



Fundamentos de Dados Abertos

Aplicação de conceitos, entendimento do histórico e dos aspectos técnicos, além das principais fontes e iniciativas consolidadas de dados abertos.

Claudia Cappelli

Propósito

Compreender as principais leis, os princípios e os aspectos técnicos da gestão de dados abertos, de forma a empregá-los na organização desses conjuntos de informação e na implantação de projetos.

Preparação

Antes de iniciar a leitura deste conteúdo, consulte o Código Civil e a Lei Complementar nº 155/2016.

Objetivos

- Descrever os conceitos e a evolução dos dados abertos.
- Identificar os aspectos técnicos de dados abertos.
- Reconhecer as fontes de dados abertos.
- Nomear as principais iniciativas de dados abertos.

Introdução

A temática de dados abertos trata do acesso livre, da utilização e do compartilhamento de dados para a finalidade que for de interesse dos usuários, dando conta apenas de saber sua origem e seu nível de segurança.

Para isso, é necessário o aprendizado sobre uma série de conceitos e aspectos técnicos que estão disponíveis nas mais diversas literaturas. Porém, é muito importante que, além de conhecer os principais conceitos envolvidos, as leis e os princípios sejam seguidos, e as políticas, estabelecidas, criando regras para compartilhar, disseminar, adquirir, padronizar, classificar e inventariar esses dados.

Outro ponto bastante importante é entender sua história e sua evolução, bem como tudo o que há disponível para os acessos e tratamentos.

Neste conteúdo, entenderemos o conceito de dados abertos, seus diversos aspectos técnicos, suas políticas, suas leis e seus princípios. Saber como essas práticas devem ser estabelecidas nas organizações públicas e, muitas vezes, até em organizações privadas é essencial para o uso correto e efetivo dos dados abertos.

Discutiremos um pouco sobre as bases que podem ser utilizadas, pois já estão disponíveis para todos. Também aprimoraremos nosso entendimento sobre projetos e iniciativas de dados abertos que existem atualmente.

O mais relevante neste conteúdo é que teremos conhecimento sobre como utilizar essas informações para nosso próprio benefício ou de terceiros. Além disso, teremos uma visão geral sobre o assunto, útil para quem estiver envolvido em uma iniciativa desse porte em alguma organização.

Histórico

O termo **dados abertos** apareceu pela primeira vez em 1955, em um documento de agência científica americana. Tratava da divulgação de dados geofísicos e ambientais. O grupo que criou esse documento buscava promover um intercâmbio completo e aberto de informações científicas entre diferentes países, um pré-requisito para a análise e compreensão desses fenômenos globais.

A ideia do bem comum aplicado ao conhecimento já havia sido teorizada, um pouco antes, em 1942. Nela, era enfatizada a importância de que os resultados da pesquisa deveriam ser livremente acessíveis a todos. Cada pesquisador deveria contribuir e renunciar aos direitos de propriedade intelectual para permitir que o conhecimento avançasse.

Mas o conceito inicial do compartilhamento de conhecimento pode ser também atribuído ao pensador irlandês George Bernard Shaw (1856-1950), pela célebre frase atribuída a ele, que remonta ao início do século XX:



Se você tem uma maçã e eu tenho outra, e nós trocamos as maçãs, então, cada um terá sua maçã. Mas se você tem uma ideia, e eu tenho outra, e nós as trocamos, então, cada um terá duas ideias.

George Bernard Shaw (1856-1950)



As Tecnologias da Informação também deram um novo fôlego a essa filosofia de bens comuns. Em sua pesquisa, **Elinor Ostrom** mostrou a especificidade dos bens comuns da informação. São muito semelhantes aos bens públicos, pois seu uso por uma pessoa não impede que outras os utilizem também. No entanto, trata-se de bens públicos de um novo tipo: seu uso não só esgota o estoque comum, mas também o enriquece.

Elinor Ostrom

1933-2009 Economista norte-americana, ganhadora do Prêmio Nobel de Economia de 2009.

Muito antes de ser um objeto técnico ou movimento político, os dados abertos estavam enraizados na prática da comunidade científica. Os pesquisadores foram os primeiros a perceber o benefício da abertura e do compartilhamento de dados.

2002

Acesso aberto

Em 2002, iniciativas ocorridas em Budapeste e Bethesda formataram o conceito de acesso aberto como atualmente se reconhece. Foi recomendada a disponibilização on-line e sem limitações dos resultados dos trabalhos científicos. Também foram descritos os tipos de documentos que devem ser de acesso aberto, como, por exemplo: dados não processados, metadados, fontes originais etc.



2004

Princípios e Diretrizes da OCDE para o acesso a dados

Em 2004, os ministros da Ciência de todas as nações da Organização para a Cooperação e o Desenvolvimento Econômico (OCDE) assinaram uma declaração afirmando que todos os dados de arquivo com financiamento público devem ser disponibilizados à população. Em 2007, foi publicado o documento Princípios e Diretrizes da OCDE para o acesso aos dados (OECD, 2007).



2007

Dados públicos, propriedade comum

Trinta ativistas da internet (inspirados diretamente pela abordagem e prática do código de programa aberto) se reuniram em Sebastopol, na Califórnia, com o objetivo de definir o conceito de dados públicos abertos e fazer com que fosse adotado pelos candidatos presidenciais dos Estados Unidos. Os participantes vinham dos movimentos de software livre. O mote era que os dados deveriam ser uma propriedade comum, assim como as ideias científicas, isso através da partilha e utilização desse bem comum.



2008

Cultura do governo aberto

Em 2008, o então presidente dos Estados Unidos, Barack Obama, ao assumir o cargo na Casa Branca, assinou três memorandos presidenciais. Dois deles dizem respeito ao governo aberto, do qual os dados abertos são um dos pilares. Esses memorandos presidenciais definem explicitamente a cultura do código aberto no centro da ação pública ao reivindicar seus princípios fundamentais: transparência, participação e colaboração.



Definições oficiais

De acordo com a Open Definition, dados abertos são:

Open Definition

Define os conceitos de "dados abertos", "conteúdo aberto" e "conhecimento aberto", através de um conjunto de princípios.



[...] dados que podem ser livremente utilizados, reutilizados e redistribuídos por qualquer pessoa – sujeitos, no máximo, à exigência de atribuição à fonte original e ao compartilhamento pelas mesmas licenças em que as informações foram apresentadas.

(ISOTANI; BITTENCOURT, 2015, p. 19)



A abertura de dados faz com que mecanismos de restrição possam ser aplicados para ocultá-los, impedindo que tanto pessoas físicas quanto jurídicas possam utilizá-los livremente. Segundo a definição da Open Knowledge Internacional:

Open Knowledge Internacional

É uma organização cuja missão é promover “um mundo aberto, onde todas as informações não pessoais sejam abertas, para que todos possam usar, desenvolver e compartilhar; e os criadores e inovadores sejam bastante reconhecidos e recompensados” (OKFN, 2021, tradução livre).



Dados são abertos quando qualquer pessoa pode livremente acessá-los, utilizá-los, modificá-los e compartilhá-los para qualquer finalidade, estando sujeito, no máximo, a exigências que visem preservar sua proveniência e sua abertura.

(BRASIL, 2018)

Leis e princípios

Diversos organismos nacionais e internacionais citam as leis de dados abertos, criadas por David Eaves, que representam um conjunto de testes para avaliar se um dado pode ou não ser considerado aberto. São elas:

David Eaves

Especialista em políticas públicas.

1

Se o dado não pode ser encontrado e indexado na Web, ele não existe.

2

Se o dado não estiver aberto e disponível em formato compreensível por máquina, ele não poderá ser reaproveitado.

3

Se algum dispositivo legal não permitir sua replicação, ele não será útil.

Essas leis podem ser aplicadas tanto a dados abertos governamentais quanto a dados de empresas privadas, organizações da sociedade civil e organismos internacionais.

Já os **princípios de dados abertos** foram criados, em 2007, por um grupo de 30 pensadores e ativistas da internet, que se reuniram em Sebastopol, como mencionamos. Naquele encontro, eles chegaram à conclusão e ao consenso sobre oito princípios (BRASIL, 2018). Segundo o Portal Brasileiro de Dados Abertos, são eles:

Completos

“Todos os dados públicos são disponibilizados. Dados públicos são dados que não estão sujeitos a limitações válidas de privacidade, segurança ou controle de acesso.”

Primários

“Os dados são publicados na forma coletada na fonte, com [o mais alto nível de] granularidade possível, [isto é, dados brutos], e não de forma agregada ou transformada.”

Atuais

“Os dados são disponibilizados o [mais] rapidamente para preservar o seu valor.”

Acessíveis

“Os dados devem estar disponíveis para a mais ampla gama de usuários e para as mais diversas finalidades.”

Processáveis por máquina

“Os dados são razoavelmente estruturados para possibilitar o seu processamento automatizado.”

Acesso não discriminatório

“Os dados estão disponíveis a todos, sem [necessidade de] identificação ou registro.”

Formatos não proprietários

“Os dados estão disponíveis em um formato sobre o qual nenhum ente tenha controle exclusivo.”

Licenças livres

“Os dados não estão sujeitos a restrições por 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.”

No Brasil, além das três leis e dos oito princípios de dados abertos, também foram definidos pelo Tribunal de Contas da União (TCU) **5 motivos para a abertura de dados na administração pública** (BRASIL, 2015). As razões para que as instituições públicas invistam em iniciativas de abertura de dados governamentais incluem (BRASIL, 2018)

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

Legislação

O Portal Brasileiro de Dados Abertos também apresenta os principais recortes da **Lei de Acesso à Informação (LAI)** – Lei nº 12.527/2011 – quanto a dados abertos. Vejamos a seguir:

Art. 8º

Art. 8º É dever dos órgãos e [das] entidades públicas promover, independentemente de requerimentos, a divulgação em local de fácil acesso, no âmbito de suas competências, de informações de interesse coletivo ou geral por eles produzidas ou custodiadas.

[...]

§ 2º Para cumprimento do disposto no caput, os órgãos e [as] entidades públicas deverão utilizar todos os meios e instrumentos legítimos de que dispuserem, sendo obrigatória a divulgação em sítios oficiais da rede mundial de computadores (internet).

§ 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.

A LAI também define as hipóteses de sigilo e de informações pessoais, que são consideradas exceções à regra geral de transparência, ou seja, que os dados devem ser abertos.

Conceitos e evolução dos dados abertos

No vídeo deste módulo, veremos a apresentação do conceito de dados abertos, leis e princípios, histórico e estado atual da arte.



Conteúdo interativo

Acesse a versão digital para assistir ao vídeo.

Verificando o aprendizado

Questão 1

Os dados abertos têm relação com os seguintes tópicos:

A

Livre acesso e licenças proprietárias.

B

Reuso e direitos reservados.

C

Direitos reservados e distribuição por qualquer pessoa.

D

Fonte original e reuso.

E

Licenças proprietárias e fonte original.



A alternativa D está correta.

Segundo o conceito de dados abertos da Open Definition, eles devem ter explicitados suas fontes originais e o tipo de reuso. As licenças proprietárias e os direitos reservados não são compatíveis com esse conceito.

Questão 2

Sobre as leis e os princípios de dados abertos, é correto afirmar que:

A

dado aberto é o dado disponível em formato compreensível por máquina.

B

o dado é considerado aberto, mesmo que não possa ser replicado em algum dispositivo.

C

as leis de dados abertos são oito.

D

o dado aberto na Web não precisa ser indexado.

E

os princípios de dados abertos são três.



A alternativa A está correta.

Há oito princípios e três leis de dados abertos. Uma dessas três leis estabelece que o dado precisa estar aberto e disponível em formato compreensível por máquina. Do contrário, seu reaproveitamento será impossível.

Manual técnico brasileiro

Para orientar as organizações do governo em relação às melhores práticas de publicação de dados, alguns países criaram guias e cartilhas. Em 2012, apoiado em uma norma técnica do governo do Chile (*Propuesta de Norma Técnica para Publicación de Datos en Chile*), o Ministério do Planejamento, Orçamento e Gestão disponibilizou a *Cartilha técnica para publicação de dados abertos no Brasil* (BRASIL, 2012b).



Nesse mesmo ano, foi criada a Infraestrutura Nacional de Dados Abertos (INDA), que define um “conjunto de padrões, tecnologias, procedimentos e mecanismos de controle necessários para atender à disseminação e [ao] compartilhamento de dados e informações públicas”, seguindo padrões internacionais.

(BRASIL, 2019).

O Portal Brasileiro de Dados Abertos é sua principal forma de organização. Seu objetivo é centralizar toda publicação, toda busca e todo acesso a dados públicos no Brasil. Para isso, definiu uma série de padrões, metodologias e processos para publicação de dados. Ele foi construído para ser a arquitetura padrão de todos os governos (federal, estadual e municipal).

A *Cartilha técnica para publicação de dados abertos no Brasil* (BRASIL, 2012b) possui os requisitos mínimos para que uma organização pública consiga organizar e preparar os conjuntos de dados, seja capaz de estabelecer um processo de publicação e faça isso por meio do uso do Portal Brasileiro de Dados Abertos, que é um serviço de apresentação desses dados.

Cada conjunto de dados (*dataset*) disponibilizado tem de estar referenciado na Web para que possa ser encontrado de forma única. A infraestrutura criada tem alguns requisitos técnicos básicos (BRASIL, 2012b). São eles:

Controle de acesso

Somente as pessoas autorizadas têm acesso para incluir e alterar os dados.

Integridade

Os dados não podem ser adulterados durante qualquer movimentação.

Autenticidade

Os dados provêm de fonte legítima.

Há dois documentos que servem como mecanismos para auxiliar todas as organizações governamentais que queiram publicar seus dados: *Guia de abertura de dados* (BRASIL, 2012a) e *Arquitetura técnica referencial para abertura de dados* (BRASIL, 2011). Nesses documentos, existem as seguintes condições gerais (BRASIL, 2012b):

1 Condição 1

“Os dados devem estar em seu formato mais bruto possível, ou seja, antes de qualquer cruzamento ou agregação.”

2

Condição 2

“Os dados devem estar em formato aberto, não proprietário, estável e de amplo uso.”

3

Condição 3

“Não deve existir nenhum instrumento jurídico que impeça sua reutilização e redistribuição por qualquer parte da sociedade.”

4

Condição 4

“Para os dados que são estruturados ou estão em planilhas na sua fonte, deve-se preservar ao máximo a estrutura original.”

5

Condição 5

“É recomendável a disponibilização dos dados em diversos formatos.”

6

Condição 6

“Cada conjunto de dados deve possuir um identificador único e persistente, seguindo uma padronização de URL.”

7

Condição 7

“É recomendável a utilização de considerações semânticas na definição de URLs, de forma que seja possível deduzir o conteúdo de um conjunto de dados apenas lendo seu identificador.”

8

Condição 8

“É extremamente desaconselhável a utilização de mecanismos antirrobôs, como *captchas*, para acesso aos conjuntos de dados.”

9

Condição 9

“É recomendável que os nomes dos arquivos sigam as boas práticas de formação de um *slug*, tal como é realizado em muitas soluções de software para blog ou gerenciadores de conteúdo.”

10 Condição 10

“Cada conjunto de dados deve ter informações sobre seus dados e metadados.”

11

Condição 11

“Para conjunto de dados muito grande, recomenda-se a divisão em conjuntos menores, permitindo uma fácil manipulação.”

12

Condição 12

“É desejável que o repositório dos dados possibilite a composição de filtros dentro da URL, seguindo algum padrão de API.”



Saiba mais

Vamos aproveitar para conhecer melhor alguns termos que permeiam as condições acima? O termo Captchas, do inglês Completely Automated Public Turing test to tell Computers and Humans Apart, refere-se a um teste de Turing realizado de forma pública e automatizada, a fim de diferenciar pessoas de computadores. No teste de Turing, um indivíduo controla o teste. No captcha, no entanto, uma máquina gerencia todo o processo (CONCEITO.DE, 2019). Slug é parte de uma URL que identifica uma página com palavras-chave legíveis por humanos. E a sigla API, vem do inglês: Application Programming Interface, que significa Interface de Programação de Aplicações.

Formatos

Não existe uma lista completa de formatos permitidos para apresentação de dados. O que sabemos é que um dado, para ser considerado aberto, deve estar disponível em um formato não proprietário, conhecido e estruturado, de modo que possa ser usado sem restrições e de forma automatizada.

Na *Cartilha técnica para publicação de dados abertos no Brasil* (BRASIL, 2012b), existe uma lista dos cinco principais formatos de dados abertos. São eles:

JSON (*JavaScript Object Notation*)

“Padrão aberto de estruturação de dados baseado em texto e legível por humano. Possibilita serialização de estrutura de objetos complexos, como listas e subpropriedades. Está se tornando o padrão mais utilizado para integração de dados entre repositórios e frameworks, e o padrão nativo de armazenamento em alguns bancos de dados modernos.”

XML (*Extensible Markup Language*)

“Conjunto de regras para codificar documentos com estrutura hierárquica e em um formato legível por máquina. É baseado em texto e [muito] utilizado como formato de troca de dados nos clássicos Web Services SOAP. [Não tem sido muito usado] para integração de aplicações na Web, por utilizar mais recursos para transmissão e processamento dos dados.”

CSV (*Comma-Separated Values*)

“Formato para armazenamento de dados tabulares em texto. Cada linha do arquivo representa uma linha na tabela, e as colunas são separadas por vírgula [ou outro caractere específico]. Recomendado para representação de estrutura de dados mais simples.”

ODS (*Open Document Spreadsheet*)

“Formato [aberto] de arquivo baseado em XML. [Substitui os softwares do Microsoft Office Excel, pois é aberto.] Possibilita manipulação e mistura de diversos tipos de dados, como imagens e textos formatados.”

RDF (*Resource Description Framework*)

“Modelo de dados estruturado em grafos que possui diversos formatos de serialização, tais como RDF/XML, Notation 3 e Turtle. [Esse formato tem] dados descritos em vocabulários na Web.”

Embora não sejam os mais utilizados, os formatos listados a seguir também estão disponíveis na Web (BRASIL, 2012b):

- “SVG (*Scalable Vector Graphics*) – para dados vetoriais e geográficos;
- GML (*Geography Markup Language*) – para exportação de dados geográficos;
- HTML/RDFa – para marcações semânticas com uso de RDF sobre as tags HTML de uma página”.

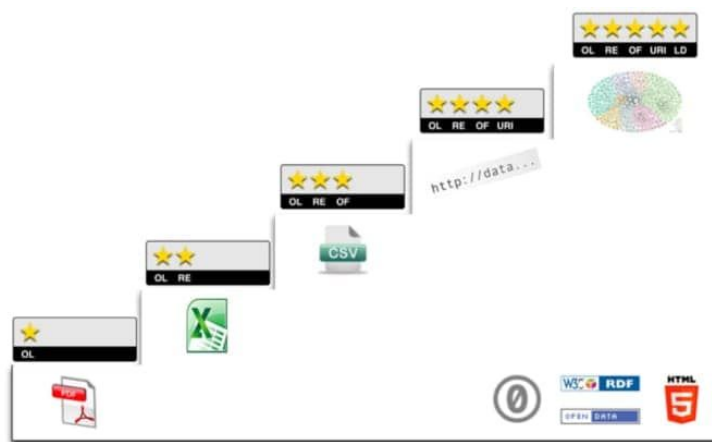
Modelo de maturidade

Tim Berners-Lee elaborou um padrão para definir um modelo de maturidade de dados abertos que visa qualificar as publicações de dados, seguindo uma escala de cinco estrelas. Cada nova estrela alcançada torna os dados cada vez mais fáceis de serem processados e utilizados.

Tim Berners-Lee

Físico britânico, criador da World Wide Web.

Veja o esquema de implementação do padrão cinco estrelas:



Primeiro nível

Basta que o dado exista na Web, “independentemente de formato, sob uma licença aberta” (BRASIL, 2013). No esquema, podemos ver um arquivo em PDF representando esse nível. Apesar de ser um formato que não permite reutilização, ainda assim, é aberto.

Segundo nível

Além das características do primeiro, já há “dados estruturados legíveis por máquina” (BRASIL, 2013). Não importa que estejam em formato proprietário, desde que respeitem a licença livre.

Terceiro nível

Temos os requisitos dos níveis anteriores, mas, agora, com a obrigatoriedade do uso de um formato não proprietário. No esquema, podemos ver um formato CSV.

Quarto nível

Assim como nos demais, atendemos a todas as características dos níveis anteriores, mas, além disso, utilizamos a URI (Universal Resource Identifier) para identificar as coisas, permitindo que possam ser referenciadas.

Quinto nível

Também mantemos a regra de atender a todas as características dos níveis anteriores, só que agora acrescentando a possibilidade de ligar dados para prover contexto.

Metadados

Metadados ou metainformação são dados sobre os dados, ou seja, informações que permitem entender o conteúdo de um conjunto de dados. A *Cartilha técnica para publicação de dados abertos no Brasil* (BRASIL, 2012b) apresenta a lista de um conjunto de metadados obrigatórios e desejáveis. São eles:

Metadados obrigatórios

“Título: Nome do conjunto de dados.

Descrição: Uma breve explicação sobre os dados.

Catálogo origem: Página (URL) do órgão onde está publicado o conjunto de dados.

Órgão responsável: Nome e sigla do órgão ou [da] entidade responsável pela publicação do conjunto de dados.

Categorias no VCGE: O Vocabulário Controlado de Governo Eletrônico é uma lista hierarquizada de assuntos do governo que utiliza termos comuns e é voltada para a sociedade.”

Metadados desejáveis

“Etiquetas: Lista de palavras-chave relacionadas ao conjunto de dados.

Autoria: Instituição ou pessoa responsável pela produção do recurso.

Documentação: URL de documento que expõe detalhes sobre o conjunto de dados.

Cobertura geográfica: Localização ou região geográfica a que se referem os dados.

Cobertura temporal: Data ou período a que se referem os dados.

Granularidade geográfica: Precisão geográfica da cobertura geográfica.

Granularidade temporal: Precisão temporal da cobertura temporal.

Frequência de atualização: Frequência temporal com que o conjunto de dados é atualizado.

Referências: Relações com outros conjuntos de dados.

Metodologia: Processo de criação dos dados.

Vocabulário/ontologia: Documentos estruturados com metadados específicos do conjunto de dados.”

Catálogos

Um catálogo de dados “tem o objetivo de simplificar a busca e o acesso aos conjuntos de dados” (BRASIL, 2012b). Trata-se de um tipo de organização para que se tenha acesso aos metadados dos repositórios.

Segundo a *Cartilha técnica para publicação de dados abertos no Brasil* (BRASIL, 2012b), um catálogo deve:

- Ter uma lista completa dos conjuntos de dados do repositório.
- Permitir a navegação e a busca dos dados por meio de diferentes mecanismos.
- Utilizar o conjunto de metadados obrigatórios e permitir que sejam estendidos.
- Estar integrado com o catálogo central.



Dica

Na construção dos catálogos, é importante pensar em seu processo de manutenção. Deve ser assegurado que as URLs utilizadas permaneçam sempre ativas e persistentes. Além disso, é necessário se certificar que o processo de atualização dos metadados esteja alinhado com o processo de publicação de dados, a fim de “garantir que a informação mais nova e correta estará [sempre] disponível”. Por fim, a organização responsável pelos dados deverá instituir um processo de publicação que preserve o valor do dado (BRASIL, 2012b).

Aspectos técnicos de dados abertos

No vídeo deste módulo, o especialista apresenta o manual brasileiro de regras técnicas, as regras dos formatos de dados abertos, o conceito de metadados e catálogos.



Conteúdo interativo

Acesse a versão digital para assistir ao vídeo.

Verificando o aprendizado

Questão 1

Assinale a alternativa que apresenta um metadado obrigatório no padrão brasileiro:

A

Autoria.

B

Referências.

C

Descrição.

D

Metodologia.

E

Etiquetas.



A alternativa C está correta.

São metadados obrigatórios o título, a descrição, o catálogo origem, o órgão responsável pela publicação e as categorias no VCGE. Autoria, referências, metodologia e etiquetas fazem parte do conjunto de metadados desejáveis.

Questão 2

Que padrão aberto está se tornando modelo para integração de dados na Web e armazenamento em bancos de dados?

A

ODS.

B

RDF.

C

CSV.

D

JSON.

E

XML.



A alternativa D está correta.

JSON é o padrão classificado na literatura como aquele mais usado para integração de dados na Web, devido à linguagem que utiliza.

Fontes

Existem diversas fontes de dados abertos. Elas podem ser utilizadas com os mais diferentes propósitos e se dividem em 4 categorias:

1. Dados científicos;
2. Dados de governo;
3. Dados de empresas privadas;
4. Dados de universidades.

O governo é protagonista nesse cenário, pois é a organização da sociedade que mais coleta dados, os quais são considerados públicos: um direito garantido em Constituições de vários países. No Brasil, a Constituição Federal estabelece, em seu artigo 5º, os seguintes direitos e garantias fundamentais:



XIV - é assegurado a todos o acesso à informação e resguardado o sigilo da fonte, quando necessário ao exercício profissional; [...] XXXIII - todos têm direito a receber dos órgãos públicos informações de seu interesse particular, ou de interesse coletivo ou geral, que serão prestadas no prazo da lei, sob pena de responsabilidade, ressalvadas aquelas cujo sigilo seja imprescindível à segurança da sociedade e do Estado.

(CF, 1988)

Segundo a Wikipedia, podemos notar várias áreas e atividades em que os dados abertos geram valor:

- Participação popular;
- Controle da democracia;
- Inovação;
- Melhora e criação de novos serviços do governo;
- Novos conhecimentos ao combinar fontes de dados.

Universidades como a Open University e as de Oxford, Münster, Waterloo, Toronto, Pensilvânia, entre outras no mundo, têm apresentado diversas iniciativas de abertura de dados por meio de projetos de pesquisa reconhecidos mundialmente.

Principais fontes de dados

Aqui, apresentaremos algumas das principais fontes de dados abertos existentes no Brasil e no mundo:

Dados Abertos do Banco Mundial

Repositório de dados sobre o que está acontecendo em diferentes países acerca de vários assuntos. O acesso é bastante simples. Basta elencar os nomes dos indicadores, países ou tópicos e ter acesso a séries

estatísticas e dados georreferenciados. Todos os dados podem ser baixados em diversos formatos, como CSV, Excel e XML. Possui ferramentas que auxiliam a análise e a visualização desses conjuntos de dados. A intenção é que tenhamos uma visão em nível global de dados sobre um mesmo assunto. Podemos obter acesso a uma API, que ajuda a criar as visualizações de dados.

WHO (World Health Organization)

Acompanha as estatísticas de saúde dos 194 Estados membros da Organização Mundial da Saúde (OMS). Os conjuntos de dados podem ser acessados por assuntos, categorias, indicadores e país. É permitido o download em formato Excel. Também possui uma API que permite o acesso direto por meio de passagem de parâmetros.



Google Public Data Explorer

Analisa grandes quantidades de conjuntos de dados públicos. Funciona como um grande *hub* que permite e organiza o acesso a diferentes agências e fontes. Nele, podemos encontrar dados do Banco Mundial, do Bureau de Estatísticas do Trabalho dos Estados Unidos e do US Bureau, da OCDE, do Fundo Monetário Internacional (FMI) etc. Por permitir juntar diferentes bases de dados em uma única pesquisa, é acessado para diversas finalidades por diferentes tipos de usuários, desde um estudante ou pesquisador até algum órgão do governo.



Curiosidade

Hub é a peça central que realiza a conexão entre computadores em uma rede local, transmitindo e retransmitindo informações de uma máquina para as demais. Aqui, o termo hub é usado no sentido de concentrador, um processo pelo qual se difunde informações de diversas fontes para diversos receptores.

Registro de dados abertos no AWS/RODA (Registry of Open Data on Amazon Web Services)

É um repositório que contém conjuntos de dados públicos. Permite a busca por palavras-chave e tags. Possui uma interface Web bastante simplificada e de fácil entendimento. Tem um metamodelo muito interessante de apresentação de informação. Para cada busca feita, quando apresenta o resultado, traz consigo o conjunto de dados, a página de detalhes sobre esses dados, os exemplos de uso dos dados, as informações de licença e os tutoriais ou aplicativos que usam os dados. Também permite que sejam feitas análises com os dados abertos e que sejam criados serviços utilizando esses dados. Todos os dados apresentados estão em nuvem própria de cada uma das agências, das organizações governamentais, dos pesquisadores, das empresas e dos indivíduos que os compartilham.

Portal dos Dados Abertos da União Europeia



É o repositório de dados abertos da União Europeia. Inclui dados sobre economia, emprego, ciência, meio ambiente e educação. Tem dados da Agência Europeia do Ambiente, do Centro Comum de Investigação e de outras Direções-Gerais da Comissão Europeia e agências que tornaram públicos seus conjuntos de dados. Os dados podem ser baixados e reutilizados, além de analisados.

Para isso, utiliza um catálogo de metadados que pode ser pesquisado por meio da linguagem **SPARQL**. Os dados

disponíveis nesse repositório podem ser usados para fins comerciais ou não.

SPARQL

“Acrônimo recursivo do inglês SPARQL, Protocol and RDF Query Language. Trata-se de uma linguagem padronizada para a consulta de grafos RDF.” Fonte: Wikipedia.

Departamento de Recenseamento dos Estados Unidos

É a maior agência estatística do governo federal dos Estados Unidos. Possui uma base de dados com fatos sobre pessoas, lugares e economia da América. É muito utilizada para estabelecimento de novas moradias e instalações públicas. Também permite conhecimento sobre características demográficas das cidades, dos estados e do país como um todo. Os dados existentes nessa base servem para planejamento de sistemas de transporte, estradas, escolas, serviços públicos etc.

Data.gov

São os dados abertos do governo federal dos Estados Unidos. Lá, é possível encontrar dados, ferramentas e recursos para conduzir uma pesquisa, desenvolver aplicativos e até mesmo visualizações e análises de dados. Basta inserir as palavras-chave sobre o assunto que buscamos e navegar por tipos, tags, formatos, grupos, tipos de organizações, organizações e categorias. Possui, ainda, algumas APIs que permitem tratamento de dados.

DBpedia

Oferece conteúdo estruturado das informações da Wikipedia. Podemos pesquisar e explorar relacionamentos e propriedades da Wikipedia por meio de buscas semânticas. Possui links para outros conjuntos de dados relacionados. Grande parte dos dados é classificada em ontologias, que são representações do conhecimento associado aos dados. As consultas são realizadas em SPARQL. Foi muito utilizada pela Apple (via aplicativo Siri), pelo Google (via Freebase e Google Knowledge Graph) e pela IBM (via projeto Watson).

Dados Abertos FreeCodeCamp

É uma comunidade de código aberto de programas. Permite codificar e criar projetos de software. A comunidade tem acesso a grandes quantidades de dados. É possível encontrar nessa base um enorme conjunto de dados, análises e até mesmo projetos já realizados, com muitos exemplos para quem a utiliza. Também tem links para projetos externos.



Datasets abertos do Yelp



A Yelp é uma empresa multinacional, com base na Califórnia, que desenvolve aplicativos voltados à avaliação de estabelecimentos comerciais. Possui um conjunto de bases de diversos negócios e dados do usuário para uso em atividades pessoais, educacionais e acadêmicas. Nessas bases, existem dados de muitas empresas, fotos e áreas metropolitanas, que podem ser usados para diferentes propósitos.

Todos os arquivos estão disponíveis como JSON e podem ser usados para ensinar aos alunos sobre bancos de dados

e para aprender programação neurolinguística.

Conjunto de dados do United Nations Children's Fund (UNICEF)

O Fundo das Nações Unidas para a Infância possui dados sobre educação, trabalho, deficiência e mortalidade infantis, mortalidade materna, água e saneamento, baixo peso ao nascer, cuidados pré-natais, pneumonia, malária, deficiência de iodo, mutilação/corte genital feminino e adolescentes. Esses dados são extraídos

diretamente das bases de dados do UNICEF. Por serem de domínio próprio, sua atualização é controlada e feita mensalmente. Todos os dados estão disponíveis em formato CSV. Foram construídas algumas visualizações de dados para serem utilizadas pelos desenvolvedores, de modo a testarem os dados antes de baixarem.



Kaggle



É um portal na Web com diferentes formatos de publicação de *datasets* destinados à ciência de dados, em especial para aprendizado de máquina. Sua importância vem não apenas do acesso, mas também do que desejamos fazer com esses dados. É bastante didático e autoexplicativo. Define todos os formatos de arquivo que são recomendados durante o compartilhamento de dados.

Não é apenas um repositório de dados. Sua intenção é criar uma comunidade que permite discutir sobre os dados, os códigos e as técnicas públicas, além de conceituar seus

próprios projetos. Suporta a maioria dos formatos abertos, como CSV, JSON, SQLite, Archive, Big Query etc.

Linked Open Data University of Münster (LODUM)

Iniciativa de dados abertos da Universidade de Münster. Permite que qualquer pessoa acesse qualquer informação pública sobre a universidade de Münster em formatos legíveis por máquina. Utiliza o conceito de Linked Data, utilizando o editor SPARQL ou o pacote SPARQL da linguagem R para análise estatística de dados.



Saiba mais

Linked Data é o conjunto de dados publicados na Web, legíveis por máquina e ligados a outros conjuntos de dados por meio de padrões abertos.

Repositório de Aprendizado de Máquina da UCI

É um repositório mantido pelo Centro de Aprendizado de Máquina e Sistemas Inteligentes da Universidade da Califórnia em Irvine. Contém diversas bases de dados, teorias de domínio e geradores de dados para a análise empírica de algoritmos de aprendizado de máquina. Utiliza basicamente arquivos de texto ou CSV. Os conjuntos de dados podem ser classificados e pesquisados por aspectos como tipos de atributos, número de instâncias, número de atributos e ano publicado.

Localização de fontes de dados abertos

Neste vídeo, o especialista fará uma breve apresentação das principais fontes de dados abertos e suas formas de uso.



Conteúdo interativo

Acesse a versão digital para assistir ao vídeo.

Verificando o aprendizado

Questão 1

A fonte de dados abertos do Banco Mundial permite:

A

busca somente por países ou tópicos.

B

download em um único formato.

C

combinações com outras fontes de dados.

D

uso de ferramentas de cálculo.

E

listar seus metadados



A alternativa C está correta.

LoA fonte de dados abertos do Banco Mundial dá acesso à API (Application Programming Interface), que ajuda a criar as visualizações de dados e permite combinações ao vivo com outras fontes de dados.

Questão 2

Qual é a fonte de dados abertos que permite pesquisa semântica e exploração de relacionamentos do recurso da Wikipedia?

A

DBpedia.

B

LODUM.

C

Kaggle.

D

Portal dos Dados Abertos da União Europeia.

E

Data.gov.



A alternativa A está correta.

A DBpedia é uma base de dados que permite pesquisa e exploração de relacionamentos e propriedades do recurso da Wikipedia com outros repositórios de dados.

Benefícios para organizações

Seguindo todas as leis e todos os princípios, o uso de dados abertos pode aumentar muito a produtividade de uma organização. Alguns benefícios são referentes a:

1

Pedidos de acesso à informação

Reduzir os custos de produção desses pedidos, se já tiverem sido pensados e colocados à disposição ao invés de serem tratados a cada solicitação.

2

Ações duplicadas

Disponibilizar os dados uma única vez ao invés de refazer a cada solicitação.

3

Dados complementares de outros órgãos

Investigar dados complementares aos da organização, existentes em outra empresa, e linká-los.

4

Atenção positiva dos cidadãos, da mídia e de outros órgãos

Destacar o uso de transparência ativa pela organização, ao invés de somente passiva, pode resultar em uma percepção melhor do cidadão e da sociedade em geral.

5

Gerar receita e criar empregos

Fazer com que sejam gerados empregos e receitas, através do aumento do uso de apresentação de informações.

Um relatório desenvolvido pelas Nações Unidas (DATA REVOLUTION GROUP, 2014) avalia como os dados abertos podem ajudar a alcançar as metas para o desenvolvimento sustentável. Nele, conta que a publicação de dados abertos estimula a economia, na medida em que agentes econômicos privados utilizam esses dados para a criação de novos processos de negócio e a otimização dos processos existentes.

Principais iniciativas

No Brasil e no mundo, existem diversas iniciativas que apoiam e estruturam o universo de dados abertos. Algumas delas apresentam uma possível infraestrutura para que os governos se organizem. Outras apresentam trabalhos de avaliação das diversas fontes disponíveis, expondo diversas análises prontas sobre os dados. As demais são eventos que contribuem para a discussão sobre o assunto.

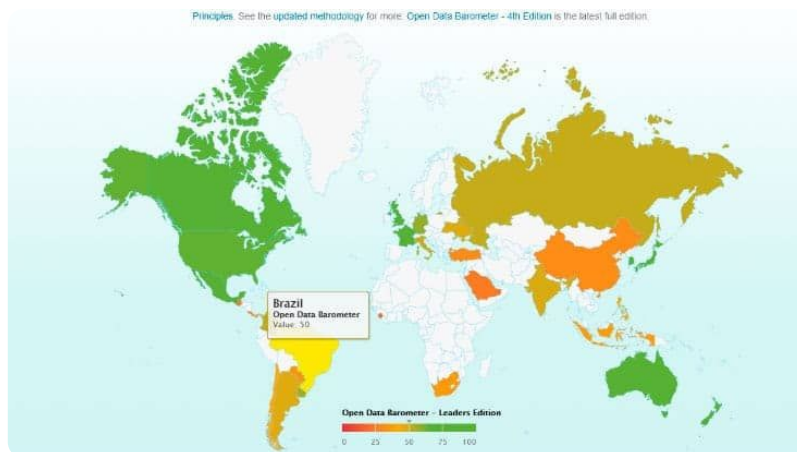
Enfim, há muitos exemplos dessas iniciativas. A seguir, apresentamos algumas das principais, com o objetivo de nomear suas funcionalidades e reconhecer sua relevância para o desenvolvimento do conceito de dados abertos:

Open Data Barometer (opendatabarometer.org)

Analisa os 30 governos que adotaram a **Carta de Dados Abertos** e os países membros do G20. Além disso, traz uma série de relatórios não só indicando a classificação, mas também mostrando como cada país evoluiu desde a última classificação e o que ainda falta ser feito. Quem acessa também pode ver todas as classificações anteriores. Por fim, há, ainda, uma funcionalidade que permite ao visitante do site fazer comparações entre os países de sua escolha.

Carta de Dados Abertos

É um sistema de colaboração entre governos e organizações baseado em um conjunto de princípios e práticas recomendadas para a liberação de dados abertos governamentais. A carta foi formalmente adotada por dezessete governos de países, estados e cidades na Cúpula Global de Parceria de Governo Aberto no México em outubro de 2015.



Base dos Dados (basedosdados.org)

Site de organização não governamental que organiza bases de dados internacionais. Classifica as informações em grupos (temas) e, para cada um desses, oferece determinado número de conjuntos de dados. Alguns dos temas são esporte, saúde, agropecuária, energia, educação, infraestrutura, história etc. Os dados também podem ser acessados por meio das organizações responsáveis por eles. Mais de 500 organizações têm alguns de seus conjuntos de dados apresentados por lá. A função principal do site é permitir que os interessados entrem e façam buscas de seus interesses nesses conjuntos de dados. Para cada conjunto disponibilizado, há uma breve descrição de seu conteúdo e de links específicos para baixar o conjunto de dados solicitado.

dadosabertos.social

Fórum aberto da comunidade de dados abertos brasileira. Reúne pessoas com diversos interesses, como desenvolvedores, servidores públicos, jornalistas, pesquisadores, empreendedores etc. As informações estão organizadas em categorias. Para incluir dados, basta que a pessoa crie um conjunto de dados #sem-categoria (no caso de a categoria não existir), que, posteriormente, os administradores criarão esse assunto ou irão direto na categoria desejada e adicionarão lá. Ao final, tudo é validado por administradores das bases. É um espaço aberto e colaborativo, no qual os interessados podem fazer perguntas, compartilhar experiências, descobrir conjuntos de dados, propor ideias, compartilhar dados etc.

Infraestrutura Nacional de Dados Abertos (INDA)

É a implementação do modelo de dados abertos brasileiro, cujo principal componente é o Portal Brasileiro de Dados Abertos (dados.gov.br) – ponto central para o acesso e uso dos dados abertos governamentais no Brasil. A INDA busca ser a referência brasileira para pesquisa, acesso, compartilhamento e uso de dados públicos. Mantém uma política de centralização feita por meio de um catálogo governamental. Permite a busca, o acesso, o reuso e o cruzamento dos dados públicos de diferentes fontes e assuntos. Também atua na padronização, no armazenamento, no acesso, no compartilhamento, na disseminação dos dados e de informações públicas do governo brasileiro, fomentando a colaboração com o cidadão.



International Open Data Day (opendataday.org)



Evento que acontece todos os anos globalmente. Organiza palestras, seminários, demonstrações, hackathons (maratonas de programação) e treinamentos sobre dados abertos. Ocorreu pela primeira vez em 2010. A ideia surgiu após discussões de um grupo muito atuante na área, composto, inclusive, por brasileiros. A data do evento é escolhida pelos membros do grupo, levando em consideração diferentes eventos culturais.

Wikidata (wikidata.org)

É um banco de dados orientado a documentos, focado em itens que representam qualquer tipo de tópico, conceito ou objeto. Cada item é identificado de forma única com um número inteiro positivo prefixado com a letra Q maiúscula, conhecido como "QID". Trata-se de um banco de dados secundário, livre, colaborativo e multilíngue que coleta dados estruturados que servem de suporte à Wikipédia. É um repositório de armazenamento central que pode ser acessado por outros, como as wikis mantidas pela Wikimedia Foundation. O conteúdo carregado dinamicamente a partir da Wikidata não precisa ser manuseado em cada wiki individual. Por exemplo, estatísticas, datas, locais e outros dados comuns podem ser centralizados na Wikidata.



Budapest Open Access Initiative (budapestopenaccessinitiative.org)

É uma declaração pública de princípios relacionados ao acesso aberto à literatura de pesquisa (Creative Commons Initiative), criada em 2002. Contém a definição de acesso aberto mais utilizada:



Por ‘acesso aberto’ a esta literatura, queremos dizer sua disponibilidade gratuita na internet pública, permitindo que qualquer usuário leia, faça download, copie, distribua, imprima, pesquise ou crie links para os textos completos desses artigos, rastreá-los para indexação, transmiti-los como dados para software ou usá-los para qualquer outra finalidade lícita, sem barreiras financeiras, legais ou técnicas, a não ser aquelas inseparáveis de obter acesso à própria internet.

(BOAI, 2002, tradução livre)

De acordo com BOAI (2002), a **única restrição à reprodução e distribuição**, e a única função dos direitos autorais neste domínio, deve ser dar aos autores o controle sobre a integridade de seu trabalho e o direito de serem devidamente reconhecidos e citados.

Creative Commons Initiative (creativecommons.org)



É uma declaração pública de princípios relacionados ao acesso aberto à literatura de pesquisa, criada em Budapeste pelo Open Society Institute, em 2001. Possui várias formas de licenças de direitos autorais públicos que permitem a distribuição gratuita de uma obra protegida por direitos autorais. É usada quando um autor deseja dar a outras pessoas o direito de compartilhar, usar e construir sobre sua obra. São vários os tipos de licença Creative Commons diferindo nos termos de distribuição. Existem várias jurisdições de Creative Commons que se adaptam às leis internacionais, inclusive o CC Brasil.

Open Knowledge Foundation (okfn.org)

É uma organização cuja missão é promover “um mundo aberto, onde todas as informações não pessoais sejam abertas, para que todos possam usar, desenvolver e compartilhar; e os criadores e inovadores sejam bastante reconhecidos e recompensados” (OKFN, 2021, tradução livre). Essa organização indica que podemos ter dois futuros diferentes: um que é aberto e outro que é fechado.

Futuro aberto

Um futuro aberto significa que o conhecimento é compartilhado por todos – disponível gratuitamente para todos, um mundo onde as pessoas são capazes de realizar seu potencial e viver uma vida feliz e saudável.



Futuro fechado

Um futuro fechado é aquele em que o conhecimento é uma propriedade e é de controle exclusivo, aumentando a desigualdade.

Grandes empresas de tecnologia têm monopolizado a era digital, e uma concentração insustentável de riqueza e poder levou a um crescimento atrofiado e a oportunidades perdidas. Assim, os consumidores, os inovadores do futuro e a sociedade são prejudicados.

Essa organização preza para que líderes globais usufruam da abertura de todas as formas de conhecimento para garantir um futuro justo, livre e aberto. Para seguir nessa missão, os membros da OKFN buscam:

1

Apoiar pessoas e organizações para criar um futuro livre, justo e aberto;

2

Estender o alcance global a novas geografias e novos setores, em particular, saúde, educação e trabalho;

3

Ter políticas e procedimentos que apoiem a visão e os tornem adequados para o propósito;

4

Trabalhar em parceria com outras pessoas que possam ajudá-los a alcançar a visão e a garantir parcerias de financiamento que lhes permitam ser sustentáveis.

Dados abertos ligados

Para finalizar, trataremos brevemente dos dados abertos ligados, não como uma iniciativa de dados abertos propriamente dita, mas como um conceito tecnológico relacionado a dados abertos. Esse conceito é conhecido pelo termo em inglês *Linked Data*. A ideia foi criada por Berners-Lee (2006, tradução livre), que definiu os quatro princípios que caracterizam os dados ligados. São eles:

1º Princípio

“Usar URIs para nomes de ‘coisas’ [isto é, qualquer recurso].

2º Princípio

Usar URIs HTTP para que as pessoas possam procurar esses nomes.

3º Princípio

Fornecer informações úteis quando alguém acessar um URI, utilizando padrões como RDF e SPARQL.

4º Princípio

Incluir links para outros URIs para que as pessoas possam encontrar mais ‘coisas’ [outros recursos]”.



Relembrando

Como comentado anteriormente: URI (Uniform Resource Identifier ou Identificador Uniforme de Recurso) é uma cadeia de caracteres que serve para denominar e identificar univocamente um recurso na Internet.

Devido ao advento do *Linked Data*, a Web passou a ser um espaço de dados global que conecta dados de diversos domínios. As URIs identificam todos os itens de interesse na Web. Todos os recursos que encontramos na Web tradicional, como documentos, imagens e outros arquivos de mídia, são recursos informacionais. Todos os “objetos do mundo real” que existem fora da Web, como pessoas, lugares, proteínas, conceitos científicos etc., são recursos não informacionais. Todos esses podem, agora, conectar-se devido ao conceito de *Linked Data*.

O projeto Linking Open Data é um esforço aberto e colaborativo apoiado pelo grupo W3C SWEO (*Semantic Web Education and Outreach Group*). Seu principal objetivo é identificar *datasets* existentes que estão disponíveis sob licenças abertas, convertê-los para RDF de acordo com os princípios de dados ligados, publicá-los na Web e interligá-los uns com os outros, formando uma nuvem de dados ligados (*Linked Open Data Cloud*).



Comentário

Navegadores genéricos para Linked Data permitem aos usuários iniciar a navegação em uma fonte de dados e navegar em links para fontes de dados relacionadas. Existem páginas de busca de Linked Data que rastreiam a Web, seguindo links entre fontes de dados e fornecendo consulta sobre dados agregados.

Principais iniciativas de dados abertos

Neste vídeo, o especialista falará sobre os benefícios das iniciativas de dados abertos, principais iniciativas e dados abertos ligados.



Conteúdo interativo

Acesse a versão digital para assistir ao vídeo.

Verificando o aprendizado

Questão 1

Sobre o conceito de Creative Commons, assinale a alternativa correta:

A

É uma declaração privada de princípios relacionados ao acesso aberto à literatura de pesquisa.

B

Foi criado pela Open Knowledge Foundation.

C

Protege as pessoas que usam ou redistribuem a obra de um autor de preocupações de violação de direitos autorais.

D

É a única licença de direitos autorais públicos.

E

Não permite a distribuição gratuita de uma “obra” protegida por direitos autorais.



A alternativa C está correta.

Creative Commons Initiative é uma declaração pública que permite o uso ou a distribuição gratuita de uma obra que possui direitos autorais, dando às pessoas a possibilidade de compartilhá-la, por exemplo.

Questão 2

O que são dados abertos ligados?

A

Utilização de um conjunto de tecnologias e melhores práticas que possibilite a publicação e interligação de dados estruturados na Web.

B

Interligação de dados em sistemas de dados abertos.

C

Utilização de uma ontologia para interligação de dados.

D

Utilização de um conjunto de tecnologias na Web.

E

Melhores práticas de publicação de dados.



A alternativa C está correta.

O termo *Linked Data* (Dados Ligados) está diretamente relacionado a um conjunto de boas práticas para publicação e interligação de dados na Web.

Considerações finais

Neste conteúdo, entendemos o conceito de dados abertos. Conhecemos a Infraestrutura Nacional de Dados Abertos (INDA), que dá suporte ao Portal de Dados Abertos Brasileiro. Apresentamos o manual técnico brasileiro e suas regras, bem como todos os tipos de formatos mais utilizados em dados abertos. Além disso, abordamos os metadados e o conceito de catálogos de dados.

Apresentamos, ainda, as quatro principais áreas de dados abertos: dados científicos, dados de governo, dados de empresas privadas e dados de universidades. Na sequência, tratamos um pouco do movimento de dados abertos e conhecemos as principais fontes existentes para acesso. Por fim, listamos as principais iniciativas de dados abertos no Brasil e no mundo, descobrimos como cada uma delas funciona e entendemos o que são os chamados dados abertos ligados.

Podcast

Neste podcast, a especialista encerra o tema fazendo um apanhado dos tópicos mais importantes, como: conceito de dados abertos, se as organizações privadas estão prontas para abrir seus dados, se as iniciativas existentes atualmente são suficientes, dentre outros.



Conteúdo interativo

Acesse a versão digital para ouvir o áudio.

Explore +

Para conhecer as principais bases de dados públicos brasileiros disponíveis para acesso, pesquise o Portal Brasileiro de Dados Abertos.

Para saber mais sobre fontes de dados do governo, pesquise e leia o artigo *Dados abertos governamentais (open government data): instrumento para exercício de cidadania pela sociedade*, de Cláudio José Silva Ribeiro e Reinaldo Figueiredo de Almeida.

Conheça mais da [Creative Commons](#).

Referências

BERNERS-LEE, T. **Linked data**. Design Issues, 2006. Consultado na internet em: 25 ago. 2021.

BRASIL. Ministério do Planejamento, Orçamento e Gestão. Secretaria de Logística e Tecnologia da Informação. Departamento de Integração de Sistemas e Informação. **Arquitetura técnica referencial para abertura de dados**: piloto de dados abertos SICAF. Brasília: MP/SLTI, 2011.

BRASIL. **Guia de abertura de dados**. Brasília: INDA, 2012a.

BRASIL. Ministério do Planejamento, Orçamento e Gestão. Secretaria de Logística e Tecnologia da Informação. **Cartilha técnica para publicação de dados abertos no Brasil v1.0**. Portal Brasileiro de Dados Abertos, Brasília, 2012b.

BRASIL. Ministério do Planejamento, Desenvolvimento e Gestão. Secretaria de Tecnologia da Informação. **Maturidade em dados abertos**: entenda as 5 estrelas. Portal Brasileiro de Dados Abertos, Brasília, 2013.

BRASIL. Tribunal de Contas da União. **5 motivos para a abertura de dados na Administração Pública**. Brasília: República Federativa do Brasil, 2015.

BRASIL. Ministério do Planejamento, Desenvolvimento e Gestão. Secretaria de Tecnologia da Informação. **O que são dados abertos?** Portal Brasileiro de Dados Abertos, Brasília, 2018.

BRASIL. **Infraestrutura Nacional de Dados Abertos**. Governo Digital, Brasília, 2019.

BUDAPEST OPEN ACCESS INITIATIVE. **Budapest Open Access Initiative**. Budapest: BOAI, 2002.

DATA REVOLUTION GROUP. **A world that counts**: mobilising the data revolution for sustainable development. [S. l.]: Data Revolution Group, 2014.

ISOTANI, S.; BITTENCOURT, I. **Dados abertos conectados**. São Paulo: Novatec, 2015.

OPEN KNOWLEDGE FOUNDATION. **A fair, free and open future**. London: OKFN, 2021.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. **Principles and guidelines for access to research data from public funding**. Paris: OECD, 2007.

RAUTENBERG, S. *et al.* **Guia prático para publicação de dados abertos conectados na web**. Curitiba: Appris, 2018.

SINGH, J. **Open Data 101**: the latest trends, challenges and research in government open data. Erindale: Cooe Press, 2017.