



ÍNDICE DE DADOS ABERTOS PARA CIDADES 2018



Ficha catalográfica elaborada pela Biblioteca Mario Henrique Simonsen/FGV

Índice de dados abertos para cidades [recurso eletrônico] / Coordenação
Marco Aurélio Ruediger e Natália Mazzotte. – Rio de Janeiro : FGV, DAPP,
2018.

1 recurso online (70 p.) : PDF, il.

Dados eletrônicos.

Em parceria com Open Knowledge Brasil.

Inclui bibliografia.

ISBN: 978-85-68823-58-3

1. Abertura de informação ao público. 2. Serviços de informação on-line. 3.
Políticas públicas. 4. Transparência na administração pública. I. Ruediger,
Marco Aurélio, 1959- . II. Mazzotte, Natália. III. Fundação Getulio Vargas.
Diretoria de Análise de Políticas Públicas. IV. Open Knowledge Brasil.

CDD – 351

Citação

RUEDIGER, M. A.; MAZOTTE, N. (Ed.). **Índice de dados abertos para cidades**.
Rio de Janeiro: FGV DAPP, 2018.

Diretoria de Análise de Políticas Públicas da Fundação Getúlio Vargas

Open Knowledge Brasil

ÍNDICE DE DADOS ABERTOS PARA CIDADES

Rio de Janeiro



2018



Fundada em 1944, a Fundação Getúlio Vargas nasceu com o objetivo de promover o desenvolvimento socioeconômico do Brasil por meio da formação de administradores qualificados, nas áreas pública e privada. Ao longo do tempo, a FGV ampliou sua atuação para outras áreas do conhecimento, como Ciências Sociais, Direito, Economia, História e, mais recentemente, Matemática Aplicada, sendo referência em qualidade e excelência, com suas oito escolas.



Criada em 2013, a Open Knowledge Brasil (OKBR) – também chamada de Rede pelo Conhecimento Livre – é, no Brasil, o braço da Open Knowledge Internacional, já presente em 66 países. É uma Organização da Sociedade Civil (OSC) sem fins lucrativos e apartidária. A Open Knowledge Brasil utiliza e desenvolve ferramentas cívicas, projetos, realiza análises de políticas públicas e jornalismo de dados, além de promover o conhecimento livre nos diversos campos da sociedade. Na esfera política, tem como objetivo tornar a relação entre governo e sociedade mais próxima e transparente.



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

DAPP

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

Coordenação do Projeto

Wagner Oliveira

Equipe

Wagner Oliveira

Revisão

Thais Lobo

Projeto Gráfico

Luís Gomes
Rodrigo Cid



Diretora-Executiva

Natália Mazotte

Open Knowledge Brasil

br.okfn.org
@okfn.org.br

Coordenação do Projeto

Ariel Kogan

Equipe

Davi Ventura

SUMÁRIO

RESUMO EXECUTIVO	04
SOBRE A FGV DAPP	06
SOBRE A OPEN KNOWLEDGE BRASIL	07
SOBRE A PARCERIA FGV DAPP E OKBR	08
O ÍNDICE DE DADOS ABERTOS PARA CIDADES	09
AGRADECIMENTOS	11
METODOLOGIA	12
RESULTADOS	41
CONCLUSÃO	62
REFERÊNCIAS	65
ANEXO	67

RESUMO EXECUTIVO

O Índice de Dados Abertos para Cidades edição 2018 é uma avaliação do **estado da arte das políticas de dados abertos das cidades** participantes, tendo por **ano-base 2017**.

O estudo avaliou **136 bases de dados, distribuídas em 17 dimensões, referentes a oito cidades** (Belo Horizonte-MG, Brasília-DF, Natal-RN, Porto Alegre-RS, Rio de Janeiro-RJ, Salvador-BA, São Paulo-SP e Uberlândia-MG).

Dessas oito, duas cidades estiveram na avaliação do ano passado. As diferenças entre os dois anos foram marginais, e o **número de bases 100% abertas é o mesmo**.

Dentre as dimensões avaliadas, quatro destacaram-se pelo maior número de gargalos possíveis nas bases de dados: **Propriedade da Terra, Registro de Empresas, Qualidade do Ar e Qualidade da Água**.

Apenas **25%** das bases de dados avaliadas estão **100% de acordo com a definição de dados abertos**.

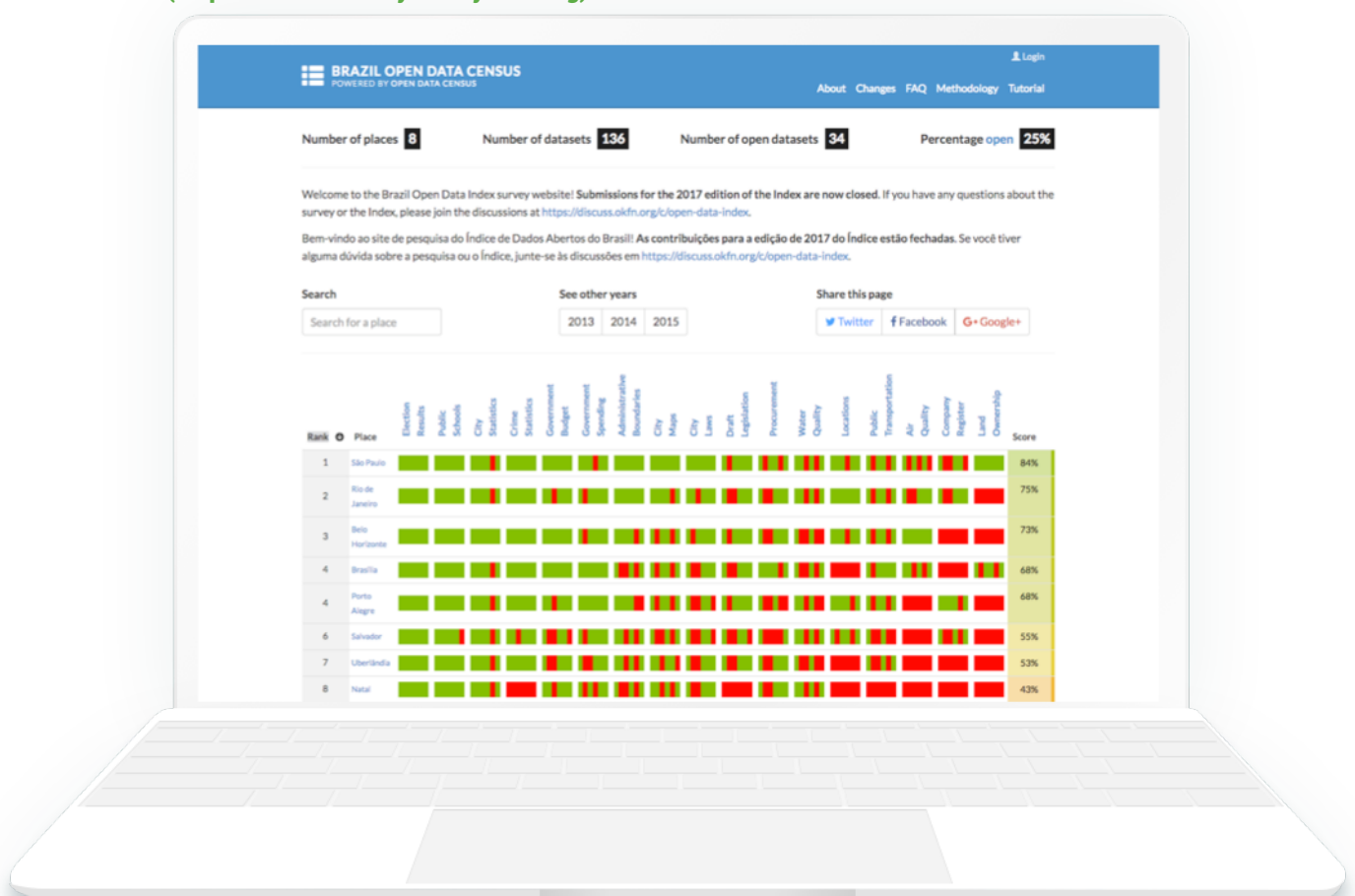
De todo o universo de obstáculos encontrados (429), **62% são problemas de usabilidade e 38% de processo**, reiterando a conclusão da avaliação de 2016 sobre a necessidade de bases de dados mais adequadas ao uso dos dados e transformação destes em informação, para além da publicização.

Os problemas mais comuns das bases de dados são: **dificuldade de trabalhar dados (incluindo os metadados insuficientes), indisponibilidade de download da base de dados completa, dataset incompleto e ausência da informação em formato aberto**.

Os dois **gargalos mais detectados entre as cidades** também foram os mais frequentes na avaliação do **Brasil** a nível federal no **Global Open Data Index 2016**, mostrando serem problemas comuns no país inteiro.

Os dados abertos das cidades (e as dimensões) apresentam diversos **problemas comuns**, que serão detalhados na seção de resultados. Mas também suscitam uma série de **boas práticas replicáveis**, também detalhadas no relatório.

Figura 1
Resultados do Índice de Dados Abertos
(Disponível em: br-city.survey.okfn.org)



Fonte e Elaboração: FGV/DAPP e OKBr.

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, oferece treinamento em produção e uso de dados abertos e promove o conhecimento livre para tornar a relação entre governo e sociedade mais transparente e promover uma participação política mais efetiva e aberta.

Durante a última década, a organização tem desempenhado um papel-chave no fortalecimento do ecossistema global de dados abertos e transparência governamental. Seu portfólio inclui: CKAN, o software livre para catálogo de dados abertos mais bem sucedido mundialmente; Escola de Dados, uma rede global comprometida com o avanço do uso de dados para resolver problemas reais em prol de sociedades mais conscientes, sustentáveis e justas; Índice Global de Dados Abertos, metodologia de avaliação que fornece uma visão abrangente do estado da publicação de dados governamentais abertos no mundo, e Serenata de Amor, um dos mais bem-sucedidos projetos brasileiros de uso de dados governamentais para controle social dos gastos públicos.

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 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 uma iniciativa pioneira 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.

O índice global, que já havia sido construído para 2013, 2014 e 2015, contou com a participação da parceria no levantamento das informações de 2016² e na elaboração do relatório “Índice de Dados Abertos para o Brasil”³, que apontou os principais gargalos encontrados nas iniciativas brasileiras de dados abertos, ressaltando também as boas práticas que podem ser replicadas pela administração pública. A principal conclusão do estudo foi que o Brasil pode ser considerado um país avançado no quesito publicidade e divulgação de informações, mas ainda precisa desenvolver aspectos no que se refere à facilidade de acesso e capacidade de gerar entendimento a partir dos dados.

Além do índice global, que compara países, a Open Knowledge partiu para um esforço, em 2016, de replicação do ODI para a esfera subnacional, mais especi-

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

² Disponível em <http://global.survey.okfn.org/>. Acesso em 05/03/2018.

³ Disponível em <http://dapp.fgv.br/wp-content/uploads/2017/04/IndiceDadosAbertosBrasil2017-1.pdf>. Acesso em 05/03/2018.

ficamente, para cidades. Em sua ramificação brasileira, a OKBr também 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 dos municípios brasileiros: transporte, criminalidade e educação.

O projeto foi aplicado de forma pioneira para São Paulo e Rio de Janeiro e gerou dois relatórios⁴, apresentados em eventos em cada uma das cidades, mobilizando a comunidade e as autoridades locais em torno da temática. Os resultados práticos da avaliação do ODI puderam ser sentidos, por exemplo, em comunicações das autoridades públicas revelando o interesse de entender como poderiam melhorar a qualidade das suas bases de dados abertas.

O sucesso do projeto deu origem a uma ampliação para outras cidades do Brasil. Para isso, foram mapeados grupos de pessoas e organizações interessadas em realizar o levantamento do Índice de Dados Abertos em outros municípios. A FGV DAPP e a OKBr escolheram, dentre esse grupo de

interessados, um conjunto inicial de dez cidades para acompanhar de perto no processo de levantamento e revisão, além das cidades de São Paulo e Rio de Janeiro, para as quais os parâmetros de avaliação seriam atualizados. Os critérios utilizados foram, principalmente, a capacidade das equipes locais para realizar as atividades previstas e a distribuição regional.

No processo é natural que algumas cidades não tenham seguido até o final, o que fez com que o resultado final apresentado neste relatório contenha um universo total de oito cidades: **Belo Horizonte-MG, Brasília-DF, Natal-RN, Porto Alegre-RS, Rio de Janeiro-RJ, Salvador-BA, São Paulo-SP e Uberlândia-MG.**

Após os agradecimentos, a seção seguinte traz um detalhamento da metodologia utilizada para esta avaliação e a subsequente os resultados consolidados dos levantamentos realizados pelos grupos locais e revisados pelas equipes da FGV DAPP e da OKBr. Tais insumos podem ser úteis para a elaboração de diretrizes para uma política de dados abertos no âmbito de cada cidade. Cada dimensão do índice será detalhada de modo a ressaltar as boas práticas encontradas e os principais pontos de atenção.

⁴ Os relatórios estão disponíveis nos links <http://dapp.fgv.br/wp-content/uploads/2017/05/Indice-DadosAbertosRio2017.pdf> e <http://dapp.fgv.br/wp-content/uploads/2017/04/RelatorioODI-SP-2017.pdf>. Acesso em 05/03/2018.

AGRADECIMENTOS

Gostaríamos de agradecer a todos aqueles que estiveram envolvidos na elaboração do Índice de Dados Abertos para Cidades edição 2018, ano-base 2017, pois, sem o esforço dessas pessoas, a ampliação do projeto teria sido inviável. A dedicação das equipes locais, que se voluntariaram para levantar informações sobre suas cidades, mostra a relevância que a questão da transparência de dados públicos tem ganhado no Brasil.

Agradecemos especialmente a:

Belo Horizonte

Carlos Castro
Lucas Parreiras
Lucas Resende
Raquel Paiva

Brasília

Laura Boeira

Natal

Elias Menezes
George Maia

Porto Alegre

Ana Júlia Possamai
Frederico Cosentino
Vitória Gonzatti

Salvador

Iracema Marques
Juan Torres
Luciano Ataíde
Marcelo de Almeida
Pablo Florentino
Taiane Rodrigues
Tiago Basto

Uberlândia

Camila Jardim
Daniel Martins
Gessica do Carmo
Jhyenne Santana
Mayara Borges
Paloma Campos

Além disso, agradecemos aqueles que contribuíram com seus conhecimentos, seja pessoalmente, seja via fórum de discussões da Open Knowledge⁵, para dirimir dúvidas e refinar os resultados do levantamento. Em especial, agradecemos aos pesquisadores da FGV DAPP, Amaro Grassi, Bárbara Barbosa, Beatriz Meirelles, Danielle Sanches e Thaís Lobo. Agradecemos também à participação sempre muito ativa nos fóruns de discussão da Open Knowledge de Augusto Herrmann, do Ministério do Planejamento, Desenvolvimento e Gestão.

⁵ Disponível em <<https://discuss.okfn.org/t/open-data-index-brasil-2017-cities-faq/6040/29>>. Acesso em 24/04/2018.

METODOLOGIA

O Índice de Dados Abertos (ODI, na sigla em inglês) é uma avaliação independente da publicação de dados abertos governamentais e se dá a partir de uma perspectiva cívica. O ODI permite que diferentes partes interessadas nos dados abertos acompanhem o progresso do governo no lançamento dos mesmos, além de possibilitar que os governos recebam feedback direto dos usuários de dados.

O índice fornece às duas partes, portanto, uma base para discussão e análise do ecossistema de dados abertos em seu país (ou cidades) e internacionalmente. Encorajamos todas as partes interessadas a participar de um diálogo aberto para permitir a apropriação dos resultados e tornar o Índice o mais relevante possível.

1. Âmbito de pesquisa

Como qualquer outra ferramenta de avaliação comparativa, o ODI tenta responder a uma pergunta. No nosso caso, a questão é a seguinte: *como os governos publicam dados abertos?* A partir desta questão, surgem outras importantes, tais como:

- ***Quais governos prontamente publicam dados abertos?***
- ***Qual é o conjunto de dados mais aberto? Qual é o conjunto de dados menos aberto?***
- ***Que governos ainda precisam melhorar a publicação de dados abertos?***
- ***Quais aspectos dos dados abertos são mais fáceis ou difíceis de implementar?***

O índice também mede aspectos de “facilidade de abertura”, como a capacidade de encontrar dados e a usabilidade. Estes também são reconhecidos pelos Princípios da Carta Internacional de Dados Abertos⁶. A informação obtida com esta avaliação é exibida nos resultados e está disponível para download. O código da plataforma está disponível online via Github⁷.

2. O que o ODI não cobre?

O ODI intencionalmente limita sua verificação à publicação de dados do governo. Não examina outros aspectos da estrutura comum de avaliação de dados abertos, como contexto, uso ou impacto. Este foco estreito permite que ele forneça uma avaliação padronizada, robusta e comparável de dados abertos para as cidades brasileiras. O ODI observa apenas as características de acesso e usabilidade, sem avançar efetivamente, ainda, na questão da qualidade de mensuração dos dados, ainda que este relatório apresente algumas inferências sobre esse quesito, analisado de forma qualitativa ao longo do processo de levantamento e revisão.

⁶ Disponível em: <http://opendatacharter.net/principles/>. Acesso em 02/05/2018.

⁷ Disponível em: <https://github.com/okfn/opendatasurvey>. Acesso em 02/05/2018.

3. Premissas da pesquisa

Esta seção apresenta as principais premissas que foram levadas em consideração ao coletar e avaliar os dados.

PREMISSA 1 OS DADOS ABERTOS SÃO DEFINIDOS CONFORME A DEFINIÇÃO DE ABERTURA (OPEN DEFINITION⁸).

Definimos dados abertos de acordo com a Definição de Abertura, um conjunto de princípios que conceitua a abertura de dados e conteúdo. Também é simples e fácil de operacionalizar. A única parte da metodologia aqui aplicada que não está completamente alinhada com a Definição de Abertura é a nossa avaliação de formatos “abertos, processáveis por máquina”. Damos a pontuação máxima aos formatos processáveis por máquina, mesmo que seu código fonte não esteja aberto. Na Definição de Abertura, os formatos devem ser utilizáveis com pelo menos um software livre e de código aberto. Assim, o índice dá preferência à facilidade de abertura sobre a abertura real de um formato.

PREMISSA 2 O GOVERNO TEM UM PAPEL CENTRAL NA PUBLICAÇÃO DE DADOS.

No passado, houve questões sobre o papel que os governos devem desempenhar para assegurar a publicação de dados abertos. Os serviços governamentais podem ser privatizados, o que significa que os dados podem ser apropriados e produzidos por uma empresa e não pelo Estado. Assumimos que, para as categorias de dados que pesquisamos, o governo tem a responsabilidade de assegurar sua publicação, mesmo que seja mantido e gerenciado por um terceiro.

⁸ Disponível em: <https://opendefinition.org/od/2.1/en/>. Acesso em 02/05/2018.

PREMISSA 3

O ÍNDICE DE DADOS ABERTOS LOCAL É UM INDICADOR PARA CIDADES.

Reconhecemos que nem todas as cidades têm a mesma estrutura política e, por isso, é possível que nem todos os governos sub-nacionais produzam os mesmos dados, já que estão potencialmente sujeitos a leis e procedimentos diferentes. O ODI municipal, portanto, pode avaliar a publicação de dados no nível da cidade que não são, necessariamente, fornecidos pela administração municipal. Desta forma, em alguns casos, a publicação nacional de dados abertos substitui os esforços no nível da cidade por três razões principais:

Os dados descrevem processos ou procedimentos do governo nacional (entidades governamentais que operam no nível administrativo mais alto).

Os dados são coletados ou produzidos pelo governo nacional ou por uma agência governamental nacional (no nível administrativo mais alto).

Os dados descrevem parâmetros nacionais e serviços públicos para todo o território nacional, mas são coletados por atores subnacionais.

Vale ressaltar que, se um determinado dado é divulgado em duplicidade, escolhemos a entidade a ser avaliada como aquela que apresenta a base mais completa possível segundo as determinações do que será avaliado em cada dimensão, como se verá a seguir.

4. Quais dados o índice avalia?

O ODI mede a abertura de determinadas categorias de dados previamente definidas ao longo dos anos pela comunidade internacional Open Knowledge. O ODI avalia os dados abertos do governo que provaram ser úteis para o público. Outros dados abertos que não se enquadram nessas categorias não foram considerados para nossa avaliação.

Além disso, o ODI é principalmente um indicador comparativo entre localidades, mas com difícil comparabilidade entre anos. No

passado, foram utilizadas categorias mais amplas e, em alguns casos, diferentes. Observam-se ainda ajustes na metodologia de ponderação ao longo do tempo. Na medida do possível, este relatório traça alguns paralelos com a avaliação feita no ano passado, que possui diferenças marginais em relação a este ano.

Ainda, há um procedimento padronizado que apoia os levantadores e revisores para reduzir a avaliação tendenciosa e o julgamento pessoal, ainda que, é sabido, não é possível eliminar completamente o viés. O processo passa pelo levantamento de dados por um grupo local, seguido pela revisão realizada em conjunto pela FGV DAPP e a OKBr. Em seguida, o resultado da revisão passa novamente pelos levantadores, chega ao conhecimento dos provedores de dados e, eventualmente, passa por uma segunda revisão, para, enfim, chegar a um resultado final.

Cada categoria de dados contém as seguintes informações:

Um mínimo de três características: as características descrevem o conteúdo obrigatório de um conjunto de dados. Normalmente, este é o principal critério para escolher uma base para se avaliar dentre um conjunto que poderia se qualificar para a avaliação.

Nível de agregação: dados estão disponíveis em diferentes níveis de agregação. Por exemplo, podem existir dados de qualidade da água para cada fonte de água individual, ou podem ser apresentados dados de poluição anual total para regiões ou o país. Na maioria dos casos, o ODI busca dados detalhados e desagregados, que amplificam os casos de uso e os insights que as pessoas podem tirar disso. A Carta Internacional de Dados Abertos⁹ também enfatiza que os dados devem ser publicados em seu formato bruto. Em geral, cada dimensão exige um determinado nível de detalhe, o que é explicitado aos levantadores para buscar o conjunto de dados correto.

Intervalos de tempo: Diferentes conjuntos de dados são atualizados em diferentes intervalos de tempo. O questionário do levantamento inclui a pergunta “Estes dados devem ser atualizados a cada [INTERVALO DE TEMPO]. Estão atualizados?”. O intervalo de tempo de cada dimensão é definido de acordo com o que é mais adequado para essa dimensão. Por exemplo, é esperado

⁹ Disponível em: <http://opendatacharter.net/>. Acesso em 02/05/2018.

que um dado de qualidade do ar esteja disponível diariamente, enquanto que um dado geográfico pode ter atualização anual. Dados não atualizados, ainda que abertos, tornam-se, muitas vezes, pouco úteis.

Os governos frequentemente publicam dados em vários sites e em muitos arquivos e formatos. Para tomar uma decisão informada e consistente sobre quais dados escolher, os revisores seguem duas abordagens:

Escolhendo um conjunto de dados de referência: Os revisores encontram um conjunto de dados ou um arquivo de referência que contém todas as características relevantes. Eles respondem a pesquisa usando este conjunto de dados. Isso pode ser um arquivo CSV, um shapefile ou dados apresentados em um site. Se os avaliadores tiverem que escolher entre dois ou mais conjuntos de dados similares, eles devem escolher primeiro o que for mais completo em suas características e, se houver mais de um que preencha esse requisito, escolher aquele que pontue mais alto e documentar sua escolha em um comentário.

Referenciando conjuntos de dados múltiplos (se um arquivo de referência não estiver disponível): Os revisores não conseguiram encontrar um conjunto de dados de referência porque os dados estão divididos em vários arquivos, formatos e locais. Nesse caso, eles referenciam a pesquisa para diferentes arquivos. É importante que a soma desses arquivos contenha todas as características de dados exigidas.

5. A lista de categorias de dados

As categorias de dados refletem informações relevantes para a sociedade civil em geral. As categorias foram desenvolvidas de forma colaborativa ao longo dos anos em parceria com especialistas temáticos, incluindo organizações que defendem dados abertos em seus respectivos campos. Em alguns casos, baseamos nossa definição em padrões internacionais de produção de dados e relatórios utilizados pelos governos em todo o mundo. Todos os anos, refinamos nossas definições para refletir as aprendizagens desses especialistas. Abaixo, segue uma breve descrição sobre cada dimensão, contendo o que procuramos em cada uma, porque procuramos e quais as características demandadas.

Tabela 1

Detalhamento das categorias de dados do ODI

ORÇAMENTO PÚBLICO

O que se busca?

Orçamento do governo municipal. Trata-se da despesa prevista do governo para o próximo ano, e não a despesa atual. Para desenvolver esta categoria, o índice baseou-se no trabalho do *Open Spending*¹⁰.

Por que se busca?

Dados de orçamento permitem identificar escolhas públicas e prioridades governamentais, informando e qualificando o debate público e facilitando mecanismos de participação.

Características

Orçamento por departamento e sub-departamento, além de descrições sobre cada seção do orçamento. Intervalo de tempo esperado: anual.

GASTOS PÚBLICOS

O que se busca?

Registros dos gastos efetivos do governo municipal de forma detalhada. Os dados devem exibir despesas em andamento por transação efetuada. Para desenvolver esta categoria, o índice baseou-se no trabalho do *Open Spending*.

Por que se busca?

Esses dados permitem mostrar se o dinheiro público é eficiente e efetivamente utilizado. Isso ajuda a entender padrões de gastos e expor corrupção, uso indevido e desperdício.

Características

Departamento do governo que efetuou a transação, data da transação, identificação do favorecido, valor nominal da transação. Essas informações devem estar disponíveis para cada transação individual. Intervalo de tempo esperado: mensal.

¹⁰ Disponível em: <https://openspending.org/>. Acesso em 02/05/2018.

COMPRAS PÚBLICAS

O que se busca?

Todo o processo de execução de uma compra pública deve ser transparente, incluindo a fase de licitações e de concessões. Essa categoria não se refere ao planejamento de compras ou à implementação (que faz parte da categoria gastos públicos). Para desenvolver esta categoria, o índice baseou-se no trabalho da *Open Contracting Partnership*¹¹.

Por que se busca?

A divulgação desses dados permite uma concorrência mais justa entre as empresas, facilita a detecção de fraudes, além de oferecer melhores serviços aos governos e aos cidadãos. Além disso, ajuda novos grupos a participarem de licitações e a aumentar a conformidade legal do governo.

Características

Na fase de licitação, deve estar disponível cada licitação por gabinete do governo, incluindo sua identificação, descrição e situação. Na fase de concessão, cada concessão por gabinete do governo deve estar disponível, incluindo sua identificação (título), descrição, valor e identificação do vencedor. Intervalo de tempo esperado: mensal.

RESULTADOS ELEITORAIS

O que se busca?

Resultados para a última eleição de prefeito e vereador. Os dados eleitorais informam sobre resultados de votação e processo de votação. Quais são as maiorias e minorias eleitorais? Quantos votos são registrados, inválidos ou nulos? O índice consultou o *Open Election Initiative*¹² e o *National Democratic Institute* (NDI)¹³ para desenvolver esta categoria de dados.

Por que se busca?

A abertura desses dados permite um exame independente de cada etapa do processo de votação e contagem. Também ajuda as partes interessadas a orientar melhor seus esforços na educação de eleitores e mobilização para as próximas eleições.

Características

Resultados para eleições de prefeito devem incluir número de votos registrados, número de votos inválidos (brancos/nulos) e urnas danificadas por seção de votação. Intervalo de tempo esperado: anual.

¹¹ Disponível em: <http://standard.open-contracting.org/latest/en/schema/>. Acesso em 02/05/2018.

¹² Disponível em: <https://openelectiondata.net/en/guide/key-categories/election-results/>. Acesso em 02/05/2018.

¹³ Disponível em: https://www.ndi.org/lea_aqui_sobre_el_ndi. Acesso em 02/05/2018.

REGISTRO DE EMPRESAS

O que se busca?

Listas de empresas registradas (de responsabilidade limitada) na cidade. Essa categoria não precisa incluir dados financeiros detalhados, como balanços. Esta categoria baseia-se no trabalho do *Open Corporates*¹⁴.

Por que se busca?

Podem ser usados para diversos fins: permitir que clientes e empresas vejam com quem eles lidam, ou para ver onde uma empresa tem sede.

Características

Nome da empresa, endereço da empresa, identificador único, com todos os registros disponíveis para a cidade. Intervalo de tempo esperado: mensal.

PROPRIEDADE DA TERRA

O que se busca?

Mapas com camada de lotes que exibem limites. Também deve haver um registro de terras com informações sobre parcelas registradas. Os critérios de avaliação foram desenvolvidos em colaboração com a *Fundação Cadasta*¹⁵.

Por que se busca?

O índice centra-se na avaliação dos dados abertos da posse de terras, descrevendo as regras e os processos de propriedade da terra. O uso responsável pode permitir a segurança da posse e aumentar a transparência das transações entre propriedades.

Características

Limites do lote, identificação do lote, valor da propriedade (preço pago pela transação ou valor fiscal/venal) e tipo de posse (público, privado etc). Intervalo de tempo esperado: anual.

¹⁴ Disponível em: <http://standard.open-contracting.org/latest/en/schema/>. Acesso em 02/05/2018.

¹⁵ Disponível em: <https://openelectiondata.net/en/guide/key-categories/election-results/>. Acesso em 02/05/2018.

MAPAS DA CIDADE

O que se busca?

Mapa geográfico da cidade, incluindo rotas de trânsito, extensões de água e marcas de relevo. O mapa deve, pelo menos, ser fornecido em uma escala de 1: 250,000 (1 cm = 2,5 km), uma escala viável para a maioria dos países. O índice desenvolveu esta categoria com base em um relatório de referência do *United Nations Committee of Experts on Global Geospatial Information Management* (UNGIM)¹⁶.

Por que se busca?

A informação geográfica é fundamental para o planejamento de deslocamentos, o mapeamento da topografia, bem como os indicadores demográficos.

Características

Marcações de rotas de trânsito, marcações de relevo, marcações de extensões de água e fronteiras da cidade. Para serem qualificados, os dados devem conter projeções geográficas que permitam interpretar as coordenadas. Intervalo de tempo esperado: anual.

LIMITES ADMINISTRATIVOS

O que se busca?

Dados sobre unidades administrativas ou áreas definidas para fins de administração por um governo (local). O desenvolvimento desta categoria baseia-se no trabalho da FAO (GAUL) (FAO Global Administrative Unit Layers project)¹⁷, bem como no *United Nations Geographic Information Working Group* (UNGIWG)¹⁸.

Por que se busca?

Dados abertos sobre as zonas administrativas têm várias possibilidades de uso, com destaque para saber quais órgãos governamentais administram uma determinada região e como a riqueza é distribuída entre regiões.

Características

Limite administrativo 1 (por exemplo: áreas administrativas), limite administrativo 2 (por exemplo: bairros), coordenadas de zonas administrativas (latitude, longitude), nome do polígono (departamento, região, cidade) e fronteiras do polígono. Para serem qualificados, os dados devem conter projeções geográficas que permitam interpretar as coordenadas. Intervalo de tempo esperado: anual.

¹⁶ Disponível em: <http://standard.open-contracting.org/latest/en/schema/>. Acesso em 02/05/2018.

¹⁷ Disponível em: <https://openelectiondata.net/en/guide/key-categories/election-results/>. Acesso em 02/05/2018.

¹⁸ Disponível em: <http://www.ungiwg.org/coreDB>. Acesso em 02/05/2018.

LOCALIZAÇÕES

O que se busca?

Banco de dados de logradouros, incluindo códigos postais e os locais correspondentes em relação à latitude e longitude (ou coordenadas semelhantes em um sistema aberto). Os dados devem estar disponíveis para toda a cidade. O índice baseou-se no trabalho da *Universal Postal Union*¹⁹.

Por que se busca?

Os dados de localização aberta mostram os endereços de edifícios públicos e privados. Embora usado principalmente para encaminhar serviços postais, esses dados têm muitas outras possibilidades de utilização: calcular o número de pessoas em um distrito da cidade, fornecer serviços para residências ou para mala direta e marketing.

Características

CEPs, endereços (obrigatório, se o código postal não incluir o endereço), coordenadas (latitude, longitude) em dados disponíveis para a cidade inteira. Nota: Para serem qualificados, os dados devem conter projeções geográficas que permitam interpretar as coordenadas. Intervalo de tempo esperado: anual.

ESTATÍSTICAS SOCIOECONÔMICAS

O que se busca?

Principais estatísticas da cidade sobre indicadores demográficos e econômicos, como o Produto Interno Bruto (PIB), estatísticas de desemprego e população. Essas estatísticas podem ser publicadas como agregados para toda a cidade. A pesquisa foi baseada no trabalho do *Open Data Watch*²⁰.

Por que se busca?

Como o Open Data Watch afirma: "As estatísticas oficiais fornecem um elemento indispensável no sistema de informação de uma sociedade democrática, servindo ao governo, à economia e ao público com dados sobre a situação econômica, demográfica, social e ambiental".

Características

População da cidade (obrigatório: informações do recenseamento ou estimativas populacionais oficiais atualizadas todos os anos), Produto Interno Bruto (medido em preços atuais ou constantes, atualizado trimestralmente - a última atualização não deve ser há mais de três meses) e desemprego na cidade (números absolutos, ou expressos em porcentagem da população, atualizados trimestralmente - a última atualização não deve ser há mais de três meses).

¹⁹ Disponível em: <http://www.upu.int/fileadmin/documentsFiles/activities/addressingAssistance/manualAddressingAndPostcodeManualEn.pdf>. Acesso em 02/05/2018.

²⁰ Disponível em: <https://opendatawatch.com/>. Acesso em 02/05/2018.

ATIVIDADE LEGISLATIVA

O que se busca?

Dados sobre as propostas de leis discutidas no parlamento municipal, incluindo os votos dos parlamentares. Não deve ser confundido com as leis aprovadas, que é o banco de dados seguinte. Os dados devem estar disponíveis para o período atual da legislatura. Esta categoria de dados baseia-se no trabalho do *National Democratic Institute* (NDI) e da *Declaration of Parliamentary Openness*²¹.

Por que se busca?

Os dados sobre o processo legislativo são cruciais para a transparência parlamentar, para que os cidadãos tenham ciência do que está sendo discutido em termos de propostas para a cidade, quem apresenta tais propostas, quem vota a favor ou contra, qual o status das tramitações etc. Isso facilita mecanismos de participação direta e de cobrança dos políticos eleitos.

Características

Conteúdo da proposta, autor da proposta, situação da tramitação, votos na proposta por membro do parlamento municipal, transcrições dos debates sobre a proposta, todos disponíveis para a legislatura atual. Intervalo de tempo esperado: semanal.

LEIS DA CIDADE

O que se busca?

Dados sobre todas as leis e estatutos vigentes no âmbito da cidade (não se trata de informações sobre atividade legislativa). Esta categoria de dados baseia-se no trabalho do *National Democratic Institute* (NDI) e da *Declaration of Parliamentary Openness*.

Por que se busca?

O acesso a dados abertos do código legal de uma cidade apoia o cumprimento da lei, permite acompanhar as mudanças legais e a deliberação pública em torno de uma lei.

Características

Conteúdo da lei e sua situação, data da última alteração e emendas à lei, se aplicáveis. Devem existir informações sobre todas as leis vigentes no âmbito municipal. Intervalo de tempo esperado: trimestral.

²¹ Disponível em: <https://openingparliament.org/declaration/>. Acesso em 02/05/2018.

QUALIDADE DO AR

O que se busca?

Dados sobre a concentração média diária de poluentes do ar, especialmente aqueles potencialmente prejudiciais para a saúde humana. Os dados devem estar disponíveis para todas as estações ou zonas de monitoramento do ar em uma cidade. O índice avalia a abertura de poluentes-chave, conforme definido pela *World Health Organization* (WHO)²².

Por que se busca?

A qualidade do ar é um fator chave para a saúde humana e o meio ambiente. É importante que os cidadãos possam ter acesso à informação sobre a situação do ar em sua cidade de acordo com sua localização.

Características

Concentração dos seguintes poluentes: material particulado (MP), óxidos de enxofre (SO_x), óxidos de nitrogênio (NO_x), monóxido de carbono (CO), ozônio (O₃), e compostos orgânicos voláteis (VOCs). Os dados devem estar disponíveis por estação/zona de monitoramento de ar. Intervalo de tempo esperado: diário.

QUALIDADE DA ÁGUA

O que se busca?

Dados de qualidade da água por fonte de água. A categoria de dados considera a qualidade das fontes de água potável designadas. Se tais informações não estiverem disponíveis, a categoria refere-se a fontes de água ambientais (lagos, rios, águas subterrâneas). A principal referência utilizada foram as definições da *World Health Organization*²³ utilizadas para regulação de padrões.

Por que se busca?

Esta informação é essencial tanto para a prestação de serviços como para a prevenção de doenças transmissíveis pela água.

Características

Concentração das seguintes substâncias: coliformes fecais, arsênico, flúor, nitratos e sólidos dissolvidos totais. Os dados devem estar disponíveis por fonte de água para toda a cidade. Intervalo de tempo esperado: semanal.

²² Disponível em: http://www.who.int/phe/health_topics/outdoorair/outdoorair_aqg/en/. Acesso em 02/05/2018.

²³ Disponível em: http://www.who.int/water_sanitation_health/water-quality/guidelines/en/. Acesso em 02/05/2018.

ESCOLAS PÚBLICAS

O que se busca?

Dados sobre matrícula e localização de escolas públicas para toda a cidade, incluindo todos os níveis (ensino infantil, fundamental e médio) e todo tipo de administração (federal, estadual e municipal). Baseou-se em estudos da FGV DAPP²³ construídos a partir de dados públicos sobre educação.

Por que se busca?

Esta é uma das dimensões específicas para o índice ao nível de cidades. A educação é um aspecto muito importante do desenvolvimento e não deve estar fora do índice. Os dados requeridos são o mínimo de informações em escolas públicas em uma cidade.

Características

Número de matrículas por escola, coordenadas (endereço ou latitude / longitude). Dados disponíveis por escola pública e todos os níveis da educação básica. Intervalo de tempo esperado: anual.

TRANSPORTE PÚBLICO

O que se busca?

Informações sobre o horário dos transportes públicos, itinerários, estações/paradas e ciclofaixas para toda a cidade. Devem ser apresentados para cada meio de transporte (ônibus, metrô, trem etc.) e para cada linha e parada/estação. Baseou-se em estudos da FGV DAPP²⁵ construídos a partir de dados públicos sobre mobilidade urbana.

Por que se busca?

Também é uma das dimensões específicas para o índice ao nível de cidades. O transporte público não deve estar fora de uma discussão sobre dados abertos para cidades, uma vez que a mobilidade urbana é uma questão discutida globalmente. Os dados necessários são substanciais para a informação do cidadão e para o desenvolvimento de aplicativos.

Características

Paradas/estações para cada modo de transporte, cronograma de linha para cada parada/estação, itinerários para cada linha de todos os modos de transporte público, cartografia de ciclofaixas/ciclovias para toda a cidade. Os dados devem mostrar conexões entre diferentes modais de transporte. Intervalo de tempo esperado: mensal.

²⁴ Disponível em: <http://dapp.fgv.br/educacao/>. Acesso em 02/05/2018.

²⁵ Disponível em: <http://dapp.fgv.br/cidades/>. Acesso em 02/05/2018.

ESTATÍSTICAS DO CRIME

O que se busca?

Indicadores básicos sobre crime como registros de assaltos, homicídios, estupros e apreensão de armas de fogo, disponíveis para toda a cidade por bairro ou divisão equivalente. Para as armas de fogo, os dados devem incluir a classificação da apreensão - tipo de arma e calibre - e se é legal ou ilegal. Baseou-se em estudos da FGV DAPP²⁶ na linha de pesquisa aplicada Segurança e Cidadania.

Por que se busca?

Também é uma das dimensões específicas para o índice ao nível de cidades. As estatísticas do crime são importantes em geral, mas no nível da cidade e, em especial, para os países em desenvolvimento, é crucial. Esta nova dimensão baseia-se nos indicadores mais utilizados para medir a insegurança, reconhecendo que este é um fenômeno muito difícil de medir.

Características

Número de assaltos, número de homicídios, número de estupros e apreensão de armas. Dados disponíveis por bairro ou equivalente. Intervalo de tempo esperado: semestral.

Fonte e Elaboração: FGV DAPP e OKBr.

²⁶ Disponível em: <http://dapp.fgv.br/seguranca-e-cidadania/>. Acesso em 02/05/2018.

6. Perguntas do questionário

Cada conjunto de dados em cada local é avaliado usando uma série de questões que examinam a abertura dos conjuntos de dados com base na Definição de Abertura — *Open Definition* — e na Carta de Dados Abertos — *Open Data Charter*. Cada pergunta de pesquisa mede um aspecto legal, técnico ou prático crucial da abertura de dados. Com essa abordagem, buscamos reduzir potenciais avaliações tendenciosas em relação a aspectos únicos de abertura.

Tabela 2 - Detalhamento das categorias de dados do ODI

SEÇÃO A - SOBRE O LEVANTADOR

A1

Dê uma nota de 1 a 3 para o seu conhecimento sobre a categoria (1 = nem um pouco familiar, 3 = conhecimento avançado).

Descrição: Trata-se de uma pergunta subjetiva, **não pontuada**, que visa a captar uma percepção do conhecimento do levantador sobre a categoria de dados analisada.

Fundamentação: O objetivo dessa pergunta é captar uma informação adicional para conhecimento do público que participou dos surveys, mas não diz respeito aos dados em si.

A2

Dê uma nota de 1 a 3 para o seu conhecimento sobre dados abertos (1 = nem um pouco familiar, 3 = conhecimento avançado)

Descrição: Trata-se de uma pergunta subjetiva, **não pontuada**, que visa a captar uma percepção do conhecimento do levantador sobre dados abertos em geral.

Fundamentação: O objetivo dessa pergunta é captar uma informação adicional para conhecimento do público que participou dos surveys, mas não diz respeito aos dados em si.

SEÇÃO B - SOBRE OS DADOS

B1

Os dados são coletados pelo governo (ou por terceiros que respondem pelo governo)?

Descrição: Essa pergunta delimita o universo de dados considerados, deixando de lado i) dados coletados por organizações que não representam o governo; ii) dados coletados, mas não para o nível governamental relevante; iii) dados não coletados. O caso em que um terceiro responde pelo governo ocorre por exemplo com empresas estatais ou empresas contratadas que prestam serviços públicos para o governo.

Fundamentação: A coleta de dados por si só não é uma característica que define a abertura de dados, mas é uma condição para que a avaliação seja possível. Por essa razão, essa é uma questão **não pontuada**, mas que **zera o resultado final caso a resposta seja “não”**.

B2

Os dados estão disponíveis online sem necessidade de se registrar ou solicitar acesso aos dados? (em caso afirmativo, colocar o (s) link (s) onde a informação foi encontrada)

Descrição: Essa pergunta visa a identificar se os dados são disponibilizados em um site público. O “não” cabe nos casos em que os dados não estiverem disponíveis online ou estiverem disponíveis somente após um cadastro, solicitando os dados de um funcionário público por e-mail, preenchendo um formulário de contato ou outro processo administrativo similar.

Fundamentação: A disponibilidade online é um requisito para a abertura: todos têm que ter acesso online a dados específicos. Além disso, é uma condição para todas as perguntas a seguir. O registro obrigatório pode dificultar e em muitos casos impedir as pessoas de obterem e utilizarem os dados. **Uma resposta afirmativa conta 15 pontos na nota final.**

B2.1

Caso a resposta anterior tenha sido negativa, os dados estão disponíveis online após o cadastro?

Descrição: Essa pergunta serve para determinar a continuidade do levantamento. Dados indisponíveis online, mesmo após registro, não poderão ser avaliados. Dados disponíveis após registro serão avaliados, mas perderão a pontuação da questão anterior.

Fundamentação: Uma resposta negativa na questão anterior, mesmo que após o registro o levantador tenha obtido acesso aos dados, **não pontua**. Se mesmo após o registro os dados não foram obtidos, a categoria **recebe zero na avaliação final**.

B2.2

Onde foram encontrados os dados?

Descrição: O usuário indica uma URL e a descrição da URL. Às vezes é possível encontrar dados em muitos lugares na rede, por isso é possível submeter vários links aqui. A resposta serve para divulgação dos resultados a fim de que todos possam localizar as informações avaliadas com facilidade.

Fundamentação: **Questão não pontuada.** Aqui o levantador insere um ou mais links onde foram encontradas as informações.

B2.3

O quanto você concorda com a seguinte afirmação? “foi fácil para mim encontrar os dados”. Atribua uma nota de 1 a 4, sendo 1 = discordo fortemente e 4 = concordo fortemente.

Descrição: Os levantadores respondem com base em uma escala Likert. Trata-se de uma pergunta subjetiva, **não pontuada**, que visa a captar a percepção de facilidade de acesso aos dados pelo levantador.

Fundamentação: O objetivo dessa pergunta é captar uma informação adicional para aferir, ainda que qualitativamente, a facilidade de acesso aos dados avaliados. Nos casos em que as notas foram 1 ou 2 aqui, contabilizamos, para fins de análise, o indício de problemas de acesso na categoria.

B2.4

Conte-nos os passos necessários para encontrar os dados.

Descrição: O levantador relata o passo a passo para obter os dados. Exemplo: usei a pesquisa no Google, cliquei no primeiro link, acessei a seção “x” do site etc.

Fundamentação: Essa resposta ajuda a qualificar o resultado da questão anterior. **Não é pontuada.**

B3

Confirme se a base de dados avaliada contém as seguintes características.

Descrição: Nessa questão, o levantador assinala quais características da base de dados foram encontradas (essas características estão destacadas na quarta coluna do quadro anterior).

Fundamentação: A questão **não é pontuada**, mas é utilizada para definir se há uma base de dados que qualifica para a avaliação. Além disso, se houver mais de uma base de dados que qualifica, pode ser utilizada como critério de completude: a mais completa é a que será avaliada.

B4

Os dados estão disponíveis gratuitamente?

Descrição: Os dados são gratuitos se você não precisa pagar para ter acesso. Não se espera que os usuários paguem por conjuntos de dados para avaliá-los, então, uma resposta negativa nesta questão, usualmente, implicaria num zero para a categoria.

Fundamentação: Os dados devem ser gratuitos para serem acessíveis a todos e, por isso, constam como critério para abertura de dados. Por ser uma restrição forte aos dados quando são pagos, **uma resposta afirmativa conta 20 pontos na nota final.**

B5

É possível baixar os dados de uma só vez?

Descrição: A resposta será afirmativa se for possível baixar todos os dados de uma vez a partir da URL encontrada. Caso os arquivos de dados para baixar sejam muito grandes, os downloads também poderão ser organizados por intervalo de tempo ou divididos em sub-arquivos. A resposta será “não” se for necessário conduzir muitas etapas manuais para baixar os dados, ou se só for possível recuperar poucas partes de um grande conjunto de dados ao mesmo tempo (por exemplo, através de uma interface de busca).

Fundamentação: Esta questão recompensa a possibilidade técnica de recuperar todos os dados da internet sem ter que baixar dezenas de pequenas informações, ter acesso a dados apenas através de uma interface de busca, enviar solicitações, ter captchas ou outros limites para baixar. Trata-se de um critério para abertura de dados importante na questão da usabilidade e facilidade de trabalhar os dados: **uma resposta afirmativa conta 20 pontos na nota final.**

B6

Os dados devem ser atualizados a cada [intervalo de tempo]: os dados estão atualizados?

Descrição: A resposta é baseada na data em que o levantamento é feito. Leva em consideração as especificidades de cada banco de dados (em alguns casos, a exigência será anual, em outras, mensal, diário etc). A resposta será "não" se não for possível determinar uma data ou se os dados estão efetivamente desatualizados.

Fundamentação: A provisão oportuna desses dados é crucial, mas sabemos que alguns dados não são tão sensíveis ao tempo quanto outros. Por essa razão, cada dimensão tem uma periodicidade esperada, conforme detalhado no quadro anterior. Trata-se de um critério para abertura de dados: **uma resposta afirmativa conta 15 pontos na nota final.**

B7

Os dados são abertamente licenciados / em domínio público?

Descrição: Esta pergunta mede se alguém legalmente pode usar, modificar e redistribuir dados para qualquer finalidade. Só então os dados são considerados verdadeiramente abertos. A resposta será afirmativa se uma licença explícita tiver sido encontrada (exemplo: *Creative Commons*), se há uma declaração sobre os termos de uso dos dados, deixando claro seu domínio público ou se é feita referência explícita a algum dispositivo legal que fornece as bases para a divulgação deste tipo de informação, incluindo critérios de utilização. A resposta negativa ocorre se a base de dados for protegida por direitos autorais, patentes ou restrições semelhantes.

Fundamentação: A usabilidade jurídica dos dados é um requisito básico da definição de abertura. É um pré-requisito para a usabilidade irrestrita para todos. **Uma resposta afirmativa conta 10 pontos na nota final.**

B7 .1

Coloque uma URL da licença aberta ou declaração do domínio público dos dados.

Descrição: O usuário indica uma URL. A resposta serve para divulgação dos resultados a fim de que todos possam localizar as informações avaliadas com facilidade.

Fundamentação: Questão não pontuada. Aqui o levantador deve colocar um link caso a resposta anterior tenha sido positiva.

B8

Os dados estão em formato de arquivo aberto e processável por máquina?

Descrição: O usuário marca quais formatos foram encontrados dentro de uma lista de formatos abertos considerados processáveis por máquina. Um formato de arquivo é chamado processável por máquina se o seu computador puder acessar e modificar elementos únicos em um arquivo de dados. O índice considera que os formatos serão abertos se eles puderem ser processados completamente por pelo menos um software livre e ferramenta de código aberto. Potencialmente, esses formatos permitem que mais pessoas usem os dados porque as pessoas não precisam comprar um software específico para abri-lo.

Fundamentação: A legibilidade por máquina é um grande aprimoramento da usabilidade técnica. No entanto, se um arquivo só é utilizável com software proprietário (como ArcGIS), há uma limitação na usabilidade. Os formatos abertos não colocam direitos autorais, restrições monetárias ou outras restrições quanto à sua utilização (importante para as pessoas que não podem ou não querem pagar por software proprietário). Trata-se de um critério-chave para abertura de dados importante na questão da usabilidade. **Uma resposta afirmativa conta 20 pontos na nota final.**

B9

Quanto esforço humano é necessário para usar os dados. (1 = pouco ou nenhum esforço, 3 = esforço extensivo)

Descrição: Os levantadores atribuem uma nota à capacidade de transformação dos dados em informação. Trata-se de uma pergunta subjetiva, **não pontuada**. A usabilidade depende do contexto e dos propósitos para os quais uma pessoa deseja usar os dados.

Fundamentação: O objetivo dessa pergunta é captar uma informação adicional para aferir, ainda que qualitativamente, a usabilidade dos dados. Nos casos em que as notas foram 1 ou 2 aqui, contabilizamos, para fins de análise, o indício de problemas de usabilidade na categoria.

Fonte e Elaboração: FGV/DAPP e OKBr.

Uma observação importante no caso desta avaliação é que, em geral, muita ambiguidade é gerada em torno da apresentação de uma licença de dados abertos. Enquanto alguns conjuntos de dados apresentam licenças explícitas (ex: *Creative Commons*), outros apresentam declarações textuais sobre o domínio público ou possibilidade de utilização dos dados, enquanto que alguns não apresentam nada muito específico, mas indicam de alguma forma que a publicação daqueles dados está em conformidade com algum aparato legal.

Como o Brasil possui dispositivos como a Lei de Acesso à Informação (Lei 12.257/2011), a Lei de Responsabilidade Fiscal (Lei Complementar 101/2000), entre outros, muitos provedores de informação eximem-se da necessidade de esclarecer melhor a possibilidade de utilização dos dados, por entenderem que isso já é feito nos dispositivos legais. Por essa razão, a avaliação focou nos casos em que havia uma restrição de uso muito clara (proteção por patente ou declaração de limitação da republicação do dado) para atribuir uma resposta negativa à esta pergunta, ainda que, como se verá nos resultados, esse não foi um problema frequente.

Outra nota importante se refere à distribuição de pesos utilizada nesta avaliação. Ela é ligeiramente diferente da que foi realizada no ano passado para Rio de Janeiro e São Paulo, bem como da avaliação de países. As mudanças foram feitas após a realização do piloto no ano passado, que detectou algumas incompatibilidades com a realidade das cidades brasileiras. Por essa razão, exercícios de comparação no tempo devem ser realizados com cuidado. Em todos os momentos em que isso foi feito neste documento, recalculamos o índice do ano passado usando os pesos deste ano, para ter alguma noção de como evoluiu o status dos dados abertos das duas cidades que participaram do piloto.

A tabela a seguir mostra como era a distribuição de pesos original do ODI global, criado pela Open Knowledge International, que foi utilizada para a realização da avaliação piloto no ano passado para São Paulo e Rio de Janeiro na coluna da esquerda. Na coluna do meio, são mostrados os pesos aplicados neste ano, enquanto que na coluna da direita há uma explicação do motivo da mudança. Vale ressaltar que a soma da pontuação na metodologia antiga resultava em 90 e não em 100. Entendemos que seria mais intuitivo utilizar uma pontuação que somasse 100 e, para chegar a este valor, foram atribuídos pontos adicionais a certas perguntas conforme descrito abaixo.

Tabela 3

Mudanças de ponderação aplicadas no ODI Cidades edição 2018, ano-base 2017

Disponível sem registro ou solicitação de acesso		
B2	Metodologia original	Metodologia ODI Cidades 2018
	15	15
Justificativa: Sem alteração. Mantivemos uma pontuação menor por considerar que existem aspectos mais limitantes na definição de abertura.		
Disponível sem custo financeiro		
B4	Metodologia original	Metodologia ODI Cidades 2018
	15	20
Justificativa: Atribuímos pontos adicionais para esta categoria por entender que uma restrição nesse quesito é muito limitante (mais do que outros critérios) à abertura de dados.		
Disponível para download de uma só vez		
B5	Metodologia original	Metodologia ODI Cidades 2018
	15	20
Justificativa: Atribuímos pontos adicionais para esta categoria pela questão da usabilidade, que fica muito limitada caso o download não esteja disponível de forma única.		

Disponível de forma atualizada

B6

Metodologia original	Metodologia ODI Cidades 2018
15	15

Justificativa: Sem alteração. Mantivemos uma pontuação menor por considerar que existem aspectos mais limitantes na definição de abertura.

Disponível com licença aberta ou declaração de domínio público

B7

Metodologia original	Metodologia ODI Cidades 2018
20	10

Justificativa: Conforme discussão na página 33, a existência de certos dispositivos legais gera ambiguidades sobre a necessidade de publicar uma licença para dados públicos no Brasil. Por isso, entendemos que essa questão deveria receber um peso menor.

Disponível em formato aberto/processável por máquina

B8

Metodologia original	Metodologia ODI Cidades 2018
10	20

Justificativa: O formato aberto é uma condição muito importante para definir abertura. Por essa razão, entendemos que deveria estar entre os atributos mais pontuados. Uma limitação aqui representa uma forte restrição à usabilidade e aplicabilidade dos dados.

Fonte e Elaboração: FGV/DAPP e OKBr.

7. Como ler os resultados finais

Conforme explicado nas seções acima, o índice examina dados específicos usando questões específicas de pesquisa. O resultado é uma pontuação final que deve ser lida com cuidado. Em primeiro lugar, refere-se exclusivamente a dados com características obrigatórias. Se nenhum conjunto de dados pode ser encontrado online combinando essas características, os dados não serão considerados disponíveis (igualando uma pontuação de 0%). Além disso, as perguntas da pesquisa conferem diferentes aspectos do acesso e da usabilidade de dados (veja a tabela a seguir). Isso significa que, por trás de pontuações bastante altas, muitas vezes não encontramos dados abertos, mas dados de acesso controlados ou dados públicos em formatos mal estruturados ou não processáveis por máquina. A pontuação, portanto, não mostra um aumento linear de abertura. Em vez disso, destaca áreas onde o governo pode melhorar a publicação de dados abertos.

Um exemplo: podemos avaliar os dados do orçamento em formato PDF, que podem estar em domínio público, disponíveis online gratuitamente, mas em um formato praticamente inutilizável. Estes dados são apresentados como 80% abertos. A pontuação sugere um grau bastante alto de abertura, mas, na verdade, os dados não estão abertos. Consideramos que um dado é de fato aberto se a dimensão obtém 100%. Com essa abordagem, atribuindo notas intermediárias a bases de dados não plenamente abertas, o índice procura demonstrar quais dados já estão disponíveis e como isso pode ser melhorado. É, portanto, importante ler cuidadosamente como os dados são publicados.

Dependendo de quais itens da pesquisa são verificados, encontramos:

Tabela 4

Leitura dos resultados do ODI Cidades edição 2018, ano-base 2017

Dados abertos

Descrição: Dados considerados abertos podem ser usados, modificados e compartilhados gratuitamente por qualquer pessoa para qualquer propósito. Critérios principais: legíveis por máquina, formatos abertos, pleno acesso, disponíveis de forma completa, sem custo e juridicamente em domínio público.

Pontuação: 100%

Dados públicos, mas não plenamente abertos

Descrição: Dados públicos, porém com alguma limitação que não os fazem plenamente abertos. Os dados são públicos se podem ser vistos pelo público online sem restrições. Podem estar disponíveis apenas em formato fechado para edição (ex: PDF) ou desatualizados, com alguma limitação de acesso (cadastro ou proteção por patente ou algo do gênero), disponíveis apenas parcialmente, ou sem possibilidade de download.

Pontuação: Notas intermediárias entre 0 e 100%.

Ausência de dados

Descrição: A ausência de dados significa que os governos não produzem dados sobre um fenômeno ou produzem, mas não divulgam. Eles mostram que alguns governos ainda têm um longo caminho a percorrer antes de se prepararem para produzir dados.

Pontuação: 0%

Fonte e Elaboração: FGV/DAPP e OKBr.

8. Nossa coleta e análise de dados

A coleta de dados do índice é feita em três fases:

- **Fase de levantamento**
- **Fase de revisão**
- **Garantia de qualidade da revisão dos resultados**

Fase de levantamento

O índice obtém seus dados através de colaboração em massa. Para o índice de cidades, foram formados grupos locais nas cidades interessadas, que passaram por um processo de seleção com base em critérios de regionalização e capacidade do grupo. A seleção deu origem a uma amostra de 12 cidades. No entanto, ao longo do processo, alguns não foram capazes de concluir o levantamento, restando uma amostra final de oito municípios.

Cada grupo foi formado por contribuintes interessados em dados abertos do governo e que podem avaliar a disponibilidade e a qualidade dos conjuntos de dados abertos em suas respectivas localidades. Os grupos tomaram conhecimento do levantamento através de divulgação nas mídias sociais, comunicações regulares nos Fóruns de Dados Abertos Governamentais²⁷ e ativamente em rede, em conferências ou eventos.

A capacidade de busca de dados também tem um impacto na qualidade dos dados que coletamos. Os colaboradores têm conhecimentos e cenários diversos em dados abertos e às vezes precisam de ajuda para encontrar os dados. A seção a seguir explica como tentamos lidar com esse problema.

É importante destacar o **caráter contextual** da avaliação do ODI. Os dados foram levantados pelos grupos locais entre novembro e dezembro de 2017, sendo posteriormente revisados em 2018. Foi levado em consideração o estado de dados abertos em 2017, para

²⁷ Disponível em: <https://discuss.okfn.org/>. Acesso em 02/05/2018.

manter a consistência de comparação entre as cidades. Por essa razão, bases de dados que foram criadas ou divulgadas em 2018 não foram consideradas nesta avaliação.

Fase de revisão

Para fornecer resultados confiáveis e válidos, cada levantamento deve ser revisado. Para isso, a OKBr, em parceria com a FGV DAPP, já com a experiência adquirida de outros anos com o ODI, realizou uma extensa revisão dos conteúdos submetidos, consultando, sempre que necessário, especialistas de cada área para tirar dúvidas e especialistas em dados abertos por meio do fórum de discussão.

A revisão passou pelo diálogo com os grupos locais de levantadores, identificando eventuais inconsistências, o que originou na re-submissão de algumas bases de dados, bem como na modificação de alguns parâmetros. Esse processo foi pautado pelo respeito às orientações da *Open Definition* e da metodologia do índice. Com isso, os levantamentos finalmente aceitos na plataforma contemplam o resultado de uma revisão sistemática e do retorno aos levantadores locais.

A informação oriunda desse processo é usada para refinar nossas categorias de dados e como orientação para futuras edições. Para isso, documentamos nossas descobertas em diários de revisão.

Garantia de qualidade da revisão

Para garantir a qualidade da revisão, foram empreendidos os seguintes passos:

Verificação periódica dos fóruns para incorporar e responder comentários da comunidade Open Knowledge, incluindo os próprios levantadores.

Comparação com os resultados do ano passado, quando possível, verificando se as mesmas URLs de origem são usadas, se há alguma iniciativa nova de dados abertos ou se algo do ano passado foi descontinuado, identificando os motivos por trás de cada mudança. Isso foi feito para São Paulo e Rio de Janeiro, que estiveram no piloto do ano passado.

Checagem junto aos levantadores para refinar informações apresentadas. Nesse processo, respostas foram alteradas e bases de dados foram submetidas novamente.

9. Fase de diálogo público

Uma vez que os resultados são publicados, a sociedade civil e o governo são convidados a fornecer opiniões sobre os resultados e dizer como eles pensam que poderíamos fortalecer a avaliação. Esta fase de diálogo estará aberta por um mês após a divulgação dos resultados e as autoridades locais podem submeter pedidos de alteração com base em dados disponíveis no ano de avaliação, no caso 2017. No ano passado, a título de exemplo, esse processo rendeu a alteração de uma base de dados da cidade de São Paulo. Isso é possível pois, em alguns casos, nem os levantadores nem os revisores foram capazes de encontrar uma determinada informação, ainda que ela existisse em meio online.

Em geral, os formuladores de políticas públicas e os responsáveis pela divulgação de dados abertos fazem referência aos resultados e destacam publicamente seu avanço no índice. Além disso, os resultados apontam como cada governo pode avançar na publicação de dados abertos.

O ODI opera num equilíbrio entre a fundamentação nas realidades e prioridades dos governos para que eles possam melhorar suas pontuações, e o destaque às demandas de dados da sociedade civil. Através de um diálogo aberto, o índice torna-se uma ferramenta útil para ambas as partes, diagnosticando como a oferta e demanda de dados abertos podem corresponder e estimular uma captação mais construtiva para melhorar a publicação de dados abertos no nível das cidades.

Para mais informações consultar <http://br-city.survey.okfn.org/>

RESULTADOS

Aqui apresentamos os resultados da avaliação referente ao estado dos dados abertos em 2017 para as oito cidades avaliadas. Esses resultados nos permitem entender o estado da arte da política de dados abertos e transparência das cidades avaliadas bem como diagnosticam as frentes prioritárias de ação visando ao máximo de transparência.

A seguir, apresentamos análises sob diferentes espectros de avaliação dos resultados obtidos durante o levantamento. Após uma visão dos resultados gerais, apresentaremos os resultados observados por dimensão, analisando quais conjuntos de dados exibem maior necessidade de esforço por parte das autoridades governamentais. Em seguida, apresenta-se a categorização dos problemas observados nos conjuntos de dados. Logo após, identificamos os problemas mais frequentes em cada dimensão. Por fim, analisamos a forma como os gargalos de abertura e transparência de dados são mais frequentes em cada cidade.

Visão geral

A pontuação do Índice de Dados Abertos pode ser visualizada através do *escore*, que avalia a adequação dos dados disponibilizados pelo governo aos critérios de transparência utilizados em diversos países do mundo, e através do *%open*, que calcula qual o percentual dos conjuntos de dados avaliados que atende a todos os critérios da metodologia. A tabela a seguir mostra os resultados das duas métricas para o conjunto de cidades participantes.

Tabela 5

Pontuação das cidades no ODI segundo *escore* e *%Open*

Cidades	Escore	Cidades	%Open
São Paulo	84%	São Paulo	47%
Rio de Janeiro	75%	Belo Horizonte	35%
Belo Horizonte	73%	Rio de Janeiro	29%
Porto Alegre	68%	Brasília	29%
Brasília	68%	Porto Alegre	23%
Salvador	55%	Uberlândia	17%
Uberlândia	53%	Natal	11%
Natal	43%	Salvador	5%

Fonte e Elaboração: FGV/DAPP e OKBr.

A média geral entre as cidades foi de 65%, com resultados individuais variando entre 43% e 84%. Neste ano, o estudo avaliou 17 dimensões, em contraste com 18 no ano passado. Das 136 bases de dados avaliadas (17 por cidade, 8 cidades), 25% (ou 34) estavam plenamente abertas.

São Paulo foi a cidade, dentre as oito avaliadas, que apresentou o maior *escore* e também o maior *%open*. Isso significa que a cidade foi a que obteve maior êxito tanto em divulgar bases 100% abertas quanto em aproximar suas bases públicas aos critérios da *Open Definition*. Isso não significa, por outro lado, que todos

os desafios já foram atendidos, pois mesmo assim, menos da metade das bases de dados avaliadas atendia a plenitude dos critérios do ODI. Oito dos 17 datasets avaliados para São Paulo atendem a todos os critérios da metodologia, mas não necessariamente apresentam todas as características²⁸.

Para a capital paulista, é possível comparar com o resultado da avaliação do ano passado²⁹. Há uma pequena melhoria qualitativa em relação à última avaliação: mais características buscadas pelo índice foram disponibilizadas através do portal GeoSampa (ex: coordenadas para escolas), o que, para o cidadão comum, que não seja um profissional no tipo de dado visualizado ou algum pesquisador, é uma ótima prática de transparência e clareza na disponibilização de informações.

Contudo, em relação aos parâmetros gerais de transparência, a diferença entre o ano passado e o atual foi pouco significativa. Comparando com o ano passado, a capital paulista obteve uma diferença positiva de 2 pontos percentuais no escore³⁰, o que, dado que temos uma dimensão a menos neste ano (Previsão do Tempo), é um resultado muito semelhante. As bases com nota máxima de abertura foram diferentes. Em relação ao ano passado, a dimensão **Estatísticas Criminais** passou a atender todos os critérios metodológicos e, em **Atividade Legislativa**, houve perda de pontos devido à **impossibilidade de download da base completa**.

²⁸ Em Estatísticas Criminais, não há informações detalhadas sobre apreensões de armas. Para Propriedade da Terra, não é disponibilizado os limites das propriedades.

²⁹ Replicamos as mudanças metodológicas implementadas neste ano nos resultados do levantamento referente a 2016 para permitir comparabilidade. Isso foi feito apenas para São Paulo e Rio de Janeiro, as cidades que foram avaliadas no projeto-piloto realizado em 2017.

³⁰ O escore de 2016 foi 82% e o de 2017, 84%. O %open não sofreu alteração na prática, pois considerando 17 dimensões (retirando previsão do tempo do ano passado), São Paulo continua com oito bases 100% abertas, logo, 47% no indicador.

Na capital fluminense, a situação é semelhante. Quando comparamos com os resultados do ano passado, há uma diminuição de 1% no escore, diferença atribuível ao fato de que a dimensão Localizações ter obtido nota máxima em 2017 e 80% em 2016, enquanto que a dimensão Mapas da Cidade caiu de 100% para 85%³¹. Assim como em São Paulo, no Rio de Janeiro a quantidade de bases de dados que obtiveram abertura completa nos dois anos foi a mesma (no caso, 5 das 17, portanto 29%).

Uma pontuação alta no critério %open demonstra a intenção governamental de conceder à população dados abertos de qualidade, atendendo tanto a necessidade de pesquisadores e desenvolvedores que queiram trabalhar com essas bases quanto de cidadãos comuns, que desejem informações sobre seu governo, políticas públicas, leis, local onde vivem etc.

Podemos obter insights mais interessantes comparando, para cada cidade, seu escore e seu %open, em vez de simplesmente comparar o desempenho das cidades. É possível observar, por exemplo, que os rankings não são simétricos. A cidade do Rio de Janeiro, por exemplo, possui o segundo melhor escore, o que significa que cada dataset em particular apresenta desafios pontuais para alcançar a plena abertura, mas seu %open é menor do que o de Belo Horizonte, por exemplo, que apresenta um número ligeiramente maior de bases 100% abertas.

Não há qualquer juízo sobre uma métrica ser melhor ou pior do que a outra, mas a comparação dos resultados sugere caminhos diferentes para ampliar a abertura de dados: enquanto, numa visão geral, o Rio de Janeiro pode focar sua atuação em verificar quais os problemas pontuais de suas bases de dados que não as tornam 100% abertas, Belo Horizonte pode tentar olhar para suas bases 100% abertas e verificar como replicar para as demais.

³¹ Os mapas da cidade utilizados não sofreram atualizações no período entre as duas avaliações, o que resultou numa diminuição no escore da dimensão.

Um destaque no aspecto de abertura das bases é a cidade de Salvador, que obteve um escore de 55%, ficando em sexto lugar, mas com %open de 5% (apenas uma das bases de dados atende integralmente aos critérios), ficando em oitavo lugar. Esse tipo de resultado mostra que houve um esforço para tornar públicos e acessíveis online informações pertinentes a vida pública, contudo ainda com pouca aderência aos critérios da *Open Definition*.

Para a cidade de Natal, última colocada no escore, foram encontrados dados públicos para apenas dez dos 17 conjuntos de dados avaliados. Destes, apenas dois obtiveram uma avaliação de abertura máxima, nenhum deles disponibilizados pela prefeitura. Os resultados mostram que é necessário um esforço maior das entidades governamentais em providenciar mais informações relevantes para a vida pública dos seus cidadãos, aderindo às boas práticas de transparência e acesso à informação adotadas em outras cidades brasileiras e no mundo.

Os resultados permitem constatar que há uma intenção dos mecanismos governamentais de trazer ao cidadão informações pertinentes à vida pública, com a criação de portais de transparência aderentes à Lei de Acesso à Informação, tornando os dados públicos. Contudo, ainda há muito a avançar no sentido desses dados públicos se tornarem, de fato, abertos, além de desafios na qualidade dos dados, nos formatos disponibilizados e na estabilidade dos sites governamentais. Durante o processo de pesquisa, diversos sites utilizados apresentaram falhas ou estiveram fora do ar durante algumas horas ou até mesmo dias.

Resultados por dimensão

Nesta subseção, apresentamos os resultados consolidados por dimensão, de modo a identificar quais são as frentes prioritárias de ação dos municípios avaliados e onde estão os principais desafios para o avanço das políticas de dados abertos.

Como se pode ver na tabela a seguir, os quatro conjuntos de dados com mais gargalos foram os de: **Propriedade da Terra, Registro de Empresas, Qualidade do Ar e Qualidade da Água**. Os dados de Propriedade da Terra, tidos como um gargalo internacionalmente, só estavam publicados para São Paulo e Brasília. Já os dados de **Qualidade do Ar e Registro de Empresas** só estavam disponíveis em metade das cidades avaliadas. Os dados disponibilizados sobre monitoramento da água, ainda que disponíveis em todas as cidades, traziam muitos problemas de usabilidade, apesar de poucas dificuldades na obtenção das informações. Ou seja, são informações fáceis de obter, mas de pouca qualidade.

Na avaliação do %open percebemos que cinco dimensões não apresentaram nenhuma base de dados 100% aberta: **Registro de Empresas, Qualidade da Água, Transporte Público, Compras Públicas e Atividade Legislativa**. Cruzando as duas informações, percebemos que as dimensões de Registro de Empresas e Propriedade da Terra são especialmente críticas entre as cidades.

Tabela 6
Avaliação média das dimensões do ODI

Datasets	Score	Datasets	%open
Resultados Eleitorais	100%	Resultados Eleitorais	100%
Escolas Públicas	99%	Escolas Públicas	87%
Estatísticas Socioeconômicas	87%	Estatísticas Criminais	75%
Estatísticas Criminais	85%	Orçamento Público	37%
Gastos Públicos	83%	Gastos Públicos	25%
Orçamento Público	81%	Limites Administrativos	25%
Limites Administrativos	76%	Estatísticas Socioeconômicas	12%
Leis em Vigor	73%	Leis em Vigor	12%
Atividade Legislativa	70%	Mapas da Cidade	12%
Mapas da Cidade	68%	Localizações	12%
Compras Públicas	67%	Qualidade do Ar	12%
Transporte Público	57%	Propriedade da Terra	12%
Localizações	53%	Atividade Legislativa	0%
Qualidade da Água	53%	Compras Públicas	0%
Qualidade do Ar	35%	Transporte Público	0%
Registro de Empresas	30%	Qualidade da Água	0%
Propriedade da Terra	21%	Registro de Empresas	0%
Mediana	70%	Média	25%

Fonte e Elaboração: FGV/DAPP e OKBr.

Durante a pesquisa, foi observado que parte dos dados em âmbito municipal é disponibilizada no Brasil de forma consolidada na esfera federal, como é o caso dos dados de **Resultados Eleitorais** (o que explica o fato de todo o universo de cidades ter obtido nota máxima). Em **Estatísticas Socioeconômicas**, este também frequentemente é o caso. Em relação à presente avaliação, observamos que estes dados estão em sua maioria adequados aos parâmetros de transparência internacionais, à exceção de informações sobre o PIB, disponibilizadas pelo IBGE, que são desatualizadas quando comparadas com o requerido pela metodologia do ODI.

Através dos resultados do estudo, destaca-se a relevância de uma base de dados estruturada na forma de um mapa online que disponibiliza informações referentes à estrutura geográfica e urbana da cidade. Frequentemente, diversos conjuntos de dados foram extraídos de plataformas centralizadas (congregam diferentes conjuntos de dados de forma integrada em um único portal) como **Limites Administrativos, Mapas da Cidade e Localizações**. Iniciativas como estas são um facilitador ao acesso à informação dos cidadãos e consideradas boas práticas de transparência.

Os conjuntos de dados de **Transporte Público** apresentaram uma falha recorrente ao passo que em nenhuma das cidades avaliadas havia informações adequadas sobre conexões entre diferentes tipos de transporte. Outro problema frequente foi a falta de dados atualizados. Fontes privadas, como o Google, disponibilizam informações de melhor qualidade e clareza sobre o assunto, sendo um ponto de atenção importante para os municípios avaliados.

Já em relação aos dados sobre **Atividade Legislativa**, foram localizados problemas entre todas as cidades. Informações sobre os votos nas sessões plenárias dificilmente são disponibilizadas junto aos projetos de lei, e para nenhum dos casos a transcrição dos debates era associada a uma lei em votação. Ambas informações devem ser obtidas manualmente procurando as leis por data em que foram debatidas em sessão plenária, ou número do projeto, separadamente das leis. Essa prática dificulta o acompanhamento dos debates legislativos e, assim, nenhum dos municípios obteve nota máxima na avaliação deste conjunto de dados.

Detalhamento dos gargalos

Em linha com os relatórios da última avaliação, uma das metodologias aqui empregadas para análise dos dados abertos para as cidades avaliadas é a contagem dos problemas encontrados em cada dimensão e, então, a categorização em dois tipos: **usabilidade e processo**. Na categoria usabilidade, estão listados problemas relacionados à dificuldade de uso dos dados e à capacidade de transformação destes em informação, enquanto que na categoria processo, identificamos dificuldades de acesso aos dados e omissão/ausência de licença ou declaração de que os dados estejam em domínio público.

Tabela 7
Categorização dos problemas em usabilidade e processo

Usabilidade	Processo
Dataset incompleto	Acesso restrito
Desatualizado	Dificuldade de localizar dados
Indisponibilidade de formato aberto	Download da base completa indisponível
Dificuldade de trabalhar os dados	Licença não transparente

Fonte e Elaboração: FGV/DAPP e OKBr.

Cada um desses gargalos está relacionado a alguma etapa de preenchimento dos surveys de coleta de informação (como descrito na etapa de metodologia). Em usabilidade, temos:

1

O *dataset incompleto* refere-se à ausência de alguma das características demandadas para a avaliação, mas que não necessariamente invalida a dimensão como um todo³².

³² Como convenção, adotamos a ideia de que se o dataset não estava disponível (recebeu nota 0%), então ele possui todos os problemas possíveis (8).

2

Para cada conjunto de dados avaliado é necessária uma periodicidade de atualização específica. A pergunta no questionário relacionada a esse tipo de problema assume intervalos variando entre um dia de atualização (para **Qualidade do Ar**, por exemplo) até um ano (como é o caso dos **Mapas da Cidade**). Quando não está atualizado de acordo com este parâmetro, contamos como um problema de *dataset desatualizado*.

3

Uma lista de formatos abertos possíveis é disponibilizada, onde se escolhe os que foram encontrados. Caso nenhum seja selecionado, entende-se que há um problema de *indisponibilidade de formato aberto*.

4

A *dificuldade de trabalhar os dados* é um quesito não diretamente relacionado com uma pergunta do questionário, mas com características que dificultam transformar os dados brutos em informações. Quando há necessidade de muito trabalho para reunir os dados, dificuldade de visualização de informações ou problemas nos metadados (documentação insuficiente, codificação não explicada etc), contabilizamos um gargalo nessa categoria.

Já em processo, temos:

5

Alguns dados governamentais são disponibilizados apenas mediante criação de uma conta ou provimento de CPF, e-mail etc. Quando há necessidade de registro ou requisição de acesso aos dados, dizemos que há um problema de *acesso restrito*.

6

O levantador responde uma pergunta com base na sua experiência sobre encontrar os dados, atribuindo um número de 1 a 4, sendo 1 o mais difícil. Entendemos que há um problema de *dificuldade de localizar dados* quando a resposta é 1 ou 2.

7

Pergunta-se no questionário se é possível fazer download dos dados todos de uma vez, o que, quando a resposta é negativa, caracteriza um problema de *download da base completa indisponível*.

8

Se os dados não estão claramente licenciados por meio de uma licença aberta ou declarados como sendo de domínio público, entende-se que há um gargalo de *licença não transparente*.

Com base nessa metodologia de identificação de gargalos, foram detectados **429 problemas**. Destes, **267 (62%) são problemas de usabilidade e 162 (38%), de processo**. Isto é, os problemas de usa-

bilidade são mais frequentes, o que reitera a conclusão da avaliação-piloto do ano passado de que a dimensão em que mais avançamos no Brasil, em termos de transparência, é a publicização dos dados, mas não necessariamente o entendimento e a transformação dos dados em algo útil.

Dentre as 17 dimensões avaliadas, a única que não apresentou qualquer gargalo em relação às categorias de processo e usabilidade foi **Resultados Eleitorais**, cujos dados são consolidados e disponibilizados pelo TSE, como já ressaltado anteriormente, atendendo a todos os critérios avaliados.

Tabela 8
Frequência de gargalos encontrados por tipo (usabilidade e processo)

Usabilidade	Nº Gargalos e % do total	Processo	Nº Gargalos e % do total
Dificuldade de Trabalhar os Dados	72 (16,8%)	Download da Base Completa Indisponível	70 (16,3%)
Dataset Incompleto	67 (15,6%)	Licença Não Transparente	34 (7,9%)
Indisponibilidade de Formato Aberto	65 (15,2%)	Dificuldade de Localizar Dados	32 (7,4%)
Desatualizado	63 (14,7%)	Acesso Restrito	26 (6,1%)
Total	267 (62,2%)	Total	162 (37,8%)

Fonte e Elaboração: FGV/DAPP e OKBr.

O problema mais frequente foi **dificuldade de trabalhar dados**. Este problema também foi detectado como o mais frequente durante a avaliação federal de 2016 e surge como consequência dos seguintes fatores: dificuldade de utilizar/visualizar as informações, problemas nos metadados (documentação insuficiente ou codificação incompreensível), impossibilidade de baixar uma base de dados completa ou grandes conjuntos de dados ou indisponibilidade de formato aberto. Tais problemas trazem dificuldades ao manuseio dos dados e cria entraves, por exemplo, em análises estatísticas e científicas a partir de tais dados.

O segundo problema com maior incidência é a **indisponibilidade de download da base de dados completa**, também detectado como o segundo gargalo mais frequente no Brasil durante o Global Open Data Index 2016. Isso indica que ainda não há uma compreensão dos órgãos responsáveis pelo provimento das informações sobre a facilitação do processo de edição e manipulação dos dados, que é importante para promoção da transparência de fato, além de ser parte fundamental da aderência aos critérios de transparência internacionais.

Outro problema frequente é a falta de características em relação ao que o índice busca em cada dimensão, categorizado como **dataset incompleto**. Tomemos como exemplo a quantidade de arsênio em relatórios de **Qualidade da Água**, uma característica mensurada em apenas uma das cidades avaliadas. Os conjuntos de dados sobre **Qualidade da Água** possuíam, em média, apenas 4 das 7 características buscadas pelo índice, à exceção do dataset do Rio de Janeiro, que contempla todas as características desta dimensão.

Para metade das cidades avaliadas, o problema mais frequente é a **indisponibilidade de formato aberto**, que é um dos principais limitantes para atingir plena abertura de dados. Apresentaremos os principais gargalos de cada cidade entre os conjuntos de dados avaliados, permitindo uma análise mais clara sobre onde cada município deve focar seus esforços. O gargalo categorizado como **dificuldade de trabalhar dados** não será considerado nesta análise mais específica por sua transversalidade entre todas as cidades e pelo fato de ser consequência de outros problemas detectados. A tabela a seguir é uma matriz do número de gargalos encontrados por tipo e cidade.

Tabela 9
Frequência de dimensões com gargalos por tipo de problema e cidade

Gargalos / Cidades	SP	RJ	BH	BSB	POA	UDI	SSA	NAT	Total
Dificuldade de trabalhar dados	7	9	6	5	7	12	12	14	72
Download da base completa indisponível	5	7	9	9	7	11	11	11	70
Dataset incompleto	2	5	7	10	10	12	10	11	67
Indisponibilidade de formato aberto	2	8	4	7	6	13	12	13	65
Desatualizado	4	5	6	9	10	8	10	11	63
Licença não transparente	2	1	3	2	6	6	7	7	34
Dificuldade de localizar dados	2	4	4	2	4	4	5	7	32
Acesso restrito	3	1	3	2	2	4	3	8	26
Total Gargalos	27	40	42	46	52	70	70	82	429

Fonte e Elaboração: FGV/DAPP e OKBr.

São Paulo

A cidade de São Paulo, no geral, apresentou um bom desempenho de acordo com os critérios de abertura e transparência almejados. O município apresentou boas práticas em múltiplos conjuntos de dados através do seu geoportal, que reuniu diversos segmentos de informação avaliados. A plataforma é de fácil uso além de permitir o download de bases completas com formato aberto. Os problemas mais frequentes observados foram a falta de uma base de dados completa: seis dos 17 conjuntos de dados falharam nesse aspecto, além de dados desatualizados. As dimensões mais problemáticas foram **Qualidade do Ar** e **Registro de Empresas**, ambas apresentando quatro gargalos cada.

Rio de Janeiro

No Rio de Janeiro, o portal de dados geográficos da prefeitura disponibiliza diversos conjuntos de dados para as dimensões avaliadas, de fácil visualização e uso, uma boa prática de divulgação de informações reunindo diversos grupos de dados em uma mesma plataforma. Outro destaque na cidade foram os dados sobre **Estatísticas Criminais**, sendo o único município a não apresentar problemas na dimensão. O maior número de problemas observados foi referente à indisponibilidade de formato aberto, uma falha observada em quase metade das bases de dados avaliadas. Detectou-se também que em 41% dos casos não era possível baixar uma base de dados completa. As dimensões mais problemáticas foram **Qualidade do Ar**, **Registro de Empresas** e **Transporte Público**, todas apresentando quatro gargalos cada. Não foi encontrada uma base de dados que se qualifique para os critérios buscados na dimensão **Propriedade da Terra**.

Belo Horizonte

A capital mineira foi a única a apresentar dados atualizados para todas as características em **Estatísticas Socioeconômicas** e a obter nota máxima em **Qualidade do Ar**. Uma boa prática adotada pela prefeitura é a reunião de diversos dados relevantes sobre a vida pública em seu portal, facilitando o acesso à informação. Contudo, mais da metade dos conjuntos de dados no município apresentou problemas no que se refere ao download de uma base completa. Também foi detectado que em sete desses conjuntos os dados estavam incompletos, não apresentando todas as características buscadas. A dimensão mais problemática foi **Qualidade da Água**, sendo detectados seis dentre oito possíveis problemas. Não foram encontradas bases de dados que se qualifiquem para os critérios buscados nas dimensões **Propriedade da Terra** e **Registro de Empresas**.

Porto Alegre

Em Porto Alegre, observamos a facilidade de se obterem dados cartográficos em formato aberto de várias características da cidade através do site observaPOA. Além disso, o portal de transparência da cidade mostra a intenção da prefeitura em mostrar aos seus cidadãos como está sendo utilizado o dinheiro público de forma simples e fácil, atendendo a todas as exigências de abertura de dados e transparência buscadas. A dimensão mais problemática foi **Compras Públicas**, apresentando seis dos oito possíveis problemas categorizados. Não foram encontradas bases de dados que se qualifiquem para os critérios buscados nas dimensões **Propriedade da Terra** e **Qualidade do Ar**.

Brasília

A capital brasileira apresentou um desempenho mediano, ficando em quarto lugar junto com Porto Alegre. O portal de transparência do Distrito Federal disponibiliza diversos conjuntos de dados buscados pelo índice, atendendo a critérios como formato aberto e download da base de completa, uma boa prática de transparência. Download da base completa indisponível, dataset desatualizado e dataset incompleto foram os três problemas mais frequentes, aparecendo em 53% dos datasets avaliados. A dimensão com maior quantidade de gargalos foi **Qualidade da Água**, onde detectamos cinco gargalos. Não foram encontradas bases de dados que se qualifiquem para os critérios buscados nas dimensões **Registro de Empresas** e **Localizações**.

Uberlândia

Na cidade mineira, a única base de dados que obteve nota máxima, com exceção das bases divulgadas por entes federais, foi **Estatísticas Criminais**, que é disponibilizada por um órgão estadual. A prefeitura de Uberlândia tem feito esforços no sentido de ser transparente com os gastos governamentais disponibilizando essa base de dados de forma simples de ser visualizada e encontrada. É recomendado que haja uma maior adesão aos critérios de transparência atuais, disponibilizando os dados em formato aberto e com a possibilidade de baixar a base completa integralmente. Os problemas mais frequentes foram indisponibilidade de formato aberto e dataset incompleto, aparecendo em 76% e 70%, respectivamente, dos datasets avaliados. A dimensão mais problemática foi **Qualidade da Água**, apresentando seis gargalos. Não foram encontradas bases de dados que se qualifiquem para os critérios buscados nas dimensões **Localizações**, **Qualidade do Ar**, **Registro de Empresas** e **Propriedade da Terra**.

Salvador

Uma boa prática adotada pela cidade é a instauração de um portal de transparência contendo diversos tipos de informação úteis aos cidadãos, o que mostra a intenção da prefeitura de ser transparente com a administração dos recursos públicos. Por outro lado, à exceção da base de **Resultados Eleitorais**, divulgada pelo TSE, a capital baiana não obteve nota máxima em quaisquer dos conjuntos de dados. Os problemas mais frequentes foram indisponibilidade de formato aberto, em 70% dos datasets, e download da base completa indisponível, em 64%. A dimensão mais problemática foi **Transporte Público**, apresentando sete dos oito gargalos categorizados. **Mapas da Cidade**, **Registro de Empresas** e **Compras Públicas** apresentaram seis gargalos, o que as fazem também prioritárias na necessidade de abertura de dados. Não foram encontradas bases de dados que se qualifiquem para os critérios buscados nas dimensões **Qualidade do Ar** e **Propriedade da Terra**.

Natal

Natal apresenta boas práticas de abertura de dados com a implementação do portal de transparência da prefeitura que permite uma visualização fácil e acessível a todos de como tem sido utilizado o dinheiro público. Outro ponto positivo é o portal de compras, onde é possível obter informações sobre as licitações em andamento e passadas. Contudo, a cidade obteve uma nota de 43%, a mais baixa dentre as cidades avaliadas. Nenhuma das bases de dados municipais ou estaduais obteve escore de 100%. Para somente dez das 17 dimensões avaliadas foram encontradas bases de dados que se qualificaram para o levantamento. Os problemas mais frequentes foram indisponibilidade de formato aberto, desatualização, dataset incompleto e download da base completa indisponível. Todos os problemas mais frequentes foram detectados em 60% ou mais das bases de dados avaliadas. As dimensões com maior quantidade de gargalos foram **Qualidade da água**, **Leis em Vigor** e **Limites Administrativos**. Não foram encontradas bases de dados que se qualifiquem para os critérios buscados nas dimensões Estatísticas Criminais, Atividade Legislativa, Transporte Público, Localizações, Registro de Empresas, Qualidade do Ar e Propriedade da Terra.

As tabelas a seguir invertem a lógica adotada até aqui. Apresentamos a frequência de cidades por dimensão do índice e problemas diagnosticados, além das categorias de usabilidade e processo, de modo a aferir quais problemas são mais relevantes em cada dimensão. Pode-se observar que os datasets menos problemáticos são **Resultados Eleitorais**, **Escolas Públicas** e **Estatísticas Socioeconômicas**. Com exceção de alguns casos específicos³³, para todas as cidades avaliadas as principais fontes de dados desses datasets foram entidades da esfera federal, que apresentavam boas práticas de transparência e abertura de dados.

Tabela 10

Frequência de cidades com gargalos por tipo de problema e dimensão

	Acesso restrito	Dificuldade de localizar dados	Licença não transparente	Desatualizado	Indisponibilidade de formato aberto	Dataset incompleto	Download da base completa indisponível	Dificuldade de trabalhar dados	Total de problemas encontrados
Resultados Eleitorais	0	0	0	0	0	0	0	0	0
Estatísticas Socioeconômicas	0	0	0	7	0	0	0	0	7
Escolas Públicas	0	0	1	0	0	5	0	4	10
Gastos Públicos	2	0	0	0	1	0	5	4	12
Orçamento Público	0	0	1	0	5	1	2	5	14
Estatísticas criminais	1	2	1	1	2	7	1	2	17
Limites Administrativos	0	2	1	6	4	1	2	2	18
Leis em Vigor	0	0	2	0	6	2	6	5	21
Mapas da Cidade	0	1	1	6	3	6	4	2	23
Atividade Legislativa	1	1	2	1	5	3	8	6	27
Compras Públicas	1	0	1	4	6	3	7	7	29
Localizações	5	6	3	5	3	6	4	4	36
Transporte Público	1	4	2	7	3	7	8	4	36
Qualidade da água	0	0	3	8	8	6	3	8	36
Qualidade do Ar	5	4	5	5	6	5	6	6	42
Registro de Empresas	4	6	5	6	7	7	7	7	49
Propriedade da Terra	6	6	6	7	6	8	7	6	52
Total	26	32	34	63	65	67	70	72	429

Fonte e Elaboração: FGV/DAPP e OKBr.

³³ Belo Horizonte em Resultados Eleitorais, Porto Alegre em Estatísticas Socioeconômicas e Salvador e Brasília em Escolas Públicas.

Tabela 11

Frequência de cidades com gargalos por categoria e dimensão

	Usabilidade	Processo	Total de problemas encontrados
Resultados Eleitorais	0	0	0
Estatísticas Socioeconômicas	7	0	7
Escolas Públicas	9	1	10
Gastos Públicos	5	7	12
Orçamento Público	11	3	14
Estatísticas criminais	12	5	17
Limites Administrativos	13	5	18
Leis em Vigor	13	8	23
Mapas da Cidade	17	7	24
Atividade Legislativa	15	12	27
Compras Públicas	20	9	29
Localizações	18	17	36
Transporte Público	21	15	36
Qualidade da água	30	6	36
Qualidade do Ar	22	20	42
Registro de Empresas	27	22	49
Propriedade da Terra	27	25	52
Total	267	162	429

Fonte e Elaboração: FGV/DAPP e OKBr.

Dentre os datasets foram encontrados gargalos que se repetiam ao longo dos diferentes municípios avaliados, possuindo ou não a mesma fonte de dados (casos em que um ente federal era utilizado como fonte). A análise feita sobre essas repetições trouxe as seguintes conclusões:

Entre as **Estatísticas Socioeconômicas** o problema mais comum era falta de atualização, decorrido dos dados disponibilizados pelo IBGE serem desatualizados para PIB municipal.

Escolas Públicas apresentaram bases de dados frequentemente incompletas, consequência, em especial, do fato de que o INEP não divulga endereços e coordenadas das escolas. Em muitos casos, essas informações também não eram disponibilizadas pelos entes municipais ou estaduais responsáveis pelos dados escolares.

As bases de dados sobre **Gastos Públicos** não possibilitam download de uma base completa em 62,5% dos casos.

A indisponibilidade de formato aberto é um problema frequente em datasets de **Orçamento Público**, frequentemente disponibilizados em formato não processável por máquina, como o PDF.

Estatísticas criminais apresentaram, para quase todas as cidades, datasets incompletos. Em sete das oito avaliações não foram encontrados dados detalhados sobre apreensão de armas.

Limites Administrativos e Mapas da Cidade foram bases de dados frequentemente disponibilizadas na mesma plataforma, apresentando, na maioria dos casos, problemas idênticos. Em seis das oito cidades os conjuntos de dados avaliados apresentavam problemas de atualização: suas últimas atualizações haviam ocorrido há mais de um ano em relação a data em que foram feitos os levantamentos. A principal distinção entre as duas dimensões se deve ao fato de várias cidades não disponibilizarem algumas das características buscadas pelo índice para os mapas municipais.

As bases de **Leis em Vigor** são na maioria avaliadas como simples de serem encontradas e visualizadas. No entanto, quando há a intenção de obter uma grande quantidade de leis simultaneamente ou dados em formato aberto, os problemas relacionados à dimensão são encontrados. Todas as cidades, exceto São Paulo, apresentaram problemas de formato aberto ou falta de uma base completa disponível, sendo que na maioria dos casos ambos os problemas foram detectados.

Em **Atividade Legislativa**, todas as bases de dados avaliadas foram diagnosticadas com problemas para fazer download de uma base completa já que nenhuma das fontes disponibiliza as votações ou transcrição dos debates junto aos projetos de lei. Dessa forma, é necessária a obtenção destes separadamente através da data em que o projeto foi debatido.

Para **Compras Públicas**, notamos que em quase todos os casos (sete de oito cidades) não foi possível baixar uma base completa de dados, e em seis dos oito municípios avaliados não havia disponibilidade de formato aberto.

Em todas as cidades, com exceção do Rio de Janeiro, os conjuntos de dados referentes à dimensão **Localizações** receberam nota inferior a 100%. Notou-se também que apesar de todos os municípios possuírem essa base de dados, apenas cinco a disponibilizam, dos quais apenas dois apresentam informações relativas aos CEPs dos logradouros³⁴.

³⁴ Há uma controvérsia sobre o compromisso de divulgação destes dados por parte dos Correios (para todo o território nacional). Para mais informações, ver <https://discuss.okfn.org/t/entry-for-locations-brazil/4533>. Há também um paper que analisa a apropriação do CEP pela empresa Correios. Disponível em <https://www.reciis.iciet.fiocruz.br/index.php/reciis/article/view/1353/pdf1353>. Ambos documentos acessados em 19/04/2018.

Nenhum dos datasets de **Transporte Público** atendeu aos critérios de download da base completa. Para informações relativas à conexão entre diferentes tipos de transporte, a única cidade que não falhou nesse quesito (Uberlândia-MG) foi a que não tinha essa possibilidade, por haver apenas um modal de transporte disponível.

O dataset de **Qualidade da Água** foi o que apresentou maior frequência de problemas de usabilidade, mesmo que todas as cidades tenham apresentado informações públicas sobre o tema. Os oito conjuntos de dados apresentavam problemas de atualização — que deveria ser no mínimo semanal — e não disponibilizavam formato aberto. Foi observado também que nenhuma das cidades avaliadas, à exceção do Rio de Janeiro, apresentou todos os parâmetros de avaliação da água almejados pela metodologia do ODI.

Dentre as bases de dados sobre **Qualidade do Ar**, das oito cidades avaliadas, apenas quatro disponibilizaram dados sobre o assunto. A cidade de Belo Horizonte foi a única a obter um escore de 100%, enquanto que as outras cidades avaliadas apresentaram falhas relevantes na apresentação dos dados. Em Brasília, os dados eram bastante defasados em relação ao que o ODI busca: as últimas informações disponíveis eram de quatro meses antes da data do levantamento, enquanto que se esperava uma atualização diária. Em Porto Alegre, de acordo com o boletim de qualidade do ar, as estações da cidade encontravam-se em manutenção e não havia dados suficientes para fazer o levantamento. Para a cidade de Salvador, o órgão governamental INEMA (Instituto do Meio Ambiente e Recursos Hídricos) relata que os dados são coletados pela CETREL (Central de Tratamento de Efluentes Líquidos) em cooperação técnica com o governo do estado da Bahia, mas não havia informações disponibilizadas sobre a capital baiana.

Registros de empresas também são coletados pelos municípios em todas as cidades avaliadas, porém os dados são públicos em apenas quatro destas. Na maioria das outras cidades avaliadas, havia a possibilidade de comprar essa base de dados com as respectivas juntas comerciais estaduais, um grande entrave à transparência desses conjuntos de dados. Em Porto Alegre, a divulgação dessas informações é feita através da Secretaria Municipal da Indústria e Comércio, que disponibiliza uma base de dados completa e em formato aberto, mostrando o intuito de ser transparente com os registros de empresas na cidade. No Rio de Janeiro e em Salvador, só é possível obter os dados através do conhecimento prévio de uma informação, o CNPJ ou nome da empresa, algo que limita o acesso aos dados completos. Mais especificamente no caso do Rio, é possível obter uma lista com as empresas registradas no sistema da nota carioca, que não é o caso de todas as empresas registradas do município, visto que tal cadastro é necessário apenas para emissão de nota eletrônica.

Registros de **Propriedade da Terra** são dados que apresentam problemas em qualquer parte do mundo³⁵. No Brasil não é diferente, apesar de os dados serem coletados — é a administração municipal o ente responsável pela cobrança de impostos sobre propriedade urbana (IPTU) por exemplo. No entanto, nada mudou em relação ao ano passado no que diz respeito à transparência dessa informação, ao menos nos casos comparáveis (Rio de Janeiro e São Paulo). Em relação às outras seis cidades, apenas uma (Brasília) apresentou uma base de dados qualificável. Em São Paulo, destaque por possuir uma base 100% adequada, os dados divulgados não disponibilizam os limites das propriedades, apenas uma identificação, o tipo de propriedade e o valor.

³⁵ Pode-se ver isso nos resultados do ODI global de 2016 disponível em <http://global.survey.okfn.org/>. Acesso em 19/04/2018.

CONCLUSÃO

O índice de dados abertos para cidades brasileiras neste ano avaliou o estado da arte dos dados abertos, tendo por ano-base 2017, **a partir de 136 bases de dados: foram 17 dimensões avaliadas em oito cidades participantes**, um aumento considerável em relação a 2016, em que foram avaliadas 36 bases de dados distribuídas em duas cidades. Isso, por si só, já revela um fato positivo: cresce o número de interessados na temática da transparência e dados abertos nas cidades brasileiras.

O processo de levantamento, revisão e consolidação é pioneiro no país e contou com a participação de grupos locais em seis das oito cidades avaliadas: **Belo Horizonte-MG, Brasília-DF, Natal-RN, Porto Alegre-RS, Salvador-BA e Uberlândia-MG**. Além disso, a FGV DAPP e a OKBr atualizaram a avaliação das cidades que participaram do piloto em 2017, **São Paulo-SP e Rio de Janeiro-RJ**

O objetivo deste relatório é apontar os caminhos para que as cidades participantes possam aprimorar suas políticas de dados abertos. Nesse sentido, o principal método de procedimento das análises dos resultados foi a identificação dos gargalos por dimensão e por cidade.

Em relação ao ano passado, para as duas cidades em que foi feita a avaliação nos dois momentos, não há uma evolução visível. O número de bases 100% abertas é o mesmo e o avanço no escore foi marginal. Isso significa que há ainda o mesmo espaço para melhoria que havia no ano passado, para estas cidades.

³⁶ O Brasil obteve 68% de escore na avaliação que teve por ano-base 2016. Disponível em: <https://index.okfn.org/place/>. Acesso em 19/04/2018.

Dentre as dimensões avaliadas quatro se destacaram por apresentarem as maiores quantidades de gargalos: **Propriedade da Terra, Registro de Empresas, Qualidade do Ar e Qualidade da Água**. Dentre todos os tipos de gargalos detectados, 62% são problemas de usabilidade e 38% de processo, reiterando a conclusão do ODI do ano passado de que os avanços são mais visíveis na publicização dos dados do que na facilitação do entendimento e usabilidade dos mesmos. Há ainda, uma necessidade latente de bases de dados mais adequadas ao uso e à transformação destes em informação.

A média das dimensões nas cidades avaliadas ficou em 65%, similar à média nacional no Índice de Dados Abertos global 2016³⁶. Esse resultado mostra que ainda há muito espaço para aprimoramento em todos os municípios avaliados. Somente 25% das bases de dados analisadas no universo das oito cidades obtiveram nota máxima, sendo que em duas dimensões os dados são disponibilizados para todas as cidades através de órgãos federais.

Cinco das 17 dimensões avaliadas não obtiveram nota máxima em quaisquer das cidades avaliadas (**Registro de Empresas, Transporte Público, Qualidade da Água, Compras Públicas e Atividade Legislativa**). Para três dos conjuntos de dados avaliados — **Qualidade do Ar, Registro de Empresas e Propriedade da Terra** — não havia informações públicas para metade ou mais das cidades avaliadas, mostrando que muitos municípios ainda carecem de disponibilizar dados aos seus moradores.

Muitas melhorias podem ser feitas na forma como os dados são disponibilizados pelas prefeituras ou órgãos governamentais, e diferentes dimensões requerem formas distintas de melhoria. Este documento traz quais conjuntos de dados precisam de atenção prioritária e onde se encontram os problemas mais comuns, trazendo diretrizes para melhoria da política de dados abertos nas cidades.

As boas práticas replicáveis, nesse sentido, focam nos esforços de centralização das informações em um portal único e de fácil compreensão, privilegiando a experiência do usuário. Percebe-se que os esforços municipais nesse sentido são, em geral, incompletos. Por exemplo, a cidade de São Paulo apresenta um bom exemplo de consolidação das informações geográficas (GeoSampa), mas o portal não diz respeito aos outros tipos de dados. Os esforços de disponibilização de APIs, por exemplo, comum a muitas das cidades avaliadas, também podem avançar para uma amplitude maior de dimensões, além de utilizar boas práticas de documentação para auxiliar no processo de utilização dos dados.

Com uma avaliação mais poderosa do que a do ano passado, chegamos a uma conclusão semelhante, que também é comum ao resultado da avaliação federal: o Brasil é um ator importante na questão da transparência e isso se reflete em certa medida nas suas cidades, mas há ainda um foco na publicização das informações como suficiente para o compromisso com a transparência. Contudo é preciso fazer mais: disponibilizar bases em formato aberto, amigável, com metadados adequados, claramente destinados ao domínio público e que facilitem o uso e o entendimento para uma população cada vez maior. Com isso, a agenda da transparência e dos dados abertos pode, de fato, contribuir para o avanço da democracia.

REFERÊNCIAS

CADASTA FOUNDATION. Overview of property rights data. Disponível em <<http://cadasta.org/open-data/overview-of-property-rights-data/>> Acesso em 15 mar 2018.

FGV DAPP. Estudos sobre dados públicos sobre mobilidade urbana. Disponível em <<http://dapp.fgv.br/cidades/>> Acesso em 20 mar 2018

FGV DAPP. Estudos sobre dados públicos de educação. Disponível em <<http://dapp.fgv.br/educacao/>> Acesso em 20 mar 2018

FGV DAPP. Estudos sobre segurança e cidadania. Disponível em <<http://dapp.fgv.br/seguranca-e-cidadania/>> Acesso em 26 mar 2018.

FGV DAPP; OKBR. Relatório ODI Rio de Janeiro 2017 ano-base 2016. Disponível em <<http://dapp.fgv.br/wp-content/uploads/2017/05/IndiceDadosAbertosRio2017.pdf>> Acesso em 05 mar 2018.

FGV DAPP; OKBR. Relatório ODI São Paulo 2017 ano-base 2016. Disponível em <<http://dapp.fgv.br/wp-content/uploads/2017/04/RelatorioODI-SP-2017.pdf>> Acesso em 05 mar 2018.

FGV DAPP; OKBR. Relatório ODI Brasil 2017 ano-base 2016. Disponível em <<http://dapp.fgv.br/wp-content/uploads/2017/04/IndiceDadosAbertosBrasil2017-1.pdf>> Acesso em 5 mar 2018.

GEO NETWORK. Global administrative unit layers project. Disponível em <<http://www.fao.org/geonetwork/srv/en/metadata.show?id=12691>> Acesso em 16 mar 2018.

ISPRS FOUNDATION. United Nations committee of experts on global geospatial information management. Disponível em <http://www.isprs.org/documents/reports/The_Status_of_Topographic_Mapping_in_the_World.pdf> Acesso em 16 mar 2018.

NATIONAL DEMOCRATIC INSTITUTE. Disponível em <https://www.ndi.org/lea_aqui_sobre_el_ndi> Acesso em 15 mar 2018.

OPEN CONTRACTING PARTNERSHIP. Open contracting data standard Disponível em <<http://standard.open-contracting.org/latest/en/schema/>> Acesso em 14 mar 2018.

OPEN CORPORATES. Api reference. Disponível em <<http://api.opencorporates.com/documentation/API-Reference>> Acesso em 15 mar 2018.

OPEN DATA CHARTER. Open data principles. Disponível em <<http://opendatacharter.net/principles/>> Acesso em 6 mar 2018.

OPEN DATA WATCH. Disponível em <<https://opendatawatch.com/>> Acesso em 19 mar 2018.

OPEN DEFINITION. Disponível em < <http://open-definition.org/>> Acesso em 24 mar 2018.

OPEN ELECTION DATA INITIATIVE. Key election process categories: election results. Disponível em <<https://openelectiondata.net/en/guide/key-categories/election-results/>> Acesso em 15 mar 2018.

OKFN. Código plataforma ODI. Disponível em <<https://github.com/okfn/opendatasurvey>> Acesso em 06 mar 2018.

OKFN. Fórum OKFN sobre dados abertos governamentais. Disponível em <<https://discuss.okfn.org/>> Acesso em 26 mar 2018.

OKFN. Fórum OKFN, discussão sobre logradouros no Brasil. Disponível em <<https://discuss.okfn.org/t/entry-for-locations-brazil/4533>> Acesso em 19 abril 2018.

OKFN. Índice global. Disponível em <<http://global.survey.okfn.org/>> Acesso em 5 mar 2018.

OPENING PARLIAMENT. Declaration of parliamentary openness. Disponível em <<https://openingparliament.org/declaration/>> Acesso em 19 mar 2018

OPEN SPENDING. Disponível em <<https://openspending.org/>>. Acesso em 27 mar 2018.

RIBEIRO, Ana Maria Almeida; OLIVEIRA, Eliane Braga de. Código de endereçamento postal (CEP), um obstáculo aos dados abertos no Brasil. Revista eletrônica de comunicação, informação e inovação em saúde. Volume 11, novembro 2017 . Disponível em <<https://www.reciis.iciict.fiocruz.br/index.php/reciis/article/view/1353/pdf1353>> Acesso em 19 abril 2018.

UNITED NATIONS GEOGRAPHIC INFORMATION WORKING GROUP. Core Geo-Database. Disponível em < <http://www.unguiwg.org/coreDB>> Acesso em 16 mar 2018.

UNIVERSAL POSTAL UNION. Addressing and Postcode Manual. Disponível em <<http://www.upu.int/fileadmin/documentsFiles/activities/addressingAssistance/manualAddressingAndPostcodeManualEn.pdf>> Acesso em 16 mar 2018.

WORLD HEALTH ORGANIZATION. Air quality guidelines. Disponível em <http://www.who.int/phe/health_topics/outdoorair/outdoorair_aqg/en/> Acesso em 20 mar 2018.

WORLD HEALTH ORGANIZATION. Water quality guidelines. Disponível em <http://www.who.int/water_sanitation_health/water-quality/guidelines/en/> Acesso em 20 mar 2018

ANEXO

Aqui, apresentamos um resumo por cidade com alguns destaques de dimensões-chave, boas práticas e pontos de atenção.

São Paulo

A cidade apresentou o melhor desempenho dentre os municípios avaliados, sendo a única a possuir bases de dados para todas as dimensões avaliadas no ODI deste ano. Com uma nota de 84%, apresenta algumas boas práticas replicáveis para outras cidades avaliadas. Ainda assim, os pontos a seguir devem ser priorizados para se atingir grau de excelência em transparência e abertura de dados: disponibilização de bases de dados completas e ampliação da frequência de atualização das bases.

Rio de Janeiro

Na capital fluminense, destacamos as dimensões **Estatísticas Criminais** e **Qualidade da Água**. Na primeira, encontramos todas as características de transparência e abertura de dados desejadas, sendo a única cidade que atendeu a todos os critérios da dimensão. Para a segunda, apesar de não obter nota máxima na avaliação, o ente carioca responsável por avaliar a água da cidade foi o único que mensurou todos os compostos químicos que o índice busca. A pesquisa mostrou que as principais mudanças a serem efetuadas são na direção de aumentar disponibilidade de bases de dados completas para download e incluir formatos abertos para os conjuntos de dados já disponíveis.

Belo Horizonte

A cidade apresentou um bom desempenho na avaliação, destacando-se nas bases de **Estatísticas Socioeconômicas, Qualidade do Ar e Leis em Vigor**, onde disponibiliza uma API para download da legislação municipal e estadual. Seus maiores esforços devem ser direcionados para disponibilização de bases completas para download, problema mais frequentemente detectado nas bases de dados da cidade.

Brasília

A capital brasileira foi a única a obter 100% de abertura nas dimensões **Orçamento Público e Gastos Governamentais**. Seus esforços em prol de obter um maior grau de transparência devem ser orientados em disponibilizar mais bases de dados completas e atualizar os dados já disponibilizados com maior frequência.

Porto Alegre

Porto Alegre atingiu um desempenho igual ao de Brasília, sendo essas duas as únicas cidades que atingiram nota máxima em **Gastos Públicos**. A capital gaúcha obteve, contudo, um grau de abertura inferior, disponibilizando uma base de dados totalmente aberta a menos que Brasília. Observou-se que mais da metade das dimensões avaliadas apresentou problemas de atualização, assim como datasets incompletos. Ambos os problemas foram encontrados em dez das 17 dimensões avaliadas.

Salvador

Com um resultado mediano, a cidade possui muitos aspectos onde melhorar suas iniciativas de transparência. Há, no entanto, uma singularidade em relação aos outros municípios avaliados: em Salvador os dados disponibilizados pela Secretaria da Educação da Bahia permitem acompanhar os recursos destinados às escolas da rede estadual e os gastos efetuados pelos gestores. Os principais problemas detectados foram relacionados à indisponibilidade de dados em formato aberto, falta de uma base de dados completa para download e desatualização.

Uberlândia

A única cidade do interior dentre as avaliadas obteve um desempenho mediano e comparável a de algumas capitais. Apresentou boas práticas de transparência através do Portal do Cidadão onde é possível fazer pesquisas sobre processos licitatórios, contas públicas, consultas de leis, processo legislativo (dados relativos às dimensões **Compras Públicas, Gastos Públicos, Leis em Vigor e Atividade Legislativa**, respectivamente) entre outras informações úteis à vida pública. Acreditamos que a intenção das autoridades locais deve ser mantida e ampliada para uma maior participação da população no controle fiscal do município. O estudo revelou que o principal problema entre as bases de dados avaliadas foi a indisponibilidade de um formato aberto, um entrave a trabalhos científicos e análises técnicas sobre dados municipais.

Natal

Natal apresentou um fraco desempenho no ODI, contudo apresentou iniciativas e boas práticas em consonância com a abertura de dados governamentais. O portal de transparência municipal é o início da implementação de políticas de abertura dos dados municipais. Deve, portanto, ser estimulado e ampliado a fim de aumentar a clareza das informações governamentais que serão úteis aos cidadãos, permitindo, assim, um avanço na temática da transparência. Os principais problemas detectados giram em torno da não disponibilização de bases de dados, não existindo bases adequadas para o estudo em sete das 17 das dimensões avaliadas.

INOVAÇÃO E POLÍTICAS PÚBLICAS

dapp.fgv.br | dapp@fgv.br

+ 5 5 2 1 3 7 9 9 . 4 3 0 0