

Dados Abertos

Caderno 1

Definições de Arquitetura

SAS/GDAS

Janeiro/2020

Prefácio

Um dos segmentos de tecnologia em maior efervescência na atualidade é o da *accountability* da administração pública. Em todo o mundo os governos estão se comprometendo com a iniciativa *Open Government Partnership* que visa garantir compromissos concretos de governos nacionais e subnacionais para promover um governo aberto, capacitar cidadãos, combater a corrupção e aproveitar novas tecnologias para fortalecer a governança. O Brasil, um dos oito países cofundadores da iniciativa, é exemplo para a América Latina e Caribe, pela implementação de políticas de governo aberto.

A inegável força propulsora exercida pela abertura de dados governamentais invade o cenário da participação da sociedade em seu governo e transforma o detalhamento do funcionamento da gestão da administração pública em uma transparência inédita em relação ao desempenho do governo na entrega de políticas e serviços à sociedade.

Em torno dessa rica e potencialmente útil massa de dados e informações gravitam diversos atores, tais como as próprias instituições públicas, empreendedores, organizações da sociedade civil, ativistas engajados nas causas sociais, a mídia e o próprio cidadão. Um conjunto interligado de princípios, políticas, padrões e tecnologias estruturam as relações entre esses atores, formando uma espécie de ecossistema que ainda carece de maturidade e que ainda busca modelos para geração de valor e indução de avanços sociais com sustentabilidade.

Recentemente publiquei o livro *Dados Abertos – Transparência Governamental: perspectivas, cenários e planejamento* objetivando tornar a pesquisa do assunto mais palatável e amigável para interessados no tema.

Entretanto, como o tema apresenta elevada complexidade e baixa disponibilidade de material para consulta e atualização decidi detalhar o livro em 5 textos que denominei Cadernos de tecnologia sobre o tema Dados Abertos. Cadernos numerados de 1 a 5. Cada Caderno trata de um grande tema do universo dos dados abertos. Em cada Caderno apresento, explico e exemplifico conteúdos que deverão dar tranquilidade ao leitor para se aprofundar nesse novo caminho.

No Caderno 1 trato dos assuntos vinculados às questões de arquitetura estrutural para o processo de abertura de dados. No Caderno 2 apresento e analiso exemplos de Planos de Dados Abertos, PDAs, para que o leitor tome conhecimento de casos reais selecionados na Internet e que podem elucidar o leitor sobre como elaborar o PDA de sua instituição. O Caderno 3 é um roteiro de elaboração de PDA ampliado com exemplos. O Caderno 4 é um roteiro de elaboração de Planos de Ação de abertura de dados. No Caderno 5 a geração de datasets para publicação é analisada e enriquecida com o vocabulário controlado do Portal de Dados Abertos da Prefeitura de Belo Horizonte.

Indico o livro e os Cadernos para profissionais de Ciência da Informação, da administração pública, de Ciências Sociais, ONGs, institutos de pesquisa, mídia e demais profissionais que almejam desempenhar funções ou interagir com o seguimento dos dados abertos governamentais.

Boa Leitura!

– *Marcão - Marcus Vinicius Pinto*
Belo Horizonte-Minas Gerais-Brasil

Sumário

1	<u>PLANEJAMENTO DA ABERTURA DE DADOS</u>	6
2	<u>PARÂMETROS INICIAIS</u>	6
3	<u>AUTORIZAÇÃO PARA PUBLICAÇÃO</u>	8
4	<u>ATRIBUIÇÃO DE LICENÇA DE USO</u>	9
5	<u>ELABORAR OS METADADOS</u>	12
6	<u>PRIORIZAR DADOS A SEREM ABERTOS</u>	14
7	<u>SUSTENTAÇÃO: UM DESAFIO DE LONGO PRAZO</u>	16
8	<u>GOVERNANÇA</u>	17
9	<u>USABILIDADE DO PORTAL E DE SEU CONTEÚDO</u>	18
10	<u>PROTOCOLO DE ACESSO</u>	21
11	<u>COBERTURA DOS CONJUNTOS DE DADOS</u>	21
12	<u>CONSIDERAÇÕES FINAIS</u>	23
13	<u>GLOSSÁRIO</u>	24
14	<u>PERGUNTAS FREQUENTES</u>	29

14.1	O QUE SÃO DADOS ABERTOS?	29
14.2	POR QUE ABRIR OS DADOS?	30
14.3	COMO EU POSSO COMPARTILHAR MEUS APLICATIVOS E VISUALIZAÇÕES?	31
14.4	O QUE SÃO METADADOS?	31
14.5	QUAL A DIFERENÇA DOS PORTAIS DE DADOS ABERTOS PARA OS PORTAIS DE TRANSPARÊNCIA?	31
14.6	O QUE SÃO CONJUNTOS DE DADOS E RECURSOS?	32
14.7	O PORTAL DE DADOS ABERTOS TEM ALGUM DADO PESSOAL OU SIGILOSO?	32
14.8	QUEM É RESPONSÁVEL PELOS DADOS CATALOGADOS NO PORTAL?	33
14.9	COMO ABRIR ARQUIVOS CSV?	35
14.10	COMO ABRIR ARQUIVOS JSON?	36
14.11	COMO ABRIR ARQUIVOS XML?	37
14.12	O QUE É INFRAESTRUTURA NACIONAL DE DADOS ABERTOS - INDA?	37
14.13	O QUE SÃO DADOS ABERTOS GOVERNAMENTAIS?	38
14.14	O QUE SÃO LICENÇAS? COMO POSSO DISPONIBILIZAR OS MEUS DADOS COM UMA LICENÇA ABERTA?	38
14.15	A MINHA INSTITUIÇÃO JÁ DISPONIBILIZA OS DADOS GRATUITAMENTE NA INTERNET. POSSO DIZER QUE SÃO DADOS ABERTOS?	38
14.16	O QUE SE PODE FAZER COM OS DADOS ABERTOS?	39
14.17	QUE BONS EXEMPLOS EXISTEM DE REUTILIZAÇÃO DE DADOS ABERTOS?	39
14.18	POR QUE EXISTE TANTA INFORMAÇÃO FECHADA?	40
14.19	QUE TIPO DE APLICAÇÕES DIGITAIS PODEM SER FEITAS COM ESSES DADOS?	41
16	<u>BIBLIOGRAFIA</u>	<u>42</u>
17	<u>ANEXO 1 – RECOMENDAÇÕES DE FORMATOS PARA DADOS ABERTOS</u>	<u>43</u>

1 Planejamento da abertura de dados

Preparar os dados para abertura implica em identificá-los, formatá-los e obter permissão dos patrocinadores no órgão para a publicação. Em teoria a instituição da administração pública deve publicar o conjunto total de dados sob sua guarda o mais rápido possível. Entretanto, na prática, com as limitações financeiras e técnicas presentes em toda a estrutura governamental os órgãos detentores dos dados não estão organizados para as ações de publicação dos mesmos nos moldes preconizados pelo Governo Federal e pelas instituições internacionais de governo aberto.

Este caderno objetiva apresentar ao leitor os conhecimentos necessários à preparação para a abertura de dados. Ao final estão algumas perguntas frequentes dos leitores.

2 Parâmetros iniciais

Sugere-se como boa prática que a instituição publique as informações conforme dispõe a Lei de Acesso à Informação¹ - LAI - em seu Art. 8º, §1º e informações adicionais de interesse público. Um bom conjunto de dados a ser publicado seria o seguinte:

- Dados sobre desempenho de programas, projetos e ações finalísticas do órgão publicador.
- Endereços, informações para contato, horários de funcionamento e estrutura organizacional do órgão publicador.
- Dados referentes ao relacionamento com o cidadão.
- Dados referentes a licitações, editais e contratos.
- Registro de repasses, transferências de recursos financeiros e despesas.

¹ Lei 12.527 de 18 de novembro de 2011.

A publicação dos dados no Portal de Dados Abertos necessita dos seguintes elementos de tecnologia e definições:

- *Dataset*². Conjunto de dados. É o arquivo gerado a partir da base de dados original do sistema de informação. Precisa seguir normas e padrões de dados abertos para que sua publicação seja aceita no Portal.
- Dicionário de dados. O dicionário de dados corresponde às informações que traduzem cada campo do *dataset* e informa suas propriedades.
- Metadados. Os metadados são informações que detalham o *dataset* e são necessárias à sua carga no CKAN.
- Área temática. Informação que organiza os *datasets* em temas de acordo com a lógica de organização dos serviços.
- Órgão. Informação que identifica o órgão da estrutura administrativa da instituição publicadora responsável pelo *dataset*.

Para que os dados de um órgão sejam abertos em sua totalidade é essencial que o trabalho seja coordenado por um profissional que tenha a função de Referencial Técnico do projeto com total conhecimento estrutural e operacional dos dados produzidos e custodiados pelo órgão. Esse profissional deve analisar os processos de negócio, identificar todos os grupos de dados produzidos nesses processos e gerá-los no melhor formato para que sejam acessados no Portal de Dados Abertos.

² Um conjunto de dados ou "*dataset*" é uma coleção de dados normalmente tabulados. Por cada elemento se indicam várias características. Cada coluna representa uma variável particular. Cada linha corresponde a um determinado membro do conjunto de dados em questão. Cada valor é conhecido como um dado. O conjunto de dados pode incluir dados para um ou mais membros, correspondentes ao número de linhas.

3 Autorização para publicação

É importante que o órgão tenha segurança em relação à liberação do conjunto de dados. Para tanto é necessário considerar as seguintes recomendações:

- Deve-se fazer uma análise jurídica para se ter uma permissão de publicação formal por parte da instância responsável.
- Deve-se ter certeza de que não há instrumentos jurídicos, portarias, instruções normativas ou afins que proíbam a publicação de certas informações presentes no conjunto de dados em questão.
- Deve-se avaliar se a publicação desses dados pode comprometer algum dos itens a seguir em seu estado natural: segurança nacional, segredos estratégicos de governo, direito de privacidade do cidadão ou legislação.
- Deve-se submeter o Plano de Dados Abertos³ ao Gestor de Dados Abertos da instituição para que o mesmo seja analisado e tenha autorização para sua publicação e execução.

Caso a publicação do conjunto de dados seja impedida por alguma razão será necessário que se façam as adequações necessárias para que seja possível publicar os dados. Uma adequação relativamente corriqueira é a anonimização, na qual são retirados os dados motivadores da restrição à publicação. Observe-se que a anonimização não é apenas um mascaramento do dado.

Tomando como exemplo o caso dos dados de uma pessoa em uma determinada base de dados a anonimização não se dá apenas alterando o nome da pessoa para algo genérico, pois outros dados da mesma pessoa podem permitir o cruzamento de dados ou a filtragem do dado resultando em sua identificação.

Este processo precisa seguir protocolos reconhecidamente bem-sucedidos pelas equipes de Tecnologia da Informação para não se correr o risco do dado ser desanonimizado. Entretanto, como tudo tem seu custo, é necessário se observar

³ Artefato que organiza o planejamento da abertura de dados.

que quanto mais anônimo o dado maior será a redução do valor agregado ao mesmo, podendo se tornar irrelevante ao ponto de restringir a transparência da instituição. Outra restrição que pode impedir a publicação de um conjunto de informação é a sua caracterização como dado sensível.

Apesar destas exceções mantém-se a orientação de que todos os dados da instituição devem ser publicados. Cabe notar ainda que, mesmo não publicado integralmente, um grupo de dados sempre agrega algum valor para a sociedade.

4 Atribuição de licença de uso

Tem-se como premissa para dados abertos a definição de que os dados publicados estão liberados para qualquer finalidade de uso. Entretanto, objetivando a melhor utilização dos dados as organizações internacionais responsáveis pelos padrões para dados abertos propõem atribuir licenças de acordo com o assunto e as condições de uso dos dados. Essa atribuição implica disponibilizar as informações de forma completa na Web, em formatos não proprietários, com acesso irrestrito e livre para qualquer cidadão e de modo que qualquer um possa reutilizá-los para qualquer fim, sem restrições. 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.

Assim, a atribuição da "licença aberta" é importante para que se proteja o que tem que ser protegido e para que a Lei de Acesso à Informação seja respeitada. A Infraestrutura Nacional de Dados Abertos - INDA, do governo federal, ainda não concluiu se o ordenamento jurídico brasileiro é adequado 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.

Pode-se citar como exemplo, no caso da legislação brasileira, a Lei de Direitos Autorais. A lei prevê a proteção a banco de dados que constituam "criação intelectual". Entretanto, por seu turno, a Lei de Acesso à Informação – LAI - afirma que o governo deve proporcionar o acesso a toda informação que produz ou detém, desde que não seja considerada sigilosa. A interação conflitante dessas leis ainda é objeto de estudo.

A LAI não trata da determinação de qual licença será aplicada ao se publicar dados abertos, mas lista princípios que tendem a direcionar a determinação de uma licença aberta ou termos de utilização compatíveis com os conceitos de abertura de dados. Este direcionamento tem como premissa que os dados governamentais são públicos, configurando o sigilo como exceção.

A compatibilidade de licenças entre as diversos instâncias de governo é outra questão que não pode ser desconsiderada ao se atribuir um licenciamento a um conjunto de dados. Pois, por exemplo, se o governo estadual usa uma licença incompatível com a licença usada por uma de suas prefeituras para *datasets* de conteúdo similar poderá haver conflitos legais em seu uso por parte da sociedade.

A materialização desta hipótese poderia ser o caso em que não seja possível, legalmente, cruzar as bases de dados de escolas, por exemplo, para construir um aplicativo que exhibe as instalações municipais e estaduais. Uma licença ou termos de uso de dados abertos devem ser atribuída 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 governo.

No Brasil, a maior parte dos órgãos públicos não atribui qualquer licença ao publicar dados na Internet. Há exceções, contudo, com a utilização das licenças ODbLe DbCL.

As licenças recomendadas são as seguintes (Brasil, 2012d):

- *Open Database License (ODbL)*⁴: Qualquer pessoa é livre para copiar, distribuir, usar, criar novas publicações a partir da abertura original, modificar, construir ou transformar os dados (ou o banco de dados). Usuários têm que atribuir a fonte, da forma que for definida na licença e garantir que reusos de seus trabalhos com os dados sigam a mesma regra de atribuição. Nos casos de usos dos dados para fins comerciais, deve ser disponibilizada cópia pública.
- *Database Content License (DbCL)*: mesmas orientações da ODbL, mas aplicável ao conteúdo dos bancos de dados - os dados em si;
- *Creative Commons Zero (CC0)*⁵: aplicável a contextos em que o conjunto de dados se caracteriza como de domínio público. Usada quando a análise do gestor aponta para a inexistência de direito autoral e passa a ser “natural” a renúncia permanente aos direitos de copyright e outros relacionados. Nenhuma atribuição à instituição gestora dos dados é esperada por parte dos usuários.
- *Creative Commons Attribution 4.0*⁶: permite ao usuário (qualquer usuário) realizar livremente atividades de compartilhamento (cópia e redistribuição dos dados em qualquer formato ou mídia), de adaptação (incluindo remixagem, transformação e construção de novos produtos) para quaisquer propósitos inclusive comerciais. Ficam os usuários sujeitos às seguintes condições:
 - Atribuir crédito apropriado, fornecendo URL da licença e indicando eventuais mudanças;
 - Ao atribuir crédito não é necessário seguir um formato padrão, mas é exigido que não se leve o leitor a supor que a instituição que

⁴ Conteúdo integral da licença disponível em <https://opendatacommons.org/licenses/odbl/>

⁵ Conteúdo integral da licença disponível em <https://creativecommons.org/share-your-work/public-domain/cc0/>.

⁶ Conteúdo integral da licença disponível em <https://creativecommons.org/licenses/by/4.0/>

licenciou o conteúdo endossa o usuário ou os trabalhos produzidos com os dados licenciados.

É essencial que os cidadãos interessados em usar dados do Portal de Dados Abertos estejam cientes de que as condições que regem cada conjunto de dados podem ser atualizadas, corrigidas e/ou substituídas a qualquer tempo, a critério da instituição publicadora.

5 Elaborar os metadados

A abertura de dados tem por objetivo permitir o acesso aos dados da administração pública por parte da sociedade. Este objetivo considera que as instituições devem publicar toda informação sobre sua tutela e que esta publicação seja a mais automatizada possível.

No caso em que um conjunto de dados publicado não está produzindo conclusões ou não está sendo utilizado para algum serviço, então ele não satisfaz seu objetivo inicial como dado aberto. Uma parte importante desse sucesso é o dicionário de dados, os metadados, que devem ser suficientes para que um usuário externo à instituição publicadora compreenda a definição, a validade e o escopo dos dados em um *dataset*, o intervalo de tempo entre as publicações, a data de publicação, direitos de uso etc.

Propõe-se para o Portal de Dados Abertos, como padrão os conjuntos de metadados obrigatórios e desejáveis listados nas tabelas 1 e 2.

	Nome do Metadado	Descrição
1	Título	Nome do conjunto de dados.
2	Descrição	Breve explicação sobre os dados.
3	Etiquetas	Palavras-chave relacionadas ao conjunto de dados que são úteis na classificação e busca dele. Ex.: economia, aposentadoria, ENEM.).
4	Tipo de Licença	Ver detalhamento no Anexo 1.
5	Catálogo origem	Página (URL) do órgão onde está publicado o

	Nome do Metadado	Descrição
		conjunto de dados.
6	Órgão responsável	Nome e sigla do órgão ou entidade responsável pela publicação do conjunto de dados.
7	Identificador	URL persistente que aponta para o recurso na Web.
8	Nome do arquivo	Nome do arquivo do <i>dataset</i> .
9	Formato	Formato do <i>dataset</i> . Ex.: XML, JSON, CSV etc.
10	Data de Publicação	Data de publicação no Portal.
11	Cobertura temporal	Data ou período à que referem os dados do <i>dataset</i> . Ex.: 09/2017.
12	Nome do mantenedor	Nome do Mantenedor dos dados do <i>dataset</i> .
13	E-mail do mantenedor	E-mail do Mantenedor dos dados do <i>dataset</i> .
14	Versão do arquivo	Número da versão do arquivo do <i>dataset</i> .
15	Frequência de atualização	Frequência temporal com que o recurso é atualizado.
16	Melhor horário	Horário menos sobrecarregado para extração no do sistema de origem (apenas para publicações de <i>Datasets</i> de forma automática): Ex.: 23:00 (vinte e três horas).
17	Cobertura geográfica	Localização ou região geográfica a que se refere o <i>dataset</i> . Caso esta informação se aplique. Ex.: Barreiro.
18	Granularidade geográfica	Precisão geográfica da cobertura geográfica. Caso esta informação se aplique. Ex.: regional.
19	Granularidade temporal	Precisão temporal da cobertura temporal. Caso esta informação se aplique. Ex.: mês.

Tabela 1 - Metadados obrigatórios.

	Nome do Metadado	Descrição
	Documentação	URL de documento que expõe detalhes sobre o conjunto de dados.
	Referências	Relações com outros conjuntos de dados.
	Metodologia	Material com explicação sobre o processo de criação dos <i>datasets</i> .

Tabela 2 - Metadados desejáveis.

O vocabulário utilizado deve ser submetido ao *checklist* proposto pelo W3C (W3C, 2014) e deve seguir vocabulários controlados e ontologias que sejam de conhecimento público. Essa iniciativa deve ser parte das diretrizes da instituição publicadora para composição dos metadados dos dados a serem abertos objetivando facilitar seu entendimento, evitar estrangeirismos, regionalismos e neologismos, além de maximizar a possível conexão com outros conjuntos de dados.

Recomenda-se utilizar o VCGE – Vocabulário Controlado do Governo Eletrônico – (Brasil, 2014b) devido à qualidade de sua estrutura e a ser o vocabulário padrão para dados governamentais no Brasil.

6 Priorizar dados a serem abertos

Os defensores do conceito de dados abertos vislumbram o cenário em que todos os dados de uma instituição são publicados o mais rápido possível e mantidos sempre atualizados. Entretanto, a realidade impõe adequações face ao que é de fato possível. Assim, é uma boa prática analisar o potencial de valor dos dados para a sociedade e elaborar um planejamento de publicação priorizando os mais importantes.

Dentre as boas práticas para seleção de dados a serem abertos temos as seguintes orientações:

- Dados catalogados ou criados prioritariamente pela instituição publicadora na estrutura administrativa da sua unidade federativa e que podem ser liberados ao público. Ex.: dados da arrecadação tributária do município.
- Dados que ao serem combinados com outros dados produzam grande valor. Ex.: dados de evasão escolar combinados com dados de adoecimento em determinada região do município.
- Dados que tenham potencial de reuso. Ex.: dados do fluxo de trânsito nas vias preferenciais.

- Dados com potencial para compor séries históricas. Ex.: quantificação de serviços prestados.
- Dados muito requisitados através dos meios de solicitação de acesso às informações públicas pela sociedade. Ex.: número de vagas por tipo de ensino por escola.
- Dados de interesse público frequentemente entregues à imprensa. Ex.: dados de contaminação por dengue.
- Dados não sigilosos regularmente solicitadas por outras instituições. Ex.: dados relativos à qualidade de vida.
- Dados anonimizados que continuam tendo valor para a sociedade. Ex.: quantificações referentes ao atendimento da população em vulnerabilidade social.
- Dados com grau de saneamento que não comprometam o conjunto de dados publicado. Ex.: indicadores de desempenho dos alunos da rede municipal.
- Dados de interesse público em mídia que esteja fora dos padrões de dados abertos, mas que podem ser facilmente convertidas e liberadas. Ex.: mapeamento georreferenciado de infestação de mosquitos transmissores da dengue.
- Dados estáveis com baixo grau de correção após sua publicação para a sociedade. Ex.: série histórica do desempenho da rede hoteleira no município.
- Dados que podem compor rotinas automatizadas. Ex.: dados referentes à posição atual dos ônibus municipais.

Cabe lembrar que após sua publicação o conjunto de dados não deve ter seu conteúdo alterado sem ampla divulgação. A alteração pode provocar resultados inesperados e indesejáveis em sua cadeia de utilização.

A clareza do papel de cada pessoa atuante na abertura de dados potencializa sua melhor participação. Os papéis são os seguintes:

1. Cidadão. É o usuário dos dados disponibilizados e das aplicações. O planejamento de abertura de dados deve conter ações informativas do processo de abertura como um todo, esclarecendo a priorização adotada para a seleção dos dados a serem publicados. Também deve

incentivar a utilização das aplicações que utilizam os dados disponibilizados.

2. Grupos organizados. É uma boa prática incluir esses grupos nas etapas da abertura dos dados, pois podem contribuir para uma melhor elicitação dos requisitos técnicos e formas de acesso aos dados. Eles têm a possibilidade de atuar como entidades facilitadoras do acesso à informação presente nos dados abertos.
3. Acadêmicos. Universidades e centros de pesquisa têm capacidade técnica e são parceiros importantes no aprimoramento de práticas e de aplicativos baseados nos dados abertos. O Governo pode se beneficiar das análises multidisciplinares realizadas por especialistas deste segmento.
4. Gestores da administração pública. O público-alvo principal dos dados abertos é o cidadão e suas formas de organização. Entretanto, a abertura de dados das instituições do governo constitui uma iniciativa de interoperabilidade e pode ser utilizada como ferramenta de gestão. Os dados abertos reduzem o trabalho de geração de informação de instituição para instituição e aumentam a utilização de dados publicados, pois o dado publicado pensado em uma instituição pode ser utilizado por toda a máquina pública.
5. Imprensa. A imprensa tem o potencial de aumentar a visualização dos dados pela sociedade, pois já é uma fonte de informação utilizada por ela.

7 Sustentação: um desafio de longo prazo

O maior desafio das instituições públicas que integram o grupo de publicadores de dados abertos é tornar a adesão em um projeto contínuo, sustentado e com manutenção atualizada dos dados. Assim, é recomendável se pensar em um plano de sustentação no início da ação de abrir os dados, pois caso não seja possível mantê-los atualizados, com séries históricas íntegras e confiáveis, a iniciativa perde seu propósito.

O plano deve então considerar dois aspectos. O primeiro aspecto é a continuidade da abertura. Já no início do processo de abertura de dados identifica-

se a necessidade de priorizar os dados que serão abertos face à quantidade finita de recursos. Entretanto, a partir do momento que se considera a abertura como ação imposta por leis é salutar incorporar o valor estratégico desta ação para a instituição inserindo o processo de geração e publicação de dados abertos no conjunto de procedimentos de persistência de dados nos sistemas de informação origem. Nesta ação deve-se prever melhorias na geração e publicação dos dados abertos através da troca de experiências com outras instituições da administração pública e a satisfação dos usuários com o acervo publicado.

O segundo aspecto trata da manutenção dos dados já disponibilizados. Sempre que uma instituição publica um conjunto de dados em formato aberto ela assume o compromisso de mantê-lo disponível e íntegro. Considerando que a maioria dos dados têm perfil de séries históricas, gerados ininterruptamente, o plano de sustentação deve prever a manutenção da série com publicações em intervalos de tempo regulares, com versões mais recentes sempre que possível e preservando sempre seu valor agregado. A frequência de atualização e seus custos operacionais devem ser considerados quando da definição da estrutura da série temporal.

8 Governança

O termo governança, neste texto, é o adotado pelo Comitê de Governança da Internet (KURBALIJA 2016): a governança dos assuntos de qualquer instituição, incluindo instituições não governamentais. Situando a governança no cenário específico dos dados abertos o Comitê Gestor para dados abertos na instituição deve ter como finalidade a gestão da política de dados abertos na mesma. Sua finalidade e composição devem ser definidas em regimento interno a ser definido pela instituição e deve ter representatividade nos diversos setores produtores de dados com potencial de serem abertos na instituição.

A título de referência sugere-se consultar a Instrução Normativa SLTI nº 4/2012⁷, que institui o comitê Gestor da Infraestrutura Nacional de Dados Abertos - INDA.

Os princípios da administração pública impregnam os eixos sustentadores do Governo Aberto potencializando a importância dos dados abertos nesse cenário. Os eixos sustentadores são:

- **Transparência.** Trata da promoção da responsabilidade de informar os cidadãos sobre o que o governo está fazendo e que ações pretende tomar.
- **Prestação de serviços.** Aprimora a eficácia do Estado, encorajando a cooperação entre a sociedade, os diferentes níveis de governo e a iniciativa privada.
- **Participação.** Este é o eixo que apresenta maior relação com os dados abertos. Permite aos cidadãos contribuir com suas ideias e competências, auxiliando o poder público a elaborar políticas eficazes e abrangentes, graças à informação amplamente dispersa em nossa sociedade.

Considerando-se os princípios e os eixos sustentadores é necessário que se tenha instâncias de controle do processo de abertura de modo a garantir que eles estão sendo cumpridos da melhor forma possível. Uma proposta já institucionalizada é a figura do Comitê Gestor.

9 Usabilidade do Portal e de seu conteúdo

A usabilidade do Portal de Dados Abertos de uma instituição se traduz no quanto é fácil e intuitivo o uso deste portal como interface de acesso aos conjuntos

⁷ Disponível em <http://wiki.dados.gov.br/IN-SLTI-MP-04-INDA.ashx>. Acesso em 16/12/2019.

de dados. Esta usabilidade é fator determinante no sucesso da iniciativa de abrir os dados e deve minimizar impactos negativos de uma infraestrutura tecnológica limitada e da necessidade de se publicar dados em quantidade e com a maior qualidade possível.

O ponto inicial do projeto do Portal é a especificação, com a máxima precisão, de quais são os recursos disponíveis para a plataforma dos dados abertos da instituição e qual o volume de dados a ser disponibilizado. Tais informações farão parte das variáveis que definirão a arquitetura de dados abertos da instituição, tais como:

- O escopo da hospedagem dos dados.
- As ferramentas de acesso.
- O espaço em servidores de dados.
- O formato dos *datasets*.
- A disponibilidade dos servidores de aplicação.
- A banda de rede.
- A estabilidade dos sistemas de informação origem.

Os propositores das teorias de dados abertos não elencam os formatos permitidos ou proibidos, mas determinam que, para um dado ser considerado como dado aberto, o conjunto de dados deve estar disponível em um formato de especificação aberta, não proprietário, e estruturado, ou seja, que possibilite seu uso irrestrito e automatizado através da *Web*. Cabe observar que a publicação de séries temporais e muitos downloads simultâneos, no caso de arquivos muito grandes, podem causar sobrecarga nos servidores.

É imprescindível que seja utilizado um formato amplamente conhecido. Recomenda-se, como parte das boas práticas, alguns formatos não proprietários listados na Cartilha Técnica para Publicação de Dados Abertos no Brasil (Brasil, 2012c): JSON, XML, CSV, ODS, RDF, SVG, GML, HTML/RDFa e GTFS. Estes formatos estão detalhados no anexo 1.

A instituição, ao escolher um padrão de publicação, deve analisar sua capacidade tecnológica atual e as limitações advindas das perspectivas futuras. A tendência, neste tipo de plataforma, é de crescimento constante na ocupação do espaço em disco e tráfego de rede intenso.

Tem-se como recomendação e uma boa prática não se publicar conjuntos de dados compactados. Entretanto, em casos de muitos arquivos ou de arquivos com grande capacidade de compressão recomenda-se a utilização de formatos abertos de compactação como o 7Z, TAR/GZIP ou ZIP.

Quanto às responsabilidades advindas do acesso para reuso dos dados, disponíveis no portal de dados abertos de uma instituição da administração pública, os usuários devem observar que (Brasil, 2012d):

- A geração de serviços secundários a partir do acesso ou manipulação dos dados gera para si a responsabilidade integral sobre as informações ou serviços oferecidos. O mesmo se aplica a quaisquer outras atividades baseadas nestes dados e exime a instituição publicadora de toda e qualquer responsabilidade referente às atividades, informações e serviços por ele desenvolvidos.
- Os usuários dos dados do Portal não devem apresentar os dados obtidos de forma incorreta ou enganosa, nem os utilizar para amparar ou sustentar atividades criminosas ou ilícitas ou de favorecimento ostensivo de atividades de igual caráter.
- A instituição publicadora poderá exigir que o usuário cancele a utilização, distribuição e/ou serviço de todo e qualquer conjunto de dados por ele utilizado se for verificada má-fé sem prejuízo da adoção das demais providências legais cabíveis.
- Os usuários dos dados do Portal ao gerarem produtos/serviços com os respectivos dados deverão se responsabilizar pelos conteúdos gerados e prestarem as devidas informações sobre os aplicativos, sistemas, estatísticas ou qualquer informação sobre o que foi produzido.

Por seu turno, a entidade publicadora, enquanto provedora dos dados disponíveis no portal tem limitações naturais, tais como:

- Não garante o fornecimento continuado dos dados nem a sua publicação tempestiva.
- Disponibiliza os dados conforme estão persistidos nos sistemas de informação origem.
- Exime-se de qualquer garantia e obrigação em relação aos dados e à utilização dos mesmos.
- Não se responsabiliza pela utilização feita por pessoas e/ou instituições que transformem e/ou usem direta ou indiretamente os dados e que porventura venham a gerar prejuízos, danos econômicos ou materiais advindos desse uso e/ou transformação dos dados.

10 Protocolo de acesso

Um protocolo de acesso, ou protocolo de rede, é um conjunto de normas que permitem que qualquer computador conectado à internet possa se comunicar com outro também já conectado na rede. A definição do protocolo de acesso, do mesmo modo que a escolha do formato de publicação, deve ser precedida de uma análise cautelosa do cenário tecnológico da instituição, pois o esforço por parte do servidor e a banda de transferência de dados são fatores críticos.

Quando o conjunto de dados é muito grande uma boa saída é gerar um pequeno arquivo acessível on-line no portal e o uso de arquivos compactados. Uma alternativa é o uso de Torrents, baseado no compartilhamento distribuído dos arquivos em que o esforço de acesso ao servidor é distribuído com outras pessoas que já fizeram download desses arquivos.

São exemplos de protocolos: HTTP, FTP, Endpoint SPARQL e Bit Torrent.

11 Cobertura dos conjuntos de dados

Ao se projetar a cobertura geográfica ou temporal de um conjunto de dados deve-se considerar a eficiência vinculada ao uso de um arquivo do tamanho que

será gerado. Caso o uso seja sempre de parte reduzida do todo há que se considerar a opção por arquivos menores compondo o todo.

Um exemplo é a busca por uma escola em uma região de uma cidade. Se o arquivo trata o estado inteiro a pesquisa terá complexidade desnecessária, pois será mais fácil processar a busca em um segmento de arquivo que contenha a região metropolitana desejada, a cidade ou mesmo arquivos segmentados por bairro. Por outro lado, pode ser trabalhoso para a instituição gerar arquivos de bairro para busca de escolas. Ainda mais considerando-se que esta geração dará origem a inúmeros arquivos em série histórica.

Assim, a decisão da cobertura considerada em um *dataset* é uma decisão estratégica e influi diretamente no tamanho dos conjuntos de dados, na complexidade da série histórica, sua organização no portal e na usabilidade por parte do usuário. Cada instituição deve fazer uma análise individualizada e encontrar as melhores dimensões relacionadas à natureza do dado gerado ou manipulado.

Exemplos de dimensões de cobertura:

- Geográfica no país: UF, estado, município, regional, bairro.
- Temporal: anual, trimestral, mensal, semanal, diário.

12 Considerações finais

A disponibilização de dados governamentais abertos tem o potencial de promover a inclusão digital ao fornecer dados em formatos padronizados e abertos que podem ser transformados e colocados à disposição em formato que respeite os padrões de acessibilidade possibilitando a sua utilização por pessoas com deficiência.

Os próximos cadernos referentes ao tema dos dados abertos trarão mais informações para o leitor que esteja participando do processo de abertura de dados de sua instituição. Caso necessite de apoio ou mais informações sobre os assuntos aqui tratados entre em contato com Marcão (Marcus Vinícius Pinto) no e-mail marcao.dick@gmail.com.

13 Glossário

1. Anonimação. Tornar um dado anônimo não sendo possível associá-lo a uma pessoa. Esta prática é usual quando se trata de dado sigiloso.
2. Área temática. Informação que organiza os datasets em temas de acordo com a lógica de organização dos serviços da instituição responsável pelo dataset.
3. Cidadão. É o usuário dos dados disponibilizados no Portal de Dados Abertos e das aplicações.
4. Cobertura. Abrangência geográfica ou temporal de um dado.
5. Comitê Gestor. Grupo de profissionais da instituição publicadora com a missão de administrar a publicação de dados no Portal da instituição.
6. Dados Abertos. São os dados que estão disponíveis ao cidadão, sem custo ou restrições de qualquer ordem proveniente de qualquer origem, seja privado ou público.
7. Dados Governamentais Abertos (DGA). São os dados abertos que instituições governamentais disponibilizam ao cidadão.
8. Dados sigilosos. São os dados submetidos temporariamente à restrição de acesso público em razão de sua imprescindibilidade para a segurança da sociedade e do Estado”.
9. Dataset. é uma coleção de dados normalmente tabulados. Por cada elemento se indicam várias características. Cada coluna representa uma variável particular. Cada linha corresponde a um determinado membro do conjunto de dados em questão. Cada valor é conhecido como um dado. O conjunto de dados pode incluir dados para um ou mais membros, correspondentes ao número de linhas.
10. Dicionário de dados. Acervo que descreve e define o significado de todos os dados que compõem o universo de informação de um sistema e permite fazer a verificação de consistência entre os vários modelos.

11. Direito autoral. Direito de propriedade sobre uma criação.
12. ePing. Concebida como uma estrutura básica para a estratégia de governo eletrônico, aplicada ao governo federal – Poder Executivo, não restringindo a participação, por adesão voluntária, de outros Poderes e esferas de governo.
13. Ferramenta de acesso. Programa ou aplicativo utilizado para acessar os dados em um servidor de dados.
14. Formato. Estrutura do dado que permite que seja compreendido por um programa leitor ou interpretador.
15. Gestor de dados. Profissional da instituição publicadora responsável por estabelecer as diretrizes do processo de abertura na instituição.
16. Governança. A governança dos assuntos de qualquer instituição, incluindo instituições não governamentais.
17. Hospedagem de dados. Estrutura de tecnologia responsável por armazenar dados. Um banco de dados exerce esta função.
18. INDA. A Infraestrutura Nacional de Dados Abertos (INDA) é um conjunto de padrões, tecnologias, procedimentos e mecanismos de controle necessários para atender às condições de disseminação e compartilhamento de dados e informações públicas no modelo de Dados Abertos, em conformidade com o disposto na ePING.
19. LAI. Lei de Acesso à Informação. Lei nº 12.527, de 18 de novembro de 2011, é uma lei ordinária federal que regulamenta o art. 5º, XXXIII, art. 37, §3º, II e art. 216, §2º da Constituição Federal de 1988, que asseguram o direito fundamental de acesso às informações produzidas ou armazenadas por órgãos e entidades da União, Estados, Distrito Federal e Municípios.
20. Lei de Acesso à Informação. Ver LAI.
21. Licença aberta. É uma forma de o titular de direitos de autor (criador ou outro titular de direitos) conceder ao público em geral autorização para utilizar a sua obra.

22. Licença de uso. É o documento que define os limites de uso que um usuário pode ter em relação a um produto de terceiros.
23. Metadado. Os dados, especialmente os dados digitais, assumem muitas formas. As conversas por voz, as mensagens de texto ou as redes sociais comunicam dados. As transações digitais bancárias ou comerciais envolvem a transferência de dados. Conteúdo na web, entretenimento digitalizado e transferido, bancos de dados e repositórios de informações de todos os tipos são exemplos de publicações de dados. Os metadados descrevem o que são esses dados: eles fornecem informações sobre esses dados. Isso é bastante simples. No entanto, se analisarmos isso em detalhe, descobrimos que "descrever" os dados é um exercício técnico rigoroso e um problema carregado de implicações sócio-políticas (PISCITELLO, 2016).
24. Órgão. Integram a estrutura do Estado e das demais pessoas jurídicas como partes desses corpos vivos, dotados de vontade e capazes de exercer direitos e contrair obrigações para a consecução de seus fins institucionais.
25. Origem dos dados. Todo elemento de dado precisa identificar sua origem ou o processo que o gera. Esta identificação é muito importante para que se possa ter informações sobre a fonte geradora do dado. Esta informação deve ser única, ou seja, cada dado deve ter uma e somente uma fonte de origem.
26. Papel. É uma definição precisa e bem delimitada de uma função e responsabilidades a serem desempenhadas por uma ou mais pessoas. Um papel pode ser desempenhado por mais de uma pessoa simultaneamente, por exemplo. um arquiteto (papel) pode ser interpretado por um grupo de desenvolvedores com habilidades complementares. Para desempenhar um determinado papel de maneira adequada o desenvolvedor deve possuir as habilidades necessárias à sua interpretação.
27. Participação. Um dos três eixos sustentadores do Governo Aberto. Permite aos cidadãos contribuir com suas ideias e competências, auxiliando o poder público a elaborar políticas eficazes e abrangentes, graças à informação amplamente dispersa em nossa sociedade.

- 28.PDA. Plano de Dados Abertos. É o documento orientador para as ações de implementação e promoção de abertura de dados, inclusive geoespacializados obedecendo a padrões mínimos de qualidade, de forma a facilitar o entendimento e a reutilização das informações da instituição da administração pública. Sua elaboração vem ao encontro do disposto na Lei de Acesso à Informação (LAI), na Instrução Normativa SLTI nº 4, de 13 de abril de 2012 (que institui a Infraestrutura Nacional de Dados Abertos), o Decreto Presidencial nº 6.666, de 27 de novembro de 2008 (que institui a Infraestrutura Nacional de Dados Espaciais), bem como dos compromissos assumidos pelo Brasil no âmbito do 2º Plano de Ação Nacional sobre Governo Aberto, entre outros normativos que abordam o tema de transparência.
- 29.Planejamento. Fase do ciclo de vida de um projeto composta por processos para planejar e administrar com sucesso um projeto; para isto, desenvolvem o Plano de gestão, que inclui principalmente o alcance do projeto, seus custos, as atividades a serem desenvolvidas, os recursos humanos e materiais necessários, as comunicações necessárias, os riscos a serem enfrentados, a qualidade a ser garantida e as necessidades de aquisições e compras.
- 30.Plano de Dados Abertos. Ver PDA.
- 31.Prestação de serviço. Um dos três eixos sustentadores do Governo Aberto. Aprimora a eficácia do Estado, encorajando a cooperação entre a sociedade, os diferentes níveis de governo e a iniciativa privada.
- 32.Produto. Um artigo produzido que é quantificável e que pode ser um elemento terminado ou um componente.
- 33.Protocolo de acesso.
- 34.Recursos: Recursos humanos especializados, equipes, serviços, fornecimentos, matérias primas, materiais, orçamentos ou recursos necessários para realizar as atividades do cronograma.

35. Servidor de aplicação. É um computador que disponibiliza um ambiente para a instalação e execução de certas aplicações, centralizando e dispensando a instalação nos computadores clientes.
36. Sistema de informação. É a expressão utilizada para descrever um Sistema seja ele automatizado (que pode ser denominado como Sistema Informacional Computadorizado), seja manual, que abrange pessoas, máquinas e/ou métodos organizados para coletar, armazenar, processar, transmitir e disseminar dados que representam informação para o usuário e/ou cliente.
37. Torrent. Maneira de compartilhar arquivos entre usuários, sem que o arquivo precise estar em um servidor (computador central de alto desempenho).
38. Transparência. Um dos três eixos sustentadores do Governo Aberto. Trata da promoção da responsabilidade de informar os cidadãos sobre o que o governo está fazendo e que ações pretende tomar.
39. Usabilidade. Termo usado para definir a facilidade com que as pessoas podem empregar uma ferramenta ou objeto a fim de realizar uma tarefa específica e importante.

14 Perguntas frequentes

14.1 O que são Dados Abertos?

Dados Abertos são dados que estão livremente disponíveis para todos utilizarem e redistribuírem como desejarem, sem restrição de licenças, patentes ou mecanismos de controle. Na prática, a filosofia de dados abertos estipula algumas restrições tecnológicas para que os dados sejam legíveis por máquina.

Todo dado público tem vocação para ser dado aberto. Como praticamente todo dado governamental é público, é fundamental que os governos implementem políticas para disponibilizá-los. Para saber mais, conheça as leis e os princípios dos dados abertos

Segundo a definição da Open Knowledge Internacional, dados são abertos quando qualquer pessoa pode livremente acessá-los, utilizá-los, modificá-los e compartilhá-los para qualquer finalidade, estando sujeito a, no máximo, a exigências que visem preservar sua proveniência e sua abertura. Isso geralmente é satisfeito pela publicação dos dados em formato aberto e sob uma licença de uso aberta. Os dados abertos também são pautados por três leis e oito princípios.

O especialista em políticas públicas e ativista dos dados abertos David Eaves propôs as seguintes “leis”:

1. Se o dado não pode ser encontrado e indexado na Web ele não existe;
2. Se não estiver aberto e disponível em formato compreensível por máquina ele não pode ser reaproveitado; e
3. Se algum dispositivo legal não permitir sua replicação ele não é útil.

As leis foram propostas para os Dados Abertos Governamentais, mas pode-se dizer que elas se aplicam aos Dados Abertos de forma geral. Em 2007 um grupo de trabalho de 30 pessoas reuniu-se na Califórnia, Estados Unidos da América, para definir os princípios dos Dados Abertos Governamentais.

Chegaram a um consenso sobre os 8 princípios para o dado ser considerado aberto:

1. Completos. Todos os dados públicos são disponibilizados. Dados são informações eletronicamente gravadas, incluindo, mas não se limitando a, documentos, bancos de dados, transcrições e gravações audiovisuais. Dados públicos são dados que não estão sujeitos a limitações válidas de privacidade, segurança ou controle de acesso, reguladas por estatutos.
2. Primários. Os dados são publicados na forma coletada na fonte, com a mais fina granularidade possível, e não de forma agregada ou transformada.
3. Atuais. Os dados são disponibilizados o quão rapidamente seja necessário para preservar o seu valor.
4. Acessíveis. Os dados são disponibilizados para o público mais amplo possível e para os propósitos mais variados possíveis.
5. Processáveis por máquina. Os dados são razoavelmente estruturados para possibilitar o seu processamento automatizado.
6. Acesso não discriminatório. Os dados estão disponíveis a todos, sem que seja necessária identificação ou registro.
7. Formatos não proprietários. Os dados estão disponíveis em um formato sobre o qual nenhum ente tenha controle exclusivo.
8. Livres de licenças. Os dados não estão sujeitos a regulações de direitos autorais, marcas, patentes ou segredo industrial. Restrições razoáveis de privacidade, segurança e controle de acesso podem ser permitidas na forma regulada por estatutos.

14.2 Por que abrir os dados?

A publicação “5 Motivos Para a Abertura de Dados na Administração Pública” (TCU, 2015), elaborada pelo Tribunal de Contas da União, apresenta razões para que as organizações públicas invistam em iniciativas de abertura de dados governamentais. Os cinco motivos para a abertura dos dados são:

1. Transparência na gestão pública;
2. Contribuição da sociedade com serviços inovadores ao cidadão;

3. Aprimoramento na qualidade dos dados governamentais;
4. Viabilização de novos negócios;
5. Obrigatoriedade por lei.

14.3 Como eu posso compartilhar meus aplicativos e visualizações?

A Internet é o melhor ambiente para o compartilhamento de conteúdo digital. A forma mais rápida para você compartilhar é publicando em um blog ou qualquer outro ambiente acessível publicamente na Internet, e assim espalhar o endereço, a URL, do seu aplicativo. Além disso, tem-se a intenção de ordenar e democratizar o acesso a serviços e aplicativos construídos sobre os dados abertos disponibilizando-os em um portal.

14.4 O que são metadados?

Metadados são geralmente descritos como dados sobre dados. Metadados são essenciais na organização e classificação de *datasets*. Todo conjunto de dados estruturado faz uso de metadado para descrever a estrutura e os tipos de dados contidos no conjunto.

14.5 Qual a diferença dos portais de dados abertos para os portais de transparência?

A Lei Complementar 131, de 27 de maio de 2009, que alterou a redação da Lei de Responsabilidade Fiscal no que se refere à transparência da gestão fiscal, determina a disponibilização de informações pormenorizadas sobre a execução orçamentária e financeira da União, dos Estados, do Distrito Federal e dos Municípios. Em outras palavras, os portais de transparências têm o objetivo de aumentar o controle das despesas e receitas no governo.

O portal de dados abertos tem o objetivo de ser o referencial para a busca e o acesso à dados públicos. Ele é um serviço simplificado que organiza e

padroniza o acesso aos dados públicos, primando pelo reuso dos dados e o uso de tecnologias modernas. Os portais de dados abertos disponibilizam acesso às bases de dados do governo, sem filtros, completas e seriadas.

14.6 O que são conjuntos de dados e recursos?

Os dados catalogados no portal estão organizados utilizando as estruturas de conjuntos de dados e recursos. Os conjuntos de dados, *datasets*, são as entidades principais retornadas a partir das buscas. Cada conjunto de dados possui uma descrição, um ou mais recursos, e uma série de outros metadados, como periodicidade de atualização e órgão responsável. Exemplos de conjuntos de dados são uma coleção de tabelas relacionadas entre si, dados extraídos de um mesmo sistema de informações, ou ainda uma API de dados abertos. Um conjunto de dados deve possuir pelo menos um recurso que seja dado aberto.

Cada recurso compreende uma fonte de dados. Pode ser um arquivo como uma planilha, um método de webservice ou um documento.

A princípio cada recurso poderia ser catalogado como um conjunto de dados separado, porém é preferível que os recursos sejam agrupados, sempre que forem oriundos da mesma base de dados, ou possuírem metadados em comum, para facilitar a busca e o entendimento de seu conteúdo.

Simplificando, o conjunto de dados é o agregador dos recursos, que são os arquivos.

14.7 O portal de dados abertos tem algum dado pessoal ou sigiloso?

Os dados que estiverem sujeitos a restrições de sigilo (art. 25) ou de privacidade (art. 31), conforme definidos na Lei de Acesso à Informação, precisam ser previamente classificados pelos seus responsáveis como tais, a partir de sua origem (art. 27) e por definição não são dados abertos. Assim, eles não

passam pelo processo de abertura e não devem ser catalogados no Portal Brasileiro de Dados Abertos, podendo, em caso de descumprimento, o órgão ou entidade ser responsabilizado nos termos do art. 34.

14.8 Quem é responsável pelos dados catalogados no portal?

O Portal de Dados Abertos organiza os dados abertos em um catálogo para fácil localização pelo cidadão. Entretanto, os responsáveis pelos dados são as organizações públicas que os publicam. Por isso cada uma dessas organizações responde pelos seus próprios dados.

Para mais informações ou esclarecimentos sobre um conjunto de dados específico, procure na própria página do conjunto de dados a seção “Informações Adicionais”. Nela há um campo “Autor que se pode contatar”.

Exemplo:



The screenshot shows the 'Dados Consumidor.gov.br - Janeiro/2015' page. At the top right is an 'Explorar' button. Below it are filters for 'Defesa do consumidor' and 'Justiça'. The main section is titled 'Informações Adicionais' and contains a table with the following data:

Campo	Valor
Fonte	https://www.consumidor.gov.br/pages/dadosabertos/externo/
Autor	Secretaria Nacional do Consumidor - Senacon/MJ
Mantenedor	Secretaria Nacional do Consumidor - Senacon/MJ
Estado	active
Cobertura geográfica	Brasil
Cobertura temporal	2014 e 2015
Periodicidade	Mensal

At the bottom of the page is a green navigation bar with three sections: 'Comunidade' (with a sub-link 'Comunidades'), 'Manuais' (with a sub-link 'Cartilha de publicação de dados'), and 'Saiba mais' (with a sub-link 'O que são dados abertos?').

Manter os dados *on-line* também é responsabilidade do órgão que o publica. Em alguns casos, essa pode ser uma instituição diferente daquela responsável pelo conteúdo dos dados. Caso um conjunto de dados esteja *off-line* ar ou apresentando problemas técnicos, procure na própria página do conjunto de dados a seção “Informações Adicionais”. Nela há um campo “Mantenedor”. Procure o nome do órgão que consta no valor desse campo para entrar em contato.

Exemplo:



The screenshot shows the 'Dados Consumidor.gov.br - Janeiro/2015' page. At the top, there is a search bar with 'Defesa do consumidor' and 'Justiça' filters. Below this is the 'Informações Adicionais' section, which contains a table with the following data:

Campo	Valor
Fonte	https://www.consumidor.gov.br/pages/dadosabertos/externo/
Autor	Secretaria Nacional do Consumidor - Senacon/MJ
Mantenedor	Secretaria Nacional do Consumidor - Senacon/MJ
Estado	active
Cobertura geográfica	Brasil
Cobertura temporal	2014 e 2015
Periodicidade	Mensal

At the bottom of the page, there is a green navigation bar with three main sections: 'Comunidade', 'Manuais', and 'Saiba mais'. Under 'Comunidade' is 'Comunidades'. Under 'Manuais' is 'Cartilha de publicação de dados'. Under 'Saiba mais' is 'O que são dados abertos?'.

No caso de não constar e-mail de contato nesses campos, sugere-se procurar o site do próprio órgão e localizar as opções “Institucional”, “Quem é quem” e/ou “Contato”.

14.9 Como abrir arquivos CSV?

Os arquivos do tipo CSV constituem-se em arquivos texto, onde os valores são delimitados por um caractere, normalmente a vírgula, ponto e vírgula ou tabulação. Menos frequentemente, é possível também que os campos sejam delimitados por uma largura fixa de caracteres.

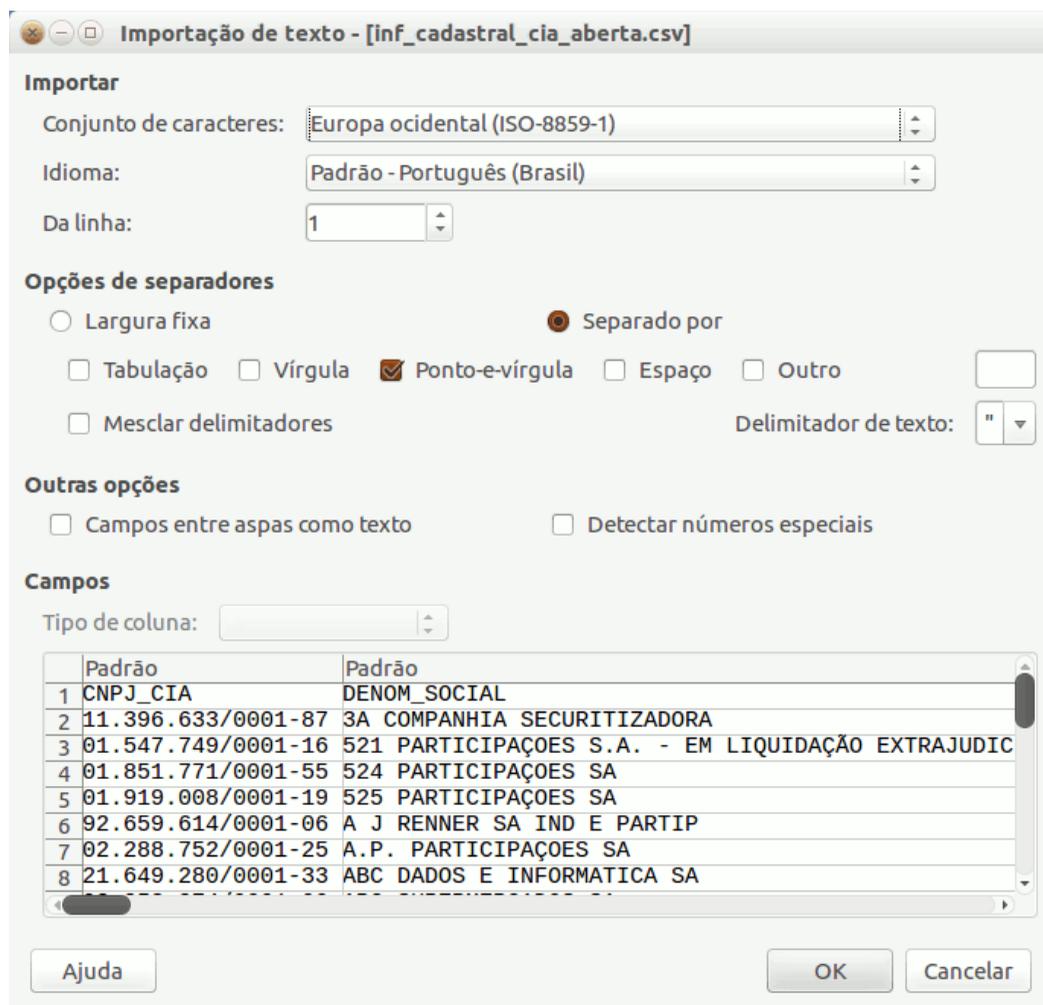
Podem ser abertos em qualquer editor de planilhas, como o LibreOffice/OpenOffice ou Excel. Nesse caso, dependendo da versão do Excel, pode ser necessário executar um passo adicional para converter texto em colunas.

Como o formato CSV não segue um padrão único, geralmente é necessário ajustar parâmetros, tais como o caractere delimitador (ex.: vírgula, ponto e vírgula ou tabulação) e a codificação de caracteres (geralmente UTF-8 ou ISO-8859-1), de acordo com a forma em que o arquivo tiver sido gravado. Por exemplo, o aparecimento de caracteres estranhos no lugar da acentuação das palavras é um sinal de que a codificação de caracteres selecionada ao abrir o arquivo deve ser substituída.

Caso o arquivo CSV seja muito grande, pode não ser viável abri-lo em um editor de planilhas, como o Excel ou LibreOffice, pois pode não haver memória suficiente para fazê-lo. Nesses casos, recomenda-se utilizar ferramentas especializadas, frequentemente utilizadas em projetos de ciência de dados, mineração de dados ou big data. Outra possibilidade é utilizar uma linguagem de programação, tal como Python ou R, para fazer operações com os dados.

Arquivos CSV também podem ser importados em sistemas gerenciadores de bancos de dados (SGBD). Consulte o manual do seu SGBD⁸ caso deseje fazer a importação de uma tabela a partir de um arquivo CSV.

⁸ Sistema Gerenciador de Banco de Dados



Exemplo de janela de abertura de arquivo CSV na ferramenta LibreOffice.

14.10 Como abrir arquivos JSON?

Arquivos JSON, a depender do seu tamanho, podem ser visualizados diretamente no navegador. Entretanto, caso o arquivo seja grande, o uso intensivo de memória e processamento podem causar travamento da tela.

Para fazer algo de útil com os arquivos, é necessário utilizar ferramentas especializadas, frequentemente utilizadas em projetos de ciência de

dados, mineração de dados ou big data. Outra possibilidade é utilizar uma linguagem de programação, tal como Python ou R, para fazer operações com os dados.

A estrutura de dados do formato JSON foi especialmente concebida para ser de leve processamento e fácil manipulação em linguagens de programação.

14.11 Como abrir arquivos XML?

Arquivos XML, a depender do seu tamanho, podem ser visualizados diretamente no navegador. Entretanto, caso o arquivo seja grande, o uso intensivo de memória e processamento podem causar travamento da tela.

Para fazer algo de útil com os arquivos, é necessário utilizar ferramentas especializadas, frequentemente utilizadas em projetos de ciência de dados, mineração de dados ou big data. Outra possibilidade é utilizar uma linguagem de programação, tal como Python ou R, para fazer operações com os dados.

A estrutura de dados do formato XML é especialmente favorável nas situações em que os dados precisam ser mesclados com conteúdo de texto. Por isso, é considerada uma linguagem de markup.

14.12 O que é Infraestrutura Nacional de Dados Abertos - INDA?

A Infraestrutura Nacional de Dados Abertos (INDA) é um conjunto de padrões, tecnologias, procedimentos e mecanismos de controle necessários para atender às condições de disseminação e compartilhamento de dados e informações públicas no modelo de Dados Abertos, em conformidade com o disposto na ePING. A INDA é a política do governo brasileiro para dados abertos.

A INDA é regulamentada através da Instrução Normativa da Infraestrutura Nacional de Dados Abertos.

14.13 O que são Dados Abertos Governamentais?

De acordo com o Manual de Dados Abertos do Governo Federal: “são dados produzidos pelo governo e colocados à disposição das pessoas de forma a tornar possível não apenas sua leitura e acompanhamento, mas também sua reutilização em novos projetos, sítios e aplicativos; seu cruzamento com outros dados de diferentes fontes; e sua disposição em visualizações interessantes e esclarecedoras”.

14.14 O que são licenças? Como posso disponibilizar os meus dados com uma licença aberta?

Uma licença, como o nome sugere, é um conjunto de autorizações que o proprietário dos dados proporciona explicitamente a quem a eles a cede.

Para especificar uma licença para os dados, basta indicar de forma visível qual a licença e um link para a mesma.

14.15 A minha instituição já disponibiliza os dados gratuitamente na Internet. Posso dizer que são dados abertos?

A gratuidade não significa abertura. É fundamental que sejam explicitamente garantidos os princípios mencionados a seguir para um recurso ser considerado aberto.

- Deve ser permitida a reutilização e redistribuição dos dados. Qualquer restrição à sua edição, modificação e redistribuição impede que sejam considerados abertos.
- Não pode haver restrições a áreas específicas. Por exemplo, um dataset não é aberto quando é apenas permitido o uso pessoal, não-comercial e/ou acadêmico.
- É preciso que estejam disponíveis em formatos abertos de modo a não obrigar o público a utilizar ferramentas de software específicas para a eles poder aceder.

14.16 O que se pode fazer com os dados abertos?

Há inúmeras áreas nas quais podemos usar dados abertos para melhorar a vida das pessoas. Podemos encontrar vários casos de sucesso no envolvimento da sociedade civil no processo democrático, na colaboração e participação cívica, na melhoria de serviços públicos, na criação de produtos e serviços inovadores, na investigação científica ou na constituição de modelos de negócio eficaz. Os dados abertos podem constituir uma preciosa matéria-prima para novos projetos por parte da sociedade civil, de iniciativas comunitárias, do setor público ou do setor privado.

14.17 Que bons exemplos existem de reutilização de dados abertos?

A Informação pública, na forma de dados abertos, pode ajudar o cidadão a tomar decisões na sua vida pessoal, ou possibilitar-lhe ser mais ativamente interveniente na sociedade.

- Na Dinamarca, o Findtoilet.dk mapeou todos os sanitários públicos dinamarqueses, para que pessoas com problemas de incontinência pudessem ter confiança para sair de casa mais frequentemente.
- O *Mapumental* no Reino Unido e o *Mapnificent* na Alemanha permitem encontrar locais para comprar ou alugar casa, tendo em conta fatores como a proximidade de transportes públicos ou a duração da deslocação para o emprego.

No campo da economia, os dados abertos também potencializam o aparecimento de novas abordagens e produtos por parte de empresas e organizações.

- Na Dinamarca, o husetsweb.dk ajuda os cidadãos a encontrar meios de melhorar a eficiência energética da sua casa, incluindo o planeamento financeiro e ajudar proprietários a implementar melhorias que levem à poupança de eletricidade. Este projeto baseia-se na reutilização de

informações cadastrais e informação relativa a subsídios governamentais, assim como o cadastro local de empresas.

- Ferramentas de tradução online como o *Linguee* ou o *Google Translate* recorrem ao enorme volume de documentos da União Europeia, traduzidos em todos os idiomas europeus e publicados sob termos abertos, para afinar e melhorar os seus algoritmos de tradução.

Os dados abertos também têm valor para o próprio governo e instituições públicas para, entre outras benesses, facilitar a prestação de serviço à sociedade e aos cidadãos.

- O Ministério da Educação holandês publicou na internet os seus dados relacionados com a educação, o que fez diminuir o número de perguntas recebidas, reduziu a carga de trabalho e os custos associados a dar respostas; agora, os funcionários podem também responder às perguntas mais rapidamente, uma vez que a informação está centralizada num portal estruturado.
- No Reino Unido, o *Where Does My Money Go?* mostra como o dinheiro dos impostos é distribuído pelo governo, permitindo um debate mais ampliado e consequente sobre a despesa pública.
- No Canadá, o *OpenParliament* publica versões acessíveis dos debates parlamentares, permitindo aos cidadãos uma visão muito mais profunda e significativa do processo legislativo.

Todos estes exemplos existem porque os governos e instituições públicas desses países disponibilizam a informação necessária usando licenças abertas.

14.18 Por que existe tanta informação fechada?

Apesar de todo este potencial, as fontes de dados abertos são ainda escassas. Podemos apontar duas razões principais:

- O desconhecimento da existência de licenças abertas, e o uso de termos de utilização padrão que, por vezes sem intenção por parte das entidades, colocam restrições que tornam os seus dados efetivamente fechados.

- A falta de incentivo à abertura por parte das entidades detentoras da informação. Para resolver este problema, é fundamental a implementação de legislação e medidas claras que consagrem a abertura de dados públicos como procedimento normal e necessário.

14.19 Que tipo de aplicações digitais podem ser feitas com esses dados?

As mais variadas. A sociedade poderá, por exemplo, desenvolver aplicações que auxiliem a análise das propostas de leis em tramitação na Câmara dos Deputados; aplicações que forneçam os resultados das votações de uma comissão; ou que comparem, em tabelas interativas, como os parlamentares usam sua verba indenizatória; ou ainda, que permitam conhecer todos os deputados envolvidos em um determinado tema – direitos das crianças, por exemplo.

16 Bibliografia

- BRASIL. 2012b. Ministério do Planejamento, Orçamento e Gestão. Secretaria de Logística e Tecnologia da Informação. Instrução Normativa nº 4, de 12 de abril de 2012. Disponível em <http://dados.gov.br/pagina/instrucao-normativa-da-inda>. Acesso em 21 out. 2019.
- BRASIL. 2012c. Governo Federal. Cartilha Técnica para Publicações de Dados Abertos no Brasil v.1.0. Brasília, DF. Disponível em: <http://dados.gov.br/pagina/cartilha-publicacao-dados-abertos>. Acesso em: 13 nov. 2019.
- BRASIL. 2012d. Governo Federal. TERMO DE USO DOS DADOS ABERTOS. Brasília, DF. Disponível em: [http:// https://dados.gov.br/termo-de-uso.html](http://https://dados.gov.br/termo-de-uso.html). Acesso em: 22 nov. 2019.
- KURBALIJA, Jovan. 2016. Uma Introdução à Governança da Internet. São Paulo: Comitê Gestor da Internet no Brasil. Disponível em: https://cgi.br/media/docs/publicacoes/1/CadernoCGIbr_Uma_Introducao_a_Governanca_da_Internet.pdf. Acesso em 3 set. 2019.
- TCU. 2015. 5 Motivos Para a Abertura de Dados na Administração Pública. Disponível em <https://portal.tcu.gov.br/biblioteca-digital/cinco-motivos-para-a-abertura-de-dados-na-administracao-publica.htm>. Acesso em 20 jan. 2020.
- W3C. 2014. Best Practices for Publishing Linked Data. World Wide Web Consortium Disponível em: <https://www.w3.org/TR/ld-bp/>. Acessado em 6 dez. 2019.

17 Anexo 1 – Recomendações de formatos para dados abertos

Recomenda-se, como parte das boas práticas, alguns formatos não proprietários listados na Cartilha Técnica para Publicação de Dados Abertos no Brasil (Brasil, 2012c): JSON, XML, CSV, ODS, RDF.⁹

- JSON. É um acrônimo para *Java Script Object Notation*. É um padrão aberto de estruturação de dados baseado em texto e legível por humano. A especificação é a RFC 4627. JSON ganhou maior utilização com o advento do Ajax. A serialização em JSON é muito simples e resulta em uma estrutura pouco verbosa o que se mostra uma ótima alternativa ao XML. JSON possibilita serialização de estrutura de objetos complexos, como listas e subpropriedades. JSON está se tornando o padrão mais utilizado para integração de dados entre repositórios e frameworks, também está se tornando o padrão nativo de armazenamento em alguns bancos de dados modernos.
- XML. É um acrônimo para *Extensible Markup Language*. É um conjunto de regras para codificar documentos com estrutura hierárquica e em um formato legível por máquina. É baseado em texto e tem como principais objetivos simplicidade, extensibilidade e usabilidade. XML é largamente utilizado como formato de troca de dados nos clássicos Web Services SOAP. Possui uma ampla gama de ferramentas associadas, tais como o padrão XSLT que permite transformar para outra estrutura XML ou outro formato. Apesar de sua ampla utilização, tem sido menos encorajada a utilização desse formato para integração de aplicações na Web, por utilizar mais recursos para transmissão e para o processamento do dados. Em substituição, recomenda-se utilizar JSON.
- CSV. É um acrônimo para *Comma-Separated Values*. São valores separados por vírgula. É um formato para armazenamento de dados tabulares em texto. A codificação é muito simples: cada linha do arquivo representa uma linha na tabela, e as colunas são separadas por vírgula. Campos que podem conter vírgula devem ser delimitados por

⁹ Os formatos estão melhor explicados no anexo 1.

aspas. CSV é recomendado para representação de estrutura de dados mais simples, de natureza tabular, onde não existem subpropriedades ou listas, gerando um arquivo menor e mais leve para processamento. Arquivos CSV são processáveis diretamente por editores de planilhas, como o OpenOffice e o MS Excel.

- ODS. É um acrônimo para *Open Document Spreadsheet*. Formato não proprietário de arquivo baseado em XML, padronizado pela ABNT sob a norma NBR ISO/IEC 26300:2006. É comumente chamado de planilha, similar ao XLS do MS Office Excel, porém aberto, por isso deve ser utilizado em substituição ao XLS. Planilhas são largamente utilizadas, sendo de fácil utilização e manipuláveis por diversos aplicativos. Apesar de ser um formato estruturado, é muito flexível, possibilitando manipulação e mistura de diversos tipos de dados, como imagens e textos formatados. Entretanto, para a publicação de dados abertos tabulares, é recomendável a utilização de CSV, pela sua simplicidade e padronização.
- RDF. É um acrônimo para *Resource Description Framework*. É um modelo de dados estruturado em grafos e possui diversos formatos de serialização, tais como RDF/XML, Notation 3 e Turtle. Os formatos baseados em RDF têm seus dados descritos em vocabulários disponíveis na Web. Apesar da grande qualidade dos dados disponibilizados em RDF, a construção de vocabulários para seu uso não é trivial. Numa escala de níveis de qualidade/complexidade de dados abertos, o RDF está no último nível, onde se constituirá a Web semântica.

Outras recomendações sobre formatos constam na Cartilha (Brasil, 2012c) na qual ainda são considerados como formatos válidos: SVG, GML, HTML/RDFa e GTFS. Estes formatos estão esclarecidos no anexo 1.

- SVG - *Scalable Vector Graphics*. Utilizado para dados vetoriais e geográficos.
- GML - *Geography Markup Language*. Utilizado para exprimir características e exportação de dados geográficos.

- HTML/RDFa. É a incorporação de marcações semânticas com uso de RDF sobre as tags HTML de uma página.
- GTFS - *General Transit Feed Specification*. Utilizado para dados de mobilidade e trânsito. Ver detalhes em <http://gtfs.org/>.

Para saber mais sobre formatos abertos consulte a e-PING (Padrões de Interoperabilidade de Governo Eletrônico) em <http://eping.governoeletronico.gov.br/>. Para a publicação de dados cartográficos e espaciais, consulte padronização na Infraestrutura Nacional de Dados Espaciais – INDE em <http://inde.gov.br/>.