

manual dos

dados abertos: governo

traduzido e adaptado de opendatamanual.org
(<http://opendatamanual.org/>)



**DADOS ABERTOS
GOVERNAMENTAIS**



laboratório brasileiro
de cultura digital

W3C® WORLD WIDE WEB
consortium
Escritório Brasil

cgi.br

Comitê Gestor da Internet
no Brasil

Este manual foi projetado para **governos que querem abrir dados**, mas pode ser usado por qualquer pessoa que queira saber mais dos aspectos técnicos, sociais e políticos dos dados abertos. Aqui se discute por que abrir dados, o que são dados abertos e como abrir dados de governo.

Essa é a primeira publicação resultante do acordo de cooperação técnico-científica entre o Laboratório Brasileiro de Cultura Digital e o Núcleo de Informação e Coordenação do Ponto BR (NIC.br).

Créditos e licenças

Original disponível em Open Data Manual (<http://opendatamanual.org/>), produzido pela Open Knowledge Foundation (<http://okfn.org/>) e seus colaboradores (<http://opendatamanual.org/introduction/credits/>), versão 1.0b, de outubro de 2010, revisada em janeiro de 2011:

Jonathan Gray (<http://jonathangray.org/>)

Daniel Dietrich (<http://ddie.me/>)

Tim McNamara (<http://timcnamara.co.nz/>)

Antti Poikola (<http://apoikola.wordpress.com/>)

Rufus Pollock (<http://rufuspollock.org/>)

Julian Tait (<http://www.littlestar.tv/>)

Ton Zijlstra (<http://www.zylstra.org/>)

Fontes do material original

Fontes usadas diretamente

ACCESS INFO AND THE OPEN KNOWLEDGE FOUNDATION. *Beyond Access Report*.

POIKOLA, Antti. *Finnish Manual*.

POLLOCK, Rufus; GRAY, Jonathan; PARRISH, Simon; HATCHER, Jordan. *Unlocking the Potential of Aid Information*.

THE IATI TECHNICAL ADVISORY GROUP (liderado por Simon Parrish). *Technical Proposal for how IATI is Implemented*.

Outras

W3C PUBLISHING GOVERNMENT DATA. 2009. Disponível em: <http://www.w3.org/TR/gov-data/>.

Tradução e adaptação para o contexto brasileiro produzidas pela Comunidade Transparência Hacker (<http://thacker.com.br/>), versão 1.0port, de abril de 2011, com a colaboração de: Esfera.mobi (<http://blog.esfera.mobi/>) e de Everton Zanella Alvarenga.



Creative Commons 3.0 – Atribuição

Você pode usar, copiar, modificar, remixar, imprimir, distribuir e traduzir esse material para outras línguas, citando a fonte original.

<http://creativecommons.org/licenses/by/3.0/legalcode>

sumário

Introdução, 3

Público-alvo, 4

Por que dados abertos?, 7

Exemplos no Brasil, 11

Afinal, o que significa “aberto”?, 13

De que dados se está falando?, 16

Como abrir os dados?, 17

Escolha os conjunto de dados, 19

Disponibilize os dados, 20

Torne-os mais fáceis de encontrar, 21

Abrimos alguns dados. E agora?, 25

Conte ao mundo, 25

Conecte pessoas: desconferências, *meetups* e *barcamps*, 26

Faça coisas: *hackdays*, prêmios e protótipos, 28

Transparência Hacker, 32

Algumas informações técnicas, 35

APIs, web services e bancos de dados, 35

Formatos, 36

Anexos, 43

Acesso a dados públicos no Brasil, 43

Como (des)organizar um *hackday*, 49



introdução

O leitor sabe exatamente quanto dos seus impostos é gasto com iluminação das ruas ou pesquisas contra o câncer? Qual é o caminho mais curto, mais seguro e com a melhor vista da sua casa até o seu trabalho? E qual é a qualidade do ar que respira ao longo desse trajeto?

O leitor sabe onde, na sua região, pode encontrar as melhores oportunidades de emprego e os locais mais arborizados? Sabe quando e como influenciar leis ou decisões públicas sobre temas com os quais se preocupa?

Novas tecnologias tornam possível a construção de serviços para responder automaticamente a essas perguntas. Muitas pessoas, e não apenas os governos, seriam capazes de construir serviços assim. Mas, infelizmente, os dados necessários para a criação de projetos que atuem nesse sentido não estão disponíveis ou não são liberados em formato que torne possível o seu uso pela sociedade.

Este manual busca explicar como aproveitar o potencial de informações oficiais para permitir a existência de novos serviços

de informação, melhorando a vida dos cidadãos e fazendo com que governo e sociedade trabalhem juntos e melhor. Além de abordar os conceitos básicos de dados abertos, serve ainda para os funcionários do governo, indicando maneiras de utilizar esses dados para criar mais valor e impacto em diversas áreas e dando informações concretas sobre como abrir dados governamentais.

Dados abertos governamentais são dados produzidos pelo governo e colocados à disposição das pessoas de forma a tornar possível não apenas sua leitura e acompanhamento, mas também sua reutilização em novos projetos, sítios e aplicativos; seu cruzamento com outros dados de diferentes fontes; e sua disposição em visualizações interessantes e esclarecedoras.

Já há alguns anos ocorre o debate sobre dados governamentais abertos – sobre informação pública livre para qualquer um acessar e reutilizar, com qualquer finalidade. Em 2009, o tema começou a aparecer na grande mídia, com governos de vários países (como Estados Unidos, Reino Unido, Canadá e Nova Zelândia) anunciando iniciativas voltadas a abrir sua informação pública.

público-alvo

O manual é voltado para um público amplo: serve para quem nunca ouviu falar de dados abertos; para quem se considera um experiente “profissional dos dados”; para especialistas de dados, como *hackers* (ver quadro a seguir) e desenvolvedores; e para quem nunca ouviu falar de uma API (o leitor já ouviu?). É um guia voltado especialmente para aqueles que trabalham em governos que querem, podem e devem liberar dados para o público – mesmo que tenham pouco ou nenhum conhecimento sobre a questão.

É importante verificar as informações técnicas adicionais e os anexos que se encontram no final do manual para saber mais sobre formatos de disponibilização (algumas siglas que vão aparecer durante o texto se repetirão lá) e sobre questões legais em torno dos dados públicos no Brasil.

De acordo com a RFC 1.392, da Internet Engineering Task Force (IETF), de 1993, *hacker* é “uma pessoa que sente prazer em ter uma compreensão profunda dos processos internos de um sistema, de um computador ou de uma rede informática em específico. O termo é frequentemente colocado em um contexto pejorativo, em que ‘*cracker*’¹ seria o nome correto”.

Algumas definições da própria comunidade:

“*Hacker* é a pessoa que conhece profundamente qualquer um dos vários níveis de abstração relativos a tecnologia e que tenta ampliar e aprofundar ao máximo o conhecimento das pessoas leigas” (Leandro Salvador).

“*Hacker* é quem, devido a uma experiência (de trabalho) repetitiva, inventa soluções que facilitam sua vida e permitem lidar com o problema em um nível diferente da sua mera execução, mas nas diferentes abordagens de sua solução. *Hacker* é uma identidade dentro de uma cultura *hacker*” (Capi Etheriel).

¹ Disponível em <http://www.ietf.org/rfc/rfc1392.txt>



por que dados abertos?

Dados abertos, especialmente os governamentais, são um ótimo recurso ainda muito pouco explorado. Muitos indivíduos e organizações coletam uma ampla gama de diferentes tipos de dados para executar suas tarefas. O governo é particularmente importante nesse contexto, tanto por causa da quantidade e da centralidade dos dados que coleta quanto pelo fato de que tais dados são públicos, *um direito garantido no artigo 5º da Constituição Federal brasileira* (e a regulamentação desse direito está em fase de tramitação no Senado Federal, no Projeto de Lei nº 41/2010 – veja mais nos Anexos).

Os dados de governo, portanto, podem ser abertos e disponibilizados para outros usarem. Mas qual é a importância disso?

Há muitas áreas de atuação para as quais são importantes dados abertos, e também vários grupos de indivíduos e organizações que podem se beneficiar dessa disponibilidade, inclusive o próprio governo. Ao mesmo tempo, é impossível saber exatamente como e onde os dados abertos serão mais valorizados, pois, ao

permitir que haja inovação em diversas áreas, novos jeitos de utilizá-los surgirão de locais e situações não muito previsíveis.

Já é possível notar um grande número de áreas e atividades em que os dados abertos estão gerando valor, e há potencial para muito mais. Entre essas áreas estão:

- Transparência e controle democrático;
- Participação popular;
- Empoderamento dos cidadãos;
- Melhores ou novos produtos e serviços privados;
- Inovação;
- Melhora na eficiência de serviços governamentais;
- Melhora na efetividade de serviços governamentais;
- Medição do impacto das políticas;
- Conhecimento novo a partir da combinação de fontes de dados e padrões.

Exemplos existem para a maioria dessas atividades. Quanto a transparência, há projetos como o Tax Tree (<http://www.mindtrek.org/2009/node/127>), da Finlândia, e Where Does My Money Go (<http://wheredoesmymoneygo.org/> – para onde vai o meu dinheiro?), da Grã-Bretanha, que mostram como os recursos dos impostos estão sendo gastos pelo governo. Mais um caso: no Canadá, a abertura de dados permitiu que se economizassem 3.2 bilhões de dólares canadenses em fraudes fiscais de caridade. Outros sítios, como o Folketsting (<http://folketsting.dk/>), da Dina-

marca, acompanham as atividades parlamentares e o processo legislativo, permitindo que se veja exatamente o que está acontecendo e se saiba quais parlamentares estão envolvidos nos fatos.

Dados abertos podem também ajudar as pessoas a tomar decisões melhores nas suas vidas, ou permitir que sejam mais ativas na sociedade. Também criado na Dinamarca, o Find Toilet (<http://findtoilet.dk/>) mostra em um mapa todos os banheiros públicos do país. Pode parecer desimportante, mas esse projeto possibilita que pessoas com certos problemas de saúde saiam de casa mais vezes. Na Holanda, o sítio Vervuilings Alarm (<http://www.vervuilingsalarm.nl/>) avisa com uma mensagem se a qualidade do ar da vizinhança está perto de atingir um nível muito prejudicial. Em Nova Iorque, um serviço busca lugares onde é possível caminhar com um cachorro. Há projetos como o Mapumental (<http://mapumental.channel4.com/>), no Reino Unido, e o Mapnificent (<http://www.mapnificent.net/>), na Alemanha, que permitem encontrar locais para morar de acordo com características indicadas, como duração do trajeto até o local de trabalho, o preços da casa e a beleza da região. São todos exemplos de serviços que utilizam dados abertos.

Dados abertos também têm grande importância econômica. Estudos estimaram que seu valor monetário seja de dezenas de bilhões de euros, somente na União Europeia – novos produtos de novas empresas estão reutilizando esses dados e gerando valor com eles. O Husetsweb (<http://www.husetsweb.dk/>), da Dinamarca, por exemplo, ajuda a encontrar formas de melhorar a eficiência energética da sua residência, incluindo uma ferramenta de planejamento financeiro e uma busca por empreiteiros que possam realizar o trabalho necessário para economizar mais. O Google Translate (<http://translate.google.com.br/>) aproveita o enorme volume de publicações, principalmente nas línguas euro-

peias, para treinar seus algoritmos de tradução, culminando na melhora da qualidade do serviço.

Dados abertos também são de grande valia para o próprio governo, aumentando sua eficiência. O Ministério da Educação alemão publicou *on-line*, para que fossem reutilizados, todos os dados governamentais relacionados à educação no país. Desde então, caiu muito o número de perguntas que recebem da população, reduzindo carga de trabalho e custos. Além disso, as perguntas que continuam sendo feitas tornaram-se mais simples de serem respondidas pelos servidores públicos, já que ficou mais claro para os cidadãos onde podem ser encontradas informações relevantes sobre o tema.

Aumentar a eficiência, em última instância, reduz custos. O Departamento Alemão de Patrimônio Cultural tem publicado ativamente seus dados, permitindo, assim, que sociedades de historiadores amadores e grupos como a Wikimedia Foundation (<http://wikimediafoundation.org/>) executem suas tarefas com mais agilidade.

Isso não resulta somente na melhoria da qualidade dos dados, mas também permite que se reduza o tamanho do departamento – alocando servidores públicos para outros serviços, em vez de apenas redundar o conhecimento que já está sendo construído por colaboradores em rede.

Embora haja inúmeras instâncias em que os dados abertos já criam valores sociais e econômicos das mais diversas maneiras, ainda não se sabe que outras novidades se tornarão possíveis com essa tendência. Novas combinações de dados podem criar novos conhecimentos e inspirar novas ideias, que vão criar novos campos de aplicação – como, na Londres do século XIX, quando o médico John Snow descobriu a relação entre poluição da

água e cólera ao combinar dados sobre mortes por cólera com a localização de fontes d'água, levando à construção do sistema de esgoto londrino, um grande avanço para a saúde geral da população. É provável que se veja o mesmo acontecer quando ideias inesperadas começarem a surgir a partir da combinação de bases de dados abertas.

Esse potencial inexplorado pode ser desencadeado se forem abertos os dados governamentais públicos – de forma que não haja restrições (legais, financeiras, tecnológicas ou políticas) para que aconteça sua reutilização por grupos civis. Qualquer restrição pode excluir pessoas dessa reutilização, tornando mais difícil encontrar formas de criar valor com base nos dados.

Não adianta o poder público se responsabilizar por produzir internamente sítios e serviços como esses. O que os governos precisam fazer é redefinir seu papel como fornecedores de informação, para, assim, permitirem que a sociedade agregue novos valores com base nos dados públicos. Ou seja, esse potencial só será alcançado se os dados públicos forem abertos.

exemplos no Brasil

No Brasil, apesar de existirem programas bastante avançados de transparência pública – como, por exemplo, de transparência orçamentária –, ainda são raríssimos os órgãos ou secretarias que disponibilizam dados abertos. No melhor dos casos, há dados disponíveis para visualização, mas existem inúmeras barreiras técnicas, e até políticas, para que sejam reutilizados pela sociedade na criação de novos projetos e serviços.

Mesmo com a escassez de bases de dados governamentais à disposição para se criarem novos projetos, há exemplos de sítios similares aos que têm surgido em outros países. Esses serviços brasileiros foram criados e desenvolvidos voluntariamente por *hackers* ou programadores independentes interessados em questões públicas. O CMSP (<http://cmsp.topical.com.br/>), do desenvolvedor Mauricio Maia, traz uma nova visualização das prestações de contas disponibilizadas no sítio da Câmara Municipal de São Paulo – o que, antes, só podia ser encontrado após consulta em várias listas e tabelas, impedindo a comparação entre contas de diferentes vereadores, agora está disponível em gráficos e listagens interativas.

Maia também criou e desenvolveu o projeto Alagamentos (<http://alagamentos.topical.com.br/>), que reorganiza dados públicos sobre incidências de alagamentos na cidade de São Paulo, contribuindo para o entendimento desse problema. O sistema ainda informa, por redes sociais como o Twitter, onde podem aparecer novos pontos de alagamento, ajudando o usuário a evitar vias congestionadas ou perigosas em dias de chuva.

O SACSP (<http://sacsp.mamulti.com/>), do programador Bruno Barreto, permite visualizar e acompanhar as reclamações feitas pelos munícipes no sítio da prefeitura paulistana. O que antes era acessível apenas por meio de senha – não estando inteiramente transparente para os cidadãos – está disponível no SACSP, permitindo que se visualize cada um dos problemas relatados na cidade, tenha-se ciência dos departamentos que recebem o recorde de reclamações e ainda se verifiquem incidências e solicitações de diferentes regiões em um mapa interativo.

O Legisdados (<http://legisdados.org/>), do sociólogo Pedro Belasco, republica os dados de tramitação parlamentar na Câmara dos Deputados. Apesar dos dados já estarem disponíveis na Internet, o sítio fornece as mesmas informações em formatos estrutu-

rados e compreensíveis por máquina, de forma a facilitar a vida de outros desenvolvedores que queriam criar sistemas integrados com essas informações.

Para criar esses serviços, desenvolvedores tiveram de “raspar” informações de sítios públicos. É como se a sociedade estivesse fazendo todo o trabalho sozinha. Quando os dados governamentais forem abertos, convidando as pessoas a os reutilizarem, será possível criar mais.

afinal, o que significa “aberto”?

Neste manual, foi usada a definição da Open Definition (<http://opendefinition.org/>): “dado aberto é um dado que pode ser livremente utilizado, reutilizado e redistribuído por qualquer um”. A definição completa fornece detalhes específicos do significado do termo, que se resumem em alguns pontos importantes:

Disponibilidade e acesso: o dado precisa estar disponível por inteiro e por um custo razoável de reprodução, preferencialmente por meio de *download* na Internet; também deve estar num formato conveniente e modificável.

Reuso e redistribuição: o dado precisa ser fornecido em condições que permitam reutilização e redistribuição, incluindo o cruzamento com outros conjuntos de dados.

Participação universal: todos podem usar, reutilizar e redistribuir, não havendo discriminação contra áreas de atuação, pessoas ou grupos (não são permitidas restrições como “não comercial”, que impedem o uso comercial, e restrições de uso para certos fins, como “somente educacional”).

Além dessa definição, os dados governamentais são considerados abertos quando publicados de acordo com os seguintes princípios (elencados por ativistas do governo aberto em encontro acontecido em 2007):

Completo: Todos os dados públicos devem ser disponibilizados. Dado público é aquele que não está sujeito a restrições de privacidade, segurança ou outros privilégios.

Primários: São apresentados tal como colhidos da fonte, com o maior nível possível de granularidade, sem agregação ou modificação (por exemplo, um gráfico não é fornecido aberto, mas os dados utilizados para construir a planilha que deu origem a ele podem ser abertos).

Atuais: Devem ser publicados o mais rápido possível para preservar seu valor. Em geral, têm periodicidade: quanto mais recentes e atuais, mais úteis para seus usuários.

Acessíveis: São disponibilizados para a maior quantidade possível de pessoas, atendendo, assim, aos mais diferentes propósitos.

Compreensíveis por máquina: Devem estar estruturados de modo razoável, possibilitando que sejam processados automaticamente (por exemplo, uma tabela em PDF é muito bem compreendida por pessoas, mas para um computador é apenas uma imagem; uma tabela em formato estruturado, como CSV ou XML, é processada mais facilmente por *softwares* e sistemas).

Não discriminatórios: Devem estar disponíveis para qualquer pessoa, sem necessidade de cadastro ou qualquer outro procedimento que impeça o acesso.

Não proprietários: Nenhuma entidade ou organização deve ter controle exclusivo sobre os dados disponibilizados.

Livres de licenças: Não devem estar submetidos a *copyrights*, patentes, marcas registradas ou regulações de segredo industrial.

Restrições razoáveis quanto a privacidade, segurança e outros privilégios são aceitas, desde que transparentes e bem justificadas.

Se o leitor estiver curioso em saber por que é tão importante compreender o que realmente significa “aberto” e por que essa definição é utilizada para falar de dados governamentais, há uma resposta bem simples: interoperabilidade, ou seja, a capacidade de diversos sistemas e organizações trabalharem juntos (interoperarem) – no caso, a possibilidade de interoperar diferentes conjuntos de dados. E essa possibilidade de “plugar” componentes é essencial.

A interoperabilidade serve para construir sistemas grandes e complexos. Sem ela, é algo praticamente impossível – como no mito da Torre de Babel, quando a incapacidade de comunicar (interoperar) grupos diferentes resultou na completa paralisação do esforço de construção da torre.

Enfrenta-se uma situação parecida em relação aos dados. O núcleo de um *commons* (ou conjunto comum) de dados é constituído quando um pedaço de material “aberto” nele contido pode ser livremente misturado com outros materiais “abertos”. Essa interoperabilidade é absolutamente central para se realizar o principal benefício prático da abertura: a habilidade de combinar diferentes conjuntos de dados e, assim, desenvolver mais e melhores produtos e serviços (cujos benefícios foram discutidos mais detalhadamente na seção anterior, “Por que dados abertos”).

Com uma definição clara de abertura, ligada à ideia de interoperabilidade, garante-se que um desenvolvedor, tendo acesso a dois conjuntos de dados abertos de duas fontes diferentes, será capaz de combiná-los. Isso, por sua vez, evita que ocorra o efeito “Torre de Babel” – bem parecido com o que acontece hoje em relação a dados de governo: existem muitas bases de dados, mas

há pouca ou nenhuma possibilidade de combiná-las em sistemas maiores, em que poderia ser ressaltado seu valor real.

Um exemplo: se dois departamentos do governo indicam endereços de locais onde é possível acessar diferentes serviços públicos (como uma base de dados sobre escolas e uma sobre hospitais) de uma determinada localidade, com bases abertas e interoperáveis é possível construir um mapa que mostre todas essas informações em um só lugar. Sem essas condições, será muito difícil criar um serviço assim.

de que dados se está falando?

No processo de abertura de dados, o foco não está em dados pessoais, mas naqueles que não contêm informação sobre indivíduos específicos. Da mesma forma, para alguns tipos de dados governamentais, podem ser aplicadas restrições de segurança nacional. Fala-se aqui dos dados governamentais *públicos*, aqueles que já estão ou deveriam estar expostos para a sociedade e que, assim, têm o potencial de se tornarem abertos, garantindo sua reutilização em novos projetos.



como abrir os dados?

Há três normas-chave que se recomenda seguir no processo de abertura de dados:

Simplicidade. Comece com um plano pequeno, simples e rápido. Não é obrigatório que todo conjunto de dados seja aberto imediatamente. Um bom início é abrir somente um conjunto de dados, ou mesmo uma parte de um grande conjunto de dados (mas é claro que, quanto mais conjuntos se puder abrir, melhor). Trabalhar o mais rápido possível é bom, pois isso significa que se pode criar o momento e aprender com a experiência. Mas se fala aqui de inovação, e isso depende de falhas e sucessos, e nem todo conjunto de dados será imediatamente útil, mas já ajuda a avançar no processo de abertura.

Envolve-se com as comunidades logo no início do processo, e mantenha contato com elas. Dialogue com os usuários atuais e potenciais dos dados assim que puder, sejam eles cidadãos, empresas, organizações não governamentais ou desenvolvedores. É importante ter em mente que muitos dados não atingirão diretamente os usuários finais, mas isso acontecerá através de interme-

diários, que se apoderam dos dados e os transformam e recombina para, então, reapresentá-los ao público. Por exemplo: muitas pessoas não precisam de uma grande base de dados de vetores geográficos, mas estão interessadas nos mapas que podem ser criados a partir dali. É provável que os primeiros usuários a se envolver no processo, quando os dados estiverem abertos, sejam esses intermediários, que irão reutilizar e ressignificar o material.

Atenção aos medos comuns e mal-entendidos. Isso é especialmente importante quando se está trabalhando com ou dentro de grandes instituições de governo. Ao abrir dados, surgirão muitos questionamentos (e medos), e por isso é essencial identificar os principais entraves e, então, resolvê-los o mais cedo possível.

Há três passos principais na abertura dos dados, que aqui são apresentados em ordem aproximada, pois alguns deles podem acontecer simultaneamente.

- 1. Escolha os conjuntos de dados** que planeja abrir.
- 2. Disponibilize os dados** estruturados e em um formato aberto reutilizável. Deve-se também considerar métodos alternativos de disponibilização, como uma API (*application programming interface*, ou interface de programação de aplicativo – adiante, veremos o que é isso).
- 3. Torne-os mais fáceis de encontrar.** Publique os dados na rede ou organize um repositório/catálogo central para listar todos os conjuntos de dados.

escolha os conjuntos de dados

Todos os dados públicos interessam, mas é preciso começar de algum ponto. Escolher o conjunto de dados que se deseja abrir é o primeiro passo a ser tomado – mas lembre-se de que todo o processo de abertura de dados é interativo, sendo possível retornar para esta etapa caso apareçam problemas no futuro.

Importante: dados abertos idealmente não têm restrições legais de uso, com exceção apenas de uma possível atribuição da origem. Reflita sobre as licenças e restrições de propriedade intelectual que podem recair sobre os dados (há algumas informações úteis sobre isso no anexo “Acesso a dados públicos no Brasil”).

Se o leitor já sabe exatamente que dados planeja abrir, pode passar para a próxima seção deste manual. Mas, em muitos casos, especialmente em grandes instituições, pode ser um desafio escolher que conjuntos de dados é possível/necessário publicar. Como proceder nesse caso?

Recomenda-se perguntar para a comunidade (usuários reais ou potenciais dos dados) o que eles querem. Como fazer isso? Há muitas formas de receber a opinião da comunidade a respeito dos dados que eles gostariam de ver abertos. A seguir indicamos um processo possível.

1. Crie uma página da web solicitando sugestões de bases de dados (que podem chegar via *e-mail* ou formulário). Algumas dicas:
 - » Sempre que possível, evite exigir cadastro.
 - » Se estiver preocupado com a possibilidade de a comunidade que pode se interessar pelos dados não ter opinião sobre que informações disponibilizar (a ponto de não conseguir fazer requisições), prepare uma pequena lista com

cinco a vinte exemplos de conjuntos de dados que podem ser liberados, para tentar identificar os mais importantes para as pessoas.

2. Circule a solicitação por listas de *e-mail*, fóruns e pessoas relevantes, apontando sempre para o endereço de *e-mail* ou para uma página de requisições.
3. Há ainda a opção de se fazer um evento de consulta, mas é bom se certificar de organizar o evento de forma que ativistas, representantes de organizações, empresários e, principalmente, *hackers* de dados possam participar.

Criar uma lista de requisições deve ser um processo rápido, que serve para identificar os conjuntos de dados que podem/devem ser abertos. Não há requisitos para se criar a lista dos primeiros conjuntos de dados abertos. Para dar o primeiro passo, o principal é considerar a viabilidade da publicação dos dados e o interesse das pessoas.

disponibilize os dados

Os dados precisam ser técnica e politicamente abertos. Ou seja, precisam estar:

Disponíveis para download, prontos para serem transferidos gratuitamente pela Internet. Em suma, sempre que possível, deve-se publicar a informação na Internet.

Completo. Um serviço de acesso automático pela *web* pode ser bem útil para quem quer acessar a versão atualizada dos dados, mas não é um substitutivo para o acesso dos dados brutos, em sua totalidade. Os dados devem estar disponíveis por completo.

Em um formato aberto e compreensível por máquina. A legibilidade por máquinas é importante, pois facilita a reutilização.

O principal quando da disponibilização dos dados é manter o projeto simples, mover-se rapidamente e ser pragmático. E, acima de tudo, é melhor liberar imediatamente dados brutos para *download*, no formato em que estiverem disponíveis, do que dados perfeitos seis meses depois.

torne-os mais fáceis de encontrar

Há muitas formas de disponibilizar os dados: eles podem ser publicados em páginas da *web*, podem ser expostos via uma interface de consulta em um *website*, ou podem ser acessados diretamente por sistemas eletrônicos via uma API (interface de programação de aplicativo).

Uma API, segundo a Wikipédia (<http://pt.wikipedia.org/>), é “um conjunto de rotinas e padrões estabelecidos por um *software* para a utilização das suas funcionalidades por programas aplicativos que não querem envolver-se em detalhes da implementação do *software*, mas apenas usar seus serviços”. Basicamente, é um intermediário entre os dados e os sistemas que vão reutilizá-los, que permite que programas e aplicativos tenham acesso automático aos dados. Disponibilizar informações via API pode garantir, por exemplo, que os dados estejam sempre atualizados – até mesmo em tempo real – nos sistemas que forem reutilizá-los. Se possível, é uma ação interessante, mas torna-se secundária quando se trata da disponibilização completa dos dados brutos para *download*.

A disponibilização de conjuntos de dados brutos, próximos do formato em que foram colhidos na fonte, com o máximo de detalhamento possível, permite que outras pessoas usem os dados não apenas para leitura (por exemplo, a conversão dos dados em um novo formato, referenciando e integrando com outros recursos e possibilitando o armazenamento dos dados em múltiplos locais).

Uma opção interessante de disponibilização é a última versão dos dados estar disponível via API, enquanto dados brutos e completos são publicados integralmente para *download*, em intervalos regulares de tempo.

O serviço de estatísticas Eurostat tem uma prática de disponibilização de dados completos para *download*, com mais de 4 mil arquivos de dados. O serviço é atualizado duas vezes ao dia, oferecendo dados em TSV (valores separados por abas), e inclui documentação sobre instalação para *download*, assim como sobre os arquivos de dados.

Outro exemplo é o conjunto de dados OCTO, do Distrito de Columbia, nos Estados Unidos, que permite o *download* de dados em formato CSV e XLS, além de ter *live feeds* dos dados – ou seja, é possível “assinar” e, assim, receber automaticamente as atualizações das bases de dados, tal qual acontece com os *feeds* de sítios de notícias.

Uma prática interessante para tornar os dados abertos mais fáceis de encontrar é criar catálogos de dados. Eles funcionam como um espaço de acesso a todos os conjuntos de dados disponibilizados por um governo e seus departamentos, e ainda servem para convidar a sociedade a reutilizar as informações. É por meio desses catálogos, como o norte-americano Data.Gov (<http://data.gov/>), nos Estados Unidos, ou o Data.Gov.Uk (<http://data.gov.uk/>), no Reino Unido, que governos mostram os formatos em que os con-

juntos de dados estão disponíveis, fornecem *links* para *feeds* ou APIs e permitem que os usuários indiquem os conjuntos de dados mais relevantes para as comunidades, num processo de gestão colaborativa da informação pública.

abrimos alguns dados. e agora?

Além de tornar a informação do governo política e tecnicamente reutilizável, assim como de fácil acesso para o público, aqueles que publicam dados abertos podem querer fazer algo mais para encorajar outros a criar coisas úteis e interessantes com esses dados.

Esta seção trata de ações adicionais para promover a reutilização de dados governamentais. São apresentadas algumas sugestões, mas o leitor não deve deixar de criar e documentar (compartilhando com o mundo) suas próprias estratégias para tornar os dados conhecidos e muito mais úteis para a sociedade.

Conte ao mundo

Em primeiro lugar, quando se abre uma série de conjuntos de dados, é válido dedicar um pouco de tempo para se certificar de que as pessoas saberão (ou pelo menos poderão descobrir) que isso foi feito. Além de *releases* de imprensa, anúncios em *website* e outras formas mais tradicionais de comunicação, vale considerar:

- » Contatar organizações ou indivíduos que trabalham com, ou estão interessados, na área dos dados abertos;
- » Contatar listas de e-mail e grupos em redes sociais relevantes (como, por exemplo, as comunidades de desenvolvedores que trabalham com tecnologias abertas);
- » Contatar diretamente prováveis usuários desse conhecimento, que podem se interessar pelos dados (como acadêmicos e jornalistas).

conecte pessoas: desconferências, *meetups* e *barcamps*

Encontros presenciais podem ser muito eficientes para encorajar outras pessoas a usar os dados. Entre as razões que podem ser consideradas para se organizar um evento estão:

- » Saber mais sobre quem são os prováveis usuários dos dados;
- » Saber mais sobre a demanda por diferentes conjuntos de dados (além dos já publicados);
- » Saber mais sobre como as pessoas querem reutilizar os dados;
- » Permitir que os prováveis usuários saibam mais sobre os dados que podem ser abertos;
- » Permitir que prováveis usuários se encontrem (para que possam colaborar entre si);
- » Expor os dados para uma ampla audiência (por meio dos *blogs* ou da cobertura de mídia gerada por conta do evento).

Há também formas diferentes de se organizar eventos e diversos tipos de eventos, que variam de acordo com o objetivo. Existem os modelos mais tradicionais de conferências, que incluem palestras formais, apresentações e demonstrações, mas há também vários tipos de eventos dirigidos pelos próprios participantes – uma ideia que combina mais com a participação e autonomia possibilitadas pelos dados abertos. Nesses eventos de formato inovador, os participantes podem (e devem):

- » Guiar ou definir a agenda do evento;
- » Apresentar-se, falar sobre o que estão interessados e no que trabalham, de uma maneira descentralizada e emergente;
- » Fazer pequenas apresentações improvisadas sobre seu trabalho atual;
- » Liderar sessões sobre temas que lhes interessam.

Há bastante documentação na *web* sobre como organizar eventos deste tipo, que pode ser encontrada ao se pesquisar *desconferência*, *barcamp*, *meetup*, *speedgeek*, *lightning talk*, etc. Também pode ser interessante contatar pessoas que organizaram esses tipos de evento, que certamente estarão dispostas a ajudar e aconselhar.²

Também pode ser de grande valor estabelecer uma parceria com outra organização (como uma organização da sociedade civil, uma organização de mídia ou um instituto educacional) para ampliar sua base de participantes e aumentar sua exposição na Internet e na mídia.

² Existe no Brasil uma comunidade ligada a um modelo de desconferência, o *barcamp* – hoje com menos atividade, já que se transformou em diversos grupos que fazem encontros descentralizados relacionados a temas tão diversos como modelos de negócios *startup*, blogosfera, mulheres na Internet, entre outros. Para acessar a documentação em português, procure no sítio (<http://www.barcamp.com.br/>) ou na lista de discussão (<http://groups.google.com/group/barcampbrasil>) do Barcamp Brasil.

Não se esqueça de que o formato e a localização do evento devem privilegiar a participação dos cidadãos, e não de outros membros de governo. Faça a divulgação e as inscrições pela rede, e facilite a presença de *hackers* e ativistas.

faça coisas: *hackdays*, prêmios e protótipos

A estrutura das competições em torno de dados abertos é a seguinte: uma série de conjuntos de dados é lançada e os programadores têm um período de tempo, de 48 horas a até algumas semanas, para desenvolver aplicativos. O melhor aplicativo construído com base nos dados é, então, premiado. Vários governos e organizações de diversos países (como Reino Unido, Estados Unidos, Noruega, Austrália, Espanha, Dinamarca e Finlândia) já realizaram competições semelhantes. Alguns exemplos:

Show Us a Better Way (<http://showusabetterway.co.uk> – Mostre-nos um jeito melhor) foi a primeira competição do tipo realizada no mundo. Criada pela Information Power Taskforce, do governo do Reino Unido, foi liderada pelo Ministro de Gabinete Tom Watson, em março de 2008. A competição perguntava “O que você criaria com uma informação pública?”, tendo sido aberta para programadores de todo o mundo, com um tentador prêmio de 80 mil libras esterlinas para os cinco melhores aplicativos.

Apps for Democracy (<http://www.appsfordemocracy.org/> – Aplicativos para democracia) foi uma das primeiras competições nos Estados Unidos, lançada em outubro de 2008 por Vivek Kundra, na ocasião Chief Technology Officer (CTO, gerente de tecnologia) do governo do Distrito de Columbia. Kundra tinha desenvol-

vido o inovador catálogo de dados abertos OCTO (<http://data.octo.dc.gov/>), que disponibilizava diversas informações, como atualizações em tempo real das incidências de crimes no distrito, notas escolares anonimizadas e indicadores de pobreza por região – na ocasião, o mais abrangente conjunto de dados abertos locais do mundo. O desafio era torná-lo útil para os cidadãos, visitantes, empresários e agências do governo de Washington, e a solução foi criar o concurso, pedindo às pessoas que criassem aplicativos usando os dados do conjunto recém-lançado. O concurso incluía uma inscrição *on-line* dos aplicativos, muitos pequenos prêmios no lugar de poucas grandes premiações e várias categorias (entre elas, a “Escolha do Público”). A competição ficou aberta por trinta dias e custou ao governo distrital 50 mil dólares. Em retorno, foi desenvolvido um total de 47 aplicativos para *web*, iPhone e Facebook, com um valor estimado superior a 2,6 milhões de dólares, então revertidos para a economia local.

O **Desafio Abre Datos** (<http://www.abredatos.es/>), organizado na Espanha em abril de 2010, convidou desenvolvedores a criar, em apenas 48 horas, aplicativos de código aberto fazendo uso de dados públicos. A competição teve 29 equipes participantes, que desenvolveram diversos programas: entres eles, um aplicativo de celular para acessar informações de tráfego no País Basco e outro para acessar dados sobre as linhas e pontos de ônibus em Madri – que levaram, respectivamente, o primeiro (3 mil euros) e segundo prêmios (2 mil euros).

Nettskap 2.0 (<http://www.regjeringen.no/en/dep/fad/kampanjer/nettskap.html?id=635971>), organizado em abril de 2010 pelo Ministério de Administração do governo norueguês. Desenvolvedores de todo o país – além de empresas, agências públicas e indivíduos – foram desafiados a apresentar ideias de projetos *web* nas áreas de desenvolvimento de serviços, processos de trabalho

mais eficientes e aumento da participação democrática. O uso de dados governamentais foi explicitamente encorajado. Embora o prazo para envio de aplicativos fosse somente de um mês após a abertura das inscrições, em 9 de maio o ministro Rigmor Aasrud anunciou um retorno “esmagador”: um total de 137 aplicativos recebidos, e nada menos do que 90 deles baseados na reutilização dos dados do governo. 2,5 milhões de noks foram distribuídos entre os dezessete vencedores, enquanto o montante financeiro gerado pelos aplicativos foi avaliado em 28,4 milhões de noks.

Mashup Australia (<http://mashupaustralia.org/>). A força-tarefa australiana de governo 2.0 convidou os cidadãos a mostrar por que o acesso aberto a informações do governo seria positivo para a economia do país e para o desenvolvimento social. O concurso, que contou com 82 aplicativos, durou de 7 de outubro a 13 de novembro de 2009, tendo publicado alguns conjuntos de dados sob uma licença aberta e em diversos formatos reutilizáveis.

Microbolsas Hacker (<http://blog.esfera.mobi/sobre-as-microbolsas-hackers/>). Organizadas no Brasil pela comunidade Transparência Hacker, têm um modelo diferente das competições tradicionais. Os desenvolvedores (ou *hackers* e ativistas de qualquer área) submetem ideias de projetos em vez de aplicativos prontos, gerando cerca de 10 ideias por edição mensal. As duas melhores propostas levam uma pequena bolsa de R\$ 2 mil para começar a se transformar em projeto, e todas as ideias recebem *feedback* completo dos avaliadores.

Na outra ponta da abertura dos dados de governo, uma das maneiras mais efetivas para as organizações da sociedade civil e comunidades demonstrarem aos governos o valor dos seus conjuntos de dados é usar as informações públicas para gerar inovações, sendo possível mostrar as múltiplas maneiras de gerenciar a informação alcançando vantagens sociais e econômicas. As

organizações e comunidades que utilizam ou que promovem a reutilização de dados públicos têm sido fundamentais em países onde há avanços políticos e na legislação para garantir que os conjuntos de dados sejam política e tecnicamente abertos. Em países que ainda não seguem essa tendência, são as comunidades formadas em torno de dados abertos que propõem e impulsionam inovações em torno do tema.

Essas organizações e comunidades normalmente realizam e participam de competições e conferências a respeito de dados abertos, além de realizarem desconferências, oficinas e *hackdays* (eventos mais “mão na massa”, em que os participantes desenvolvem projetos e aplicativos). Tais atividades são organizadas pelas comunidades de usuários com dados já publicados, seja proativamente ou por meio de requisições de acesso a informação. Em outros casos, os ativistas trabalham com funcionários públicos progressistas para conseguir a abertura de novas bases de dados, de modo que possam ser utilizadas por programadores e *hackers* para criar novos aplicativos e projetos. Outra linha de atuação são os movimentos que “raspam” ou agregam dados de sítios governamentais, montando bases de dados úteis para o desenvolvimento.

Governos também podem estimular o uso de dados abertos usando essa mesma estratégia das comunidades e organizações: eventos com foco no aprendizado de novas tecnologias de dados ou no desenvolvimento de soluções, como oficinas e *hackdays*. É importante convidar grupos da sociedade a participar ativamente dos eventos, desde o planejamento até a avaliação de resultados. Para saber mais sobre como organizar um evento como esse, veja os Anexos.

Transparência Hacker

Trata-se de uma opção de divulgação e estímulo ao uso de dados abertos para governos e departamentos públicos brasileiros. Apesar de não existirem políticas sistemáticas de abertura de dados de governo, um grupo da sociedade começou a se organizar em torno da questão a partir de outubro de 2009, quando a Esfera, empresa ligada à questão da transparência pública, da política e das tecnologias, organizou o primeiro Transparência HackDay (<http://blog.esfera.mobi/transparencia-hackday-convide-a-participacao/>).

Os *hackdays* são encontros de desenvolvedores para colaborar em projetos inovadores. A proposta do Transparência HackDay era alinhar essa prática ao interesse do grupo, trabalhando especialmente com dados de governo. Na ausência de dados, os *hackers* trabalham com as informações disponíveis em sítios ou relatórios públicos, transformando-as em conjuntos de dados abertos e criando novos serviços e sítios a partir daí. A ideia é começar uma abertura pela ação da sociedade.

Desses encontros, organizados de forma bem descentralizada, surgiu a Comunidade Transparência Hacker, grupo hoje com quase 500 membros, que se encontram para conversar e criar ações provocativas sobre o tema dos dados abertos: projetos, sítios, *hackdays*, *techtalks* (eventos que apresentam novas tecnologias abertas para grupos da sociedade não familiarizados com o tema) e, inclusive, a criação de documentos úteis como este manual.

A comunidade é aberta a quem quiser participar, e, além de desenvolvedores, conta com ativistas, representantes de organizações da sociedade civil, jornalistas, pesquisadores e diversos oficiais do governo, que se articulam com a sociedade para estimular políticas de dados abertos em seus departamentos.

Se tiver interesse em abrir dados governamentais no Brasil e não souber como fazer (ou não sabe como encontrar pessoas para utilizá-los), há diversas iniciativas das quais é possível participar livremente. Veja algumas delas:

Comunidade *Thacker*: A comunidade Transparência Hacker é um espaço aberto de proposição e articulação de ideias e projetos que utilizem a tecnologia para fins de interesse da sociedade, principalmente com dados governamentais abertos. Acima de tudo, busca evidenciar questões sociais e políticas por meio da ressignificação de informações existentes, mas que ainda são de difícil acesso para a sociedade em geral.

site:

<http://thacker.com.br/>

lista do grupo:

<http://groups.google.com/group/thackday> e
thackday@googlegroups.com

wiki:

<http://wiki.thacker.com.br>

twitter:

tag #thackday

GT Dados Abertos W3C Brasil: Grupo de trabalho para a publicação e a reutilização de dados abertos, coordenado pelo W3C Brasil.

Lista de discussão:

para se inscrever, envie um e-mail para
gt-dadosabertos@nic.br

site:

<http://www.w3c.br>

twitter:

<http://twitter.com/w3cbrasil>

GT INDA: Grupo de trabalho da Infraestrutura Nacional de Dados Abertos, coordenado pela Secretaria de Logística e Tecnologia da Informação, do Ministério do Planejamento, para estabelecer padrões, tecnologias, procedimentos e mecanismos de controle necessários para disseminação e compartilhamento de dados e informações públicas no modelo de dados abertos.

lista do grupo:

<http://groups.google.com/group/lista-inda-gt>
(lista fechada, participação mediante convite)

OpenData-BR: Grupo com interesse em desenvolver material técnico em relação a dados públicos, de forma a auxiliar o governo, na disponibilização de dados, e a sociedade, na visualização deles.

site:

<http://www.opendatabr.org>

lista do grupo:

<http://groups.google.com/group/opendata-br>

algumas informações técnicas

APIs, web services e bancos de dados

Quando um departamento precisa disponibilizar novos dados, deve-se escolher o formato que propicia o máximo equilíbrio entre custo e adequação ao propósito dos dados. Para cada formato, há alguns aspectos que devem ser levados em conta, que serão explicados nesta seção. Serão abordadas a disponibilização e a organização de dados de forma que máquinas consigam acessá-los diretamente, do modo mais simples possível.

APIs e web services

Para dados que mudam frequentemente e em que cada requisição ao servidor tem um tamanho limitado, é bastante recomendável expor os dados por meio de serviços *web*. Há várias maneiras

de criá-los, mas alguns dos mais usados são SOAP e REST – que, além de serem padrões amplamente utilizados, são relativamente simples de se desenvolver.

acesso aos bancos de dados

Assim como no caso dos *web services*, há acesso direto a bases de dados. Mas este serviço tem a vantagem de, na hora de desenvolver novas utilizações com os dados, permitir aos usuários que selecionem apenas as tabelas e informações que lhes interessam. Há algumas questões de seguranças envolvidas no acesso remoto a banco de dados, e o acesso só é realmente útil se a estrutura e a importância de tabelas, assim como os campos individuais, estiverem bem documentadas.

Em geral, é relativamente simples e barato criar *web services* que exponham os dados de um banco, o que pode ser uma maneira fácil de resolver questões de segurança.

formatos

XML

É um formato amplamente usado para troca de dados, pois possibilita que se mantenha a estrutura dos dados em operações diferentes. O modo como os arquivos XML são construídos permite aos desenvolvedores escrever parte da documentação dentro dos dados, sem interferir na sua leitura.

JSON

É um formato de arquivo bem fácil de ser interpretado por qualquer linguagem de programação, ou seja, costuma ser mais fácil para os computadores processarem JSON do que outras linguagens, como o XML.

RDF

É um formato relativamente novo, que torna possível apresentar dados de forma que as máquinas consigam compreender e interpretar informações adicionais padronizadas sobre eles. É usado parcialmente em projetos Linked Open Data, da União Europeia. Ainda não é muito difundido, mas oferece excelentes oportunidades para a interoperabilidade e o processamento automático de dados governamentais.

CSV

Pode ser um formato bem útil, pois é compacto e, portanto, adequado para transferir grandes conjuntos de dados com uma mesma estrutura. Entretanto é tão simples, que os dados chegam a ser inúteis se não houver uma documentação das suas características, uma vez que pode ser quase impossível adivinhar o significado das diferentes colunas. Assim, é importante para as tabelas separadas por vírgulas (*comma separated values* – CSV) que haja documentação precisa sobre cada campo.

Além disso, é essencial que a estrutura do arquivo seja respeitada, já que a omissão de um único campo pode atrapalhar a

leitura de todos os outros dados, sem oportunidade de retificá-lo, porque seria impossível determinar como o restante dos dados deve ser interpretado.

planilhas

Muitas organizações possuem informação em planilhas, como, por exemplo, as do *software* Microsoft Excel. Estes dados podem ser usados de forma imediata, adicionando-se apenas a descrição correta das colunas. Em alguns casos, existem macros e fórmulas nas planilhas, o que gera um trabalho mais intenso. Portanto é sempre aconselhável documentar os cálculos e fórmulas na planilha, de maneira a tornar as informações mais acessíveis ao leitor.

Importante: para disponibilizar uma planilha, pode ser necessário mudar o formato do arquivo, de forma que ela seja acessada por qualquer tipo de *software*, sem privilegiar nenhuma empresa ou organização (de XLS para CSV, por exemplo).

documentos de texto

Documentos clássicos, em formatos como DOC, ODF, OOXML, podem ser suficientes na hora de apresentar alguns tipos de dados – por exemplo, listas de *e-mail* relativamente estáticas ou informações parecidas. É barato exibir nesses formatos, já que costuma ser o formato em que os dados foram gerados. Não há suporte no que diz respeito a manter a estrutura consistente, ou seja, é difícil acessá-los por meios automatizados. Certifique-se de usar *templates* como base dos documentos que trarão os da-

dos a serem reutilizados, permitindo-se que as informações sejam extraídas dos documentos por outras ferramentas sem grandes perdas. Não é recomendado exibir os dados no formato de texto se houver informações em formatos diferentes.

PDF

Os PDFs costumam ser usados para circulação interna de documentos nos órgãos governamentais, já que não permitem modificação (são somente leitura) e por manterem a diagramação original do conteúdo – podendo-se incluir tabelas, gráficos e imagens. No entanto são completamente inúteis para a disponibilização de dados abertos, já que não armazenam as informações de maneira estruturada. Lembre-se de que, toda vez que gerar um PDF de uma tabela ou documento de texto, um desenvolvedor terá de trabalhar muito para “desfazer” essa ação e voltar para um formato que reconheça os dados de forma mais estruturada.

TXT

Documentos de texto puro (TXT) são muito fáceis de ser lidos por computadores. Entretanto esse formato de arquivo exclui os metadados estruturais de um documento, ou seja, os desenvolvedores precisam criar um analisador que interpretará cada documento na medida em que os dados forem aparecendo.

Alguns problemas podem ocorrer ao se trocar TXTs entre sistemas operacionais diferentes. MS Windows, Mac OSX e variantes do Unix têm seu próprio método de informar ao computador que se chegou ao fim do texto.

imagens digitalizadas

Provavelmente, é o formato menos adequado para a maioria dos dados, mas tanto o TIFF quanto o JPEG podem, pelo menos, ser utilizados para mostrar como imagens informações que não foram criadas eletronicamente desde o início (por exemplo, arquivos de uma igreja antiga ou outros acervos tradicionais). Afinal de contas, para a transparência, uma imagem ainda é melhor do que nada.

HTML

Atualmente, muitos dados estão disponíveis em formato HTML em vários sites governamentais. Isso pode ser suficiente, já que os dados em HTML são muito estáveis e têm escopo limitado. Em alguns casos, é preferível ter os dados em um formato mais fácil de ser baixado e manipulado, com custo baixo e a facilidade de *linkar* em *website*.

Para armazenar dados, é mais apropriado usar tabelas em documentos HTML, e é importante que os vários campos exibidos tenham uma identificação que facilite o processo de busca e manipulação.

O Yahoo! desenvolveu uma ferramenta chamada YQL (<http://developer.yahoo.com/yql/>), que extrai informações estruturadas de um *website* em formato HTML. Ferramentas assim podem fazer muito mais com os dados se eles forem devidamente nomeados e classificados.

formatos abertos

Mesmo se as informações forem fornecidas em formatos compreensíveis por máquinas e eletrônicos, ainda pode haver impedimentos em relação ao formato do arquivo.

O formato em que a informação é publicada – em outras palavras, a base digital onde a informação é armazenada – pode ser “aberto” ou “fechado”. Um formato aberto é aquele cujas especificações do *software* são abertas para qualquer pessoa ver, que é gratuito e que pode ser aberto por diferentes tipos de programas, sem limitação de reutilização imposta por direitos de propriedade intelectual. Se um formato de arquivo é “fechado”, é possível que suas especificações não estejam disponíveis publicamente, ou que, mesmo com as especificações abertas ao público, a reutilização seja limitada.

Liberar dados em um formato fechado pode causar obstáculos significativos para a reutilização da informação codificada dentro do arquivo, forçando aqueles que desejam usar a informação a comprar o *software* necessário para acessá-la.

formatos proprietários

Alguns sistemas dedicados têm seus próprios formatos de dados para guardar ou exportar dados. Se for necessário utilizá-los, deve sempre ser indicado onde encontrar mais informações sobre esses formatos proprietários (como, por exemplo, um *link* para o *website* do fornecedor).

É sempre recomendado disponibilizar dados em formatos não proprietários. O acesso a informações públicas não deve ser restrito aos interesses de companhias e organizações.

anexos

acesso a dados públicos no Brasil

(Adaptado da dissertação de mestrado *Transparência na esfera pública interconectada*,³ publicada em abril de 2010.)

A publicidade dos atos de governo é um princípio democrático, que no Brasil aparece expressa no artigo 5º da Constituição: “todos são iguais perante a lei, sem distinção de qualquer natureza, garantindo-se aos brasileiros e aos estrangeiros residentes no País a inviolabilidade do direito à vida, à liberdade, à igualdade, à segurança e à propriedade”. No inciso XXXIII deste artigo, determina-se que *“todos têm direito a receber dos órgãos públicos informações de seu interesse particular, ou de interesse coletivo ou geral, que serão prestadas no prazo da lei, sob pena de responsabilidade, ressalvadas aquelas cujo sigilo seja imprescindível à segurança da sociedade e do Estado”*. No artigo 37 da Constitui-

³ Daniela Silva, texto disponível em: http://blog.esfera.mobi/wp-content/uploads/2010/04/transparencia_esferapublica1.pdf.

ção, fica explícito que “a administração pública direta e indireta de qualquer dos Poderes da União, dos estados, do distrito federal e dos municípios obedecerá aos princípios de legalidade, impessoalidade, moralidade, *publicidade e eficiência*”.

Por causa das novas tecnologias, novas questões passam a se relacionar com esses princípios. Como garantir o direito de receber informações públicas quando a relação entre governo e cidadãos é intermediada pelas máquinas? O que significa ser público e eficiente na era da Internet?

O acesso a dados públicos no país, até o presente momento, é garantido com base na Constituição, que determina que, a não ser em casos de exceção, as informações de governo devem ser públicas. É com base nessa garantia que desenvolvedores acessam e agregam os dados públicos presentes em sítios governamentais, quando esses ainda não estão publicados em formatos abertos, para fazer cruzamentos ou para construir novos projetos, sítios e aplicativos.

Infelizmente, o direito garantido no artigo 5º da Constituição ainda não está regulado na legislação brasileira. Entre as leis que se referem ao artigo, a Lei nº 11.111/2005, por exemplo, serve mais para destacar as exceções de acesso à informação pública do que para determinar a responsabilidade dos órgãos de oferecer informações para os cidadãos. Também há regulamentações que determinam a transparência dos dados orçamentários (como a Lei de Responsabilidade Fiscal, Lei Complementar nº 101/2000) e de informações relacionadas a assuntos ambientais (Lei nº 10.650/2003), mas falta uma lei que, como prevê a Constituição, determine prazos e formas para que o poder público atenda a pedidos de informação pública por parte da sociedade e, mais importante ainda, defina que o governo deva ter uma postura ativa na publicação e disponibilização de dados e informações públicas.

Nos Estados Unidos, desde a década de 1960, é o *Freedom of Information Act* (FOIA) que determina esse acesso. Ao Brasil, falta uma legislação de acesso às informações públicas que cumpra as mesmas funções do FOIA, de acordo com a Constituição e diversas declarações de direitos humanos, já presente em mais de oitenta nações do mundo, inclusive Libéria e Bermudas, que aprovaram projetos nesse sentido em 2010.

Um projeto de lei brasileiro com essas características começou a ser articulado pelo Poder Executivo e organizações da sociedade civil em 2003. Foi proposto pela Casa Civil três anos depois, e aprovado pela Câmara dos Deputados em 2010. Desde então, o projeto da Lei de Acesso à Informação Pública está tramitando no Senado, como PL 41/2010.⁴

Alguns meses antes da aprovação do projeto de lei na Câmara, durante a passagem por uma das comissões que tinham o poder e a responsabilidade de modificá-lo, a Comunidade Transparência Hacker inseriu sugestões que adaptam o texto para garantir o acesso a dados abertos. Hoje, o projeto de lei engloba os oito princípios dos dados abertos, de forma a garantir o acesso a informações em formatos abertos, não proprietários e compreensíveis por máquina; entre outras características.

Governos e departamentos interessados em fazer a abertura de seus dados, portanto, podem seguir como princípio as determinações do projeto de Lei de Acesso à Informação Pública, que está de acordo com as possibilidades de gerar cruzamentos, visualizações e serviços garantidos pelas novas tecnologias em rede. Pode-se observar principalmente:

⁴ Disponível em http://www.senado.gov.br/atividade/materia/Consulta.asp?STR_TIPO=PLC&TXT_NUM=41&TXT_ANO=2010&Tipo_Cons=6&IND_COMPL=&FlagTot=1.

- » No capítulo 1, artigo 4, as definições de tratamento (“conjunto de ações referentes à produção, recepção, classificação, utilização, acesso, reprodução, transporte, transmissão, distribuição, arquivamento, armazenamento, eliminação, avaliação, destinação ou controle da informação”), autenticidade (“qualidade da informação que tenha sido produzida, expedida, recebida ou modificada por determinado indivíduo, equipamento ou sistema”), integridade (“qualidade da informação que não foi modificada, inclusive quanto a origem, trânsito e destino”) e primariedade das informações (“qualidade da informação coletada na fonte, com o máximo de detalhamento possível, sem modificações”);
- » No capítulo 2, artigo 6, a garantia de integridade e autenticidade da informação (“cabe aos órgãos e entidades públicas assegurar, observadas as normas e procedimentos específicos aplicáveis, a proteção da informação, garantindo-se sua disponibilidade, autenticidade e integridade”);
- » No capítulo 2, artigo 7, a inclusão das dimensões de primariedade e integridade no escopo do acesso à informação (“o acesso à informação de que trata esta Lei compreende, entre outros, os direitos de obter informação primária, íntegra, autêntica e atualizada”);
- » No capítulo 2, artigo 8, a determinação de publicação de informações por meio da Internet (“para cumprimento do disposto no caput, os órgãos e entidades públicas deverão utilizar-se de todos os meios e instrumentos legítimos de que dispuserem, preferencialmente sítios oficiais da rede mundial de computadores”) e na regulamentação de algumas características que devem ter os sítios com essas informações (“os sítios de que se trata o parágrafo 2 deverão, na forma de regulamento, atender, entre outros, aos seguintes requisitos:

conter ferramenta de pesquisa de conteúdo que permita o acesso à informação de forma objetiva, transparente, clara e em linguagem de fácil compreensão; possibilitar a gravação de relatórios em diversos formatos eletrônicos, inclusive abertos e não proprietários, tais como planilhas e texto, de modo a facilitar a análise das informações; possibilitar o acesso automatizado por sistemas externos em formatos abertos, estruturados e compreensíveis por máquina; divulgar em detalhes os formatos utilizados para estruturação da informação; garantir a autenticidade e a integridade das informações disponíveis para acesso; manter atualizadas as informações disponíveis para acesso; indicar local e instruções que permitam ao interessado comunicar-se, por via eletrônica ou telefônica, com o órgão ou entidade detentora do sítio; e adotar as medidas necessárias para garantir a acessibilidade de conteúdo para pessoas com deficiência”).

Não é preciso esperar que a Lei de Acesso à Informação seja aprovada para colocar em prática desde já essas determinações – assim como a sociedade, por sua vez, já está trabalhando com dados e informações públicas na Internet, demonstrando assim o seu valor para os cidadãos. Este manual contribui com essa ideia – de que, em nome da inovação e da democracia, a hora de disponibilizar dados abertos é agora. Para contribuir com o acompanhamento do processo legislativo e a demanda por uma rápida aprovação da Lei de Acesso, é possível fazer parte do Movimento Brasil Aberto, que une organizações e pessoas que pedem que a Lei tramite no Senado em regime de urgência. Acesse brasilaberto.org para saber mais.

Como (des)organizar um *hackday*

(Extraído e revisado do original publicado no *Jornal de Debates* na Campus Party, janeiro de 2010)

Sim, *hackers* servem para mudar o mundo. Com essa ideia na cabeça, nós, da Esfera (<http://blog.esfera.mobi>), organizamos, em novembro de 2009, o primeiro Transparência HackDay. *Nerds*, *geeks* e programadores já têm como prática corrente realizar *hackdays*, reunindo-se na madrugada em torno de código e pizza pelo puro prazer de resolver problemas. O que queríamos fazer era juntar a fome de codificar, cumprir desafios e subverter sistemas com a vontade de trabalhar por questões de interesse público. Nesse encontro, programadores, desenvolvedores e *geeks* se encontram com ativistas, comunicadores, acadêmicos e políticos, para, juntos, pensarem em como usar as tecnologias livres e a informação pública para criar aplicativos, sítios ou projetos que gerem mais participação cidadã na rede. Disso, já saíram um mapa com problemas da cidade de São Paulo (<http://sacsp.mamulti.com>); uma API extra-oficial do processo legislativo brasileiro; uma outra API, agora do banco de dados, até então fechado a sete chaves, da Transparência Brasil (<http://github.com/pvalente/thackday-opentables/>); e outras traquitanas (ver <http://blog.esfera.mobi/tag/thackday/> para saber mais).

Recentemente, a Sunlight Foundation – importante organização norte-americana preocupada com a questão da transparência pública e das novas tecnologias nos Estados

Unidos, e de onde tiramos a ideia do HackDay – lançou o artigo “Como organizar um *hackathon*” (<http://moourl.com/sunlighthackathon>), rapidamente apontado como bastante útil para nós. Nossa primeira impressão ao abrir o *link* foi: “Legal! Fazemos tudo errado”. É que nosso processo de organizar coisas é um tanto quanto, digamos, caótico. Como mesmo (ou principalmente) do meio do caos pode sair muita coisa interessante, e como é essencial que mais encontros como esse aconteçam pelo Brasil afora, achamos que vale a pena “*hackear*” a lista da Sunlight Foundation. Seguem, então, as nossas preciosas dicas de como (des)organizar um *hackday*:

Não espere acontecer. Quando fizemos o primeiro HackDay, não tínhamos dados, não tínhamos dinheiro, não tínhamos ideia se as pessoas se interessariam pela proposta. Um mês depois, tínhamos uma lista com mais de 100 participantes e interessados, algumas bases de dados cedidas diretamente pelo governo para o evento (fora um mundo de sítios governamentais para raspar) e condições básicas para fazer o que queríamos. Como processo, como experiência e como resultado, foi um sucesso. Transparência pública e participação política são conceitos altamente relacionados ao plano local. Portanto, se você tem vontade de contribuir com a sua cidade ou bairro dessa forma, organizando um HackDay, não precisa esperar as condições ideais para isso. Coloque uma data e corra atrás do resto: consiga uma locação, articule contatos, use e abuse da rede para angariar interessados. E também pode usar a lista groups.google.com/group/thackday e a *tag* #thackday à vontade.

Gaste pouquíssimo dinheiro. O primeiro HackDay teve apoio da Associação de Software Livre, mas pelo *timing*, e até pelo nosso próprio caos de todo dia, esse patrocínio nunca se efetivou. Colocando na ponta do lápis, tivemos a sorte de usar o espaço da Casa de Cultura Digital, onde trabalhamos, gentilmente cedido pelos nossos colegas de *cluster*, com custo zero. Recebemos doações de material de escritório e pegamos emprestados *flip-charts* que serviram como murais de organização. Servimos bolachinhas caseiras com o café. Pedimos que os participantes bancassem o próprio almoço. Posso dizer que o custo total da operação foi: R\$ Zero. Claro que, com mais estrutura, teria sido tudo melhor, mais confortável e mais fácil. Mas nós testamos e comprovamos: até sem grana nenhuma dá para fazer um *hackday*. Então, mesmo que você tenha grana, otimize recursos, essa também é uma forma sensacional de mudar o mundo.

Comunique-se do melhor jeito. Não, definitivamente você não precisa colocar uma propaganda no ar no horário nobre da Rede Globo para divulgar o evento (*thanks, Internet*). Pode-se dizer que boa parte da divulgação do primeiro Transparência HackDay foi feita pelo Twitter, por onde mandamos o *link* de um convite que estava no nosso *blog* (<http://moourl.com/thackdayconvite>). Além disso, contamos com a propaganda entre amigos e com as listas de desenvolvedores. Também foram enviados *e-mails* para uma série de jornalistas que cobriam política e novas tecnologias, mas, como a mídia tradicional só cobre o que é quente, as notícias do HackDay só saíram na véspera, quando o blablablá já tinha se feito na rede e a lista de

participantes já estava cheia. Para quem pensa que é importante abrir diálogo com o governo (nós achamos que sim), vale a pena bater na porta dos administradores públicos e legisladores, levantando a bandeira da transparência e pedindo dados públicos para “*mashupear*” ou a participação deles no evento. Pode surpreender muita gente, mas encontramos muitas portas abertas nesse sentido, e colhemos frutos interessantes dessa interação política (<http://moourl.com/thackdaydecreto>).

Acredite na capacidade de autogestão das pessoas.

Seria mentira dizer que tivemos de trabalhar insanamente durante o HackDay. Estávamos lá, ajudando a tocar o barco, mas o papel de organizador, depois que o evento já está acontecendo, acaba se confundindo muito com o dos outros participantes. No meio do dia, deu tempo de participar ativamente das discussões e até de colaborar com um projeto (<http://mapadaeja.com.br>). Muito do que fizemos foi dizer a que horas as discussões e atividades de desenvolvimento começavam e terminavam (o que, claro, não funcionou muito bem, fazendo com que felizmente o evento terminasse umas três horas depois do previsto). Quando não estão sentadas num auditório de conferência, enforcadas pela gravata, as pessoas são altamente capazes de se auto-organizar; e, apesar de a coisa fluir diferentemente do que quando está tudo bem produzido por alguns poucos, acreditamos que assim se faz um processo muito mais rico e divertido para todos.

Dê autonomia, cobre responsabilidade. Existem problemas. Na quarta-feira depois do fim de semana do Trans-

parência HackDay, ficamos sabendo que a porta de uma das casas da vila onde fica a Casa de Cultura Digital tinha sido depredada. Não ficou confirmado que isso aconteceu durante o evento, mas, como justamente naquele fim de semana um fluxo de 100 novos visitantes tinha passado por ali, os indícios eram altos o suficiente para comprometer a paz com nossos vizinhos. Foi um dia bem chato, e foi difícil pensar no que fazer naquela situação. Mas, dado que, de diversas formas, ficou claro que o evento era responsabilidade de todo mundo, não encontramos outra forma de agir a não ser liberar a lista de presença para a empresa que teve sua porta depredada e informar os participantes por *e-mail* do acontecido. É importante dar autonomia para as pessoas, mas também deixar claro que elas são responsáveis pelo espaço que ocupam e pelas relações que constroem durante um *hackday*.

Entenda o processo como o melhor resultado. Muitos dos projetos criados ou empreendidos durante o primeiro Transparência HackDay ainda estão em versão superbeta, alguns já caminharam um bocado, outros nem tanto. Muito do que foi produzido é linha de código, ferramentas fantásticas para que outros desenvolvedores possam produzir aplicativos e sítios baseados em informações públicas com muito mais facilidade. As discussões também são bastante válidas, mas dois dias obviamente não é tempo suficiente pra resolver as questões que ligam política e novas tecnologias no mundo contemporâneo. O efeito demonstrativo dessas coisas, para quem é leigo, pode ser pequeno, mas isso realmente não é um problema. O encontro, as con-

versas, as ideias e os desafios, tudo o que faz parte do processo do HackDay, são muito importantes. E mais: a formação de uma comunidade de gente interessada em usar a rede para *hackear* o mundo, com resultados inimagináveis em médio e longo prazo, é o principal resultado dessa iniciativa.

