limites Administrativos (Nota: 100%)



Não foram encontrados problemas neste dataset

As informações quanto aos limites administrativos da prefeitura estão disponíveis no GeoSampa hospedado pela SMUL - Secretaria Municipal de Urbanismo e Licenciamento da Prefeitura de São Paulo (antes denominada SMDU - Secretaria Municipal de Desenvolvimento Urbano). As possibilidades de download de informações quanto ao limites das prefeituras regionais e os distritos permitem que os mapas sejam plotados em escalas diferentes e estão disponíveis em formatos abertos (shapefile e dbf). Neste caso, identificamos bases completas, atualizadas, em formatos abertos licenças de uso disponíveis, fáceis de se encontrar e de transformar em análises. A consolidação das informações georreferenciadas em um portal único é considerada uma boa prática de dados abertos da cidade de São Paulo.

Escolas Públicas (Nota: 100%)



Problemas encontrados - Usabilidade

Ausência de informações requeridas pela metodologia: coordenadas (endereços, latitude e longitude)

Dificuldades na usabilidade dos dados

O Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) realiza o censo escolar a cada ano. Neste levantamento são compilados dados de todas as escolas em funcionamento em território brasileiro. Apesar da prefeitura de São Paulo disponibilizar dados de cadastros das escolas de São Paulo (no portal de dados abertos), estes dados contemplam apenas o universo das escolas municipais e não trazem informações sobre as matrículas, mas apenas sobre a capacidade das escolas. O banco de dados do INEP contém mais informações e pode ter seu download realizado de uma só vez. O maior problema deste dataset é a indisponibilidade de dados de latitude e longitude das escolas, uma vez que, apesar de haver espaço para tal informação no censo escolar, a maior parte das escolas não responde a essa pergunta, o que leva a uma falta de informação. Por sua vez, o estado de São Paulo disponibiliza um mapa colaborativo sobre as localizações das escolas mas, dado que o conteúdo é colaborativo, não há como garantir que todas as escolas estejam representadas no mapa. A prefeitura de São Paulo disponibiliza em sua página os dados de latitude e longitude das escolas municipais, porém, não de todas as escolas da cidade de São Paulo. Desta forma, existe uma dificuldade em trabalhar com os dados, já que é necessário trabalhar com mais de um banco de dados disponível - e que não apresenta informações para todo universo existente. Mesmo assim, segundo os atributos do índice, o dataset foi avaliado com nota máxima.

Gastos Públicos (Nota: 83%)

Problemas encontrados - Usabilidade

Dificuldades na usabilidade dos dados



Problemas encontrados - Processo

Restrição de acesso por cadastro de usuário

Dificuldade para localizar informações

A avaliação do banco de dados sobre gastos públicos foi feita com base na vitrine de APIs da Prefeitura de São Paulo, que foi considerado o banco de dados mais abrangente sobre o assunto no âmbito da cidade. Ainda que o Portal da Transparência, no momento do levantamento, também contivesse a maior parte das informações sem algumas dificuldades encontradas na vitrine de APIs, haviam algumas lacunas: as rubricas individuais (ordens de pagamento) e os favorecidos/fornecedores, além dos dados estarem desatualizados. No caso do banco de dados analisado, o único requisito da metodologia que não foi cumprido foi o acesso livre, uma vez que é necessário fazer um cadastro e obter um “token” para acessar o sistema de consultas da API (ainda que o processo possa ser feito por qualquer cidadão e seja facilmente guiado por uma documentação disponibilizada no próprio site). Ainda na dimensão do processo, ressalta-se a dificuldade em encontrar tais dados, que só puderam ser analisados por meio da sugestão de um especialista que conhecia bem a gestão municipal. Além disso, a necessidade de possuir habilidades de programação limita a capacidade de utilização das informações.

Estatísticas Socioeconômicas (Nota: 83%)



Problemas encontrados - Usabilidade

Dados desatualizados

Problemas encontrados - Processo

Dificuldade para localizar informações

As estatísticas solicitadas pela metodologia original do índice – PIB, população e desemprego – já são disponibilizadas em nível federal pelo Instituto Brasileiro de Geografia e Estatística (IBGE). Sendo assim, em geral, os municípios cobertos pela divulgação do IBGE não produzem dados redundantes, mas apenas, em alguns casos, dados complementares ou estimativas. Neste caso, então, os dados do IBGE foram avaliados para o município de São Paulo. Ainda que cumpram o maior parte dos requisitos da metodologia, os dados não foram encontrados com muita facilidade uma vez que só é possível obtê-los com diferentes recortes e metodologias em diferentes locais dentro do site do IBGE. O principal problema encontrado foi a data de atualização do PIB municipal, defasada em três anos.

Compras Públicas (Nota: 83%)



Problemas encontrados - Usabilidade

Dados desatualizados

Os dados relacionados às compras públicas estão disponibilizados no portal da transparência da cidade de São Paulo. As informações cumprem praticamente todos os pré-requisitos da metodologia do índice e estão disponibilizadas para download em formatos abertos (csv, xml e xls). No entanto, considerando o período de realização da avaliação, as informações não estão atualizadas conforme periodicidade mensal requisitada pela metodologia.

Estatísticas do Crime (Nota: 83%)



Problemas encontrados - Usabilidade

Dificuldades na usabilidade dos dados

Problemas encontrados - Processo

Falta de clareza sobre a licença pública de uso de dados

Download da base completa indisponível

Dificuldade para localizar informações

As estatísticas criminais são provenientes da Secretaria Estadual de Segurança Pública de São Paulo. Dados relacionados a assassinato, estupro e roubo estão em uma seção do site, enquanto as apreensões de armas de fogo estão em outra, o que dificulta a localização dos dados. Além disso, em ambos os casos é necessário baixar dados por unidades de departamento de polícia. Desta forma, é necessário baixar os arquivos separadamente para consolidar as informações necessárias em um único conjunto de dados. A permissão para a usabilidade dos dados disponíveis também não é clara.

Localizações (Nota: 78%)



Problemas encontrados - Processo

Falta de clareza sobre a licença pública de uso de dados

Dificuldade para localizar informações

Por mais que a prefeitura municipal de São Paulo divulgue em sua página de dados abertos um dataset de logradouros, a avaliação foi feita com base numa base cartográfica realizada pelo Centro de Estudos da Metrópole da Universidade de São Paulo (USP), que é estadual, por fornecer dados mais atualizados. Os problemas encontrados foram a dificuldade em encontrar a informação (ainda que não seja imperativo realizar um cadastro, é necessário inserir um e-mail para chegar à página com os links) e a ausência de uma licença clara sobre a utilização dos dados.

Transporte Público (Nota: 67%)



Problemas encontrados - Usabilidade

Ausência de informações requeridas pela metodologia: horários por linha e estação/parada, conexões entre diferentes modais de transporte

Dados desatualizados

Dificuldades na usabilidade dos dados

Problemas encontrados - Processo

Download da base completa indisponível

Em geral, os bancos de dados de transporte público são desagregados em várias fontes diferentes, o que não é diferente no caso de São Paulo, impedindo o download da base completa e limitando a usabilidade dos dados. Quatro fontes foram avaliadas, sendo três da administração municipal (Portal Geosampa, Portal de Dados Abertos e SPTrans) e uma da estadual (CPTM). A primeira fonte traz informações sobre os cor-

redores de BRT (Bus Rapid Transit), itinerários e estações de Trem e Metrô, Terminais, paradas e itinerários de ônibus e mapa cicloviários. Informações sobre horários foram encontradas de forma separada nos demais links. Não foram encontradas informações sobre horários por linha e estação/parada de cada meio de transporte público nem as conexões entre os diferentes modais de transporte em uma base de dados. Os dados estavam desatualizados no momento do levantamento segundo os metadados do Geosampa. O ponto de atenção principal se refere aos dados sobre itinerários de trem, cuja data de atualização sinalizada é “sob demanda”.

Registro de Empresas (Nota: 44%)

Problemas encontrados - Usabilidade

Ausência de informações requeridas pela metodologia: identificador único das companhias

Indisponibilidade de dados em formato aberto

Dados desatualizados

Dificuldades na usabilidade dos dados

Problemas encontrados - Processo

Falta de clareza sobre a licença pública de uso de dados

Download da base completa indisponível

Dificuldade para localizar informações

Os registros de empresas para a cidade de São Paulo constam na página da Junta Comercial do Estado de São Paulo, sob responsabilidade do governo estadual. No entanto, sua usabilidade é restrita: não é possível saber a periodicidade de atualização das informações dispostas, tampouco realizar o download do banco de dados. É impossível saber se os dados disponibilizados pela busca no site contêm o universo de todas as empresas registradas na cidade. É difícil também encontrar as informações por empresas e a licença de usabilidade dos dados não é clara, tornando árdua a capacidade de trabalhar com estes dados.

Concentração de Poluentes (Nota: 44%)

Problemas encontrados - Usabilidade

Dificuldades na usabilidade dos dados

Ausência de informações requeridas pela metodologia: compostos orgânicos voláteis

Problemas encontrados - Processo

Falta de clareza sobre a licença pública de uso de dados

Download da base completa indisponível

Restrição de acesso por cadastro de usuário

Os dados sobre qualidade do ar e concentração de poluentes na cidade de São Paulo são coletados e fornecidos por uma entidade estadual: a Companhia Ambiental do Estado de São Paulo (CETESB). No quesito usabilidade, destaca-se a dificuldade em consolidar e utilizar as informações desagregadas. Todas as informações requeridas pela metodologia foram encontradas (material particulado, óxidos de enxofre e nitrogênio, monóxido de carbono, ozônio e disponibilidade por estação de monitoramento ou zona) exceto a de compostos orgânicos voláteis. No quesito processos, destaca-se a necessidade de realizar um cadastro para acessar a base de dados: ainda que não haja limitação prévia sobre quem pode se cadastrar, isso dificulta o acesso aos dados. Além disso, não é possível baixar todos os dados numa base única, uma vez que o sistema possibilita apenas extrações parciais da base, ainda que em formato aberto. Destaca-se ainda a indisponibilidade de licença pública de uso de dados.

Qualidade da Água (Nota: 44%)

Problemas encontrados - Usabilidade

Ausência de informações requeridas pela metodologia: níveis de arsênico, fluoreto e sólidos totais dissolvidos por fonte de água

Indisponibilidade de dados em formato aberto

Dados desatualizados

Dificuldades na usabilidade dos dados

Problemas encontrados - Processo

Falta de clareza sobre a licença pública de uso de dados

Download da base completa indisponível

O Programa Municipal de Vigilância da Qualidade da Água para Consumo Humano de São Paulo (VIGIAGUA) da Secretaria Municipal de Saúde, apresenta boletins periódicos quanto à qualidade da água na cidade que contém informações quanto a presença de coliformes fecais e nitratos. As avaliações, no entanto, expõem dados apenas de conformidade ou não conformidade, não havendo maiores detalhes das informações coletadas. Outro problema quanto a disponibilidade está relacionada ao formato: os boletins estão em PDF, o que dificulta o manuseio dos dados. A atualização dos dados também é outra questão, tendo em vista que há um hiato temporal de cerca de quatro meses entre os dados disponíveis para a última análise. Além disso, a ausência de um banco de dados histórico e consolidado para download, a falta de clareza sobre a licen-

ça pública de uso de dados e a dificuldade para encontrar informações a respeito do estado da base de dados fazem com que esta dimensão seja um dos principais pontos de atenção para a cidade de São Paulo.

Previsão do Tempo (Nota: 44%)

Problemas encontrados - Usabilidade

Ausência de informações requeridas pela metodologia: previsão para pelo menos os próximos quatro dias sobre temperatura média, velocidade e direção do vento, probabilidade e quantidade de precipitação
Indisponibilidade de dados em formato aberto

Dados desatualizados

Dificuldades na usabilidade dos dados

Problemas encontrados - Processo

Falta de clareza sobre a licença pública de uso de dados

Download da base completa indisponível

Dificuldade para localizar informações

As informações sobre previsão do tempo foram encontradas no site do Centro de Gerenciamento de Emergências (CGE) da Prefeitura de São Paulo. Os dados não foram fáceis de encontrar pois foi necessário primeiro buscar por monitoramento de condições climáticas para depois chegar à previsão do tempo. Mesmo assim, a única informação fornecida é a previsão para temperaturas extremas, e mesmo assim, apenas para o dia atual e os próximos dois dias. Todas as demais (previsão para o dia atual e os próximos quatro dias contendo média de temperatura, direção e velocidade de vento, probabilidade e quantidade de chuva) não foram encontradas. A base de dados não é disponibilizada em formato aberto nem ao menos para download, o que dificulta a usabilidade dos dados. Só é possível visualizá-la na web. Não foi encontrada licença para uso público dos dados.

Propriedade da terra (Nota: 0%)

Este dataset não foi encontrado.

Não existem dados públicos relacionados à propriedade de terra em São Paulo conforme requisitado pela metodologia do índice. O portal GeoSampa, da prefeitura da cidade, disponibiliza mapas digitais de São Paulo, no entanto, neste site não é possível realizar downloads de dados relativos ao loteamento da cidade. Como a administração municipal realiza a cobrança de impostos relacionados à propriedade na cidade, sabe-se que esta informação existe, no entanto, ela não está disponível publicamente.

CONCLUSÃO

Em geral, a cidade de São Paulo teve um resultado positivo no índice, tendo obtido 75% da pontuação total. Apenas um dataset foi avaliado com 0%, outros quatro obtiveram uma nota inferior a 50%. Dos outros 13 datasets com mais que 50% de pontuação, sete receberam nota máxima. A maior parte dos problemas identificados está mais relacionada à usabilidade (completude, atualização e formato) do que a processo (meios de acesso e licença de uso dos dados).

A principal boa prática ressaltada é a consolidação de informações georreferenciadas em um portal único, que disponibiliza dados em formato aberto. O problema mais crítico, segundo a análise deste trabalho, é a incompletude da base, isto é, a não disponibilização de determinadas informações consideradas essenciais sobre o tema em questão. Esse problema aparece em seis datasets, dentre os quais quatro são disponibilizados pela própria prefeitura.

Este trabalho pode auxiliar no processo de abertura de dados na cidade de São Paulo ao direcionar a atenção para as dimensões prioritárias e problemas mais comuns de acordo com a definição da Open Knowledge Foundation. Em outras palavras, este documento traz algumas diretrizes para uma política de dados abertos para a cidade de São Paulo.

Essa iniciativa pode ter uma série de benefícios para a gestão municipal, tanto internamente quanto externamente à administração pública. Internamente, porque o próprio processo de abertura de dados cria um mecanismo de cobrança interno para resultados das diferentes unidades, o que contribui para uma maior eficiência da gestão pública. Além disso, a abertura de dados permite que os cidadãos possam contribuir de diversas formas, seja com análises, controle social, criação de aplicativos, replicação da informação, entre outras, o que acaba por tornar a gestão mais eficiente ao congregar esforços por meio da participação da sociedade. Por outro lado, há também benefícios externos: a administração municipal passa a ser vista como mais accountable aos seus cidadãos, o que implica não só em ganhos de imagem como também na possibilidade de se tornar referência nacional e internacional em dados abertos.

Com isso, a FGV/DAPP e a OKBr sinalizam recomendações, com base nos resultados preliminares do levantamento, em três frentes: em primeiro lugar, resalta-se a necessidade de sincronizar a forma de apresentação dos dados abertos de todas as dimensões, se possível consolidando-os em um mesmo repositório. O desenho de soluções específicas para os problemas relatados em cada dimensão busca estabelecer boas práticas, o que passa pela padronização de processos para formatar a divulgação dos dados abertos e o desenho de estratégias de atuação para cada tipo de unidade administrativa (uma vez que os bancos de dados são disponibilizados por entidades de naturezas diversas).

Em segundo lugar, as considerações deste relatório constituem subsídios para construção de um plano municipal de dados abertos que contemple o compromisso com metas ambiciosas de abertura de dados.

Em terceiro lugar, mas não menos importante, recomenda-se a implementação de formas inovadoras de apresentação das informações, por meio da criação de uma metodologia de visualização de dados. A ideia é ir além da simples divulgação das informações, visando à garantia de que o consumidor final dos dados possa apropriá-los de forma ágil e inteligível, fomentando o uso dos dados para políticas públicas mais eficazes.

FONTES

Aqui são elencadas as fontes utilizadas (incluindo os links acessados) para a avaliação por dimensão do índice. São destacados os órgãos provedores da informação e o tipo de administração (poder e unidade da federação).

É importante destacar que a avaliação é feita para o estado dos dados abertos para o ano de 2016. Estas fontes foram consultadas no período do levantamento que foi de 01 a 30 de dezembro de 2016. O processo de revisão e inclusão das novas dimensões do índice foi realizado entre janeiro e fevereiro de 2017.

Orçamento Público

Prefeitura Municipal de São Paulo

Executivo Municipal

Portal da Transparência da Prefeitura de São Paulo
<http://transparencia.prefeitura.sp.gov.br/contas/Paginas/LeiOrcamentoAnual.aspx>

Resultados Eleitorais

TSE - Tribunal Superior Eleitoral

Judicário Federal

Repositório de dados eleitorais
<http://www.tse.jus.br/eleicoes/estatisticas/repositorio-de-dados-eleitorais>

Mapas da Cidade

SMDU - Secretaria Municipal de Desenvolvimento Urbano da Prefeitura de São Paulo

Executivo Municipal

Portal Geosampa
http://geosampa.prefeitura.sp.gov.br/PaginasPublicas/_SBC.aspx#

Leis em Vigor

Câmara Municipal de São Paulo

Legislativo Municipal

Portal SPLegis (Sistema de Processo Legislativo)
<http://www.camara.sp.gov.br/atividade-legislativa/splegis-consulta/>

Atividade Legislativa

Câmara Municipal de São Paulo

Legislativo Municipal

Portal SPLegis (Sistema de Processo Legislativo)
<http://www.camara.sp.gov.br/atividade-legislativa/splegis-consulta/>

Limites Administrativos

SMDU - Secretaria Municipal de Desenvolvimento

Urbano da Prefeitura de São Paulo

Executivo Municipal

Portal Geosampa

http://geosampa.prefeitura.sp.gov.br/PaginasPublicas/_SBC.aspx#

Escolas Públicas

INEP Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira

Executivo Federal

Matrículas e responsabilidade administrativa por escola (Censo Escolar)

<http://inep.gov.br/microdados>

Gastos Públicos

Prefeitura Municipal de São Paulo

Executivo Municipal

Vitrine de APIs da Prefeitura de São Paulo - API da SOF
<https://api.prod.am.sp.gov.br/store/apis/info?name=SOF&version=v2.1.0&provider=admin>

Estatísticas Socioeconômicas

IBGE - Instituto Brasileiro de Geografia e Estatística

Executivo Federal

Produto Interno Bruto Municipal

http://www.ibge.gov.br/home/estatistica/pesquisas/pesquisa_resultados.php?id_pesquisa=46

Taxas de desemprego trimestrais a partir da Pesquisa Nacional por Amostra de Domicílios (PNAD) Contínua

http://www.ibge.gov.br/home/estatistica/indicadores/trabalhoerendimento/pnad_continua_mensal/default.shtm

Estimativas populacionais por município

http://www.ibge.gov.br/home/estatistica/populacao/estimativa2015/estimativa_dou.shtm

Compras Públicas

Prefeitura de São Paulo

Executivo Municipal

Portal da Transparência da Prefeitura de São Paulo
<http://transparencia.prefeitura.sp.gov.br/contas/Paginas/ComprasLicitacoes.aspx>

Portal de Dados Abertos da Prefeitura de São Paulo - Base de Compras e Licitações

<http://dados.prefeitura.sp.gov.br/dataset/base-de-compras-e-licitacoes>

Portal e-negócios da Prefeitura de São Paulo

<http://e-negocioscidadesp.prefeitura.sp.gov.br/BuscaLicitacao.aspx>

Estatísticas do Crime

Secretaria de Segurança Pública de São Paulo Executivo Estadual

Homicídio, Furto, Apreensão de Armamentos

<http://www.ssp.sp.gov.br/Estatistica/Pesquisa.aspx>

Estupros

<http://www.ssp.sp.gov.br/Estatistica/ViolenciaMullher.aspx>

Localizações

CEM/USP - Centro de Estudos da Metrópole, Universidade de São Paulo

Executivo Estadual

Acesso à base de dados de logradouros para a Região Metropolitana de São Paulo

<http://www.fflch.usp.br/centrodametropole/716>

Observação: os dados disponibilizados pelo portal de dados abertos (<http://dados.prefeitura.sp.gov.br/dataset/geocodificacao-de-logradouros-geolog>) da Prefeitura de São Paulo estavam desatualizados no momento de realização do levantamento.

Transporte Público

SMDU - Secretaria Municipal de Desenvolvimento Urbano da Prefeitura de São Paulo

Executivo Municipal

Portal Geosampa

http://geosampa.prefeitura.sp.gov.br/PaginasPublicas/_SBC.aspx

Ciclovias

<http://dados.prefeitura.sp.gov.br/dataset/ciclovias>

Horário dos Trens (CPTM - Companhia Paulista de Trens Metropolitanos)

<http://www.cptm.sp.gov.br/sua-viagem/Pages/sua-viagem.aspx>

Horários dos Ônibus

<http://itinerarios.extapps.sptrans.com.br/PlanOperWeb/>

Registro de Empresas

JUCESP - Junta Comercial do Estado de São Paulo

Executivo Estadual

Ferramenta de busca da JUCESP

<https://www.jucesponline.sp.gov.br/ResultadoBusca.aspx>

Concentração de Poluentes

CETESB - Companhia Ambiental do Estado de São Paulo

Executivo Estadual

Página principal sobre qualidade do ar

<http://ar.cetesb.sp.gov.br/>

Dados mais recentes disponibilizados por estação de monitoramento e por poluente

http://sistemasinter.cetesb.sp.gov.br/Ar/php/boletim_por_poluente.php

Mapa da Região Metropolitana de São Paulo com a situação de cada estação de monitoramento

http://sistemasinter.cetesb.sp.gov.br/Ar/php/mapa_qualidade_rmsp.php

Dados por hora e por estação de monitoramento

http://sistemasinter.cetesb.sp.gov.br/Ar/php/ar_dados_horarios.php

Sistema "Qualar" (Sistema de Informações da Qualidade do Ar) que permite download de informações por estação em formato .csv. Requer cadastro para acesso <http://qualar.cetesb.sp.gov.br/qualar/home.do>

Qualidade da Água

Secretaria Municipal de Saúde da Prefeitura de São Paulo

Executivo Municipal

Programa Municipal de Vigilância da Qualidade da Água
para Consumo Humano de São Paulo - VIGIAGUA

http://www.prefeitura.sp.gov.br/cidade/secretarias/saude/vigilancia_em_saude/saude_ambiental/agua/index.php?p=6967

Previsão do Tempo

Centro de Gerenciamento de Emergências (CGE) da Prefeitura de São Paulo

Executivo Municipal

Plataforma online de controle de emergências relacionadas a condições meteorológicas com informações de previsão do tempo para a cidade de São Paulo

http://www.cgesp.org/v3/previsao_estendida.jsp

Propriedade da terra

Dataset não encontrado

 **FGV DAPP**



Inovação e Políticas Públicas

