

GUIA DE DADOS ABERTOS

GUIA DE DADOS ABERTOS

Este Guia é parte integrante do projeto de cooperação entre o
Governo do Estado de São Paulo e o Reino Unido



SPUK



Improving business environment through transparency in São Paulo State

Melhoria do ambiente de negócios por meio da transparência no Estado de São Paulo

Parceiros

ceweb.br nic.br cgi.br

SEADE
Fundação Sistema Estadual
de Análise de Dados

Fundap



Embaixada Britânica
Brasília



Ano 2015



Este material está sob uma licença Creative Commons.
Atribuição-SemDerivações-SemDerivados
CC BY-NC-ND



SPUK



Improving business environment through transparency in São Paulo State

Melhoria do ambiente de negócios por meio da transparência no Estado de São Paulo

REALIZAÇÃO

GOVERNO DO ESTADO DE SÃO PAULO

Secretaria de Governo

- Subsecretaria de Parcerias e Inovação

Casa Civil

- Assessoria Especial para Assuntos Internacionais

Fundação do Desenvolvimento Administrativo - Fundap

Fundação Sistema Estadual de Análise de Dados – Seade

Conselho de Transparência da Administração Público

GOVERNO DO REINO UNIDO

Embaixada Britânica – Brasília

NÚCLEO DE INFORMAÇÃO E COORDENAÇÃO DO PONTO Br – NIC.br

Centro de Estudos sobre Tecnologia Web – CeWeb.br

AUTOR

Marco Túlio Pires

COORDENAÇÃO

Geral:

Roberto Agune - iGovSP

Vagner Diniz – CeWeb.br

Executiva e Editorial:

Caroline Burle dos Santos Guimarães - CeWeb.br

Helena Pchevuzinske - iGovSP

Sergio Pinto Bolliger - iGovSP

IDEALIZAÇÃO

Alvaro Gregório - iGovSP

DESIGN

Alcione de Godoy - iGovSP - e-books

Deblyn Pereira Prado - NIC.br - HTML

Ricardo Hurmus - Bulerías Arte & Design - ilustrações

INTRODUÇÃO.....06

O QUE SÃO “DADOS ABERTOS”?.....09

 Qualquer dado pode ser aberto?.....11

OS BENEFÍCIOS DA ABERTURA DE DADOS.....12

 Transporte Público (Iniciativa Privada).....14

 Legislativo & Executivo, Participação (Terceiro Setor).....15

 Educação & Pesquisa (Terceiro setor).....18

 Saúde & Gastos Públicos (Governo, Cidadãos).....19

LEI DE ACESSO À INFORMAÇÃO.....21

 O acesso à informação em São Paulo.....22

 Transparência Ativa.....23

 Quais são as exceções para abertura de dados?.....24

PLANO DE ABERTURA DE DADOS.....26

 As cinco estrelas dos dados abertos: ★★★★★27

 ★27

 ★★28

 ★★★29

 ★★★★30

 ★★★★★31

 Equipes de Abertura de Dados.....33

 Publicação.....36

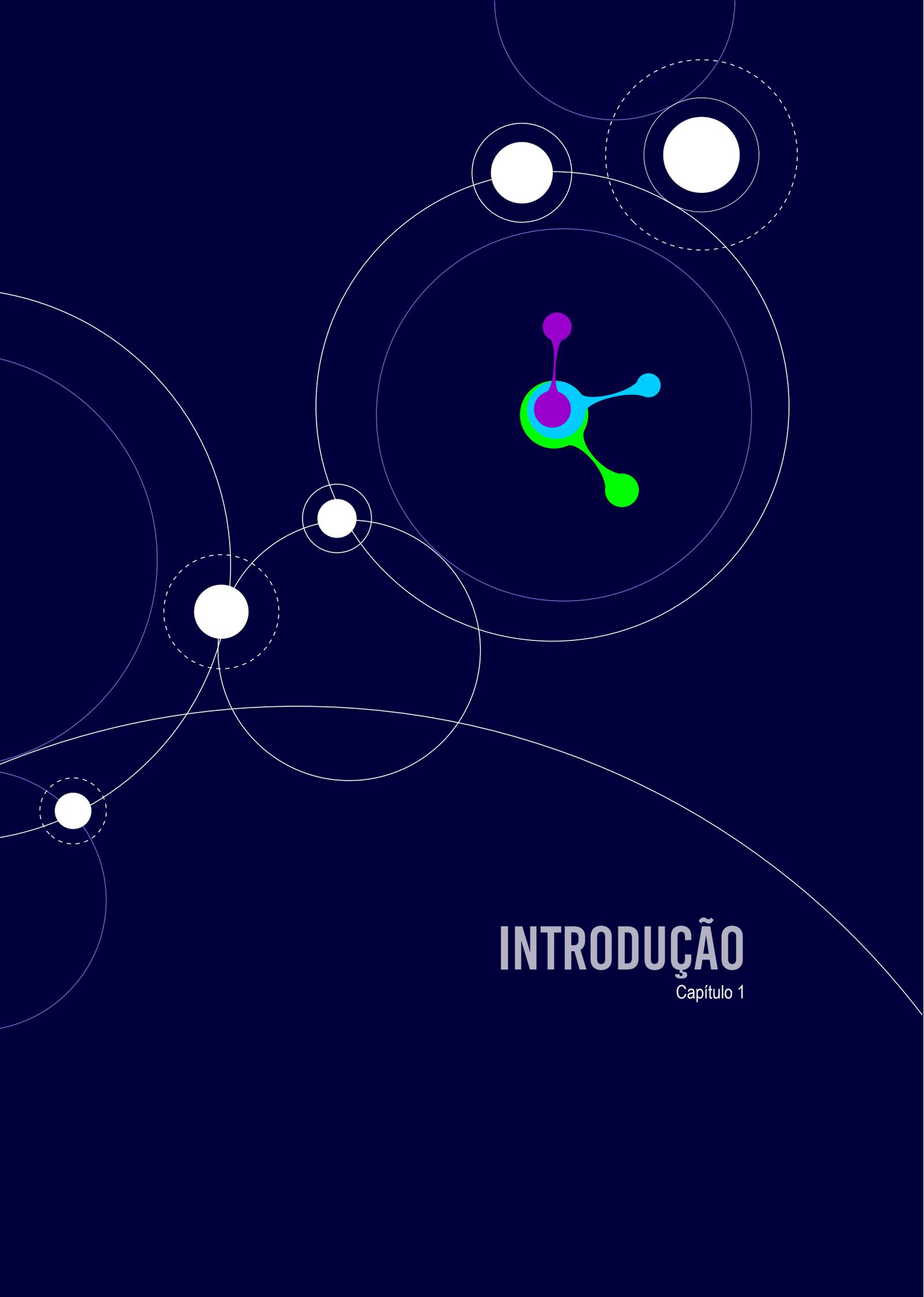
 O desenvolvimento de uma API é necessário?.....37

 Mapa de decisões tecnológicas.....40

CATÁLOGO/REPOSITÓRIO DE DADOS.....42

CENÁRIOS TÉCNICOS, OPÇÕES TECNOLÓGICAS.....45

Nível 1.....	46
Nível 2.....	48
Nível 3.....	49
LICENÇA DE UTILIZAÇÃO DOS DADOS.....	53
Regras de utilização do Governo Aberto SP.....	56
FORMATOS DAS BASES.....	57
Formatos separados por delimitadores (CSV).....	59
Formato XML.....	61
Formato KML.....	63
Formato JSON.....	64
geoJSON/topoJSON.....	66
Formato SQL (dump).....	66
Formato Shapefile.....	67
REFERÊNCIAS.....	69



INTRODUÇÃO

Capítulo 1

Este Guia de Abertura de bases foi preparado para ampliar e contribuir com a política de transparência do Estado de São Paulo. Nele, estão informações sobre os benefícios de uma política de abertura de dados, seus desafios, suas especificidades técnicas e uma série de recomendações, pautadas em padrões internacionais e experiências exitosas ao redor do mundo, para tornar a abertura de bases um processo inspirador e virtuoso.

O guia não tem a intenção de ser um recurso exaustivo, meramente técnico ou que se esgota em si mesmo. Ele foi feito para ser abrigado na Web e usa referências da Web. Muitas das informações contidas nas próximas seções podem, e devem, ser extrapoladas navegando pelos links de referências e leituras complementares. As informações reunidas aqui procuram levar em consideração, da melhor maneira possível, o complexo ecossistema da Administração Pública, com seus desafios e diferenças, seja no tamanho das equipes quanto nas condições de infraestrutura dos órgãos públicos.

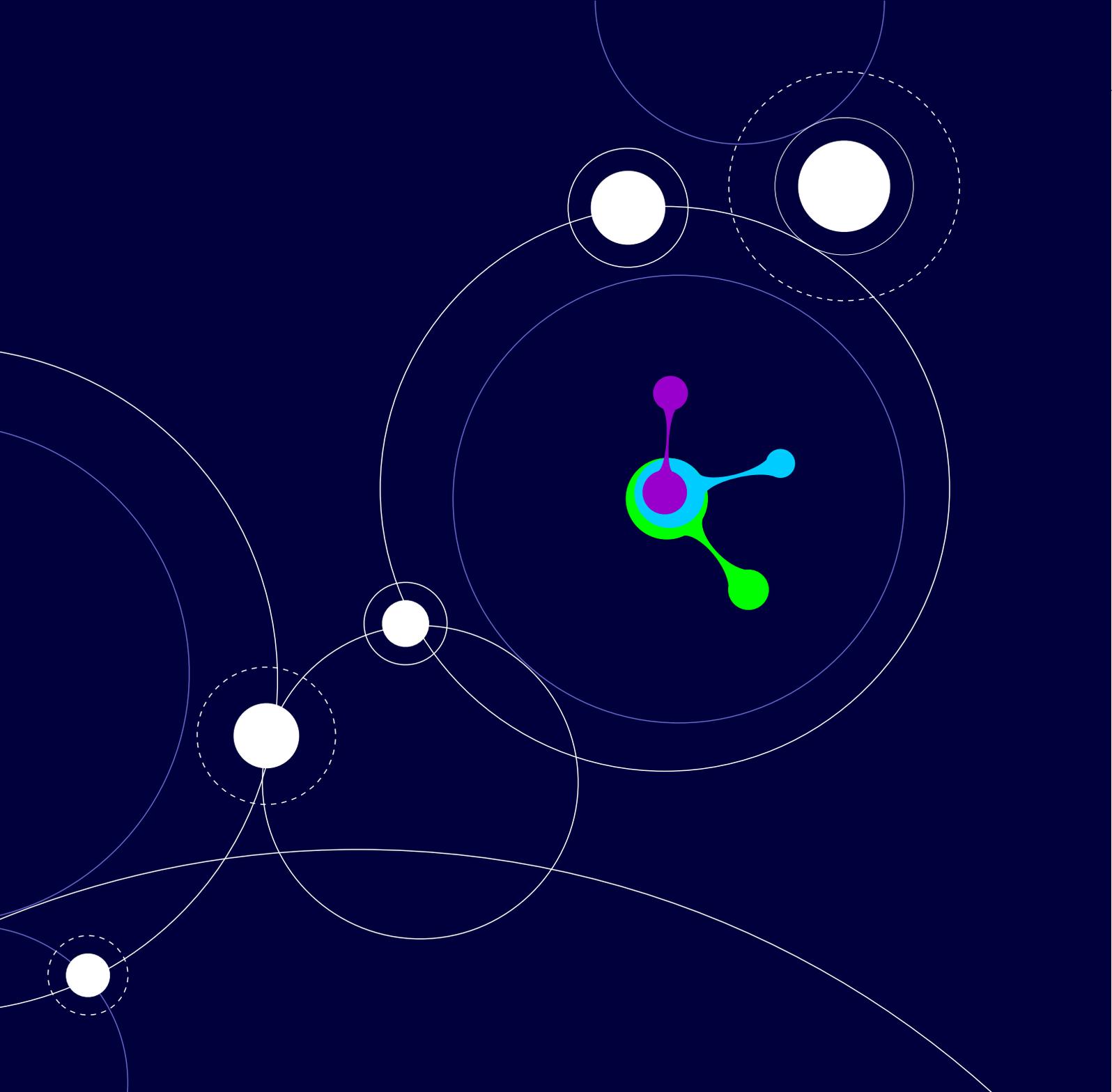
Trata-se de um guia para técnicos, servidores, gestores e executivos informarem-se sobre um movimento que vem tomando conta das democracias em todo o planeta e que ganha tração cada vez mais intensa no Brasil. A abertura de bases se apresenta como um caminho sem volta dentro do contexto de governos mais transparentes e sociedades mais justas. É apenas uma pequena parte dentro desse movimento que também abriga recursos “clássicos” de transparência, como a Lei de Acesso à Informação, e também áreas relativamente novas, como a participação e o controle social.

Uma política de abertura de bases precisa ter uma narrativa de curto, médio e longo prazo: onde estamos? Onde queremos chegar? Quais são os beneficiados diretos pela abertura de dados? Cidadãos? Jornalistas? Servidores públicos? Empresas? Cientistas? Infomediários? Como chegar até eles? Essas são algumas das perguntas que precisam de respostas claras e objetivas dentro da organização e do planejamento de qualquer esforço de abertura de dados.

Governos ao redor do mundo já estão colhendo bons frutos por terem seguido a via da abertura de bases. Um dos exemplos mais emblemáticos é o do governo britânico. Lá, foi descoberto, por causa de dados abertos, que diversos setores de informática da Administração Pública estavam comprando o mesmo serviço de consultoria de uma mesma empresa. Ao analisar os dados, a Administração percebeu que poderia economizar seis milhões de libras diminuindo o número de horas de consultoria contratada. A cifra é o equivalente a todo recurso necessário para custear o programa de transparência

do governo britânico. Isso sem contar o número de empresas e serviços que surgiram a partir da abertura de dados, possibilitando a formação de novos negócios. Alguns desses casos são contemplados neste guia.

Exemplos como esses, que representam o que há de melhor no esforço da abertura de dados, estão aguardando para serem descobertos em São Paulo e no Brasil.



O QUE SÃO “DADOS ABERTOS”?

Capítulo 2

“Dados abertos” é um termo que ganhou popularidade no movimento de transparência e governo aberto em todo mundo, mas nem sempre é tratado com clareza. A abertura de dados segue o mesmo princípio do Governo Aberto: tratar o acesso à informação pública como regra, não como exceção. No âmbito deste guia, os “dados” que aqui se referem são as informações produzidas por todos os órgãos públicos no estado de São Paulo, fruto da atividade administrativa da gestão governamental: seus contratos, suas funções, seus projetos, suas políticas e suas parcerias com outros setores. Em suma, todos os dados que estão sob a guarda do governo estadual ou de entes ligados a ele.

A abertura desses dados quer dizer que essas informações governamentais podem ser utilizadas livremente, reutilizadas e redistribuídas por qualquer pessoa, sem qualquer restrição — a exceção seria, no máximo, a indicação da fonte da base de dados e a redistribuição dessas informações sob as mesmas condições, ou licenças, em que elas foram adquiridas originalmente.

Para que um conjunto de dados possa ser considerado “aberto”, ele precisa reunir, no mínimo, as três características abaixo:

- **Disponibilidade e acesso:** os dados precisam estar disponíveis de forma completa e de uma forma que não gere custos exorbitantes para a parte interessada em copiá-los. O cenário mais favorável é disponibilizar os dados para que sejam baixados por meio da Internet. Os dados também precisam estar disponíveis num formato conveniente e modificável.
- **Reutilização e redistribuição:** os dados devem ser disponibilizados a partir de termos de utilização que permitam o reuso e a redistribuição, inclusive a mistura desses dados com outras bases.
- **Participação universal:** qualquer pessoa deve poder usar, reutilizar e redistribuir os dados. Não deve haver discriminação contra campos de atuação, indivíduos ou grupos. Por exemplo, restrições “não comerciais” que impediriam o uso “comercial” dos dados, ou restrições de uso para um fim específico (por exemplo, só para pesquisa pessoal), não são permitidas.

Essas três características são resumidas em três “leis” sugeridas pelo ativista dos dados abertos David Eaves:

- 1. Se o dado não pode ser encontrado ou indexado na Web, ele não existe;*
- 2. Se o dado não está disponível num formato aberto e legível por máquina, ele não pode ser reutilizado;*
- 3. Se dispositivos legais não permitem que ele seja compartilhado, ele não é útil.*

Existem muitos motivos para se ter tanta clareza sobre o significado de “dados abertos”. Por se tratarem de palavras com significado amplo, é preciso demarcar exatamente quais características são consideradas ideais no escopo deste guia, de modo que as informações publicadas pelo governo possam ser utilizadas por todos os setores da sociedade de forma compatível.

A abertura de bases com as características descritas acima é relevante pois dá condições para a *interoperabilidade*, ou seja, o trabalho em conjunto de diferentes bases de dados, por diferentes atores da sociedade. A construção de sistemas e soluções cada vez melhores, sejam aquelas desenvolvidas na esfera governamental, na privada, na acadêmica ou na sociedade civil, dependem da interoperabilidade das bases de dados.

A clareza sobre a definição de “dados abertos” também garante que duas ou mais bases de dados vindas de fontes diferentes possam ser combinadas sem grandes empecilhos técnicos. Evita, dentre outras coisas, que o governo seja um grande armazém de bases de dados “fechadas”, ou seja, que não servem senão para a consulta humana, inúteis para aplicação em sistemas maiores e complexos, capazes de proporcionar soluções, visualizações, serviços ou valor para qualquer cidadão ou grupo da sociedade.

Qualquer dado pode ser aberto?

Não. Todo dado que é público deve ser aberto, mas nem todo dado é público. A legislação brasileira trata como exceção à abertura dados particulares, capazes de identificar indivíduos, ferir seu direito à privacidade ou sua honra, dados considerados sigilosos ou dados que possam comprometer a segurança nacional.

The background is a dark blue gradient. It features several overlapping circles of varying sizes, some with solid white centers and others with dashed white outlines. In the center-right, there is a colorful, abstract shape resembling a molecule or a data node, composed of a central purple circle, a cyan circle, and a green circle, with thin lines connecting them to other colored dots.

OS BENEFÍCIOS DA ABERTURA DE DADOS

Capítulo 3

A abertura de bases de dados governamentais pode trazer uma série de benefícios para diferentes setores da sociedade criando ciclos virtuosos. Um dos primeiros beneficiários é o próprio órgão público que decide abrir suas bases. A simples estruturação e execução de mecanismos organizacionais que permitem a publicação e abertura sistematizada dessas bases pode contribuir para um aumento significativo na qualidade das equipes, dos serviços prestados e no conhecimento sobre gargalos e obstáculos internos. Não é possível administrar, ou conhecer, aquilo que não se mede. Abrir os dados governamentais possibilita conhecer e medir essas atividades.

A abertura também pode contribuir para o avanço da ciência. A abertura de dados governamentais permite que pesquisadores, independentes ou associados a uma instituição, tenham acesso a uma camada valiosa de informações sobre a formulação e execução de políticas públicas, sobre a forma como os recursos da sociedade são direcionados, sobre o impacto dos programas na melhoria da qualidade de vida das pessoas. Pesquisas melhores fundamentadas contribuem para elevar o nível de conhecimento que determinada sociedade tem sobre si mesma, e até sobre outras sociedades, em pesquisas internacionais, e dá condições para que seus atores tomem decisões cada vez mais justificadas.

É benéfica também para o setor privado. A abertura responsável e consistente de dados governamentais permite que qualquer empreendedor, ou grupo de empreendedores, use sua criatividade para construir ferramentas, soluções e avanços tecnológicos que muitas vezes escapam aos olhos das equipes de especialistas que trabalham com esses dados dentro da estrutura governamental. Isso quer dizer que a abertura de dados pode contribuir para a geração de empregos e riqueza, criando uma interdependência saudável entre o governo e a iniciativa privada.

Organizações da sociedade civil também podem tirar vantagem dos dados abertos. A abertura de bases pode elevar a qualidade dos serviços que organizações de cunho não governamental prestam, muitas vezes complementares aos oferecidos pelo estado. O governo possui uma infinidade de informações sobre públicos e serviços de interesse dessas organizações. O acesso livre e irrestrito a esses dados permite que essas organizações coloquem à prova seus serviços e possam medir seus resultados, aumentando cada vez mais a qualidade de suas atividades.

Finalmente, qualquer pessoa pode se beneficiar da abertura de bases ao ter acesso facilitado e livre aos dados produzidos na seara governamental. Quem

ganha com isso é a sociedade inteira. Esse benefício no nível individual não vem apenas do exercício de um direito humano fundamental, o do acesso à informação, mas também do direito à liberdade. A abertura de dados, que faz parte dos governos abertos e transparentes, é um dos pilares que sustentam sociedades que desejam ser mais livres e justas. Ela contribui para fortalecer os processos democráticos, abrindo caminhos para uma maior participação dos cidadãos na administração pública e fomenta o controle social das atividades governamentais. O acesso aos dados do governo permite que qualquer pessoa possa acompanhar a aplicação de políticas públicas e medir sua eficácia. A disponibilização de ferramentas que permitem que o cidadão conheça as ações adotadas pelo governo fortalece a legitimidade do Estado, enriquecendo seu papel como um dos principais atores na busca do bem-estar social.

Não faltam iniciativas exitosas ao redor do mundo, inclusive no Brasil, que servem de exemplo inspirador para a abertura de dados. Conheça algumas delas a seguir.

TRANSPORTE PÚBLICO (INICIATIVA PRIVADA)

- Citymapper

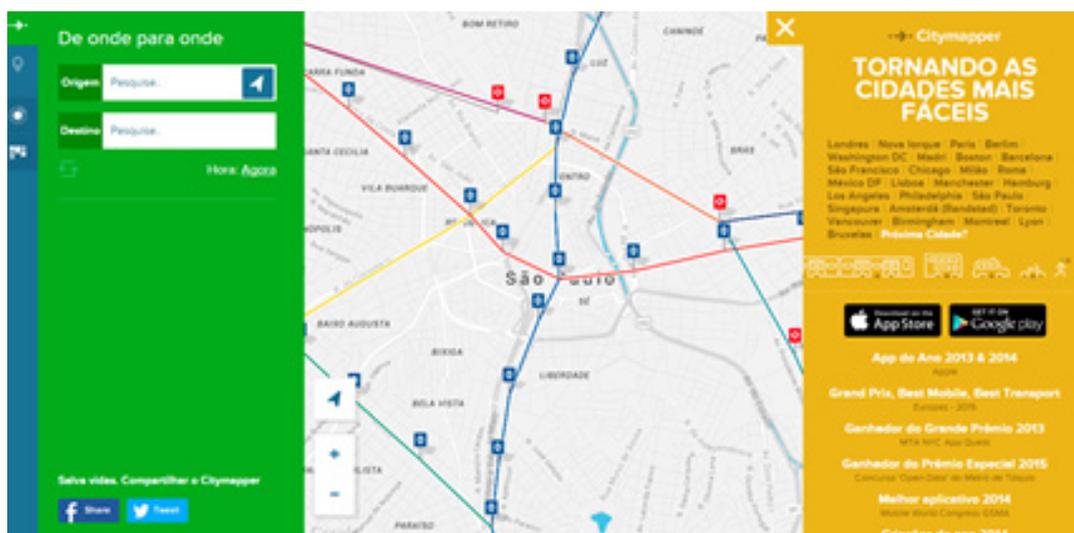


O Citymapper é um dos bons exemplos ao redor do mundo que mostram o potencial dos dados abertos para fomentar a geração de novos negócios. Aplicativo gratuito para *smartphones* criado por um ex-funcionário do Google em 2011 em Londres, com o objetivo de melhorar a experiência de milhões de pessoas na cidade, todos os dias, que dependem do transporte público.

Os desenvolvedores do Citymapper tiraram vantagem dos dados sobre ônibus, trens e metrô, publicados em tempo real pela autoridade de trânsito londrina, para formular um algoritmo que mostra sempre o melhor tempo de viagem entre um ponto e outro; quando será a próxima viagem e quantas calorias se gastou caminhando. As informações são atualizadas em tempo real e mostram

as condições de trânsito, clima e problemas técnicos nos veículos.

A receita deu certo e a empresa expandiu seus serviços para outras cidades. De 2011 a 2015 chegou a sete países e 13 cidades: Londres, Nova York, Paris, Berlim, Washington DC, Madri, Boston, Barcelona, São Francisco, Chicago, Milão, Roma e Cidade do México. O aplicativo pode ser conectado em qualquer cidade que ofereça dados sobre serviços de transporte público em tempo real.



LEGISLATIVO & EXECUTIVO, PARTICIPAÇÃO (TERCEIRO SETOR)

- Fundação Sunlight

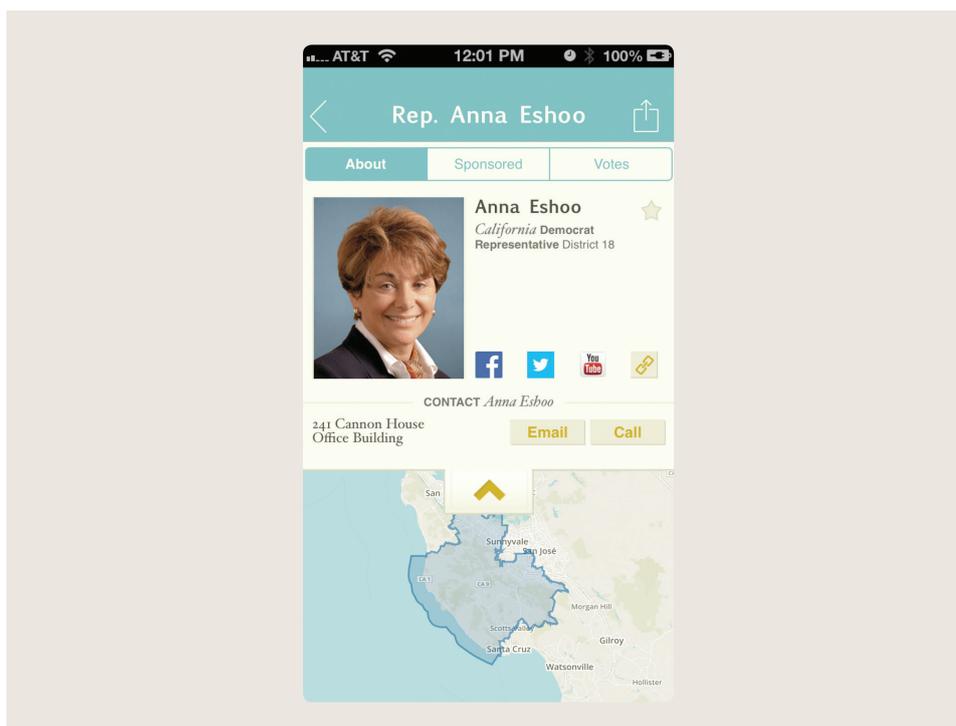


A Fundação Sunlight é um dos exemplos de sucesso que mostram como o terceiro setor pode trabalhar junto do governo fomentando a participação popular e o controle social por meio da tecnologia. Criada em abril de 2006 nos Estados Unidos, a fundação tem o objetivo de aumentar a transparência e prestação de contas no legislativo e executivo americano. Seu foco inicial foi monitorar o papel do dinheiro na política americana e propor mudanças que visam a governos abertos.

Hoje, a fundação também contribui em fóruns internacionais que discutem a transparência governamental. A Sunlight possui três frentes principais: uma equipe de política que coordena ações dentro e fora do Congresso americano para propor mudanças na legislação; uma equipe de jornalistas especialistas em dados que produzem e publicam reportagens sobre o estado da transparência americana; um laboratório de tecnologia que desenvolve ferramentas a partir de dados abertos que visam a aumentar a participação de cidadãos no governo e contribuir para que outros desenvolvedores possam construir aplicativos com informações públicas.

A Fundação Sunlight realiza a organização sistemática de várias bases de dados do governo americano. A garantia de que as bases serão publicadas de forma consistente e periódica num formato aberto permitiu que a fundação desenvolvesse uma série de ferramentas que aproximam o cidadão da administração pública:

- Congress



Aplicativo para *smartphones* que exibe informações sobre deputados e senadores, permitindo que o cidadão entre em contato com eles e acompanhe suas atividades. É possível visualizar as últimas leis aprovadas, exibir a lista de atividades no plenário, navegar pelas votações e se atualizar sobre a agenda das comissões parlamentares e suas audiências. O aplicativo foi desenvolvido a partir de bases de dados abertas pelo governo americano e é distribuído gratuitamente. Seu código fonte é aberto, qualquer pessoa pode baixá-lo.

- Influence Explorer

INFLUENCE EXPLORER

Sítio da Web que rastreia as doações políticas em nível estadual e federal, permitindo que qualquer pessoa possa acompanhar o grau de influência por legislador, empresa ou indivíduo. A ferramenta oferece uma visão geral sobre financiamentos de campanha, *lobbying*, verbas parlamentares, irregularidades em contratações e gastos públicos federais.

- Capitol Words

capitolwords

a project of the Sunlight Foundation

Ferramenta que permite explorar o conteúdo dos discursos de todos os senadores e deputados do Congresso americano, a partir de 1996. A procura pode ser feita por estado, data ou legislador. O serviço permite comparar termos e frases e exibe os resultados com gráficos e rankings dos políticos e seus partidos. O Capitol Words também serve como uma grande base de dados para que outros desenvolvedores possam criar aplicativos que dependam dele.

Confira outras ferramentas (sítio da Web em inglês) desenvolvidas pela Fundação Sunlight.

Dentre as empresas e organizações que utilizam os serviços da Fundação Sunlight estão a Fundação Wikimedia, que administra, por exemplo, a Wikipédia; o Greenpeace; o sítio If This Then That, que permite criar, dentre outras coisas, avisos instantâneos no celular sobre andamento de leis no Congresso americano; e a equipe de campanha do presidente Barack Obama.

EDUCAÇÃO & PESQUISA (TERCEIRO SETOR)

- QEdu



O QEdu é um dos bons exemplos brasileiros que demonstra os benefícios do trabalho em conjunto do terceiro setor, da academia e do governo para melhor entender a educação básica no Brasil e fornecer subsídios para políticas públicas. Trata-se de um portal gratuito desenvolvido em parceria entre a Fundação Lemann e a Meritt que permite a qualquer pessoa obter informações sobre a qualidade do aprendizado nas escolas brasileiras, com dados sobre escolas públicas e particulares. As visualizações são geradas a partir de bases do governo brasileiro, como dados do Exame Nacional do Ensino Médio (Enem), da Prova Brasil, do Censo Escolar e de indicadores especiais do Instituto Nacional de Estudos e Pesquisas (INEP).

A ferramenta mostra como foi o aprendizado de alunos dos 5º e 9º anos em matemática e português, perfil dos alunos, professores e diretores das escolas que realizaram a Prova Brasil, matrículas para cada etapa escolar, taxas de aprovação, abandono e reprovação, nível sócio econômico, infraestrutura escolar, distorções de idade em relação às séries e muito mais. É possível fazer buscas e comparações a partir de escolas, cidades ou estados. O portal também permite que qualquer pessoa possa filtrar e baixar os dados que precisar em formato aberto.



SAÚDE & GASTOS PÚBLICOS (GOVERNO, CIDADÃOS)

- Prescribing Analytics



O Prescribing Analytics mostra como agentes do governo e profissionais independentes podem trabalhar juntos para enxergar formas de economizar dinheiro público. Trata-se de uma ferramenta de análise de dados sobre os gastos do governo britânico com um grupo específico de remédios, a estatina. Esse medicamento ajuda no combate às altas taxas de colesterol em pacientes com problemas de saúde.

O serviço foi criado por um grupo de desenvolvedores independentes e médicos que trabalham no setor público britânico. O sistema de saúde britânico recebeu a missão de economizar 20 bilhões de libras até 2015, sendo que metade disso é gasto com remédios. Uma das substâncias mais prescritas pelos médicos do sistema público de saúde do Reino Unido é justamente a estatina. A ferramenta acessa a base de dados aberta sobre receitas médicas no país e analisa quais tipos de estatina foram prescritas pelos médicos no sistema público: as variantes mais caras ou as “genéricas”, normalmente mais baratas.

Os dados mostram que o governo britânico poderia ter economizado, entre os anos de 2011 e 2012, 27 milhões de libras por mês se os médicos tivessem receitado variantes genéricas da estatina que possuem a mesma eficácia das mais caras.

The background is a dark blue gradient. It features several overlapping circles of varying sizes and colors (white, light blue, purple). A central graphic element consists of a purple circle connected to a cyan circle, which is further connected to a green circle, forming a molecular-like structure. The text 'LEI DE ACESSO À INFORMAÇÃO' is written in a bold, white, sans-serif font, centered horizontally in the lower half of the page.

LEI DE ACESSO À INFORMAÇÃO

Capítulo 4

A abertura de bases governamentais encontra grande motivação na legislação brasileira e internacional. O Brasil é signatário de diversos acordos internacionais que tratam o acesso à informação como um direito de qualquer cidadão, incluindo a Declaração Universal dos Direitos Humanos, adotada pela Assembleia Geral da Organização das Nações Unidas (ONU) em 1948. O direito à informação (abaixo, em negrito) está também previsto no artigo 19 da declaração:

*“Todo ser humano tem direito à liberdade de opinião e de expressão; esse direito inclui a liberdade de ter opiniões sem sofrer interferência e de **procurar, receber e divulgar informações e ideias por quaisquer meios, sem limite de fronteiras**”*

Entende-se que qualquer informação que está sob a guarda do Estado é pública, respeitando algumas restrições. Esse é o espírito da maior parte das Leis de Acesso à Informação ao redor do mundo e, no Brasil, não é diferente. As bases de dados governamentais são consideradas abertas, via de regra. Os casos em que a abertura não é considerada apropriada são exceções. É por isso que é recomendado familiarizar-se com a Lei de Acesso à Informação do Brasil e sua regulamentação no Estado de São Paulo para promover uma abertura de bases governamentais consistente e de acordo com os ditames legais.

O ACESSO À INFORMAÇÃO EM SÃO PAULO

O Governo de São Paulo vem se preparando desde meados da década de 1980 para organizar as informações sob a guarda do Estado, com a instituição do Sistema de Arquivos do Estado de São Paulo e de uma série de leis e decretos ao longo das décadas de 1990 e 2000 que cuidam do ecossistema de responsabilidades e gerenciamento de dados em São Paulo¹. Um dos passos mais importantes foi dado em 2010, com um decreto que tratou especificamente sobre a forma como o Estado deveria publicar e organizar dados considerados abertos na rede mundial de computadores, a *Internet*.

¹Leis estaduais nº 10.177, de 30 de dezembro de 1998, que regula o processo administrativo, nº 10.294, de 20 de abril de 1999, que dispõe sobre proteção e defesa do usuário de serviços públicos; decretos estaduais nº 22.789, de 19 de outubro de 1984, que institui o Sistema de Arquivos do Estado de São Paulo - SAESP, nº 44.074, de 1º de julho de 1999, que regulamenta a composição e estabelece a competência das Ouvidorias, nº 54.276, de 27 de abril de 2009, que reorganiza a Unidade do Arquivo Público do Estado, da Casa Civil, nº 55.479, de 25 de fevereiro de 2010, que institui na Casa Civil o Comitê Gestor do Sistema Informatizado Unificado de Gestão Arquivística de Documentos e Informações - SPdoc, alterado pelo de nº 56.260 de 6 de outubro de 2010.

O decreto nº 55.559, de 12 de março de 2010, cria o portal de dados abertos de São Paulo, intitulado “Governo Aberto SP”. Nesse portal estão concentrados esforços para a publicação de bases de dados não sigilosas e de acesso irrestrito do governo paulista. O decreto prevê a publicação dessas bases em “formato aberto” e faz as devidas considerações em relação aos dados sigilosos que não devem ser publicados. De acordo com o decreto, a seleção, publicação e atualização dessas bases é de responsabilidade de cada órgão público, sob a coordenação da Secretaria de Gestão Pública, responsável pela manutenção do portal. Outro decreto que trata do acesso à informação no Estado de São Paulo é o de nº 58.052, de 16 de maio de 2012, que regulamenta a lei federal nº 12.527 de 18 de novembro de 2011, a própria Lei de Acesso à Informação em vigor no Brasil.

TRANSPARÊNCIA ATIVA

O decreto estadual nº 58.052, de 16 de maio de 2012, que regulamenta a Lei de Acesso à Informação federal no estado de São Paulo, incorpora em seu Artigo 23 os princípios da transparência ativa, que determina como dever dos órgãos e entidades da Administração Pública Estadual promover, independentemente de requerimentos, a divulgação em local de fácil acesso de documentos, dados e informações de interesse coletivo ou geral produzidas pelo órgão ou sob sua guarda. Essas informações devem ser publicadas na rede mundial de computadores obrigatoriamente.

Artigo 23 - É dever dos órgãos e entidades da Administração Pública Estadual promover, independentemente de requerimentos, a divulgação em local de fácil acesso, no âmbito de suas competências, de documentos, dados e informações de interesse coletivo ou geral por eles produzidas ou custodiadas.

§ 2º - Para o cumprimento do disposto no “caput” deste artigo, os órgãos e entidades estaduais deverão utilizar todos os meios e instrumentos legítimos de que dispuserem, sendo obrigatória a divulgação em sítios oficiais da rede mundial de computadores (internet).

A publicação de informações na Internet, segundo o decreto estadual, deve obedecer alguns critérios, como oferecer uma ferramenta de pesquisa de conteúdo que permita acesso à informação de forma objetiva, transparente, clara e em linguagem de fácil compreensão. Os órgãos também devem permitir

a gravação de relatórios em diversos formatos eletrônicos, inclusive abertos e não proprietários, tais como planilhas de texto, de modo a facilitar a análise das informações. É importante destacar que apesar de o texto do decreto não estabelecer com detalhes o que é um formato eletrônico “aberto e não proprietário”, as características desses formatos já estão bem estabelecidas. O Guia de Abertura de Bases possui uma seção dedicada especialmente a descrever uma série de formatos eletrônicos abertos, a maior parte não proprietário.

Ainda sobre a publicação ativa de informações na Internet, o decreto determina que os órgãos devem disponibilizar as informações de modo a possibilitar o acesso automatizado por sistemas externos em formatos abertos, estruturados e legíveis por máquina. Um arquivo PDF, por exemplo, não é **suficiente** para atender às determinações do decreto. O PDF não é considerado uma tecnologia aberta nos termos deste guia, estruturada e muito menos legível por máquina. Sua função é gerar documentos que serão impressos exatamente como são exibidos na tela do computador. Não se trata de uma tecnologia que tem como objetivo facilitar o acesso automatizado de sistemas externos. A seção de formatos do Guia de Abertura de Bases apresenta uma série de tecnologias que se encaixam melhor nas determinações do decreto estadual.

O decreto também determina que o órgão deve divulgar, em detalhes, os formatos utilizados para estruturação da informação publicada, além de garantir sua autenticidade e integridade e mantê-la atualizada. Deve também adotar todas as medidas necessárias para garantir a acessibilidade de conteúdo para pessoas com deficiência.

QUAIS SÃO AS EXCEÇÕES PARA ABERTURA DE DADOS?

O acesso livre e irrestrito é considerado regra no âmbito da Lei de Acesso à Informação. Existem algumas exceções e, todas elas, estão descritas no texto da Lei e no decreto estadual. São elas:

Segurança da sociedade e do Estado

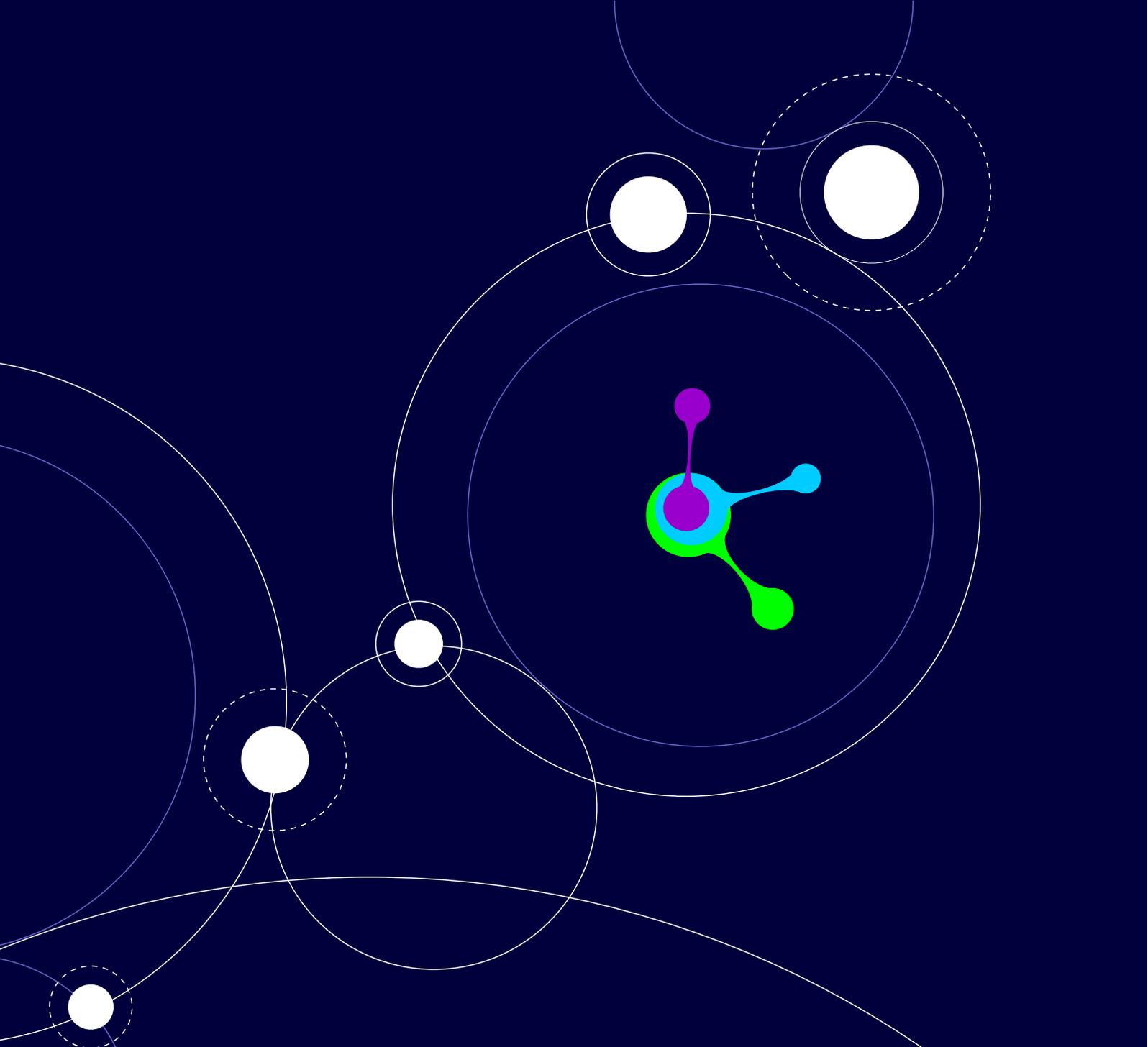
Informações cujo sigilo são considerados imprescindíveis para a segurança da sociedade e do Estado são consideradas exceções, além de pesquisas científicas e tecnológicas com este fim e segredos de Estado. São dados que, caso

divulgados, possam:

- Colocar em risco a defesa e a soberania nacional ou a integridade do território nacional;
- Prejudicar as relações internacionais do Brasil ou informações que tenham sido fornecidas ao governo por outros países em caráter sigiloso;
- Pôr em risco a vida, a segurança ou a saúde da população;
- Oferecer alto risco à estabilidade financeira, econômica ou monetária do Brasil;
- Prejudicar operações das Forças Armadas;
- Pôr em risco a segurança de autoridades estrangeiras no País;
- Comprometer atividades de investigação ou fiscalização em andamento, relacionadas com a prevenção ou repressão de infrações.

Informações pessoais

São informações que estão relacionadas à pessoa natural identificada ou identificável, relativas à intimidade, vida privada, honra e imagem das pessoas, bem como às liberdades e garantias individuais. Números de telefone, carteira de identidade ou motorista, CPF, extratos bancários, entre outros, são informações que podem identificar pessoas e, portanto, são sigilosas.



PLANO DE ABERTURA DE DADOS

Capítulo 5

A abertura de dados em qualquer órgão público é um processo que deve ser planejado com cuidado e com o envolvimento de diversos atores: guardiões das bases, profissionais de Tecnologia da Informação (TI), gestores públicos e técnicos das áreas de interesse. Via de regra, todas as bases de dados que o órgão já possui podem ser publicadas. Contudo, é preciso garantir que as bases selecionadas para publicação não se encaixam nos cenários de exceção previstos na Lei de Acesso à Informação. Se uma parte dessas bases apresentar informações sigilosas, esses dados devem ser ocultados antes de publicados.

Essa parte do guia sugere a criação de um “Plano de Abertura de Dados”. Trata-se de uma estratégia para iniciar processos de abertura de dados ou como forma de avaliação sobre a situação de abertura em determinado órgão. Contudo, cada instituição governamental tem liberdade para decidir como vai abrir suas bases, em conformidade com a legislação estadual e federal sobre o acesso à informação.

AS CINCO ESTRELAS DOS DADOS ABERTOS: ★★★★★

Em 2010, o cientista britânico Tim Berners-Lee, inventor da Web, formulou um sistema de estrelas para encorajar a sociedade, especialmente guardiões de dados governamentais, a abrirem seus dados. O sistema ajuda a diagnosticar o nível de abertura de dados dos órgãos públicos e fornece degraus alcançáveis para se chegar a níveis cada vez mais refinados de dados abertos.



O primeiro estágio do sistema de estrelas requer apenas que seus dados estejam disponíveis na Web em qualquer formato (seja um documento PDF, arquivo do processador de textos Word ou qualquer outro, proprietário ou não, aberto ou não). Além de estarem disponíveis na Web os dados devem ser oferecidos sob uma licença aberta. A licença aberta dá a autorização para que esses dados sejam utilizados por qualquer pessoa, sem restrições, e para qualquer fim, inclusive comerciais. Existem diversas licenças abertas que descrevem a utilização de dados. Este guia não faz uma análise exaustiva das licenças abertas que existem, mas aponta caminhos para que cada órgão possa formular licenças abertas e compatíveis com os princípios internacionais de abertura de dados.

Para atender as demandas da primeira estrela, basta que um arquivo seja publicado em um sítio da Web, em qualquer formato, e, junto dele, esteja especificada a licença aberta de utilização. Se o seu órgão já publica dados na Web em qualquer formato, isso quer dizer que basta acrescentar a licença aberta para receber a primeira estrela.

QUAIS SÃO OS BENEFÍCIOS DE

Tanto para o usuário dos dados quanto para quem publica, mesmo a primeira estrela traz uma série de benefícios. Enquanto usuário, é possível visualizar os dados, imprimi-los, baixá-los para um computador particular ou um pendrive, digitá-los em qualquer outro sistema, realizar análises e compartilhá-los com quem desejar, da forma que desejar.

Quem publica pode fazê-lo de forma simples, pois não precisa se ater aos formatos utilizados. Além disso, não será necessário explicar, repetidas vezes, para todos os interessados nos dados, que eles podem ser utilizados para quaisquer fins, uma vez que eles estão sendo distribuídos a partir de uma licença aberta.

Publicar dados na Web em qualquer formato sob uma licença aberta é um grande passo para a abertura de dados. Contudo, os usuários de dados podem encontrar dificuldades para extrair os dados desses documentos sem precisar digitá-los manualmente em outros sistemas.



O segundo estágio do sistema de estrelas requer que as informações sejam publicadas na Web sob uma licença aberta, mas requer que essas bases fiquem disponíveis num formato estruturado, que permita a manipulação facilitada de suas linhas e colunas.

Ou seja, enquanto a primeira estrela permite que imagens escaneadas de tabelas e arquivos PDF de relatórios sejam publicados, para se obter a segunda estrela é necessário que o arquivo seja disponibilizado de uma forma que permita a manipulação em aplicativos de análise e estruturação de dados, sem a necessidade de digitar essas informações manualmente.

Arquivos do Excel (.xlsx), por exemplo, são arquivos estruturados.

A segunda estrela não faz qualquer exigência em relação ao aplicativo que gerou o arquivo ou ao formato do documento, desde que ele seja estruturado.

QUAIS OS BENEFÍCIOS DE

Além de todos os benefícios da primeira estrela, as duas estrelas permitem que os usuários usem aplicativos proprietários, como o Excel, para agregar, fazer cálculos, visualizações e outras operações com os dados disponibilizados. Além disso, é possível exportá-los para outro formato estruturado com facilidade.

Quem publica ainda pode fazê-lo de forma simples, apenas atentando-se ao fato de que o arquivo publicado deve estar num formato estruturado, mas respeitando que ele deve estar na Web e distribuído sob uma licença aberta.

Pode parecer pouco, mas as duas estrelas representam um grande avanço na abertura de dados, uma vez que as informações estão disponíveis na Web de forma estruturada e sob uma licença aberta. Contudo, os dados ainda estão trancados dentro de um documento. Para extraí-los de lá, dependendo do formato utilizado, será preciso usar um programa proprietário (que pode custar dinheiro), o que é um obstáculo para diversas pessoas.



O terceiro estágio do sistema de estrelas é semelhante ao segundo, mas inclui uma exigência: a de que os arquivos estruturados disponibilizados na Web sob licença aberta sejam gerados em um formato não proprietário. Em vez de um documento no formato .xlsx, gerado pelo aplicativo proprietário Excel, por exemplo, a opção seria usar o CSV, um formato estruturado que não depende de programas proprietários para ser manipulado.

QUAIS OS BENEFÍCIOS DE

Os órgãos que conseguirem chegar às três estrelas da abertura de dados vão oferecer aos cidadãos todos os benefícios das duas estrelas, além de permitir que qualquer um possa baixar e manipular os dados da forma mais conveniente, sem exigir a utilização de um aplicativo específico.

Quem publica deve prestar atenção aos conversores e plug-ins necessários para exportar os dados a partir de um formato proprietário para um formato aberto. Todo o processo de publicação ainda permanece razoavelmente simples, uma vez que trata-se apenas de disponibilizar documentos em formato aberto na Web e distribuídos com uma licença aberta.

Nas três estrelas, qualquer pessoa pode utilizar os dados de forma fácil, mas essas informações podem ser disponibilizadas em formatos que permitem ainda mais interação entre sistemas e facilitam o seu compartilhamento.



Para entender melhor o funcionamento da quarta e da quinta estrela, é importante ler o Guia de Web Semântica. Enquanto as três primeiras estrelas dão diretrizes para que os dados possam ser publicados na Web na forma de documentos de formato aberto e com licença aberta, a quarta e a quinta estrela apresentam o conceito de dados conectados.

Para conseguir as quatro estrelas é necessário fazer tudo que as estrelas anteriores aconselham, só que além de usar um documento em formato aberto, os dados também são publicados na própria página da Web usando URIs para descrever cada um dos dados, de modo que qualquer pessoa possa identificar os elementos de forma padronizada no documento publicado.

QUAIS SÃO OS BENEFÍCIOS DE

Dados publicados seguindo as diretrizes das quatro estrelas podem ser conectados a partir de sistemas disponíveis em qualquer outra parte da

Web. Partes desses dados podem ser reutilizadas e combinadas com outros dados.

Quem publica passa a ter um controle fino de cada célula da base de dados e pode otimizar o acesso ao banco, equilibrando a carga de acessos, realizando cópias estratégicas e muito mais. Outros órgãos que publicam dados podem conectar as bases deles às suas usando o mesmo esquema de URIs.



O último estágio do sistema de estrelas requer que seus dados publicados no esquema de URIs estejam conectados a outras bases de dados publicadas sob as mesmas condições. Mais informações sobre como os dados conectados funcionam pode ser encontradas no Guia de Web Semântica.

QUAIS SÃO OS BENEFÍCIOS DE

As diretrizes que levam às cinco estrelas permitem que qualquer pessoa possa descobrir mais dados à medida que navega por eles. Também é possível aprender sobre o esquema de publicação dos dados diretamente, apenas estudando sua estrutura.

Quem publica permite que os dados sejam descobertos, aumentando o valor agregado dos dados. Esse benefício vale, inclusive, para o próprio órgão que publica, uma vez que os recursos estarão disponíveis para qualquer um.

As cinco estrelas dos dados abertos servem como uma diretriz sóbria e objetiva em relação aos objetivos de cada órgão público para abertura de suas bases. Em suma:



Publicar bases na Web (em qualquer formato) com licença aberta



Publicar bases em formato estruturado com licença aberta

(ex: arquivo Excel, em vez de imagem escaneada)



Usar formatos não proprietários e uma licença aberta

(ex: arquivo CSV em vez de Excel)



Usar URIs para descrever coisas, para que qualquer um possa identificá-las



Conectar seus dados a outras bases para dar contexto

Tendo em vista as recomendações do sistema desenvolvido por Tim Berners-Lee, o “Plano de Abertura de Dados” poderá ser desenvolvido levando-se em consideração...

...o escopo

Ao se preparar para abrir suas bases, o órgão deve escolher quais bases, ou suas partes, deverão ser abertas. Essa abertura, se guiada pelo esquema das cinco estrelas dos dados abertos, deve levar em consideração os desafios técnicos e humanos para se conseguir chegar a três estrelas, por exemplo, ou cinco.

É possível que algumas das bases já estejam prontas para serem publicadas na Internet em formato estruturado e aberto e com licença aberta (três estrelas), ou que outras existam apenas em formatos fechados e não estruturados, mas possam ser colocadas em sítio da Web com licença aberta (uma estrela).

O escopo vai ajudar a mapear quais bases poderão ser abertas e em qual regime de estrelas elas entrarão, de uma a cinco. O número de estrelas vai depender dos recursos técnicos e humanos disponíveis ao órgão.

...a priorização

Depois de definido o escopo de abertura de bases em relação ao número de estrelas que se deseja atingir com cada uma delas, é preciso definir qual será a

priorização para a publicação dessas bases.

- Quais bases poderão ser publicadas imediatamente? Com quantas estrelas?
- Quais bases precisarão passar por algum tratamento antes de serem publicadas?
- Quais bases serão publicadas no longo prazo?
- Qual será o mínimo de estrelas que o órgão deseja conseguir em todas as bases que forem publicadas?

Essas são algumas das perguntas a serem feitas pela equipe responsável pela abertura de bases em relação à priorização.

...as partes interessadas

A abertura de bases também deve levar em consideração quais áreas dos órgãos, sejam diretorias, coordenadorias ou subsecretarias etc, farão parte do esforço de abertura de bases, tendo em vista o número de estrelas que se deseja conquistar. Cada área deverá ser responsável pelo fornecimento dos dados que produz em formatos que estejam de acordo com o número de estrelas definido no escopo do plano de abertura.

EQUIPES DE ABERTURA DE DADOS

Para abrir as bases de dados do seu órgão, será preciso pensar em quais atores estarão envolvidos diretamente com o fluxo de trabalho de publicação dessas bases na Web, de modo que elas estejam sempre atualizadas e sigam um padrão coerente de divulgação. O número de pessoas e o perfil dos profissionais envolvidos vai depender da quantidade de estrelas que o seu órgão deseja alcançar para abrir as bases. Veja o perfil mínimo dos profissionais necessários em cada uma das estrelas:

- ★ Publicar bases na Web (em qualquer formato) com licença aberta

Equipe:

- Técnico do setor público, guardião ou responsável pela base que deverá ser aberta.
- Profissional de TI que ficará responsável por colocar os documentos em sítio da Web, junto com a licença aberta.

Fluxo de trabalho simplificado: Técnico do setor público envia base (em qualquer formato) para que profissional de TI, que disponibiliza a base no sítio da Web do órgão, com a licença aberta.

★ ★ Publicar bases em formato estruturado

Equipe:

- Técnico do setor público, guardião ou responsável pela base que deverá ser aberta, atentando-se para o fato de que a base deve estar num formato estruturado (ex: arquivo Excel, em vez de imagem escaneada).
- Profissional de TI que também vai se certificar de que a base está num formato estruturado e ficará responsável por colocar os documentos em sítio da Web, junto com a licença aberta.

Fluxo de trabalho simplificado: Técnico do setor público, guardião ou responsável pela base que deverá ser aberta, se necessário, converte a base de dados para um formato estruturado, recebendo auxílio do profissional de TI. Esse profissional envia o documento, junto com sua licença aberta, para sítio da Web do órgão público.

★ ★ ★ Usar formatos não proprietários

Equipe:

- Técnico do setor público, guardião ou responsável pela base que deverá ser aberta, atentando-se para o fato de que a base deve estar, no mínimo, num formato estruturado e não proprietário (ex: arquivo CSV em vez de Excel).

- Profissional de TI que também vai se certificar de que a base está num formato estruturado, caso necessário, converte a base para um formato aberto e ficará responsável por colocar os documentos em sítio da Web, junto com a licença aberta.

Fluxo de trabalho simplificado: Técnico do setor público, guardião ou responsável pela base que deverá ser aberta, se necessário, converte a base de dados para um formato estruturado aberto, recebendo auxílio do profissional de TI. Este profissional envia o documento, junto com sua licença aberta, para sítio da Web do órgão público.

★★★★★ Usar URIs para descrever coisas, para que qualquer um possa identificá-las

Equipe:

- Técnico do setor público, guardião ou responsável pela base que deverá ser aberta, atentando-se para o fato de que a base deve estar num formato estruturado e aberto.
- Profissional de TI que ajudará no preparo dessa base de dados para que ela possa se adequar aos padrões descritos no Guia de Web Semântica.
- Profissional de TI que ficará responsável pela manutenção dessas bases na Web, junto com a licença aberta.

Fluxo de trabalho simplificado: Técnico do setor público, guardião ou responsável pela base que deverá ser aberta recebe auxílio do profissional de TI para adequar a base de dados aos padrões descritos no Guia de Web Semântica. A base é então publicada (ou atualizada) em sítio da Web, com infraestrutura adequada para embarcar Dados Conectados.

★★★★★ Conectar seus dados a outras bases para dar contexto

A equipe e o fluxo de trabalho simplificado para as cinco estrelas são semelhantes aos de quatro estrelas.

PUBLICAÇÃO

Com os dados em mãos, num formato aberto e com uma licença que permita a sua reutilização de forma livre, chegará a hora de publicar as bases na Web. Esse é um momento importante do Plano de Abertura de Bases e deve ser planejado com calma e diligência. A publicação depende de quantas estrelas o órgão interessado deseja conquistar com a abertura dos dados. Idealmente, uma publicação que visa a conseguir cinco estrelas no longo prazo deve se preparar desde o início para que os ajustes ao longo do caminho sejam suaves e previsíveis.

O melhor caminho para conquistar a abertura e interoperabilidade dos dados, no que diz respeito à sua viabilização de acesso na Web, é por meio de repositórios estruturados e planejados. Isso não quer dizer, necessariamente, que o órgão interessado deverá adquirir sistemas complexos de armazenamento de dados. Uma página simples da Web com uma lista bem estruturada de documentos pode servir como um bom repositório de dados, ou um catálogo, desde que alguns cuidados sejam tomados. A complexidade do sistema dependerá da quantidade de bases abertas e dos recursos, técnicos e humanos, de cada órgão.

A forma como essas informações serão organizadas deve ser convencionalizada com antecedência, com a participação de todos os atores envolvidos. De preferência entre órgãos que publicam dados correlacionados e podem publicar bases que têm o potencial de “conversarem” umas com as outras. Isso é importante para que todos tenham entendimento comum do significado dos dados que serão compartilhados.

O objetivo dessa organização é que toda a sociedade que tem o acesso à informação seja capaz de interpretar os dados de maneira uniforme, utilizando sistemas e plataformas de trocas de dados. Essa padronização prévia entre as partes ganha forma quando se publica na Web os nomes e definições dos elementos usados em forma partilhável e referenciável, independentemente do grau de apoio que se obteve.

Esse planejamento abre caminho para a construção de bases de dados cinco estrelas, independente dos aplicativos utilizados para organizá-los e publicá-los. Qualquer que seja a estratégia adotada, é importante incluir no planejamento os conceitos de alguns padrões:

- URI: é um identificador de recursos que serve para identificar ou apontar para alguma coisa na Web.
 - Um URL é um URI que identifica um recurso e provê meios de agir sobre ele, obter e/ou representar este recurso, descrevendo seu mecanismo de acesso primário ou a localização na “rede”. Por exemplo, o URL <http://www.w3c.br/> é um URI que identifica um recurso (Sítio do W3c Brasil) e representa este recurso (o HTML da página por exemplo) e está disponível via HTTP de um hospedeiro de redes chamado <http://www.w3c.br>.
 - RDF/XML: O XML é um formato padrão do W3C para a criação de documentos com dados organizados de forma hierárquica, como se vê, frequentemente, em documentos de texto formatados, imagens vetoriais ou bancos de dados.
 - SPARQL: o “sparkle“, também recomendado pelo W3C e aos cuidados dos grupos de Web Semântica do W3C é utilizado para buscar a informação independente do formato dos resultados.

Existem padrões para publicações de dados em formato aberto. É imprescindível que esses padrões estejam especificados e regulamentados também em normas ou qualquer outra recomendação governamental para possibilitar um ambiente interoperável em todos os domínios e-gov.

O DESENVOLVIMENTO DE UMA API É NECESSÁRIO?²

Uma questão importante a ser levada em consideração na abertura de bases de dados é o desenvolvimento de uma API (Application Programming Interface) para servir informações na Web. Uma API, no escopo deste guia e de forma resumida, é uma camada de interação entre uma base de dados e um aplicativo que se alimenta desses dados. A API oferece ao desenvolvedor/empreendedor interessado uma série de chamadas padrões para se extrair dados de determinada base por meio de requisições na Web. O desenvolvimento de uma API requer conhecimento técnico refinado e, caso ela seja pública, a definição de padrões arbitrários, que tentam antever os casos em que o desenvolvedor/empreendedor precisará dos dados. Uma API apresenta uma

² Adaptado de: <https://www.peterkrantz.com/2012/publishing-open-data-api-design/>

série de vantagens, como o acesso facilitado e rápido às bases de dados. Em vez de baixar a base de dados inteira, basta que o programador faça uma chamada simples na Web para extrair a porção que interessa a ele naquele momento. Ela também facilita o acesso em tempo real a partes específicas da base, permitindo a criação de aplicações que dependem de dados atualizados rapidamente.

As APIs podem ser privadas, quando um desenvolvedor tem controle sobre o banco de dados e cria a API para facilitar o acesso aos dados; ou públicas, quando o guardião de uma base de dados desenvolve uma API para servir a comunidade de desenvolvedores/empreendedores, procurando antever quais tipos de requisições ao banco de dados serão úteis e genéricas o suficiente para atender o maior número de aplicações possíveis. Serviços como o Facebook e o Twitter possuem APIs públicas que permitem programadores de todo mundo interagirem, de forma limitada, com a imensidão de dados que essas empresas abrigam.

No âmbito do governo, apesar das vantagens, o desenvolvimento de uma API pode trazer situações desconfortáveis, dependendo do caso. É preciso refletir com cuidado se o desenvolvimento de uma API é o melhor caminho a ser seguido, pois há alternativas que podem funcionar melhor para ambos os lados: tanto para o desenvolvedor interessado nos dados governamentais, quanto para as equipes de servidores públicos ou terceirizados pelo estado que teriam a missão de manter as APIs funcionando de maneira estável e confiável.

UM CENÁRIO FICTÍCIO

Imagine que a Secretaria de Logística e Transportes do Estado de São Paulo tenha desenvolvido uma API pública para que qualquer desenvolvedor pudesse acessar informações sobre as condições de manutenção das estradas paulistas. Em determinado momento o servidor de API foi inundado e o servidor do banco de dados travou. Os serviços estaduais que dependem desse banco de dados pararam de funcionar. Os registros mostravam que houve um aumento grande no tráfego entre oito e nove horas da manhã com muitas requisições de API vindas de muitos lugares diferentes. Depois das nove horas o nível de acessos nos servidores diminuiu e tudo voltou ao normal.

O QUE ACONTECEU?

Seguindo com o cenário fictício, um ano antes, a Secretaria de Logística e Transportes começou a abrir seus dados como parte da política de transparência do estado. Havia pressa e, com a equipe reduzida, eles decidiram criar uma API para os dados das estradas configurando um servidor de API acessível pela Internet. A API foi desenvolvida levando em consideração potenciais situações em que os desenvolvedores de aplicativos poderiam usá-la, mas era difícil saber exatamente o que as pessoas iriam querer. Ao todo, a equipe da Secretaria estabeleceu três chamadas genéricas de API.

Passado o ocorrido com os servidores, um ano mais tarde, os servidores descobriram que um empreendedor havia desenvolvido um aplicativo de celular que fez muito sucesso, sendo usado por centenas de milhares de pessoas. Todos os dias de manhã o aplicativo anunciava, antes da pessoa ir trabalhar, qual era a situação de manutenção nas estradas paulistas. Para baixar esses dados cada aplicativo instalado em cada celular precisava realizar duas chamadas de API. Foi isso que derrubou os servidores da Secretaria, pois a infraestrutura não estava preparada para lidar com o número de acessos.

ALTERNATIVA

Uma alternativa ao modelo apresentado anteriormente é publicar “dumps” de dados na forma de arquivos. Nesse modelo, os dados da base são exportados e transformados num arquivo de formato aberto, tal como o CSV. Depois disso, recebem um nome padronizado e são armazenados num servidor de páginas da Web. Isso significa que qualquer desenvolvedor pode baixar todos os seus dados, carregá-los na infraestrutura deles e desenvolver suas próprias APIs (nesse caso, privadas) de acordo com a necessidade deles. Além disso, grandes quantidades de acesso estarão concentradas nos servidores deles, sem afetar o funcionamento de outros serviços do governo. Outra vantagem é que a publicação de arquivos “dumps” num servidor de páginas da Web é muito simples. Se os arquivos e URLs receberem nomes consistentes, será fácil para os desenvolvedores baixarem os dados ao longo do tempo (por exemplo, <http://exemplo.com/estradas/2015-01-30.csv>).

CONSIDERAÇÕES

- **Você realmente precisa de uma API?** Desenvolver uma API pode se tornar um projeto caro que vai competir com outros projetos de TI com prioridade maior. Além disso, esse tipo de projeto envolve tomar decisões sobre quais chamadas serão realizadas. Você sabe como seus usuários vão usar seus dados? Sua API vai contribuir para que os usuários utilizem os dados da melhor forma possível? Qual é seu plano para lidar com grandes quantidades de acesso?
- **Crie condições para que os desenvolvedores mantenham uma cópia local e atualizada dos seus dados.** A oferta de “dumps” de dados nomeados de forma consistente simplifica o processo de manter uma base de dados atualizada.
- **Isole sistemas internos dos efeitos da publicação externa de dados.** Tome os devidos cuidados para que os acessos vindos da Web não interfiram com bases de dados internas, afetando outros serviços do Governo.
- **Certifique-se de que você pode mudar seus sistemas sem quebrar os URLs.** Desenvolvedores vão construir aplicativos que dependem dos seus URLs. Não os force a reescrever seus programas apenas porque você vai mudar de plataforma. Sinais de que as coisas podem ser melhores planejadas incluem fragmentos que pertencem a plataformas específicas, como “aspx” ou “jsp”, nos seus URLs. Livre-se deles.

MAPA DE DECISÕES TECNOLÓGICAS

A tabela abaixo foi adaptada a partir do kit de dados abertos do governo federal. Ela mostra uma série de soluções para a publicação de dados abertos, as tecnologias mais utilizadas e o tempo médio esperado de aplicação. As estimativas dependem dos recursos tecnológicos e humanos disponíveis para cada órgão público que deseja abrir suas bases. O sistema de estrelas parte do pressuposto de que os dados são publicados com uma licença aberta.

Solução	Pré-requisitos	Prazo	Estrelas
Publicar dump da base de dados	Acesso à base de dados	Curto	★★★★
	Servidor web para arquivos		
Publicar dados em arquivos CSV	Mecanismo de ETL (caso esteja em banco relacional)	Curto	★★★★
	Servidor web para arquivos		
Publicar dados em arquivos JSON / XML	Mecanismo de ETL (caso esteja em banco relacional)	Médio	★★★★
	Serviço de desenvolvimento		
	Servidor web para arquivos		
Desenvolver módulo de dados abertos em sistema existente	Serviço de desenvolvimento	Longo	★★★★
	Servidor web para rodar nova solução		
Desenvolver API RESTful de dados abertos desacoplada da solução (você precisa mesmo de uma API?)	Mecanismo de ETL	Longo	★★★★
	Serviço de desenvolvimento		
	Servidor web para rodar nova solução		
Novo Sistema, com a gestão de dados incorporados em sua arquitetura	Mecanismo de ETL	Longo	★★★★
	Serviço de desenvolvimento		
	Servidor web para rodar nova solução		
Publicar dados em arquivos RDF	Ontologia da área do conhecimento do sistema	Longo	★★★★★
	Mecanismo de ETL		
	Servidor web para arquivos		
Disponibilizar dados por Endpoint SPARQL	Ontologia da área do conhecimento do sistema	Mais Longo	★★★★★
	Mecanismo de ETL		
	Banco de dados de triplas		
Publicar dados em API de dados ligados (Linked Data)	Ontologia da área do conhecimento do sistema	Mais Longo	★★★★★
	Banco de dados de triplas		
	Serviço de desenvolvimento		
	Mecanismo de ETL		
	Servidor web para rodar nova solução		



CATÁLOGO REPOSITÓRIO DE DADOS

Capítulo 5

São Paulo já possui um catálogo de dados centralizado chamado Governo Aberto SP. Ele representa um esforço da Administração Pública paulista em reunir, num só lugar, informações sobre suas bases públicas, seus guardiões, suas características, a forma de baixá-las pela Web e seus formatos. É recomendável que exista um repositório central de dados, para que o cidadão não tenha que passar horas “catando” bases de dados em sítios de diferentes órgãos do Estado. Iniciativas exitosas ao redor do mundo mostram que a reunião das bases de dados num catálogo central não é apenas algo recomendado sob o ponto de vista da praticidade para o cidadão, mas como forma inteligente de medir e monitorar a saúde das bases públicas disponíveis para a sociedade.

O Governo Aberto SP pode deixar de ser um produto final, que serve como ferramenta de consulta, para ser uma plataforma. É preciso refletir sobre que tipo de informação entra e quais tipos de produtos podem sair desse portal, de modo que as integrações de bastidores (a formulação de scripts para automatizar o processo periódico de publicação de bases, por exemplo) façam parte do fluxo de trabalho dos servidores, sem atrapalhar sua rotina. O portal poderia ser também uma forma de o governo rastrear quais dados abertos estão sendo publicados e mantê-los organizados e atualizados.

Dentre os produtos que podem depender de um catálogo de dados centralizado bem administrado estão os *dashboards*, ou painéis dinâmicos. Essas ferramentas poderiam mostrar o rendimento do estado a partir de diversos indicadores construídos em cima das bases de dados disponíveis no portal. Essa estratégia seria interessante, pois os dados utilizados precisam ser abertos e atualizados para que o painel dinâmico funcione, criando um ciclo virtuoso.

Os painéis podem ser criados de acordo com a demanda do executivo, dos secretários e dos cidadãos: áreas como “criminalidade”, “finanças”, “saúde”, “meio ambiente” ou “transporte”. A escolha poderia ser feita a partir de uma tempestade de ideias com servidores, gestores e a sociedade civil ou por meio de uma consulta pública. Essas informações consistiriam no argumento para o desenvolvimento, que poderia ser feito por meio de competições. O governo colocaria as bases e as APIs à disposição dos programadores e das empresas, e eles fariam o protótipo. Essa relação entre governo e iniciativa privada pode estimular a geração de novos negócios a partir da abertura de dados.

Existem diversas ferramentas no mercado que podem auxiliar a Administração Pública na implantação de um catálogo centralizado de dados. Atualmente, duas que se destacam são o Socrata, desenvolvido por uma empresa americana, e o CKAN, ferramenta gratuita e de código livre mantida pela Open Knowledge e por uma comunidade de desenvolvedores. Ambas as ferramentas são usadas ao redor do mundo em portais de dados abertos governamentais e possuem vantagens e desvantagens. Cabe ao gestor público realizar consultas técnicas para tomar a decisão mais acertada frente às necessidades e ao contexto paulistas.

The background is a dark blue gradient. It features several overlapping circles of varying sizes and colors (white, light blue, purple). A central graphic element consists of a purple circle with three colored arms (cyan, green, and blue) extending outwards, resembling a molecular structure or a network node. The text is positioned in the lower right quadrant.

CENÁRIOS TÉCNICOS, OPÇÕES TECNOLÓGICAS

Capítulo 7

Abrir bases e colocá-las disponíveis na Web significa também contemplar cenários técnicos e de infraestrutura para manter essas bases acessíveis e atualizadas. Este guia se baseia em três cenários diferentes, como sugeridos pelo kit de ferramentas do Banco Mundial para dados abertos governamentais.

Essas decisões também devem ser levadas em consideração quando comparadas aos objetivos de abertura de dados do órgão e como eles se encaixam no sistema de cinco estrelas. Para efeitos de comparação, as recomendações técnicas do Banco Mundial levam em consideração apenas bases com, no máximo, três estrelas.

São considerados três níveis de complexidade baseados no número de bases de dados disponíveis e a frequência de atualização de cada uma delas:

Nível 1: menos de 100 bases de dados com menos de 10 bases atualizadas semanalmente

Nível 2: de 100 a 1000 bases de dados, com 10-100 bases atualizadas semanalmente

Nível 3: mais de 1000 bases de dados, com 100 ou mais bases atualizadas semanalmente

Este guia apresenta as soluções técnicas apenas em caráter de recomendação, não endossando qualquer software ou solução tecnológica de uma empresa específica.

NÍVEL 1

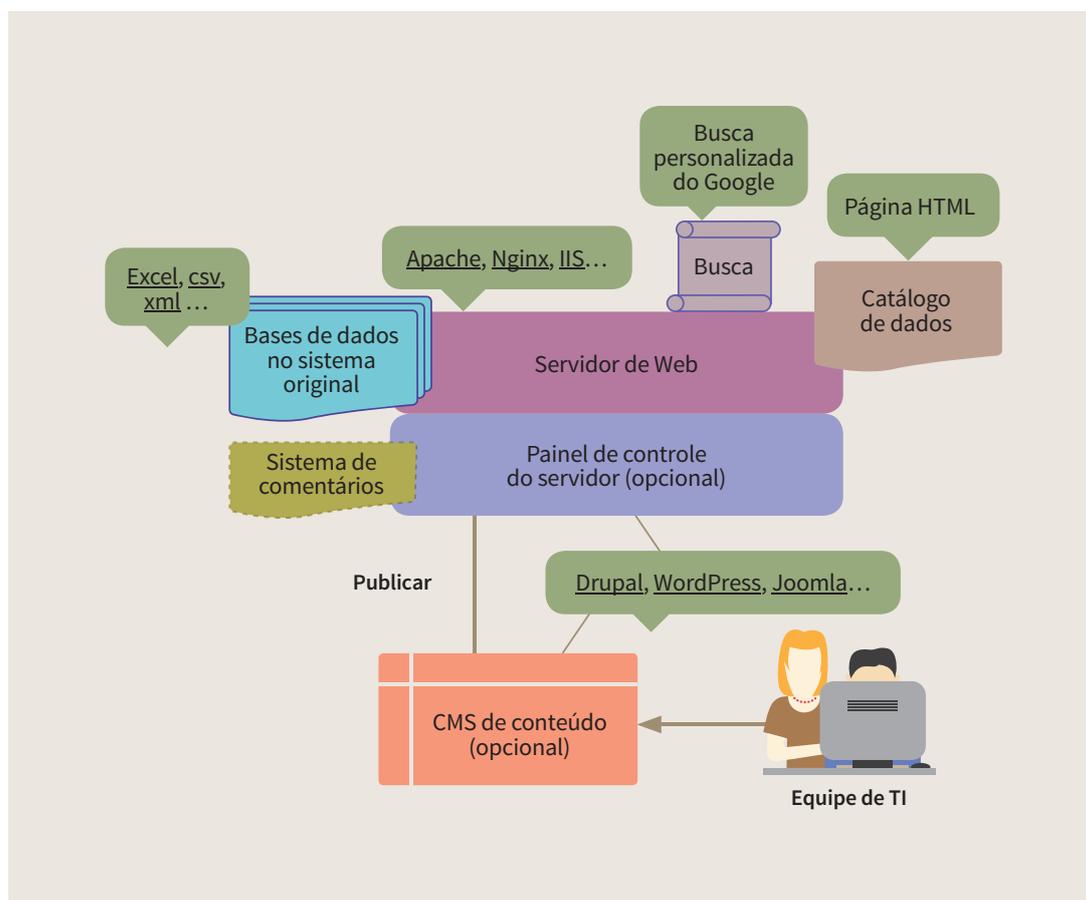
O primeiro nível de complexidade atende órgãos que desejam iniciar seus esforços em abertura de dados e que não possuam um volume muito grande de bases de dados. Nesses casos, um único profissional de TI é capaz de administrar o local onde as bases de dados serão disponibilizadas.

O órgão que tiver menos de 100 bases para publicar na Web pode configurar um servidor de Web convencional, hospedando os documentos no próprio servidor ou num serviço de armazenamento em nuvem. As bases devem ter sua descrição (também conhecida como “metadados”) embarcada na página onde são publicadas e ferramentas existentes de busca, como a busca personalizada do Google, podem ser utilizadas.

Não é necessário configurar uma rotina de atualização automática de bases: os dados podem ser enviados manualmente com apoio da equipe de TI do órgão. É recomendado incluir um sistema de comentários nas páginas das bases, de modo a estimular sugestões e críticas dos usuários que acessarem as bases.

Resumo:

- Configurar um sítio usando um servidor de Web padrão
- Hospedar dados diretamente no servidor (ou usando algum serviço de hospedagem em nuvem)
- Metadados embarcados na própria página onde as bases são exibidas
- Busca no sítio com ferramentas existentes, como a busca personalizada do Google
- Atualização manual de dados, metadados e conteúdo com apoio da equipe de TI
- Incluir sistema de comentários nas páginas das bases, para receber sugestões



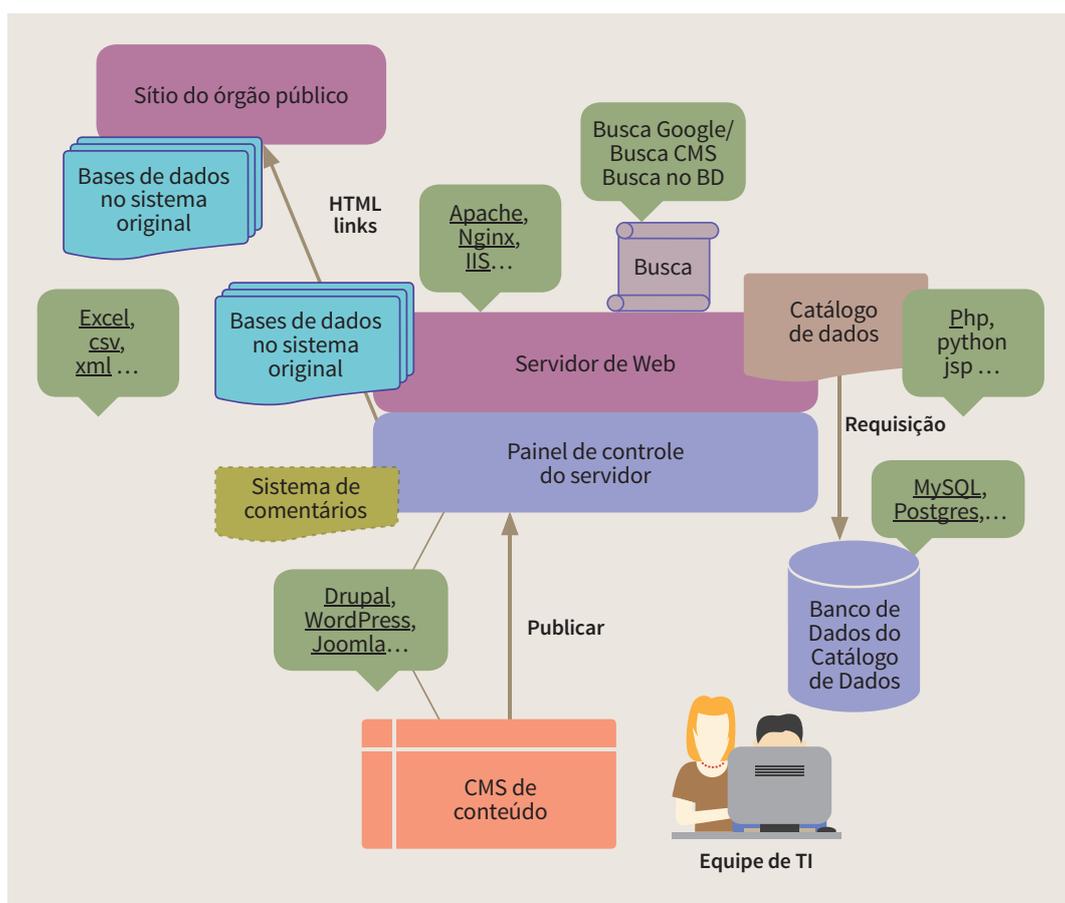
NÍVEL 2

O segundo nível de complexidade descreve um cenário em que uma quantidade razoável de bases de dados devem ser publicadas na Web. A solução apresentada pode ser compartilhada entre diferentes órgãos, dependendo da demanda. Nesse nível de complexidade, é recomendado que a equipe de TI tenha especialistas na configuração otimizada de servidores e bancos de dados.

A publicação de 100 a 1000 bases de dados poderá acontecer em um servidor central que serve de repositório de bases para o(s) órgão(s). O acesso se dá por meio de um CMS (do inglês “Content Management System”, ou “Sistema de Gerenciamento de Conteúdo”), como Wordpress ou Drupal, e os dados são gerenciados manualmente em seus formatos originais no mesmo servidor do CMS, ou armazenamento em nuvem. Os metadados das bases são armazenados em banco SQL, exibido em páginas geradas de forma dinâmica. O sistema faz verificação automática de links quebrados e a busca pode ser feita diretamente na base de metadados, ou usando as opções do CMS ou busca personalizada do Google. É recomendado o uso de *caching* para cuidar da grande quantidade de acessos e também não armazenar os metadados no CMS.

Resumo:

- *Front-end* convencional usando um CMS (Wordpress, Drupal etc);
- Bases de dados são hospedadas em formato original no próprio servidor da aplicação ou nos respectivos sítios dos órgãos, para o caso de uma atividade em conjunto;
- Metadados são hospedados em banco de dados SQL, requisições via geração dinâmica de páginas;
- Verificação automática de links quebrados;
- Busca via varredura de texto no banco SQL, busca do CMS ou busca personalizada do Google;
- Fazer *caching* para equilibrar acessos;
- Não salvar metadados no CMS.



NÍVEL 3

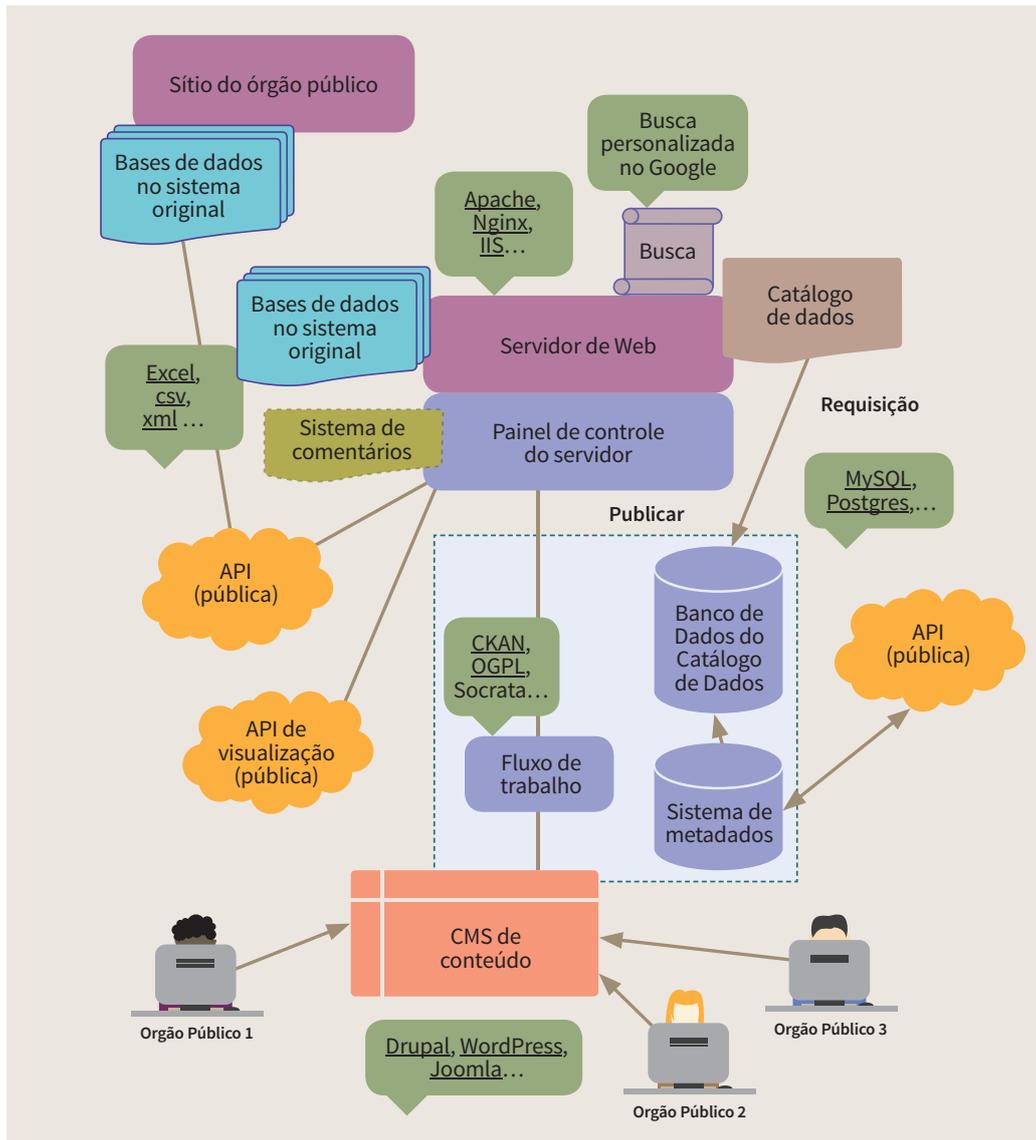
O terceiro nível de complexidade descreve um cenário onde mais de 1000 bases são publicadas e 10% delas são atualizadas frequentemente. Nesses casos, é comum que a plataforma sirva para gerenciar bases de dados de diversos órgãos, contando com uma equipe de TI especializada e que poderá atender as demandas de vários setores em tempo hábil.

Um ou mais servidores poderão ser usados, dependendo da preferência da equipe de administração da infraestrutura. A parte de *front-end* deve integrar diferentes *web services*, preferencialmente por meio de uma API. O gerenciamento dos documentos das bases de dados deve ser feito de forma automática, possivelmente em uma solução de armazenamento em nuvem. Os metadados são armazenados num repositório otimizado. O envio e a atualização das bases de dados são delegadas a cada órgão público que compartilha a plataforma com validação automática de bases e controle de

acesso em níveis. A verificação de links quebrados é automática e deverá ser oferecido um sistema para que usuários possam relatar erros com o “responsável” por cada base, via formulário web, por exemplo. A busca poderá ser implementada via pesquisa estruturada no repositório de metadados, no CMS ou usando a busca personalizada do Google no site principal.

Resumo:

- *Front-end* faz integração de vários *web services*, preferencialmente por API;
- Gerenciamento automático dos arquivos de dados originais, armazenados em solução na nuvem;
- Metadados são armazenados em repositório otimizado;
- Envio e atualização das bases de dados são delegadas a cada órgão público, com validação automática de bases e controle de acesso em níveis;
- Verificação automática de links quebrados;
- Sistema para que usuários possam relatar erros com o “responsável” por cada base, via formulário web ou similar;
- Busca estruturada no repositório de metadados, no CMS ou usando a busca personalizada do Google no site principal.



MELHORES PRÁTICAS DE SEGURANÇA DA INFORMAÇÃO

Assim como toda iniciativa que envolve tecnologia da informação, a abertura de bases deve ser feita seguindo critérios rigorosos de melhores práticas de segurança da informação. Vale lembrar:

- Governança de dados para garantir:
 - Autoridade da fonte
 - Regras de engajamento
 - Sustentabilidade
- Classificação de dados Públicos x Sigilosos para garantir:
 - Dados privados ou sigilosos não terão acesso externo
- Controles de Segurança da Informação para garantir
 - Confidencialidade e integridade dos dados
 - Proteção contra ataques de negação de serviço (DoS)



LICENÇA DE UTILIZAÇÃO DOS DADOS

Capítulo 8

A abertura de bases, nos termos sugeridos por este guia, implica disponibilizar as informações de forma completa na Web, em formatos não proprietários, com acesso irrestrito e livre para qualquer pessoa e de modo que qualquer um possa reutilizá-los para qualquer fim, sem restrições. Esse parágrafo é um breve resumo do que pode ser considerada uma “licença aberta”, ou seja, as condições que o guardião dos dados dá ao usuário para utilização dos mesmos. Sem essa licença, os dados não podem ser considerados “abertos”, pois não há garantias de que sua utilização esteja resguardada pelos princípios da abertura de dados.

Uma “licença aberta” é importante também para que a Lei de Acesso à Informação seja respeitada. A Infraestrutura Nacional de Dados Abertos, do governo federal, ainda não chegou a uma conclusão se o ordenamento jurídico brasileiro é suficiente para tratar os dados governamentais de forma aberta, sem a necessidade de licenças, se outras licenças que já existem se enquadram no contexto brasileiro, ou se será necessário criar uma licença específica para o Brasil. Por exemplo, a legislação brasileira, no âmbito da Lei de Direitos Autorais, prevê explicitamente a proteção a banco de dados que constituam “criação intelectual”. Por outro lado, a Lei de Acesso à Informação diz que o Estado deve proporcionar o acesso a toda informação que produz ou detém, desde que não seja considerada sigilosa. A interação dessas leis ainda é objeto de estudo.

A Lei de Acesso à Informação, por outro lado, não determina qual licença deve ser usada ao publicar dados, mas lista princípios que podem muito bem direcionar a formulação de uma licença aberta ou termos de utilização compatíveis com os conceitos de abertura de dados explicitados neste guia:

Art. 3º Os procedimentos previstos nesta Lei destinam-se a assegurar o direito fundamental de acesso à informação e devem ser executados em conformidade com os princípios básicos da administração pública e com as seguintes diretrizes:

I - observância da publicidade como preceito geral e do sigilo como exceção;

Art. 8º ...

...

§ 3º Os sítios de que trata o § 2º deverão, na forma de regulamento, atender, entre outros, aos seguintes requisitos:

...

II - 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;

III - possibilitar o acesso automatizado por sistemas externos em formatos abertos, estruturados e legíveis por máquina;

IV - divulgar em detalhes os formatos utilizados para estruturação da informação;

V - garantir a autenticidade e a integridade das informações disponíveis para acesso;

VI - manter atualizadas as informações disponíveis para acesso;

Em outras palavras, a lei determina que, via de regra, os dados governamentais são públicos. O sigilo é exceção. Além disso, os sítios da Web que divulgarão esses dados devem permitir a gravação de relatórios em formatos estruturados e não proprietários, divulgar em detalhes os formatos usados na estruturação da informação, garantir a integridade e autenticidade dos dados e mantê-los atualizados.

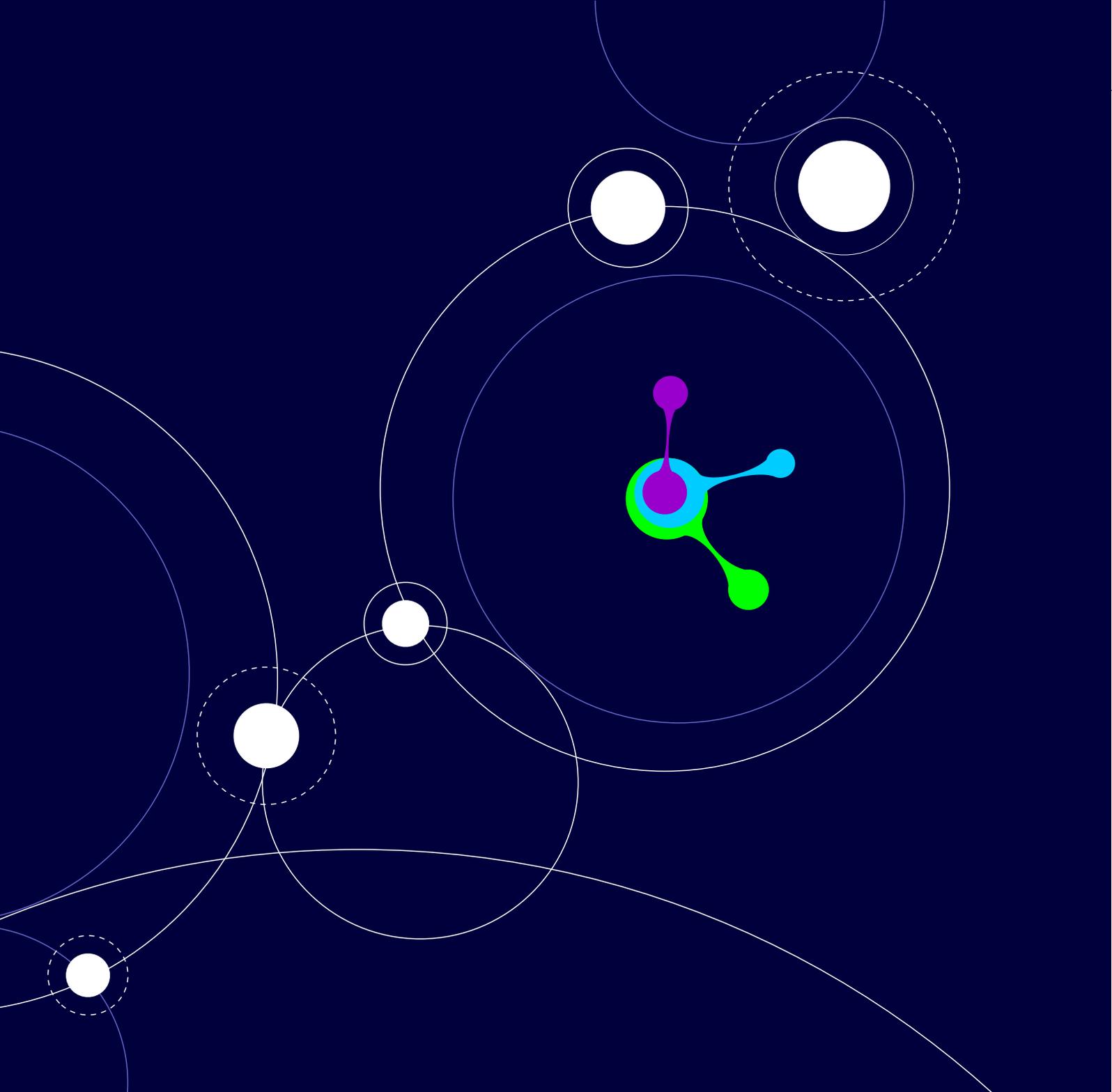
A formulação de termos de utilização ou uma licença aberta para uso de dados deve considerar com responsabilidade a compatibilidade do texto com outras licenças de dados abertos. A consideração é vital para que seja legalmente possível realizar agregações, estatísticas, análises e cruzamentos de dados de diferentes fontes. Os resultados dessas atividades é que agregam valor à abertura de dados e a torna útil para sociedade. Se o governo estadual usar uma licença incompatível com a licença usada pela administração municipal, não seria possível, legalmente, cruzar as bases de hospitais, por exemplo, para construir um aplicativo que exibe as instalações municipais e estaduais. Uma licença ou termos de uso de dados abertos devem ser formulados de modo a maximizar o grau de compatibilidade com as licenças usadas ou que possam vir a ser usadas por outras esferas e poderes do Estado.

No Brasil, a maior parte dos órgãos públicos não tem especificado qualquer licença ao publicar dados na Internet. Há exceções, contudo, com a utilização das licenças ODbL e DbCL (links em inglês).

REGRAS DE UTILIZAÇÃO DO GOVERNO ABERTO SP

O governo de São Paulo já dispõe de um portal central de dados chamado Governo Aberto SP. Nele, há regras de utilização que se inspiram em princípios abertos e que podem servir de modelo para a formulação de licenças abertas, com pequenos ajustes, dependendo da necessidade de cada órgão e sua adequação aos preceitos de abertura de dados deste guia. De forma resumida, as regras de utilização do Governo Aberto SP são:

1. Qualquer pessoa pode utilizar, física ou jurídica;
2. Não estabelece restrições relativas ao uso dos dados;
3. Governo garante autenticidade, integridade e atualidade dos dados baixados somente no portal Governo Aberto SP;
4. Determina que a reutilização dos dados está condicionada à citação da fonte original dos dados (órgão público), incluindo que os dados foram acessados por meio do Governo Aberto SP. Além disso, determina o acréscimo da declaração: “O Governo Aberto SP e o órgão ou entidade de onde foram acessados os dados não garantem sua autenticidade, qualidade, integridade e atualidade após terem sido disponibilizados para uso secundário”;
5. Usuários respondem pelo uso secundário dos dados, eximindo o Estado;
6. Condição para ter acesso aos dados é conhecer as regras e aceitá-las.



FORMATOS DAS BASES

Capítulo 9

No escopo deste guia, uma base de dados nada mais é do que um arquivo de computador construído de forma estruturada, com o objetivo de armazenar informações para consulta e análise posterior. Sua base de dados pode ser construída manualmente, desde que você defina uma estrutura para organizar esses dados e mantenha coerência. Isso é importante para garantir que as buscas realizadas nessa base encontrem aquilo que se procura. Uma base de dados pode ser um arquivo de texto, por exemplo, com a lista de todos os municípios do estado de São Paulo. Ou então, essa base pode ser uma lista de hospitais da cidade de São Paulo e, ao lado de cada nome, o bairro onde cada um se encontra:

```
Hospital Municipal Infantil Menino Jesus, Bela Vista
Hospital do Servidor Público Municipal, Aclimação
Pronto-Socorro Municipal Barra Funda, Barra Funda
Hospital Municipal Cidade Tiradentes, Cidade Tiradentes
```

Nesse caso, a estrutura é definida pela escolha de colocar dois nomes (hospital e bairro) em uma nova linha do arquivo, separados por um delimitador, a vírgula. Dois hospitais nunca apareceriam na mesma linha, por exemplo. O que define a integridade de uma base de dados, a grosso modo, são os elementos utilizados com o intuito de dar previsibilidade para as consultas realizadas nessa base: no exemplo citado, todas as linhas apresentam, primeiro, o nome de um hospital e, segundo, o nome do bairro onde ele se encontra. Se qualquer linha dessa base for diferente do modelo “Nome do hospital, Bairro”, a integridade da base ficará comprometida e ela perderá sua utilidade:

```
Hospital Municipal Infantil Menino Jesus, Bela Vista
Aclimação, Hospital do Servidor Público Municipal
Pronto-Socorro Municipal Barra Funda, Barra Funda
Hospital Municipal Cidade Tiradentes, Cidade Tiradentes
```

Na maior parte dos casos, contudo, as próprias ferramentas disponíveis nos computadores são capazes de gerar ou converter, automaticamente, arquivos estruturados que servem como bases de dados. Um dos exemplos mais comuns

é a planilha do Excel, arquivos de computador com a terminação “.xls” ou “.xlsx”. Esses documentos apresentam linhas e colunas e permitem análises e cruzamentos posteriores. Contudo, os formatos dos arquivos nativos do Excel utilizam uma tecnologia proprietária e fechada. Isso quer dizer que eles dependem de tecnologias exclusivas, muitas vezes que custam dinheiro e não estão disponíveis de forma ampla e gratuita para qualquer pessoa.

A lista abaixo sugere uma série de formatos abertos ou não proprietários que se encaixam melhor nos princípios de abertura de dados apresentados neste guia e apresenta uma breve introdução sobre cada um deles. Não há um formato mais recomendado do que outro. Cada equipe deve refletir sobre quais os formatos das bases já existentes (em arquivos do Excel, por exemplo) e se existe alguma forma de convertê-las para algum dos formatos sugeridos abaixo, dependendo da aplicação.

FORMATOS SEPARADOS POR DELIMITADORES (CSV)

Os arquivos CSV (do inglês “Character-separated values” ou “valores separados por um delimitador”) servem para armazenar dados tabulares (números e texto) em texto simples. O “texto simples” significa que o arquivo é uma sequência de caracteres puros, sem qualquer informação escondida que o computador tenha que processar.

Um arquivo CSV abriga um sem número de “registros”, separados por quebras de linha (cada “registro” permanece numa linha do arquivo) e cada registro possui um ou mais “campos”, separados por um delimitador, os mais comuns sendo a vírgula (“,”), o ponto e vírgula (“;”) e o caractere “invisível” que surge ao se pressionar a tecla “tab”. Arquivos separados por vírgula e ponto e vírgula normalmente recebem a extensão “CSV” e arquivos separados por “tab” a extensão “TSV”. Há também bases de dados nesses formatos que recebem a extensão “TXT”. Arquivos CSV são simples e funcionam na maior parte das aplicações que lidam com dados estruturados.

Fazendo uma comparação com linhas e colunas numa planilha, os “registros” de um arquivo CSV são as linhas e os “campos” são as colunas. Os valores dos “campos” do primeiro “registro”, ou seja, da primeira linha, normalmente são os nomes das colunas. Apesar de não existir um padrão internacional para o CSV, suas variações são simples o suficiente para que os aplicativos compatíveis

possam consertar facilmente as diferenças. Tipicamente, é assim que um arquivo CSV é exibido quando aberto num editor de textos:

```
Continente;País;Capital
África;Angola;Luanda
América do Norte;Estados Unidos;Washington DC
América Central;México;Cidade do México
América do Sul;Brasil;Brasília
Europa;Espanha;Madri
Europa;Alemanha;Berlim
Oceania;Austrália;Camberra
Ásia;Japão;Tóquio
```

Esse arquivo possui três colunas separadas pelo delimitador ponto e vírgula (“;”): Continente, País e Capital, como descrito na primeira linha. Ao todo, são oito registros. O primeiro é a tríade África-Angola-Luanda e o último Ásia-Japão-Tóquio. Não há limite prático para o número de linhas ou colunas em um arquivo CSV. Esse número pode chegar a milhões ou dezenas de milhões, dependendo exclusivamente da capacidade de processamento do computador que vai ser utilizado na consulta. Se o mesmo arquivo CSV fosse aberto num processador de planilhas, ele seria exibido assim:

Continente	País	Capital
África	Angola	Luanda
América do Norte	Estados Unidos	Washington DC
América Central	México	Cidade do México
América do Sul	Brasil	Brasília
Europa	Espanha	Madri
Europa	Alemanha	Berlim
Oceania	Austrália	Camberra
Ásia	Japão	Tóquio

FORMATO XML

O XML é uma linguagem de marcação, assim como o HTML (usado para construir páginas da Web), definido e mantido pelo World Wide Web Consortium (W3C). O objetivo do XML é o foco na simplicidade, generalidade e usabilidade por toda a Internet. Embora o XML enfatize a geração de documentos, ele também é usado para representar estruturas de dados arbitrarias, para integração entre sistemas de computadores. Um típico arquivo XML tem a seguinte estrutura:

```
<?xml version="1.0" encoding="UTF-8"?>
<Exemplo>
  <Localidade número="1">
    <Continente>África</Continente>
    <País>Angola</País>
    <Capital>Luanda</Capital>
  </Localidade>
  <Localidade número="2">
    <Continente>América do Norte</Continente>
    <País>Estados Unidos</País>
    <Capital>Washington DC</Capital>
  </Localidade>
  <Localidade número="3">
    <Continente>América Central</Continente>
    <País>México</País>
    <Capital>Cidade do México</Capital>
  </Localidade>
  <Localidade número="4">
    <Continente>América do Sul</Continente>
    <País>Brasil</País>
    <Capital>Brasília</Capital>
  </Localidade>
  <Localidade número="5">
    <Continente>Europa</Continente>
    <País>Espanha</País>
    <Capital>Madri</Capital>
  </Localidade>
  <Localidade número="6">
    <Continente>Europa</Continente>
    <País>Alemanha</País>
    <Capital>Berlim</Capital>
  </Localidade>
  <Localidade número="7">
    <Continente>Oceania</Continente>
    <País>Austrália</País>
    <Capital>Camberra</Capital>
  </Localidade>
  <Localidade número="8">
    <Continente>Ásia</Continente>
    <País>Japão</País>
    <Capital>Tóquio</Capital>
  </Localidade>
</Exemplo>
```

Marcadores e conteúdo

Um arquivo XML possui duas características principais: marcadores e conteúdo. Geralmente, as sequências de caracteres que começam com < e terminam com > ou começam com & e terminam com ; são chamadas de “marcadores”. Sequências de caracteres que não são marcadores são consideradas “conteúdo”. No exemplo acima, <Localidade> e <Continente>, por exemplo, são marcadores. Os nomes dos países, continentes e capitais, são “conteúdo”.

Tags

As tags são os marcadores que começam com < e terminam com >. São três os tipos de tags:

- tags de início; por exemplo: <Localidade>
- tags de término; por exemplo: </Localidade>
- tags de elemento vazia; por exemplo: <line break />

Elementos

Os elementos são componentes do XML que começam com uma tag de início e terminam com uma tag de término correspondente, ou consistem apenas de uma tag de elemento vazia. A sequência de caracteres entre as tags de início e término, se existirem, são o conteúdo do elemento e podem incluir marcadores, incluindo outros elementos, que são chamados de “filhos”. No exemplo acima, um elemento seria:

```
<País>Brasil</País>.
```

Atributos

Os atributos são pares de “nome/valor” que existem dentro da tag de início ou tag de elemento vazia. No exemplo acima, o elemento <Localidade> possui um atributo “número” e um valor correspondente:

```
<Localidade número="8">
```

O nome do atributo é “número” e seu valor é “8”. Os atributos só podem apresentar um valor entre aspas e cada atributo não pode aparecer mais de uma vez em cada elemento.

Declaração XML

Os documentos XML devem começar declarando alguma informação sobre si mesmos, como no exemplo:

```
<?xml version="1.0" encoding="UTF-8"?>
```

FORMATO KML

O Keyhole Markup Language (KML) é uma derivação do XML para expressar dados e visualizações geográficas em navegadores de mapas bidimensionais ou tridimensionais desenvolvidos para a Internet. O formato foi adquirido pelo Google em 2004 e tornou-se o padrão usado no aplicativo Google Earth. Em 2008, o formato se tornou um padrão internacional do Consórcio Geoespacial Aberto.

O formato KML tem uma estrutura semelhante ao do XML, mas define uma série de características, como marcadores de lugares, imagens, polígonos, modelos 3D e descrições textuais. Cada localidade sempre apresenta uma longitude e uma latitude. Os arquivos são distribuídos em pacotes KMZ, que são arquivos KML compactados no padrão zip com uma extensão .kmz. O conteúdo do pacote compactado compreende em um único documento KML (“doc.kml”) e, de forma opcional, subdiretórios contendo as imagens e outros arquivos referenciados no KML. Um típico documento KML se apresenta da seguinte maneira:

```
<?xml version="1.0" encoding="UTF-8"?>
<kml xmlns="http://www.opengis.net/kml/2.2">
<Document>
  <Placemark>
    <name>São Paulo</name>
    <description>Cidade de São Paulo</description>
    <Point>
      <coordinates>-23.5476258,-
46.6360159</coordinates>
    </Point>
  </Placemark>
</Document>
</kml>
```

FORMATO JSON

O formato JSON (JavaScript Object Notation) é um formato aberto usado como alternativa ao XML para a transferência de dados estruturados entre um servidor de Web e uma aplicação Web. Sua lógica de organização tem semelhanças com o XML, mas possui notação diferente. O formato ganhou popularidade em serviços da Web, como clientes de email e sítios de compras, pois consegue transmitir uma grande quantidade de informações entre o cliente e o servidor usando uma quantidade menor de caracteres.

Os arquivos JSON também trabalham com pares de atributos e valores e em vez de marcadores, como no XML, utilizam delimitadores em cadeias: {}, [], e ""'. Um típico arquivo JSON é estruturado da seguinte forma:

```
{
  "localidade 1": {
    "Continente": "África",
    "País": "Angola",
    "Capital": "Luanda"
  },
  "localidade 2": {
    "Continente": "América do Norte",
    "País": "Estados Unidos",
    "Capital": "Washington DC"
  },
  "localidade 3": {
    "Continente": "América Central",
    "País": "México",
    "Capital": "Cidade do México"
  },
  "localidade 4": {
    "Continente": "América do Sul",
    "País": "Brasil",
    "Capital": "Brasília"
  },
  "localidade 5": {
    "Continente": "Europa",
    "País": "Espanha",
    "Capital": "Madri"
  },
  "localidade 6": {
    "Continente": "Europa",
    "País": "Alemanha",
    "Capital": "Berlim"
  },
  "localidade 7": {
    "Continente": "Oceania",
    "País": "Austrália",
    "Capital": "Camberra"
  },
  "localidade 8": {
    "Continente": "Ásia",
    "País": "Japão",
    "Capital": "Tóquio"
  }
}
```

O delimitador { marca o início de uma seção e o } marca seu fim. Os pares de valor e atributo são separados por : e seus valores, quando texto, ficam entre aspas (números, por exemplo, não recebem aspas). No exemplo abaixo, a “localidade 6” é um atributo que recebe uma série de valores (Continente, País e Capital):

```
"localidade 6": {
  "Continente": "Europa",
  "País": "Alemanha",
  "Capital": "Berlim"
},
```

Repare que o valor da “localidade 6” é um novo conjunto de pares atributo-valor. Esse novo conjunto é iniciado com o delimitador { e finalizado com }. Essa lógica de encadear conjuntos de pares pode ser repetida inúmeras vezes, criando diversos níveis para a estrutura de dados desejada.

GEOJSON/TOPOJSON

Os formatos geoJSON e topoJSON são derivados do JSON para representação de coleções de características geográficas simples, junto com seus atributos não-espaciais. Dentre as características possíveis de serem armazenadas no padrão geoJSON/topoJSON estão “points”, incluindo endereços e localidades; “line strings”, incluindo ruas, rodovias e limites; “polygons”, incluindo países, estados ou demarcações de terra; e coleções misturadas desses tipos. O diferencial do topoJSON em relação ao geoJSON é que ele armazena topologia geoespacial, gerando um arquivo final com tamanho frequentemente menor.

FORMATO SQL (DUMP)

O SQL (do inglês, “Structured Query Language” ou “Linguagem Estruturada de Requisição”) é uma linguagem de programação especialmente desenvolvida para gerenciar dados em sistemas de bancos de dados relacionais. Os comandos possíveis em SQL incluem a inserção, requisição, atualização e remoção de dados, a criação e alteração de esquemas de bancos e o controle de dados. O “dump” de uma base de dados normalmente resulta numa lista de comandos SQL e permite que qualquer pessoa possa reconstruir essa base a partir do seu esquema de dados e dos valores contidos nela. Um arquivo “dump” típico se apresenta da seguinte maneira:

```

-- Base de dados
CREATE DATABASE `ex_localidades`;
USE `Exemplos de Localidades`;

-- Estrutura da tabela para a tabela `localidades`
CREATE TABLE `localidades` (
  `id` INT(8) UNSIGNED NOT NULL AUTO_INCREMENT,
  `nome de usuário` VARCHAR(16) NOT NULL,
  `senha` VARCHAR(16) NOT NULL,
  PRIMARY KEY (`id`)
);

-- Dados da tabela `localidades`
INSERT INTO `localidades` VALUES ('Continente', 'País',
'Capital'), ('África', 'Angola', 'Luanda'), ('América
do Norte', 'Estados Unidos', 'Washington DC'), ('América
Central', 'México', 'Cidade do México'), ('América do Sul',
'Brasil', 'Brasília'), ('Europa', 'Espanha', 'Madri'),
('Europa', 'Alemanha', 'Berlim'), ('Oceania', 'Austrália',
'Camberra'), ('Ásia', 'Japão', 'Tóquio');

```

As bases de dados em SQL normalmente são criadas e administradas utilizando ferramentas direcionadas a profissionais da área de TI. Enquanto os formatos CSV, XML e JSON podem ser criados facilmente em editores de texto comuns, uma base SQL requer maior refinamento e conhecimento técnico.

FORMATO SHAPEFILE

O shapefile é um formato para bases de dados geoespaciais e vetoriais em sistemas de informação geográfica (no inglês, GIS - Geographic Information System). Ele foi desenvolvido e é regulado pela empresa Esri. É considerado um formato aberto, apesar de proprietário. Por ser aberto, o formato recebe suporte de diversos aplicativos de processamento de mapas gratuitos e de código livre. O shapefile pode descrever espacialmente qualidades de vetores (pontos, linhas e polígonos, para rios, lagos e poços, por exemplo) e cada item normalmente possui atributos que o descrevem, como nome e temperatura.

Apesar de ser um termo no singular, o formato shapefile consiste numa coleção de arquivos de mesmo nome e terminações diferentes, armazenados no mesmo diretório. Existem três arquivos obrigatórios para o funcionamento correto de um shapefile: .shp, .shx e .dbf. O arquivo shapefile propriamente dito é o .shp, mas se distribuído sozinho não será capaz de exibir os dados armazenados. A distribuição deve ser feita juntamente com os outros dois arquivos.

Os arquivos obrigatórios são:

- .shp — formato shape; as características da geometria propriamente dita
- .shx — formato índice de shape; um índice com as características das geometrias para permitir buscas mais rápidas
- .dbf — formato de atributos; atributos apresentados em colunas para cada “shape”

Em cada um dos arquivos obrigatórios, cada “shape” em cada arquivo corresponde aos outros em sequência (o primeiro registro no arquivo .shp corresponde ao primeiro registro nos arquivos .shx e .dbf, e assim por diante).

Não é viável gerar arquivos shapefile manualmente, como seria possível gerar arquivos CSV, XML e JSON, pois eles possuem natureza binária. Normalmente, esses arquivos são gerados a partir da manipulação de pontos e características em mapas nos próprios aplicativos de geoprocessamento.

The image features a dark blue background with several white geometric elements. On the left side, there are three overlapping circles of varying sizes. The top-left circle is solid white. The middle-left circle is dashed white. The bottom-left circle is solid white. A large, thin white arc curves across the middle of the page. In the upper right quadrant, there is a small solid white circle surrounded by a thin white ring. The word "REFERÊNCIAS" is written in a bold, white, sans-serif font in the lower right quadrant.

REFERÊNCIAS

Além de apresentar referências no próprio corpo do texto do guia, por meio de links externos, abaixo é apresentada uma lista de leituras complementares que serviram de base para a construção dos conceitos aqui apresentados. As referências também contribuem para ampliar e extrapolar a construção de processos que visam a abertura de dados, seus desafios, desdobramentos e detalhes técnicos.

CSV:

https://en.wikipedia.org/wiki/Comma-separated_values

Código de Conduta para uso de dados abertos do governo (Reino Unido):

<http://data.gov.uk/code-conduct>

Exemplos de Licenças Abertas (Governo dos Estados Unidos):

<https://project-open-data.cio.gov/license-examples/>

geoJSON:

<http://geojson.org/>

Guia de Implantação de um Portal de Transparência (CGU):

http://www.cgu.gov.br/Publicacoes/transparencia-publica/brasil-transparente/arquivos/guia_portaltransparencia.pdf

Guia sobre Informações Classificadas (CGU):

http://www.acessoinformacao.gov.br/lai-para-sic/sic-apoio-orientacoes/guias-e-orientacoes/guia_informacoesclassificadas.pdf/@@download/file/Guia_InformacoesClassificadas.pdf

Hampshire County Open Licence (Reino Unido):

<http://www3.hants.gov.uk/opendata/licence.htm>

JSON:

<http://www.json.org/>

Kit de Dados Abertos (Infraestrutura Nacional de Dados Abertos):

<http://kit.dados.gov.br/>

KML:

<https://developers.google.com/kml/documentation/?hl=pt-br>

Manual de Lei de Acesso à Informação para Estados e Municípios (CGU):

http://www.cgu.gov.br/Publicacoes/transparencia-publica/brasil-transparente/arquivos/manual_lai_estadosmunicipios.pdf

Open Definition (Open Knowledge):
<http://opendefinition.org/od/index.html>

Open Government Data (book):
<https://opengovdata.io/>

Open Government Guide (Open Government Data - Sunlight Foundation & Open Knowledge):
<http://www.opengovguide.com/topics/open-government-data/>

Open Government Licence for Public Sector Information (Reino Unido):
<http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/>

Open Government Toolkit (Banco Mundial):
<http://data.worldbank.org/open-government-data-toolkit>

Publishing Open Data: do you really need an API?:
<https://www.peterkrantz.com/2012/publishing-open-data-api-design/>

Shapefile:
<http://doc.arcgis.com/pt-br/arcgis-online/reference/shapefiles.htm>

SQL (dump):
https://en.wikipedia.org/wiki/Database_dump

Texto Lei de Acesso à Informação - Lei nº 12.527, de 18 de novembro de 2011. (Governo Federal):
http://www.planalto.gov.br/ccivil_03/ato2011-2014/2011/lei/l12527.htm

Tool Kit para publicação de dados em formato aberto:
<http://platform.od4d.org/article?locale=pt&uri=http%3A%2F%2Fplatform.od4d.org%2Fposts%2F58>

topoJSON:
<https://en.wikipedia.org/wiki/GeoJSON#TopoJSON>

Uso e reuso de dados governamentais:
<http://br.okfn.org/2013/08/28/dados-meio-abertos-sobre-o-uso-e-reuso-dos-dados-governamentais-brasileiros/>

XML:
<http://www.w3.org/XML>

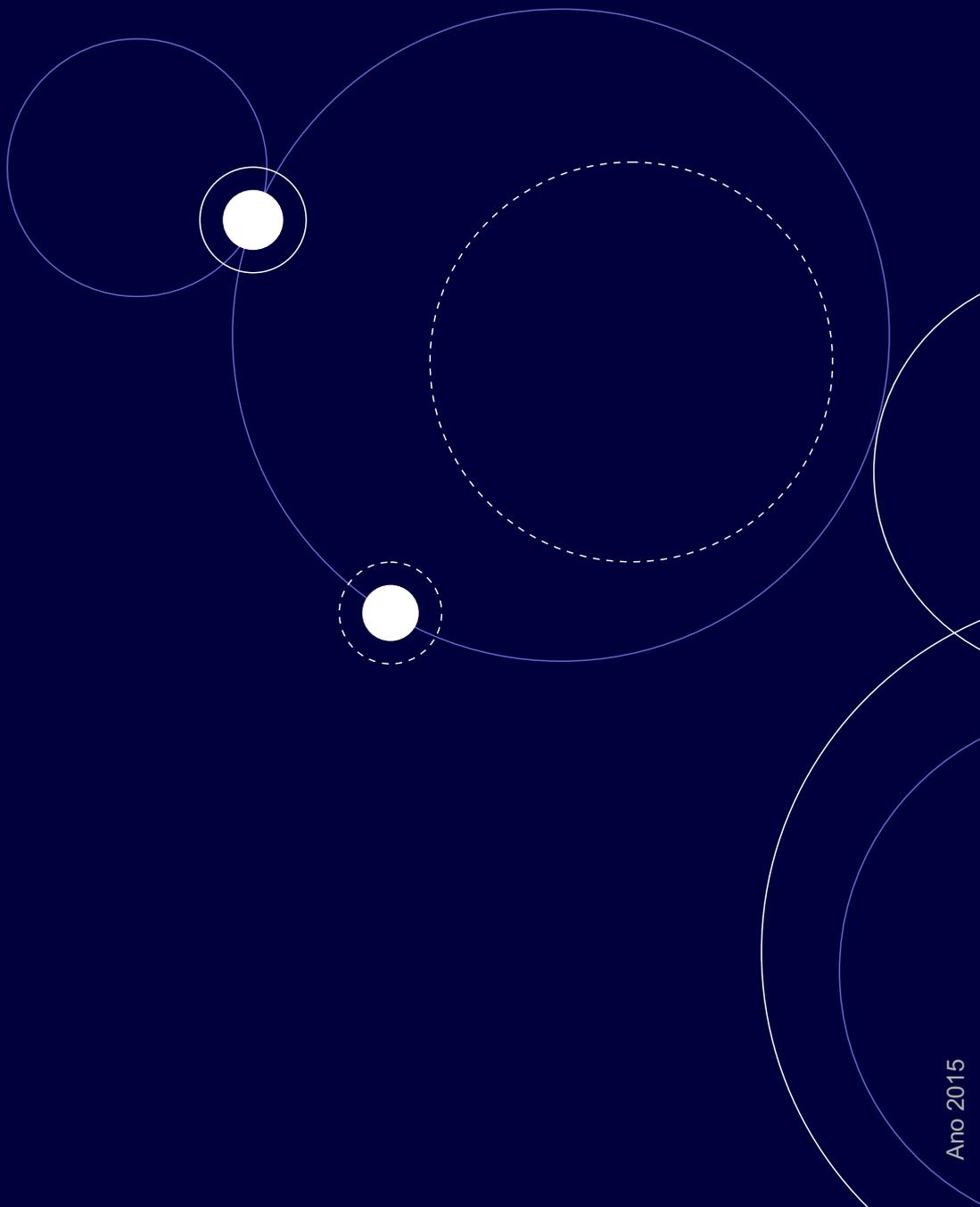


SPUK



Improving business environment through transparency in São Paulo State

Melhoria do ambiente de negócios por meio da transparência no Estado de São Paulo



Ano 2015

ceweb.br nic.br cgi.br

SEADE
Fundação Sistema Estadual
de Análise de Dados

Fundap



Embaixada Britânica
Brasília



Este material está sob uma licença Creative Commons.
Atribuição-SemDerivações-SemDerivados
CC BY-NC-ND