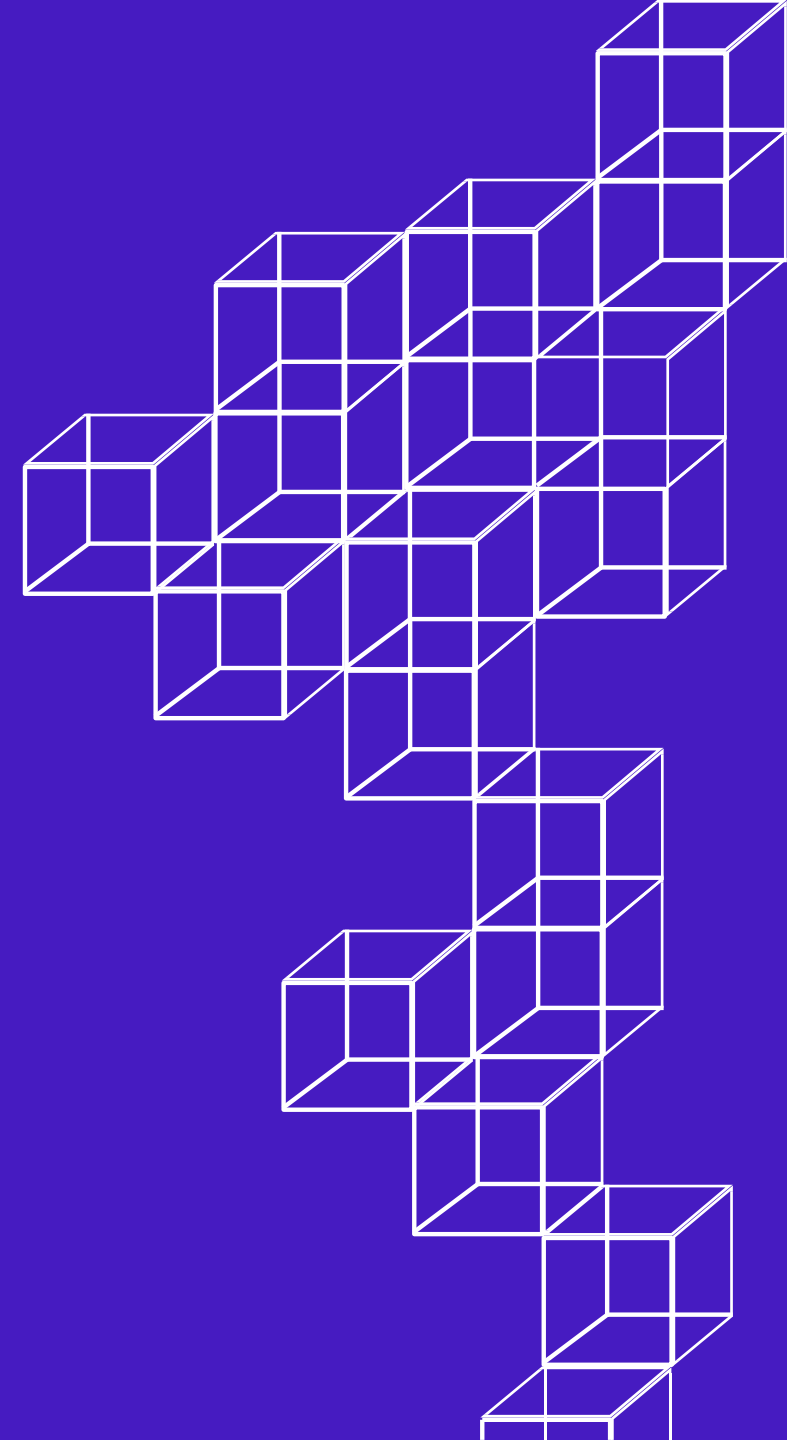


Boas Práticas para Dados na Web

Caroline Burle

São Paulo, 10 de agosto de 2021



Caroline Burle



Mestre em Relações Internacionais pelo San Tiago Dantas, da PUC-SP, UNESP e UNICAMP. Fellow do Programa da OEA de Governo Aberto nas Américas. Pesquisadora do Centro de Estudos sobre Tecnologias Web (Ceweb.br), do NIC.br. Pesquisa sobre governo aberto, relações internacionais e política externa, com experiência na atuação de governos subnacionais, na área de Web, dados abertos e governança da Internet.



ceweb.br nic.br cgi.br

membros e ex-membros do CGI.br
(somente os atuais membros têm direito a voto) ➔

ASSEMBLEIA GERAL

7 membros eleitos pela Assembleia Geral ➔

CONSELHO DE
ADMINISTRAÇÃO

CONSELHO
FISCAL

ADMINISTRAÇÃO
.....
JURÍDICO
.....
COMUNICAÇÃO
.....
ASSESSORIAS:
CGI.br e PRESIDÊNCIA

DIRETORIA
EXECUTIVA

1 2 3 4 5

registro.br

Domínios

cert.br

Segurança

cetic.br

Indicadores

ceptro.br

Redes e Operações

ceweb.br

Tecnologias Web

ix.br

Troca de Tráfego

W3C
Brasil

Padrões Web

Comitê Gestor da Internet no Brasil (CGI.br)

Criado pela Portaria Interministerial 183 de julho de 1995



Composição atual estabelecida pelo decreto presidencial 4.829 de 2003

<https://www.cgi.br/>

O que é a revolução dos dados para o desenvolvimento sustentável?

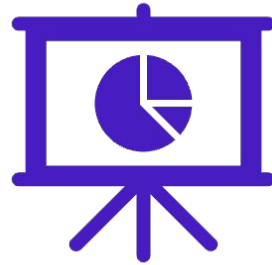
“Dados são a alma da tomada de decisões e a matéria-prima para a prestação de contas. é quase impossível a concepção, o acompanhamento e a avaliação de políticas eficazes sem dados de alta qualidade que fornecem as informações corretas sobre as coisas certas no momento certo.”

ONU. Data Revolution Group, 2014.

Governo Aberto



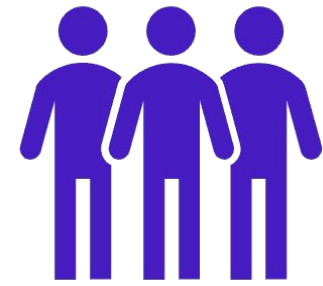
Transparência



Dados Abertos



Participação



Controle Social



Lei de Acesso à Informação (LAI)

Art. 8º É dever dos órgãos e 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.

Política de Dados Abertos do Poder Executivo federal



**Disponibilizar:
Planos de
Dados Abertos**

**Governança:
Infraestrutura
Nacional de
Dados
Abertos -
INDA**

Transparência Governamental

Ministério da Saúde

**CORONAVÍRUS
COVID-19**



**Painel de leitos e insumos
do Ministério da Saúde**

acesse >

O QUE VOCÊ PRECISA SABER

O QUE É COVID-19

QUAIS SÃO OS
SINTOMAS

COMO É TRANSMITIDO

DIAGNÓSTICO

COMO SE PROTEGER

SE EU FICAR DOENTE

SERVIÇO DE SAÚDE

FAKE NEWS

PERGUNTAS
FREQUENTES

PAINEL COVID-19

APP CORONAVIRUS

DISQUE SAÚDE 136

ÚLTIMAS NOTÍCIAS

23/04/2020 20h09

Coronavírus: Brasil registra 49.492 casos e 3.313 mortes

23/04/2020 18h10

Segunda turma de profissionais da saúde chega a Manaus (AM)

23/04/2020 15h10

Ministério da Saúde convoca empresas para venda de equipamentos e insumos

Acesse mais notícias



**ESTUDANTE FAÇA
O SEU CADASTRO**
O BRASIL CONTA COM VOCÊ



Veja quais as medidas já adotadas pelo Governo de SP para o combate ao novo coronavírus no Estado

- Tire suas dúvidas sobre o decreto da quarentena

CASOS EM SÃO PAULO

CONFIRMADOS

16.740

ÓBITOS

1.345

atualizado 23/4/2020 às 14h

Veja os dados completos

ÍNDICE DE ISOLAMENTO

NO ESTADO

48%

NA CAPITAL

48%

dados referentes a 22/4

Veja o índice dos municípios

CAMPANHA DE VACINAÇÃO CONTRA GRIPE

2ª fase

a partir de 16/4

- doentes crônicos
- sistema prisional e indígenas
- caminhoneiros e motoristas

Saiba mais

DOAÇÕES

Saiba como ajudar os cidadãos de SP

Dados Abertos (Governo)

BRASIL CORONAVÍRUS (COVID-19) Simplifique! Participe Acesso à informação Legislação Canais

dados.gov.br PORTAL BRASILEIRO DE DADOS ABERTOS

Dados | Organizações | Aplicativos | Inventários | Concursos | INDA | Perguntas frequentes | Contato | Sobre o portal

dados.gov.br PORTAL BRASILEIRO DE DADOS ABERTOS

Dados | Organizações | Aplicativos | Inventários | Concursos | INDA | Perguntas frequentes | Contato | Sobre o portal

Pesquisar conjuntos de dados... PESQUISAR

em 7862 conjuntos de dados com 41254 recursos (o que é isto?)

Conjuntos de dados em destaque

Conjuntos de dados colocados em destaque pela equipe do dados.gov.br.

Preços de Medicamentos - Consumidor

Os dados representam a lista de preços de Medicamentos, contemplando o preço Fábrica, ou preço fabricante (PF), que é o preço máximo que...

Orçamento Federal - Orçamento Fiscal e da Seguridade Social

Contém dados da execução orçamentária e financeira da despesa orçamentária aprovada na lei orçamentária e créditos adicionais (orçamentos...

Base Cartográfica Contínua do Brasil - 1:250000

A Base Cartográfica Contínua do Brasil na escala 1:250000 (BC250) é um conjunto de dados geoespaciais de referência, estruturado em bases...

Notícias

31/01/2020

PGFN lança aplicativo que lista devedores da União

A Procuradoria-Geral da Fazenda Nacional (PGFN) lançou nesta quarta-feira (29) o aplicativo Dívida Aberta, iniciativa que otimiza a Lista de Devedores. Com o

Publicações mais recentes

Conjunto de dados	Data	Responsável
Decisões da Comissão Mista de Reavaliação de Informações do ano de 2020	23/04/2020	Casa Civil da Presidência da Repúbl...
SFN - BCO DO BRASIL S.A.	18/04/2020	Banco Central do Brasil - BCB
Dados referentes ao Art. 48 da Lei Complementar nº 101/2000.	17/04/2020	Instituto Federal de Alagoas - IFAL
Planos Individuais Docentes	15/04/2020	Universidade Federal do Rio Grande d...
SFN - BCO BRADESCO S.A.	14/04/2020	Banco Central do Brasil - BCB

Dados por etiqueta

Inclusão financeira (1993)	Indicadores de crédito (832)	
INDE (681)	Cartografia (658)	Geoespacial (656)
Geoserviço (656)	dados abertos (181)	indicador (166)
PGI (163)	Alagoas (129)	finanças públicas (126)

Conjuntos de dados

- Organizações
- Banco Central do Br... (3117)
- Instituto Brasileir... (421)
- Estado de Alagoas - AL (335)
- Agência Nacional do... (238)
- Agência Nacional de... (210)
- Distrito Federal (172)
- Ministério da Fazen... (156)
- Ministério da Saúde... (148)
- Previdência Social (127)
- Agência Nacional de... (91)
- Mostrar mais Organizações
- Grupos
- Governo e Política (43)
- Saúde (27)
- Educação (19)
- Agricultura, extrat... (16)
- Conjuntos de dados ... (13)
- Economia e Finanças (11)
- Pessoa, Família e S... (10)

Buscar conjunto de dados...

7.813 conjuntos de dados encontrado(s) Ordenar por: Relevância

Programas Luz para Todos e Mais Luz para a Amazônia

INFORMAÇÕES PRELIMINARES - SUJEITAS A ALTERAÇÕES O Processo de Universalização teve seu início marcado com a edição do Decreto nº 4.873, de 11 de novembro de 2003, que...

CSV PDF

Decisões da Comissão Mista de Reavaliação de Informações do ano de 2020

Decisões da Comissão Mista de Reavaliação de Informações - CMRI sobre recursos impetrados referentes aos pedidos de acesso à informação em última instância

XML PDF

SFN - PORTOSEG S.A. CFI

Este conjunto lista todas as informações disponibilizadas pela instituição no formato de dados abertos. Atenção As informações distribuídas pelas instituições como Dados...

HTML API JSON

SFN - BCO YAMAHA MOTOR S.A.

Este conjunto lista todas as informações disponibilizadas pela instituição no formato de dados abertos. Atenção As informações distribuídas pelas instituições como Dados...

HTML API JSON

Dados Abertos (Governo)



Governo
Aberto SP

Conjunto de dados
para a sociedade

SIC.SP
Serviço de Informações
ao Cidadão



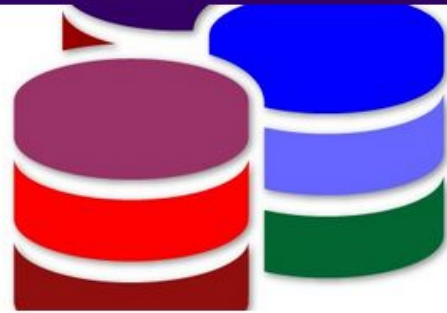
O QUE É ▾

CONJUNTOS DE DADOS

INFORMATIVOS ▾

PARTICIPAÇÃO CIDADÃ ▾

RELATÓRIOS ESTATÍSTICOS



Para obter outros conjuntos de dados, é necessário solicitar ao órgão responsável



nos conjuntos de dados

no site

OK



Arquivos, bibliotecas e centros de informação



Administração Pública



Educação



Energia



Impostos e Taxas

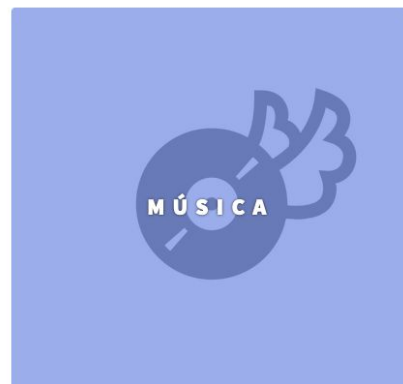


Meio Ambiente

Dados Abertos (Sociedade Civil)

 COLABORADOS

 CORONAVÍRUS



- SITES GOVERNAMENTAIS

BRASIL

- **Ministério da Saúde:** <http://plataforma.saude.gov.br/novocoronavirus/>
- **Governo do Estado de São Paulo:** <http://www.saopaulo.sp.gov.br/coronavirus/>
- **Governo do Estado de Santa Catarina:** <http://www.coronavirus.sc.gov.br/>
- **Secretaria de Saúde Pública do Pará:** <http://www.saude.pa.gov.br/coronavirus/>
- **Secretaria de Saúde de Minas Gerais:** <https://www.saude.mg.gov.br/coronavirus/>
- **Secretaria de Saúde do Espírito Santo:** <https://saude.es.gov.br/Noticias>
- **Secretaria de Saúde de Alagoas:** <http://www.saude.al.gov.br/coronavirus/>
- **Agência Alagoas, Governo do Estado de Alagoas:** <http://agenciaalagoas.al.gov.br/coronavirus>
- **Secretaria de Saúde do Amazonas:** <http://www.saude.am.gov.br/painel/corona/>

Dados Abertos (Sociedade Civil)



Sua conta ▾

Dados ▾

Sobre ▾

Blog

APOIE

O Brasil em dados libertos

Repositório de dados públicos disponibilizados em formato acessível

ESPECIAL COVID-19: DADOS POR MUNICÍPIO

Graças a uma força-tarefa de [35 voluntários](#) que, diariamente, compilam boletins epidemiológicos das 27 Secretarias Estaduais de Saúde, disponibilizamos uma base de dados com a série histórica de casos e óbitos confirmados por município. Embora essenciais para o planejamento de políticas de contenção do novo coronavírus, os dados municipais não têm sido divulgados pelo Ministério da Saúde. Ainda como parte desse esforço contínuo de fornecer dados úteis relacionados à pandemia, disponibilizamos outras bases estruturadas, como população dos municípios afetados, óbitos suspeitos registrados em cartório etc.

[VER MAIS](#)

Manifesto

Venha entender o que motivou a criação do **Brasil.IO** e quais são os objetivos do projeto!

Datasets

Confira a lista completa de todas as base de dados liberadas.

Sugira um Dataset

Conhece algum dataset bacana que gostaria de ver aqui? Sugira para nós!
=>

Colabore

Achou o projeto legal e quer contribuir? Veja aqui como.

<https://brasil.io/home/>

Uso de dados abertos

The screenshot displays the Citymapper mobile application interface. At the top, there is a navigation bar with the Citymapper logo, a location indicator for São Paulo, and links to 'CITYMAPPER MOBILITY INDEX', 'PASS', 'JOBS', and 'FEEDBACK'. Below the navigation bar, the main content area is divided into several sections:

- Location Selection:** A card for 'São Paulo' with a 'Mudar' (Change) option.
- Citymapper Branding:** A card with the Citymapper logo and the tagline 'Making cities usable'.
- Download do App:** Buttons for downloading the app on 'iPhone' and 'Android'.
- Account:** A 'Sign Up or Login' button.
- Map:** A map of São Paulo showing various transit routes in different colors (red, blue, green, purple, yellow). A red highlighted area is visible in the center of the city.
- Navigation:** A 'DE ONDE PARA ONDE' (From where to where) section with input fields for 'Start' and 'End', and a 'Now' button. A 'GO' button is also present.
- PERTO (Nearby):** A section showing icons for 'Ônibus' (Bus), 'Metrô' (Metro), 'CPTM', and 'Bicicletas' (Bicycles).
- LINHAS (Lines):** A section for displaying transit lines.
- Citymapper Mobility Index:** A purple card featuring a circular icon with a percentage sign and the text 'Citymapper Mobility Index', 'COVID-19', and 'Stats about mobility in cities'.

Uso de datos abiertos

The screenshot shows the homepage of Influence Explorer. At the top, the site name 'INFLUENCE EXPLORER' is displayed in a serif font. Below it are three navigation tabs: 'CAMPAIGN FINANCE', 'LOBBYING', and 'FOREIGN'. The main content area features a large yellow 'EXPLORE' heading, followed by the text 'COMPANIES, LAWMAKERS & PROMINENT INDIVIDUALS that matter to you see how they're influencing the political system'. A 'what's new' section is highlighted in a light blue background, containing two articles: 'Going Real-Time!' and 'Retiring TransparencyData'. The 'Going Real-Time!' article discusses the shift to real-time data streams, while the 'Retiring TransparencyData' article announces the end of the TransparencyData project and the migration of data to OpenSecrets.org and FollowTheMoney.org.

INFLUENCE EXPLORER

CAMPAIGN FINANCE

LOBBYING

FOREIGN

EXPLORE

COMPANIES, LAWMAKERS

& PROMINENT INDIVIDUALS that matter to you
see how they're influencing the political system

what's new

Going Real-Time!

We're excited to announce that we're starting a new chapter in the life of Influence Explorer. Going forward, Influence Explorer's goal will be to get crucial influence data to you as soon as IT'S available. Our mission will be to publish useful data as soon as possible to support the work of reporters, watchdogs and researchers who need to know what's happening right now.

The necessary trade-off with live streams is that data quality will be heavily influenced by the quality of releases from the government.

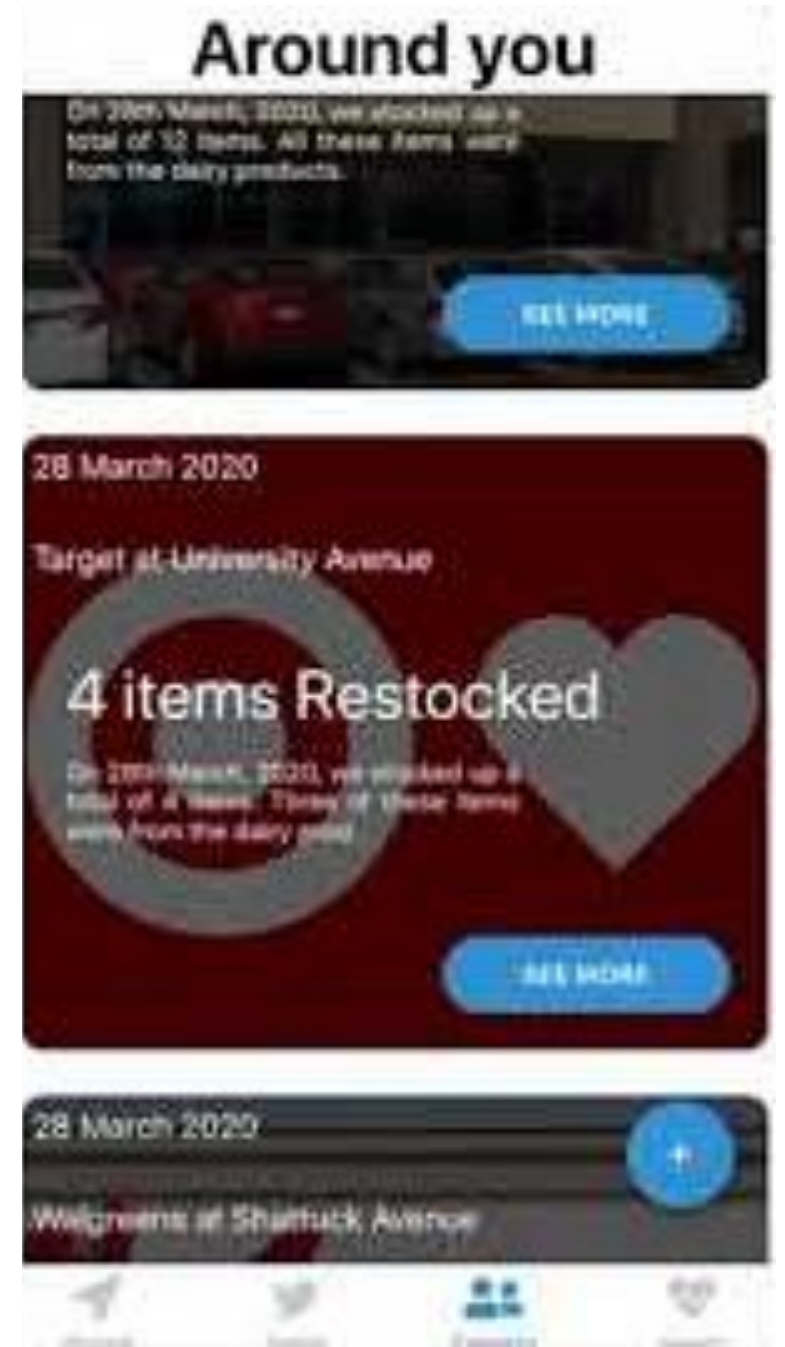
Retiring TransparencyData

Along with the changes coming to Influence Explorer, we'll also be phasing out the TransparencyData project. TransparencyData started [more than five years ago](#) as a resource for mashing up data from [OpenSecrets.org](#) and [FollowTheMoney.org](#), along with a few other sources. In the years since, both OpenSecrets and FollowTheMoney have matured and today each makes their data available both in bulk and through APIs. At Sunlight, we're thrilled to see that, and recognize the limited utility of having a third-party API that republishes these resources.

Uso de datos abiertos



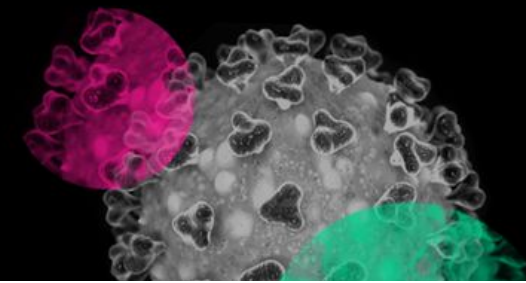
<https://devpost.com/software/coronapp>



Transparência

COVID-19

Dados abertos podem salvar vidas



ESTADOS

CAPITAIS

17 JUNHO 2021

Período de 17jun - 18ago

Avaliação por estado da federação



NÍVEIS DE TRANSPARÊNCIA



Ranking

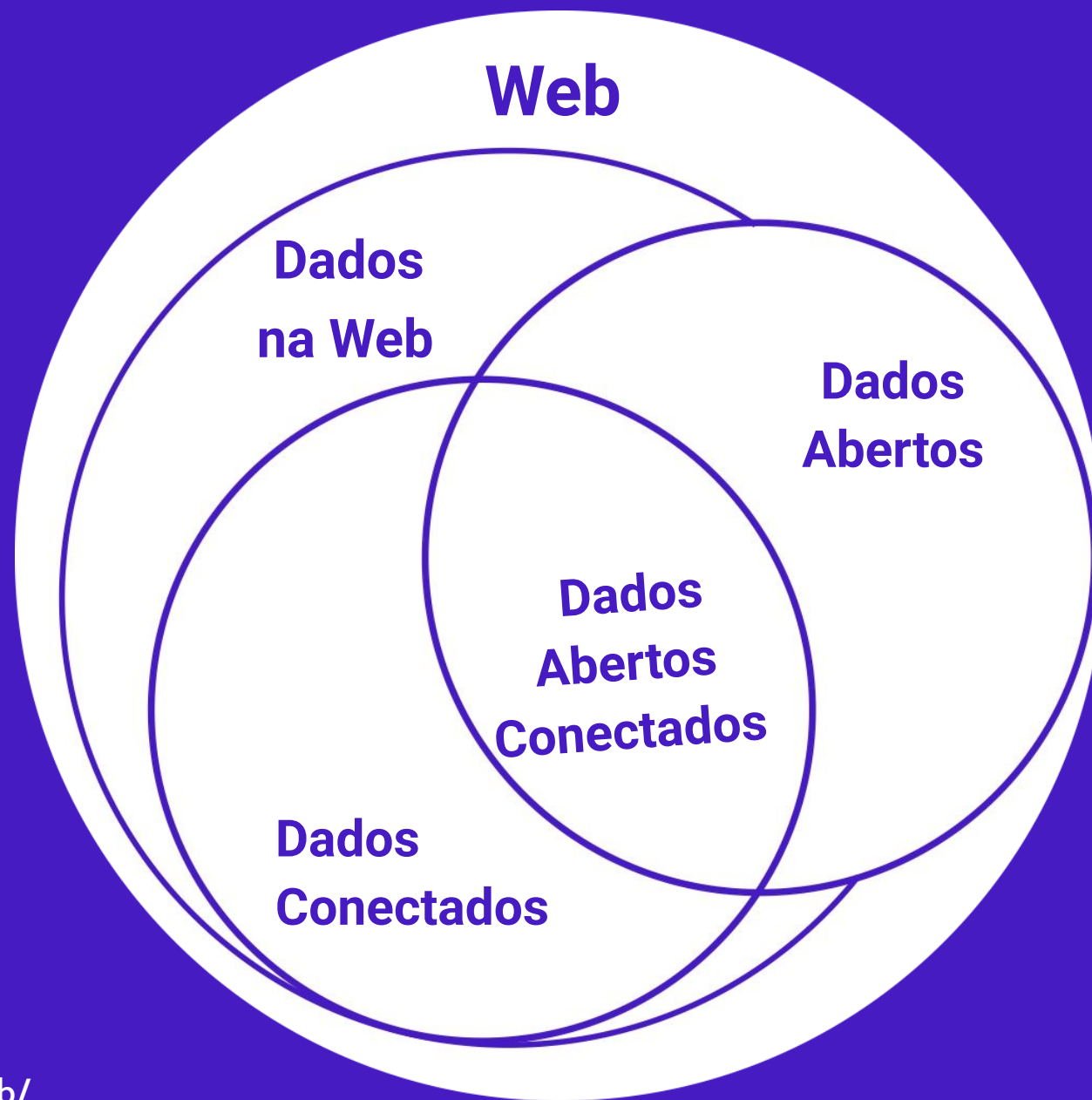
TABELA

GRÁFICO

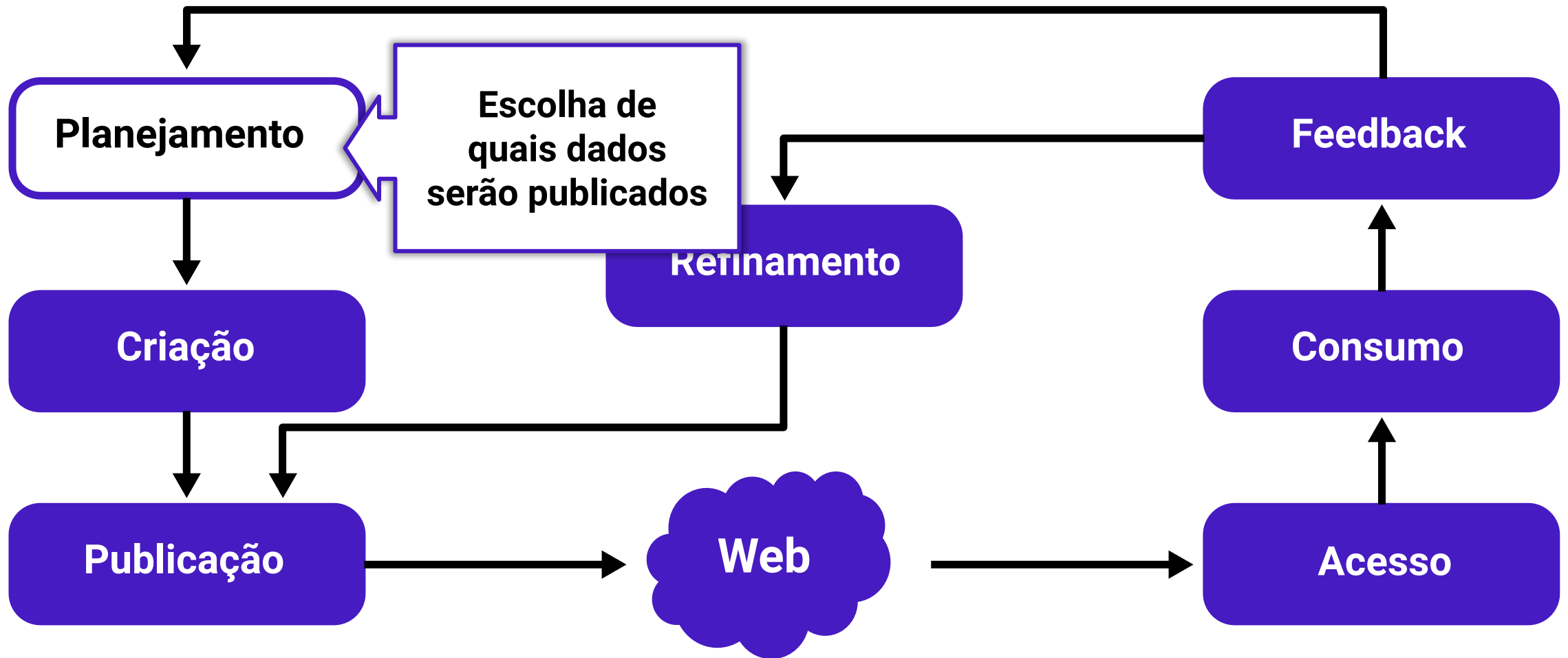
ESTADO	PONTUAÇÃO	VARIACÃO	NÍVEL				
1º Espírito Santo (ES)	93	0 0	ALTO	13º Ceará (CE)	65	0 0	BOM
2º Distrito Federal (DF)	91	0 0	ALTO	14º Bahia (BA)	64	0 0	BOM
3º Rio Grande do Sul (RS)	88	0 0	ALTO	15º Rio de Janeiro (RJ)	63	0 0	BOM
4º Minas Gerais (MG)	86	0 0	ALTO	16º Mato Grosso (MT)	61	0 0	BOM
5º Acre (AC)	81	0 0	ALTO	17º São Paulo (SP)	59	0 0	MÉDIO
6º Paraná (PR)	80	0 0	ALTO	18º Roraima (RR)	58	0 0	MÉDIO
6º Sergipe (SE)	80	0 0	ALTO	18º Tocantins (TO)	58	0 0	MÉDIO
7º Goiás (GO)	79	0 0	BOM	19º Alagoas (AL)	56	0 0	MÉDIO
7º Mato Grosso do Sul (MS)	79	0 0	BOM	20º Pará (PA)	55	0 0	MÉDIO
8º Rio Grande do Norte (RN)	78	0 0	BOM	21º Maranhão (MA)	54	0 0	MÉDIO
9º Rondônia (RO)	75	0 0	BOM	22º Piauí (PI)	53	0 0	MÉDIO
10º Governo Federal	73	0 0	BOM	23º Paraíba (PB)	52	0 0	MÉDIO
11º Amazonas (AM)	72	0 0	BOM				
11º Pernambuco (PE)	72	0 0	BOM				
12º Amapá (AP)	66	0 0	BOM				
12º Santa Catarina (SC)	66	0 0	BOM				

<https://transparenciacovid19.ok.org.br/>

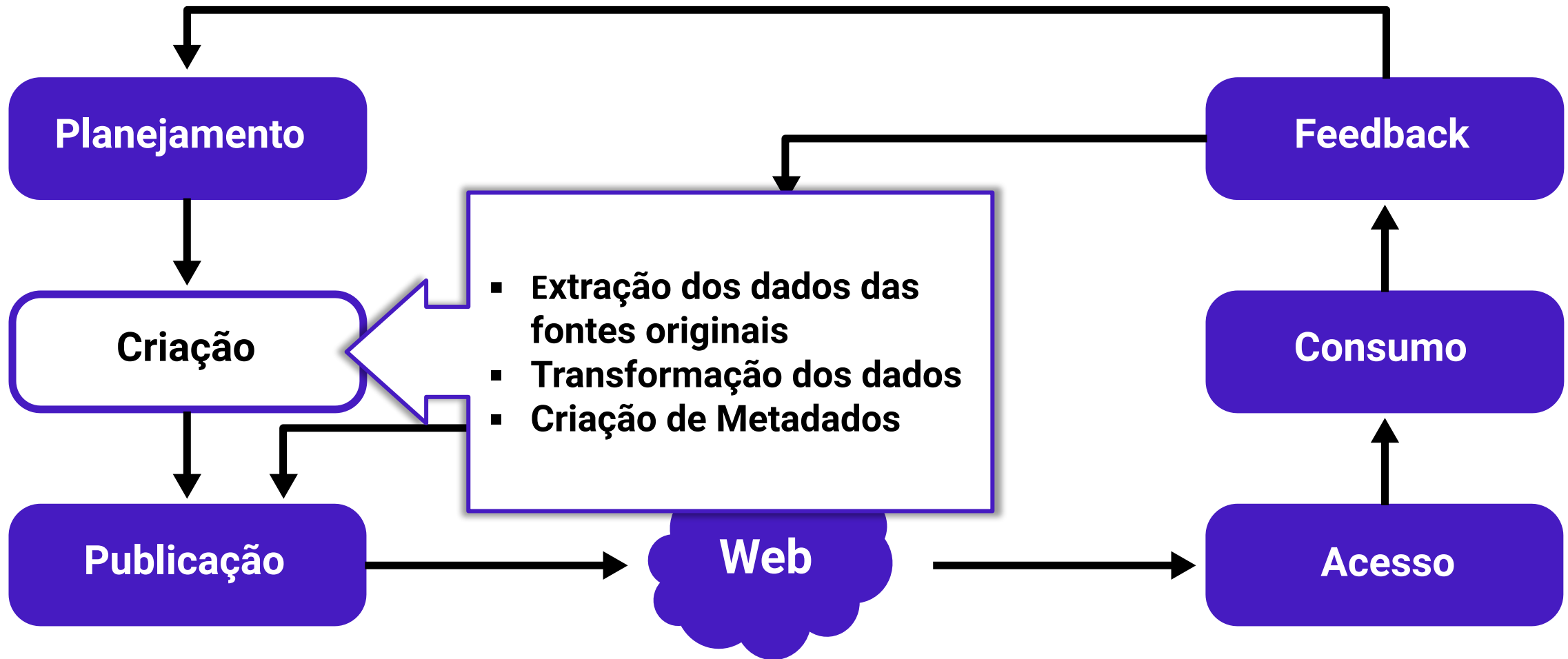
Dados na Web



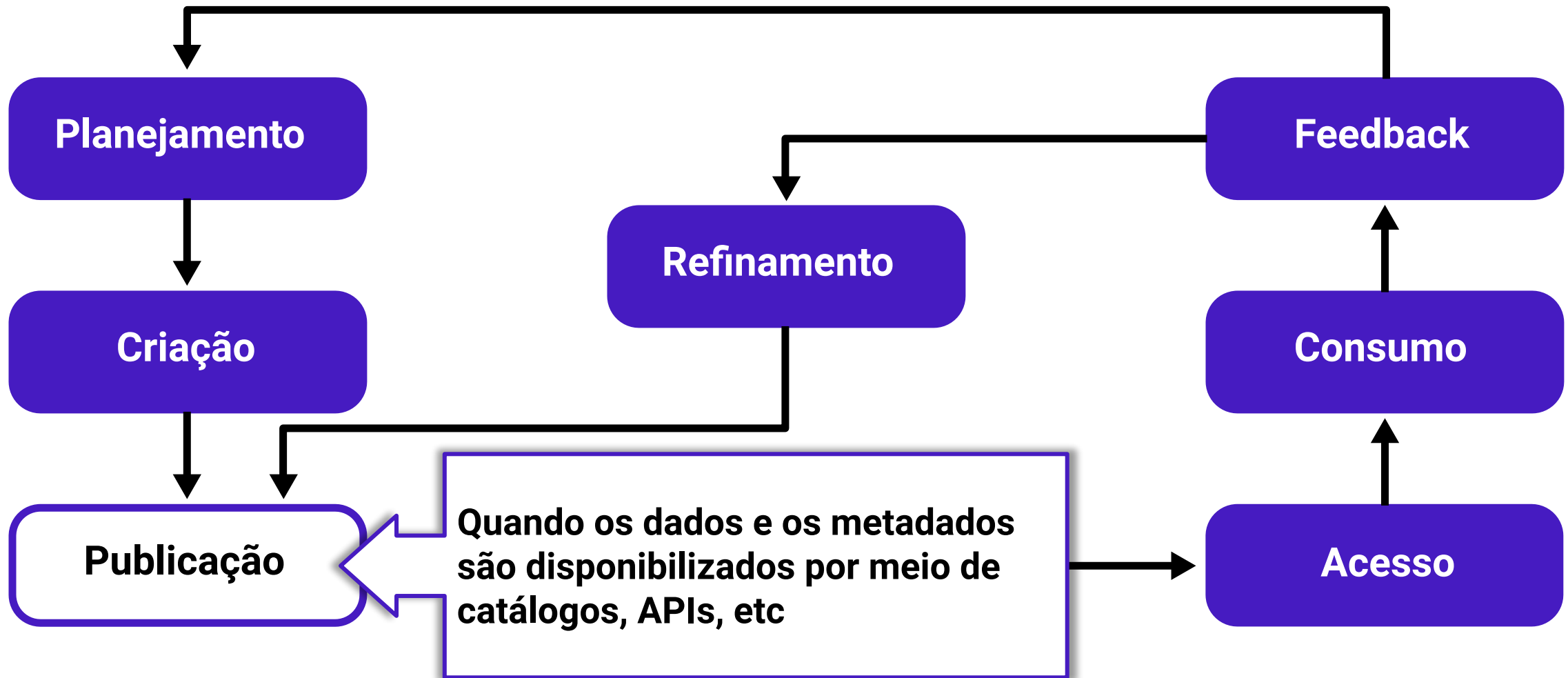
Ciclo de Vida dos Dados na Web



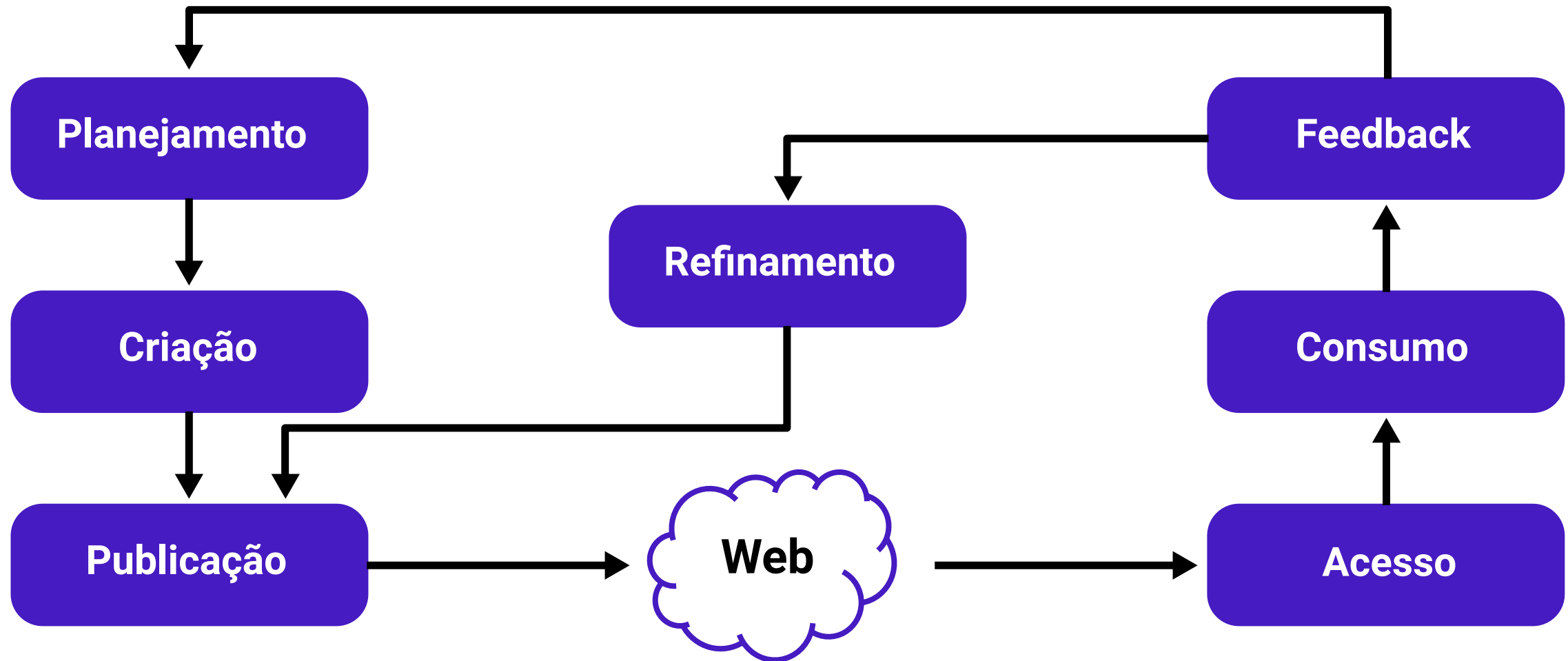
Ciclo de Vida dos Dados na Web



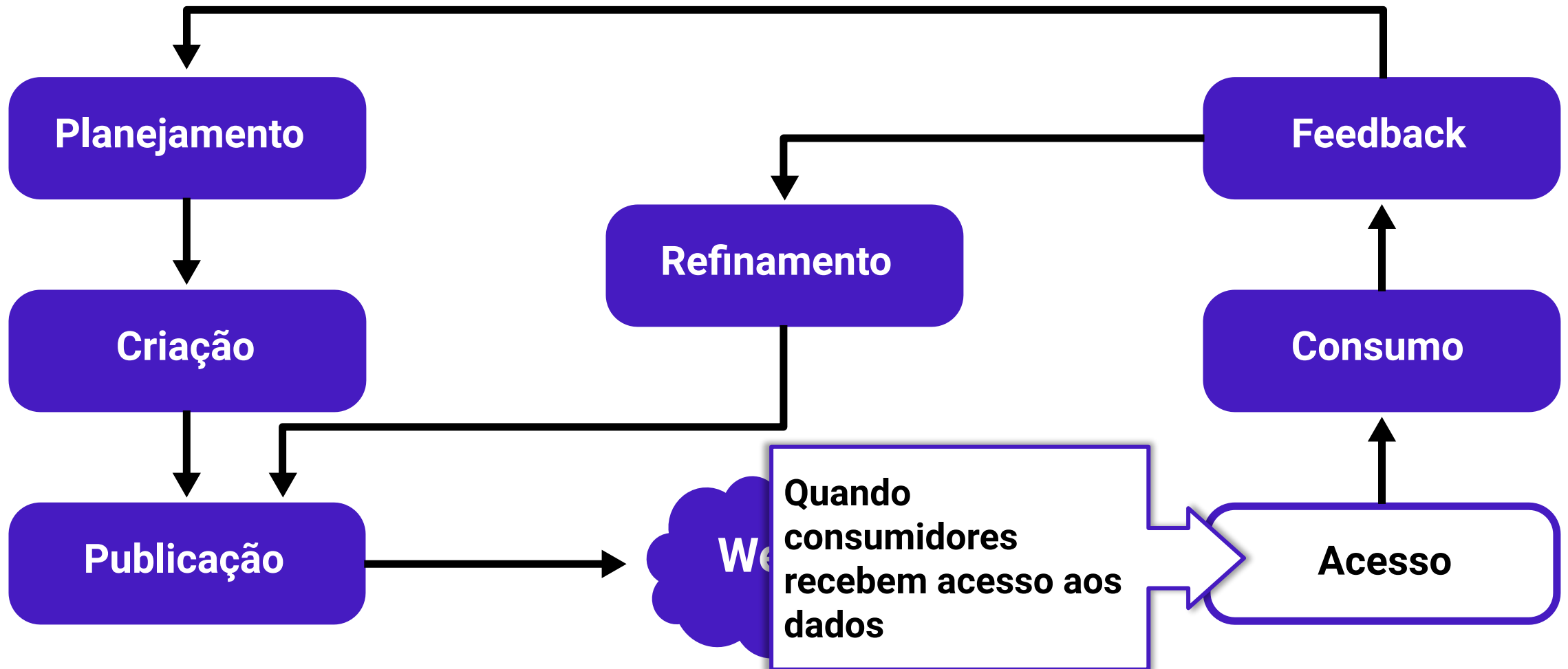
Ciclo de Vida dos Dados na Web



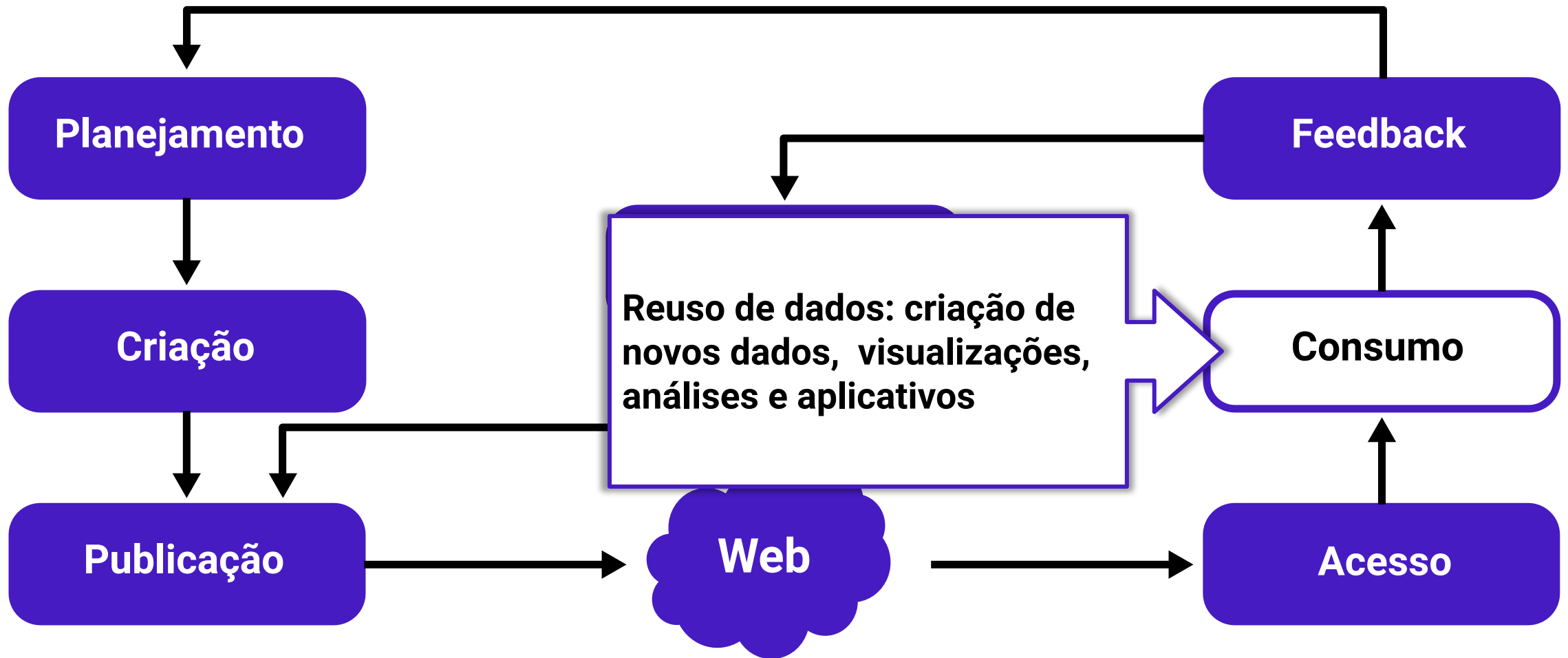
Ciclo de Vida dos Dados na Web



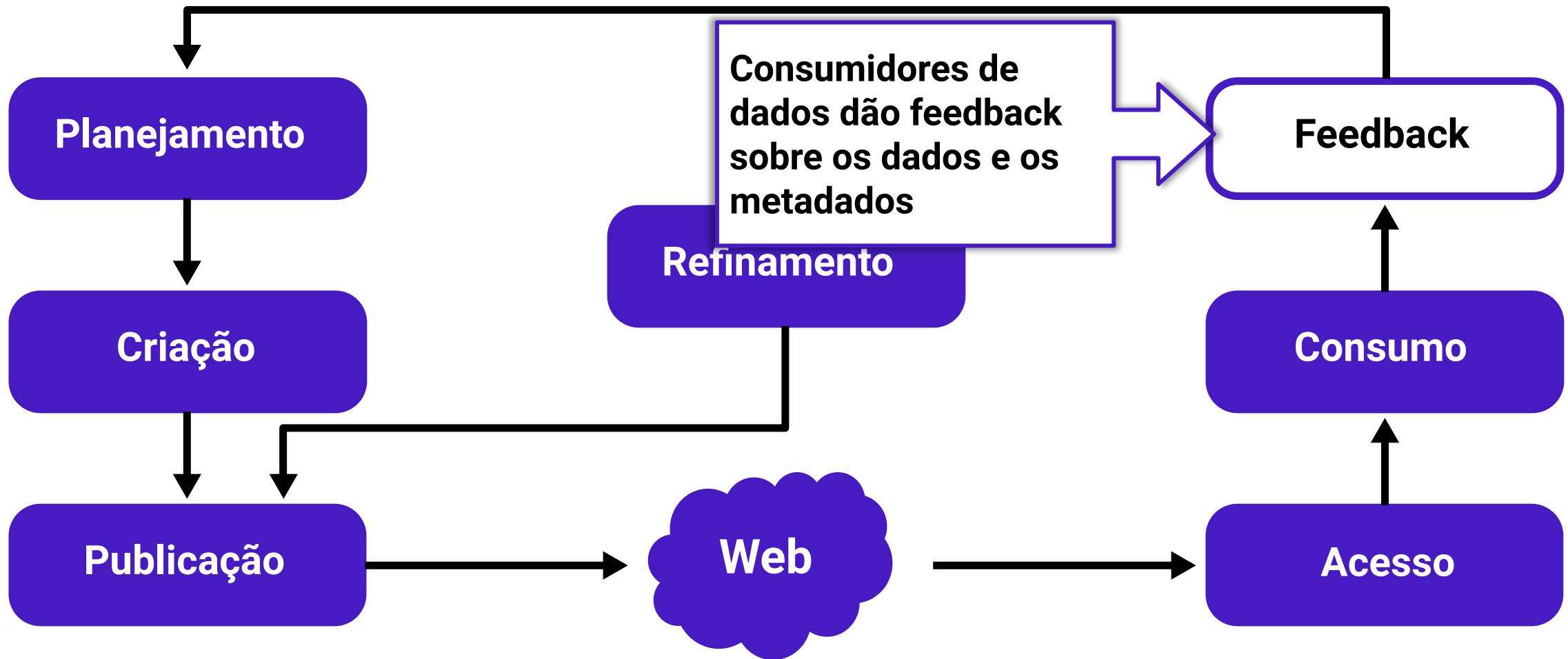
Ciclo de Vida dos Dados na Web



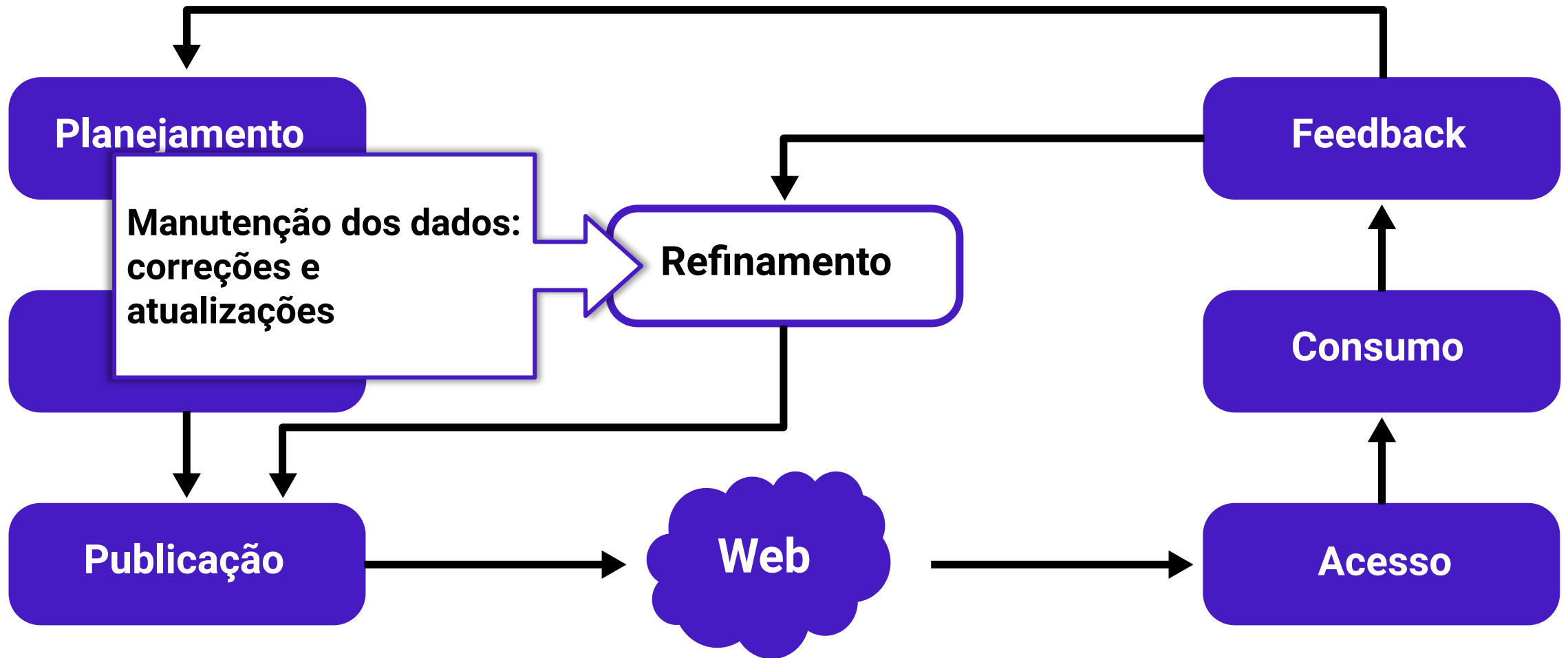
Ciclo de Vida dos Dados na Web



Ciclo de Vida dos Dados na Web



Ciclo de Vida dos Dados na Web



5 Estrelas dos Dados Abertos

<https://5stardata.info/pt-BR/>



Como possibilitar o reuso dos dados?



**Publicadores:
publicam e
compartilham
dados**

**Consumidores:
reutilizam os
dados e podem
gerar novos
dados**

View Section

All Sections

Sort By

created time

Boas Práticas para Dados na Web

Quais são os desafios para CONSUMIR dados na Web? +

Exemplo: Saber qual é a periodicidade de atualização dos dados.

+0

Quais são os desafios para PUBLICAR dados na Web? +

Exemplo: Publicar diferentes versões dos dados.

+0

- ✓ **1 dificuldade por Post-it**
- ✓ **Quantos post-its quiser**
- ✓ **5min para responder**



Data is increasingly important to society and W3C has a mature suite of Web standards with plans for further work on making it easier for average developers to work with graph data and knowledge graphs. Linked Data is about the use of URIs as names for things, the ability to dereference these URIs to get further information and to include links to other data. There are ever increasing sources of [Linked Open Data](#) on the Web, as well as data services that are restricted to the suppliers and consumers of those services.

The digital transformation of industry is seeking to exploit advanced digital technologies. This will facilitate businesses to integrate horizontally along the supply and value chains, and vertically from the factory floor to the office floor. W3C is seeking to make it easier to support enterprise wide data management and governance, reflecting the strategic importance of data to modern businesses.

<https://www.w3.org/2013/data/>





Grupo de Trabalho Boas Práticas para Dados na Web

A **Missão** do Grupo de Trabalho Boas Práticas para Dados na Web, parte da área [Data Activity](#) do W3C, foi:

1. desenvolver o **ecossistema de dados abertos**, facilitando a comunicação entre publicadores e consumidores de dados;
2. fornecer **orientação aos publicadores**, que melhorará a consistência no gerenciamento de dados e promoverá o reuso.
3. **fomentar a confiança dos consumidores** sobre os dados publicados, independente da tecnologia utilizada, aumentando o potencial para inovação.



3. **Escopo**
4. **Contexto**
5. **Espaços de Nome**
6. **Modelo de boas práticas**
7. **Sumário de Boas Práticas**
8. **As Boas Práticas**
 - 8.1 Exemplo
 - 8.2 Metadados
 - 8.3 Licença de Dados
 - 8.4 Procedência de Dados
 - 8.5 Qualidade de Dados
 - 8.6 Versionamento de Dados
 - 8.7 Identificadores de Dados
 - 8.8 Formatos de Dados
 - 8.9 Vocabulários de Dados
 - 8.10 Acesso a Dados
 - 8.10.1 APIs de Acesso a Dados
 - 8.11 Preservação de Dados
 - 8.12 *Feedback*
 - 8.13 Enriquecimento de Dados
 - 8.14 Republicação de Dados

Boas Práticas para Dados na Web

Recomendação do [W3C](https://www.w3.org/) de 31 de janeiro de 2017

Esta versão (em inglês):

<https://www.w3.org/TR/2017/REC-dwbp-20170131/>

Última versão publicada (em inglês):

<https://www.w3.org/TR/dwbp/>

Última versão do editor (em inglês):

<https://w3c.github.io/dwbp/bp.html>

Relatório de Implementação (em inglês):

<https://w3c.github.io/dwbp/dwbp-implementation-report.html>

Versão anterior (em inglês):

<https://www.w3.org/TR/2016/PR-dwbp-20161215/>

Editores:

Bernadette Farias Lóscio, [CIn - UFPE, Brazil](#)

Caroline Burle, [NIC.br, Brazil](#)

Newton Calegari, [NIC.br, Brazil](#)

Colaboradores:

Annette Greiner

Antoine Isaac

Carlos Iglesias

Carlos Laufer

Christophe Guéret

Deirdre Lee

Editores das Boas Práticas para Dados na Web

Caroline Burle, Bernadette Lóscio e Newton Calegari



DWBP: Casos de Usos



Data on the Web Best Practices Use Cases & Requirements

W3C Working Group Note 24 February 2015

This version:

<http://www.w3.org/TR/2015/NOTE-dwbp-ucr-20150224/>

Latest published version:

<http://www.w3.org/TR/dwbp-ucr/>

Latest editor's draft:

<http://w3c.github.io/dwbp/usecasesv1.html>

Previous version:

<http://www.w3.org/TR/2014/WD-dwbp-ucr-20141014/>

Editors:

[Deirdre Lee](#), [Derilinx](#) (formerly at Insight@NUIG, Ireland)

[Bernadette Farias Lóscio](#), [Centro de Informática - Universidade Federal de Pernambuco, Brazil](#)

[Phil Archer](#), [W3C/ERCIM](#)

<https://www.w3.org/TR/dwbp-ucr/>

DWBP: Casos de Usos



Table of Contents

1. Introduction
2. Use Cases
 - 2.1 ASO: Airborne Snow Observatory
 - 2.2 BBC
 - 2.3 Bio2RDF
 - 2.4 BuildingEye: SME use of public data
 - 2.5 Dados.gov.br
 - 2.6 Digital archiving of Linked Data
 - 2.7 Dutch Base Registers
 - 2.8 GS1 Digital
 - 2.9 ISO GEO Story
 - 2.10 The Land Portal
 - 2.11 LA Times' Reporting of Ron Galperin's Infographic
 - 2.12 LusTRE: Linked Thesaurus fRamework for Environment
 - 2.13 Machine-readability and Interoperability of Licenses
 - 2.14 Mass Spectrometry Imaging (MSI)
 - 2.15 OKFN Transport WG
 - 2.16 Open City Data Pipeline
 - 2.17 Open Experimental Field Studies
 - 2.18 Resource Discovery for Extreme Scale Collaboration (RDESC)
 - 2.19 Recife Open Data Portal
 - 2.20 Retrato da Violência (Violence Map)
 - 2.21 Share-PSI 2.0: Uses of Open Data Within Government for Innovation and Efficiency
 - 2.22 Tabulae - how to get value out of data
 - 2.23 UK Open Research Data Forum
 - 2.24 Uruguay Open Data Catalog
 - 2.25 Web Observatory
 - 2.26 Wind Characterization Scientific Study
3. General Challenges
 - 3.1 A Word on Open and Closed Data
 - 3.2 Requirements by Challenge
4. Requirements
 - 4.1 Requirements for Data on the Web Best Practices
 - 4.2 Requirements for Quality and Granularity Description Vocabulary
 - 4.3 Requirements for Data Usage Description Vocabulary

Como disponibilizar dados?

Como tornar os dados interoperáveis?

Como identificar recursos de dados?

Como obter *feedback*?

Publicação
de dados na
Web


Quais dados publicar?

Quais são as fontes de dados?

Quais formatos de dados utilizar?

Publicar dados na Web é mais do que apenas "publicar dados"!

Desafios de Dados na Web



Metadados (para humanos e máquinas)

Licenças de Dados (como permitir ou restringir acesso aos dados?)

Proveniência & Qualidade dos dados (como adicionar confiança aos dados?)

Versionamento dos dados (acompanhar as diferentes versões)

Identificação dos Dados (identificando datasets e distribuições)

Formatos dos Dados (quais formatos de dados utilizar?)

Desafios de Dados na Web



Vocabulários de Dados (como promover a interoperabilidade?)

Acesso aos Dados (opções de acesso aos dados)

Preservação dos Dados

Feedback (como facilitar a comunicação com usuários?)

Enriquecimento dos Dados (adicionando valor aos dados)

Republicação dos Dados (reutilizar dados com responsabilidade)

12 desafios e 42 requisitos

Desafios de Dados na Web



Metadados

Licença de Dados

Procedência de Dados

Qualidade de Dados

Versionamento de Dados

Identificadores de Dados

Formatos de Dados

Vocabulários de Dados

Acesso a Dados

Preservação de Dados

Feedback

Enriquecimento de Dados

Republicação de Dados

35 Boas Práticas para Dados na Web

[Boa Prática 1](#): Fornecer metadados

[Boa Prática 2](#): Fornecer metadados descritivos

[Boa Prática 3](#): Fornecer metadados estruturais

[Boa Prática 4](#): Fornecer informações sobre a licença de dados

[Boa Prática 5](#): Fornecer informações de procedência dos dados

[Boa Prática 6](#): Fornecer informações de qualidade de dados

[Boa Prática 7](#): Fornecer indicador de versão

[Boa Prática 8](#): Fornecer o histórico de versão

[Boa Prática 9](#): Usar URIs persistentes como identificadores de conjuntos de dados

[Boa Prática 10](#): Usar URIs persistentes como identificadores dentro de conjuntos de dados

[Boa Prática 11](#): Atribuir URIs para as versões dos conjuntos de dados e séries

[Boa Prática 12](#): Usar formatos de dados padronizados legíveis por máquinas

[Boa Prática 13](#): Usar representações de dados que sejam independentes de localidade (*locale neutral*)

[Boa Prática 14](#): Fornecer dados em formatos múltiplos

[Boa Prática 15](#): Reutilizar vocabulários, dando preferência aos padronizados

[Boa Prática 16](#): Escolher o nível de formalização adequado

[Boa Prática 17](#): Fornecer *download* em massa (*bulk download*)

[Boa Prática 18](#): Fornecer subconjuntos para conjuntos de dados extensos

[Boa Prática 19](#): Usar negociação de conteúdo para disponibilizar dados em formatos múltiplos

[Boa Prática 20](#): Fornecer acesso em tempo real

[Boa Prática 21](#): Fornecer dados atualizados

[Boa Prática 22](#): Fornecer uma explicação para os dados que não estão disponíveis

[Boa Prática 23](#): Disponibilizar dados por meio de uma API

[Boa Prática 24](#): Usar padrões Web como base para construção de APIs

[Boa Prática 25](#): Fornecer documentação completa para as APIs

[Boa Prática 26](#): Evitar alterações que afetem o funcionamento de sua API

[Boa Prática 27](#): Preservar identificadores

[Boa Prática 28](#): Avaliar a cobertura do conjunto de dados

[Boa Prática 29](#): Coletar *feedback* de consumidores de dados

[Boa Prática 30](#): Compartilhar o *feedback* disponível

[Boa Prática 31](#): Enriquecer dados por meio da geração de novos dados

[Boa Prática 32](#): Fornecer visualizações complementares

[Boa Prática 33](#): Fornecer *feedback* para o publicador original

[Boa Prática 34](#): Obedecer os termos de licença

[Boa Prática 35](#): Citar a publicação original do conjunto de dados

Benefícios das Boas Práticas

Cada benefício representa uma melhoria no modo como conjuntos de dados são disponibilizados na Web



Reuse

- BP: Provide data license information
- BP: Provide versioning information
- BP: Provide version history
- BP: Use non-proprietary data formats
- BP: Provide data in multiple formats
- BP: Use a trusted serialization format for preserved data dumps
- BP: Enrich data by generating new metadata
- BP: Provide data provenance information
- BP: Provide data quality information
- BP: Use persistent URIs as identifiers

Trustworthy

- BP: Assess dataset coverage
- BP: Assign URIs to dataset versions and series
- BP: Provide data up to date
- BP: Update the status of identifiers
- BP: Gather feedback from data consumers
- BP: Provide information about feedback
- BP: Provide data provenance information
- BP: Provide data quality information

Comprehension

- BP: Provide metadata
- BP: Provide locale parameters metadata
- BP: Provide structural metadata
- BP: Provide descriptive metadata

Accessibility

- BP: Provide bulk download
- BP: Follow REST principles when designing APIs
- BP: Provide real-time access
- BP: Maintain separate versions for a data API
- BP: Assess dataset coverage

Linkability

- BP: Use persistent URIs as identifiers
- BP: Assign URIs to dataset versions and series

Discoverability

- BP: Provide descriptive metadata
- BP: Use persistent URIs as identifiers
- BP: Assign URIs to dataset versions and series

Processability

- BP: Use machine-readable standardized data formats
- BP: Enrich data by generating new metadata

Interoperability

- BP: Use standardized terms
- BP: Re-use vocabularies



Benefícios DWBP

Cada benefício representa uma melhoria no modo como conjuntos de dados são disponibilizados na Web



Comprehension

Compreensão: os seres humanos terão um melhor entendimento sobre a estrutura e o significado dos dados, bem como dos metadados e da natureza do conjunto de dados.



Processability

Processabilidade: máquinas ou agentes de software serão capazes de processar e manipular automaticamente os dados.



Benefícios DWBP



Discoverability

Descoberta: os agentes de software serão capazes de descobrir automaticamente um conjunto de dados ou dados contidos neles.



Reuse

Reúso: as chances de reutilização do conjunto de dados por diferentes grupos de consumidores de dados tende a aumentar.



Trust

Confiança: a confiança dos consumidores em relação ao conjunto de dados tende a melhorar.



Benefícios DWBP



Linkability

Conectividade: será possível criar conexões entre conjuntos de dados e itens de dados.



Access

Facilidade de Acesso: os seres humanos e máquinas serão capazes de acessar dados atualizados em uma variedade de formas.



Interoperability

Interoperabilidade: será mais fácil chegar a um consenso entre os publicadores e consumidores de dados.



Boa Prática 1: Fornecer metadados

Fornecer metadados tanto para usuários pessoas quanto para aplicações de computadores.

Porque

Fornecer metadados é um requisito fundamental na publicação de dados na Web, porque os publicadores de dados e os consumidores de dados podem não se conhecer mutuamente. Portanto, é essencial fornecer informações que auxiliem pessoas e aplicações de computadores a compreenderem os dados, assim como outros aspectos importantes que descrevam o conjunto de dados ou a distribuição.

Resultado Pretendido

Pessoas serão capazes de compreender os metadados e as aplicações de computadores, especialmente agentes de *software* - responsáveis pela mediação do usuário com a aplicação -, serão capazes de processá-los.

Possível Abordagem para Implementação

Abordagens possíveis para fornecer *metadados legíveis por pessoas*:

- fornecer metadados como parte de uma página Web HTML
- fornecer metadados como um arquivo de texto separado

Possíveis abordagens para o fornecimento de *metadados legíveis por máquina*:

- metadados legíveis por máquinas podem ser fornecidos em um formato seriado, tal como Turtle ou JSON, ou podem ser incorporados na página de HTML utilizando [\[HTML-RDFA\]](#) ou [\[JSON-LD\]](#). No caso de múltiplos formatos serem publicados de forma separada, estes devem ser disponibilizados a partir do mesmo URL. Isto deve ser feito por meio de [negociação de conteúdo, do inglês *content negotiation* \(em inglês\)](#), disponibilizando-os sob URIs separados e distintos pelo nome de extensão do arquivo. A manutenção de múltiplos formatos é mais facilmente alcançada por meio da geração de cada formato disponível no momento, com base em uma única fonte de metadados.
- ao definir os metadados legíveis por máquinas, recomenda-se fortemente reutilizar termos oriundos de padrões e vocabulários populares já existentes. Por exemplo, os termos [\[DCTERMS\]](#) *Dublin Core Metadata* (DCMI), e o *Data Catalog Vocabulary* [\[VOCAB-DCAT\]](#) podem ser utilizados para fornecer metadados descritivos. Tais vocabulários foram elaborados para serem extremamente flexíveis, portanto geralmente é muito conveniente utilizar um *perfil* específico ou um vocabulário tal qual o [DCAT-AP \(em inglês\)](#) da Comissão Europeia.

EXEMPLO 1

Dados legíveis por pessoas

[Página de exemplo \(em inglês\)](#) apresentando uma descrição legível por pessoas de um conjunto de dados disponível.

Legível por máquina

[Arquivo de exemplo \(em inglês\)](#) apresentando uma descrição legível por máquina de um conjunto de dados disponível.

Bus stops of MyCity

This is the human-readable version of [examples](#) used in the [DWBP document](#). Please note that, as this is a fictional example, hyperlinks to the `data.mycity.example.com` domain simply link to `example.com` which defers to an explanation that the domain is designed for use in examples and serves no other function.

Dataset description

Title	Bus timetable of MyCity
URI	http://data.mycity.example.com/transport/dataset/bus/stops-2015-05-05
Keywords	transport, mobility, bus
Publication date	2015-05-05
Publisher	Transport Agency MyCity
Creator	Adrian < adrian@mycitytransport.org >
Contact point	http://data.mycity.example.com/transport/contact
Period that the dataset covers	The British calendar year of 2014
Spatial coverage	Fortaleza, Brazil
Update frequency	Annual
Theme	Mobility
Language	English, Portuguese
Date and time formats	ISO 8601
Current version	1.2

Dataset distributions

RDF Distribution

Title	RDF distribution of stops-2015-05-05 dataset
Description	RDF distribution of the stops dataset of MyCity.
Media type	text/turtle
License	CC BY-SA 3.0
Publication date	2015-05-05
Last modification	2015-05-05
Download URL	http://data.mycity.example.com/transport/dataset/bus/stops-2015-05-05.ttl

CSV Distribution

Title	CSV distribution of stops-2015-05-05 dataset
Description	CSV distribution of the bus stops dataset of MyCity.
Media type	text/csv
License	CC BY-SA 3.0
Publication date	2015-05-05
Last modification	2015-05-05
Download URL	http://data.mycity.example.com/transport/dataset/bus/stops-2015-05-05.csv

Data Quality - CSV Distribution

Completeness

Refers to the degree to which all required information is present in a particular dataset.

Metric

Ratio between the number of objects represented in the cvs and the number of objects expected to be represented according to the declared dataset scope.

Value

0.5 (Double)

Availability

Refers to the extent to which data (or some portion of it) is present, obtainable and ready for use.

Metric

dcat:downloadURL is available and its value is dereferenceable.

Value

true(Boolean)

Structural metadata - CSV Distribution

Field	Titles	Description	Datatype	Primary key	Required
stop_id	Identifier	An identifier for the bus stop.	string	true	true
stop_name	Name	The name of the bus stop.	string		
stop_desc	Description	A description for the bus stop.	string		
stop_lat	Latitude	The latitude of the bus stop.	numeric		
stop_long	Longitude	The longitude of the bus stop.	numeric		
zone_id	Zone	An identifier for the zone where the bus stop is located.	string		
stop_url	URL	URL that identifies the bus stop.	string		

Dataset version history

Version	Issue date	URI	Change history
1.1	2015-05-05	http://data.mycity.example.com/transport/dataset/bus/stops-2015-05-05	The bus stops dataset was updated to reflect the creation of a new bus stop at 1115 Pearl Street.
1.2	2016-05-05	http://data.mycity.example.com/transport/dataset/bus/stops-2016-05-05	The bus stops dataset was updated to reflect the creation of new bus stops since the last update and to follow the update frequency

Feedback

Feedback form

Choose a dataset or distribution:

CSV distribution of bus-stops-2015-05-05 dataset

How do you classify the dataset or distribution?

5

Which kind of feedback would you like to give?

Correction Question Review

Your feedback:

Is there a JSON-LD distribution available for this dataset?

Como Testar

Verifique se os metadados legíveis por pessoas estão disponíveis.

Verifique se os metadados estão disponíveis em um formato válido legível por máquina e sem erros de sintaxe.

Evidências

Requisitos Relevantes: [R-MetadataAvailable \(em inglês\)](#), [R-MetadataDocum \(em inglês\)](#), [R-MetadataMachineRead \(em inglês\)](#)

Benefícios



Reúso
(Reuse)



Compreensão
(Comprehension)



Facilidade de Descoberta
(Discoverability)



Facilidade de Processamento
(Processability)



Boa Prática 4: Fornecer informações sobre a licença de dados

Fornecer um link ou uma cópia dos termos da licença que controla a utilização dos dados.

Porque

Disponibilizar informações de licença é fundamental para que os consumidores de dados avaliem a utilidade dos dados. Os agentes de *software* podem utilizar a presença/ausência das informações de licença como um gatilho para a inclusão ou exclusão de dados apresentados a um consumidor em potencial.

Resultado Pretendido

Pessoas serão capazes de compreender a licença de dados, que descreve eventuais restrições impostas à utilização de certos dados, e agentes de *software* serão capazes de detectar automaticamente a licença de dados de uma distribuição.

Possível Abordagem para Implementação

As informações sobre a licença de dados podem ser disponibilizadas por meio de um link ou de uma cópia embutida dos termos da licença que seja legível por pessoas. Também podem ser disponibilizadas para processamento um link ou cópia embutida dos termos da licença legível por máquina.

Os seguintes vocabulários, que incluem propriedades para vincular uma licença, podem ser usados:

- Dublin Core [[DCTERMS](#)] (`dct:license`)
- *Creative Commons* [[CCREL](#)] (`cc:license`)
- schema.org [[SCHEMA-ORG](#)] (`schema:license`)
- XHTML [[XHTML-VOCAB](#)] (`xhtml:license`)

Também existe uma série de linguagens legíveis por máquinas que expressam direitos de propriedade intelectual, tais como:

- *A Creative Commons Rights Expression Language* [[CCREL](#)]
- *A Open Digital Rights Language* [[ODRL-model](#)]
- *A Open Data Rights Statement Vocabulary* [[ODRS](#)]

EXEMPLO 4

Legível por máquina

A distribuição CSV de um conjunto de dados referentes aos pontos de ônibus ([stops-2015-05-05.csv](#)) será publicada sob a licença [Creative Commons Attribution-ShareAlike 3.0 Unported \(em inglês\)](#). A licença proprietária `dct:license` é utilizada para incluir esta informação como parte da distribuição dos metadados. Neste caso a licença não é escrita em uma linguagem legível por máquinas, porém a licença proprietária `dct:license` permite que um agente de *software* detecte automaticamente a licença de dados de uma certa distribuição.

```
:stops-2015-05-05.csv
  a dcat:Distribution ;
  dct:title "CSV distribution of stops-2015-05-05 dataset" ;
  dct:description "CSV distribution of the bus stops dataset of MyCity" ;
  dcat:mediaType "text/csv;charset=UTF-8" ;
  dct:license <http://creativecommons.org/licenses/by-sa/3.0/>
  .
```

Legível por pessoas

[Página de exemplo \(em inglês\)](#) com informações de licença da distribuição de dados legíveis por pessoas.

CSV Distribution

Title	CSV distribution of stops-2015-05-05 dataset
Description	CSV distribution of the bus stops dataset of MyCity.
Media type	text/csv
License	CC BY-SA 3.0
Publication date	2015-05-05
Last modification	2015-05-05
Download URL	http://data.mycity.example.com/transport/dataset/bus/stops-2015-05-05.csv

Data Quality - CSV Distribution

Completeness	Refers to the degree to which all required information is present in a particular dataset.
Metric	Ratio between the number of objects represented in the cvs and the number of objects expected to be represented according to the declared dataset scope.
Value	0.5 (Double)
Availability	Refers to the extent to which data (or some portion of it) is present, obtainable and ready for use.
Metric	dcat:downloadURL is available and its value is dereferenceable.
Value	true(Boolean)

Como Testar

Verifique se os metadados do conjunto de dados incluem as informações de licença de dados em um formato legível por pessoas.

Verifique se um agente de *software* pode detectar ou descobrir de forma automática a licença do conjunto de dados.

Evidência

Casos de uso relevantes: [R-LicenseAvailable \(em inglês\)](#), [R-MetadataMachineRead \(em inglês\)](#), [R-LicenseLiability \(em inglês\)](#)

Benefícios



Reúso
(Reuse)



Confiança
(Trust)

Boa Prática 6: Fornecer informações de qualidade de dados

Fornecer informações sobre a qualidade dos dados e adequação para fins específicos.

Porque

A qualidade dos dados pode afetar seriamente o uso dos dados para aplicações específicas, incluindo aplicações bem diferentes do propósito para o qual estas foram originalmente criadas. Documentar a qualidade dos dados facilita significativamente o processo de seleção de conjuntos de dados, aumentando as chances de reutilização dos mesmos. Independentemente das peculiaridades específicas do domínio, a qualidade dos dados deve ser documentada e os problemas de qualidade que são conhecidos devem ser declarados de forma explícita nos metadados.

Resultado Pretendido

Pessoas e agentes de *software* serão capazes de avaliar a qualidade e, portanto, a adequação de um conjunto de dados para a sua aplicação.

Possível Abordagem para Implementação

A versão legível por máquina dos metadados de qualidade do conjunto de dados pode ser fornecida utilizando o vocabulário *Data Quality Vocabulary* desenvolvido pelo grupo de trabalho DWBP [[VOCAB-DQV](#)].

EXEMPLO 6

Legível por máquina

O exemplo abaixo demonstra os metadados para a distribuição CSV do conjunto de dados de paradas de ônibus com a inclusão de metadados de qualidade de dados. Os metadados foram definidos de acordo com o vocabulário *Data Quality Vocabulary*. Mais exemplos podem ser encontrados no documento *Data Quality Vocabulary* [[VOCAB-DQV](#)].

```
:stops-2015-05-05.csv
  a dcat:Distribution ;
  dcat:downloadURL <http://data.mycity.example.com/transport/dataset/bus/stop> ;
  dct:title "CSV distribution of stops-2015-05-05 dataset" ;
  dct:description "CSV distribution of the bus stops dataset of MyCity" ;
  dcat:mediaType "text/csv;charset=UTF-8" ;
  dct:license <http://creativecommons.org/licenses/by-sa/3.0/> ;
  dqv:hasQualityMeasurement :measure1, :measure2
  .
:measure1
  a dqv:QualityMeasurement ;
  dqv:computedOn :stops-2015-05-05.csv ;
  dqv:isMeasurementOf :downloadURLAvailabilityMetric ;
  dqv:value "true"^^xsd:boolean
  .
:measure2
  a dqv:QualityMeasurement ;
  dqv:computedOn :stops-2015-05-05.csv ;
  dqv:isMeasurementOf :csvCompletenessMetric ;
  dqv:value "0.5"^^xsd:double
```

```
#definition of dimensions and metrics
```

```
  :availability
```

```
    a dqv:Dimension ;
```

```
    skos:prefLabel "Availability"@en ;
```

```
    skos:definition "Availability of a dataset is the extent to which data (or  
    dqv:inCategory :accessibility
```

```
    .
```

```
  :completeness
```

```
    a dqv:Dimension ;
```

```
    skos:prefLabel "Completeness"@en ;
```

```
    skos:definition "Completeness refers to the degree to which all required in  
    dqv:inCategory :intrinsicDimensions
```

```
    .
```

```
  :downloadURLAvailabilityMetric
```

```
    a dqv:Metric ;
```

```
    skos:definition "It checks if dcat:downloadURL is available and if its valu  
    dqv:inDimension :availability
```

```
    .
```

```
  :csvCompletenessMetric
```

```
    a dqv:Metric ;
```

```
    skos:definition "Ratio between the number of objects represented in the cvs  
    dqv:inDimension :completeness
```

```
    .
```

CSV Distribution

Title	CSV distribution of stops-2015-05-05 dataset
Description	CSV distribution of the bus stops dataset of MyCity.
Media type	text/csv
License	CC BY-SA 3.0
Publication date	2015-05-05
Last modification	2015-05-05
Download URL	http://data.mycity.example.com/transport/dataset/bus/stops-2015-05-05.csv

Data Quality - CSV Distribution

Completeness	Refers to the degree to which all required information is present in a particular dataset.
Metric	Ratio between the number of objects represented in the cvs and the number of objects expected to be represented according to the declared dataset scope.
Value	0.5 (Double)
Availability	Refers to the extent to which data (or some portion of it) is present, obtainable and ready for use.
Metric	dcat:downloadURL is available and its value is dereferenceable.
Value	true(Boolean)

Como Testar

Verifique se os metadados do conjunto de dados incluem informações sobre a qualidade deste determinado conjunto de dados.

Verifique se uma aplicação de computador pode processar automaticamente as informações sobre a qualidade do conjunto de dados.

Evidência

Requisitos Relevantes: [R-QualityMetrics \(em inglês\)](#), [R-DataMissingIncomplete \(em inglês\)](#), [R-QualityOpinions \(em inglês\)](#)

Benefícios



Reúso
(Reuse)



Confiança
(Trust)

Boa Prática 9: Usar URIs persistentes como identificadores de conjuntos de dados

Identificar cada conjunto de dados por meio de um URI persistente e cuidadosamente escolhido.

Porque

Adotar um sistema de identificação comum permite a identificação básica dos dados e a comparação dos processos por qualquer um dos atores envolvidos de forma confiável. São pré-condições essenciais para o gerenciamento e reutilização dos dados de forma adequada.

Os desenvolvedores podem construir URIs dentro de seus códigos e, para isso, é importante que tais URIs sejam persistentes e que desreferenciem para o mesmo recurso ao longo do tempo, sem a necessidade de intervenção humana.

Resultado Pretendido

Os conjuntos de dados ou as informações sobre os conjuntos de dados serão encontrados e citados com facilidade em qualquer momento, independentemente do status, da disponibilidade ou do formato dos dados.

Possível Abordagem para Implementação

Para serem persistentes os URIs devem ser designadas como tal. Muito já foi escrito sobre este tópico, veja por exemplo o Estudo da Comissão Europeia sobre URIs Persistentes [[PURI](#)] que por sua vez propõe conexões a muitos outros recursos.

No caso de um publicador de dados não ser capaz ou estiver relutante em gerenciar um espaço de URI diretamente para persistência, uma abordagem alternativa seria utilizar um serviço de redirecionamento tal qual [Permanent Identifiers for the Web \(em inglês\)](#) ou [purl.org \(em inglês\)](#). Estes oferecem URIs persistentes que podem ser redirecionados conforme necessário de forma que a eventual localização possa ser efêmera. O [software por trás destes serviços \(em inglês\)](#) encontra-se disponível gratuitamente e, portanto, pode ser instalado e gerenciado localmente caso necessário.

Identificadores de Objetos Digitais ([DOIs, em inglês](#)) oferecem uma alternativa similar. Estes identificadores são definidos independentemente de qualquer tecnologia Web, mas podem ser anexados a um *stub* URI, uma ponta de URI. Os DOIs são parte importante da infraestrutura digital para a pesquisa de dados e bibliotecas.

EXEMPLO 9

O URI <https://data.mycity.example.com/transport/dataset/bus/stops> possui diversas características que oferecem suporte à persistência:

- Todos os nomes estão sujeitos a mudanças ao longo do tempo, no entanto ao escolher um nome de domínio é razoável que Adrian assuma que MyCity vá continuar a existir e que continuará a ter um governo. Portanto, enquanto casos como a Iugoslávia comprovam que até mesmo nomes de países mudam e que domínios de nível superior podem desaparecer (tal como .yu), um nome de domínio baseado no nome da própria cidade é tão persistente quanto qualquer nome de domínio pode ser.
- Ao inserir dados no subdomínio data.mycity.example.com, Adrian está criando um domínio específico que pode ser gerenciado independentemente de qualquer departamento em particular.
- Não é seguro assumir que um *departamento* específico vá persistir. As autoridades em MyCity podem muito bem decidir que o Departamento de Transportes deve fundir-se com outro departamento e criar, por exemplo, o Departamento de Transportes e Meio Ambiente. Está correto, portanto, não inserir o nome do Departamento de Transportes no URI, mas sim inserir a tarefa para a qual os dados são designados, neste caso o fornecimento de transporte público.

- O seguimento de caminho (do inglês *path segment*) referente ao `/dataset` é um indicativo de que o URI identifica um conjunto de dados em vez de, por exemplo, uma rota de ônibus específica.
- Da mesma forma o segmento de caminho `/bus` nos leva mais além em direção ao conjunto de dados específico pela qual Adrian é responsável.
- Finalmente, `/stops` nos conduz ao conjunto de dados que diz respeito às paradas de ônibus em MyCity.
- Nos termos do vocabulário DCAT este seria o identificador para o conjunto de dados. Distribuições específicas do conjunto de dados provavelmente serão identificadas pela adição de extensões de arquivo relevantes ao URI, tais como: `https://data.mycity.example.com/transport/dataset/bus/stops.csv`, `https://data.mycity.example.com/transport/dataset/bus/stops.json`, `https://data.mycity.example.com/transport/dataset/bus/stops.ttl` etc.

Como Testar

Verifique se cada conjunto de dados encontra-se identificado utilizando um URI que tenha sido designado para ser persistente. O ideal é que o sítio Web relevante inclua uma descrição de uma esquema de design e uma proposta plausível de persistência caso o publicador não puder manter o espaço URI por si só.

Evidência

Requisitos Relevantes: [R-UniqueIdentifier \(em inglês\)](#), [R-Citable \(em inglês\)](#)

Benefícios



Reúso
(Reuse)



Capacidade de Conexão
(Linkability)



Facilidade de Descoberta
(Discoverability)



Interoperabilidade
(Interoperability)

Boa Prática 14: Fornecer dados em formatos múltiplos

Disponibilize os dados em vários formatos quando mais de um formato for adequado ao potencial ou pretendido uso do dado.

Porque

Disponibilizar dados em mais de um formato reduz os custos decorrentes da transformação de dados e minimiza a possibilidade de erros no processo de transformação. Caso muitos usuários precisem transformar os dados em um formato específico, publicá-los neste formato desde o início poupa tempo, dinheiro e evita erros de uma forma mais eficiente. Por último, aumenta o número de ferramentas e aplicações que podem processar os dados.

Resultado Pretendido

O maior número possível de usuários será capaz de utilizar os dados sem ter que primeiramente transformá-los para seu formato de preferência.

Possível Abordagem para Implementação

Considere os formatos de dados que provavelmente serão necessários e também alternativas que possivelmente serão úteis no futuro. Os publicadores de dados devem equilibrar o esforço necessário para disponibilizar os dados em muitos formatos com relação ao custo de fazê-lo; no entanto, fornecer pelo menos uma alternativa aumentará significativamente a utilidade dos dados. Para disponibilizar dados em mais de um formato você pode utilizar negociação de conteúdo conforme descrito na [Boa Prática Usar negociação de conteúdo para disponibilizar dados em formatos múltiplos](#).

Um aviso: identificadores de localidade dentro de conjuntos de dados, que podem ser expostos como identificadores de fragmentos em URIs, devem ser consistentes nos vários formatos.

EXEMPLO 14

Para alcançar uma ampla gama de consumidores de dados, Adrian decide também fornecer uma distribuição JSON do conjunto de dados dos pontos de ônibus. No exemplo a seguir a propriedade `dcat:distribution` é utilizada para associar o conjunto de dados `stops-2015-05-05` com suas duas distribuições: `stops-2015-05-05.csv` e `stops-2015-05-05.json`.

```
:stops-2015-05-05
  a dcat:Dataset ;
  dcat:distribution :stops-2015-05-05.csv ;
  dcat:distribution :stops-2015-05-05.json
  .
:stops-2015-05-05.csv
  a dcat:Distribution ;
  dcat:downloadURL <http://data.mycity.example.com/transport/dataset/bus/stop
  dct:title "CSV distribution of stops-2015-05-05 dataset" ;
  dct:description "CSV distribution of the bus stops dataset of MyCity" ;
  dcat:mediaType "text/csv;charset=UTF-8" ;
  dct:license <http://creativecommons.org/licenses/by-sa/3.0/>
  .
:stops-2015-05-05.json
  a dcat:Distribution ;
  dcat:downloadURL <http://data.mycity.example.com/transport/dataset/bus/stop
  dct:title "JSON distribution of stops-2015-05-05 dataset" ;
  dct:description "JSON distribution of the bus stops dataset of MyCity" ;
  dcat:mediaType "application/json" ;
  dct:license <http://creativecommons.org/licenses/by-sa/3.0/>
  .
```

Como Testar

Verifique se o conjunto de dados completo está disponível em mais de um formato de dados.

Evidência

Requisitos Relevantes: [R-FormatMultiple \(em inglês\)](#)

Benefícios



Reúso
(Reuse)



Facilidade de Processamento
(Processability)

Boa Prática 15: Reutilizar vocabulários, dando preferência aos padronizados

Utilize termos oriundos de vocabulários compartilhados, preferencialmente os padronizados, para codificar dados e metadados.

Porque

A utilização de vocabulários já em uso por outros estimula e facilita o consenso em comunidades. Aumenta a interoperabilidade e reduz as redundâncias, incentivando assim a reutilização de seus próprios dados. Particularmente, a aplicação de vocabulários compartilhados para metadados (especialmente os metadados estruturais, de procedência, de qualidade e de versionamento) auxilia o processo de comparação e o processamento automático - tanto dos dados quanto dos metadados. Além disso, a referência a códigos e termos padronizados ajuda a evitar ambiguidade e conflitos entre elementos e valores similares.

Resultado Pretendido

Melhorar a interoperabilidade e o consenso entre os publicadores e consumidores de dados.

Possível Abordagem para Implementação

A seção de [Vocabulários \(em inglês\)](#) no documento *Best Practices for Publishing Linked Data* [[LD-BP](#)] do [W3C](#) fornece orientação sobre a descoberta, avaliação e seleção de vocabulários existentes.

Organizações tais como a *Open Geospatial Consortium* (OGC), [ISO](#), [W3C](#), [WMO](#), bibliotecas e serviços de pesquisa de dados, etc., disponibilizam listas de códigos, terminologias e vocabulários de Dados Conectados que podem ser utilizados por todos. Uma questão fundamental é garantir que o conjunto de dados, ou sua documentação, forneça contextualização suficiente (tanto em formatos legíveis por pessoas e por máquinas) para que os consumidores de dados possam recuperar e explorar o significado padronizado dos valores. No contexto da Web uma forma eficiente de fazer isto é utilizar identificadores (URIs) baseados na Web para recursos de vocabulário padronizado, tendo em mente que o mesmo URI pode ter rótulos multilíngues anexados para uma maior interoperabilidade em diferentes países. O glossário multilíngue da União Europeia, [Eurovoc \(em inglês\)](#), fornece um exemplo excelente.

EXEMPLO 15

1. O vocabulário DCAT expressa os metadados referentes aos conjuntos de dados [[VOCAB-DCAT](#)] e reutiliza elementos de diversos vocabulários pré-existentes: Dublin Core, FOAF, SKOS e vCard. Reutilizar as propriedades do Dublin Core, tais como `dct:title`, em vez de criar novos (`dcat:title`, por exemplo) possibilita que os metadados baseados em DCAT possam ser consumidos por qualquer aplicação que seja capaz de ler e manipular as declarações do Dublin Core.
2. No setor da cultura digital, o modelo de dados usado pela [Europeana \(EDM, em inglês\)](#) também faz uso extensivo de vocabulários compartilhados existentes, como Dublin Core, FOAF, SKOS, etc. Isto facilitou a adoção do EDM pelos fornecedores de dados da Europeana e auxiliou a posicioná-la como uma Boa Prática para iniciativas similares no mesmo setor. Por exemplo, o [perfil de aplicação de metadados \(em inglês\)](#) da [Digital Public Library of America \(em inglês\)](#) reutiliza o EDM e, portanto, emprega os diversos vocabulários em que o EDM se baseia. Como resultado, grandes quantidades de dados da cultura digital tornaram-se mais interoperáveis dentro do setor. Estes dados também são mais fáceis de reutilizar por consumidores de outras comunidades que não estejam familiarizados como os modelos e as terminologias tradicionais utilizados pelas bibliotecas, arquivos e museus.
3. A Biblioteca do Congresso dos Estados Unidos publica listagens de idiomas [ISO 639](#) como Dados Conectados (consulte [[ISO639-1-LOC](#)] para códigos de dois dígitos):

```
:stops
```

```
dct:language <http://id.loc.gov/vocabulary/iso639-1/en> .
```


4. A rede australiana [Solid Earth and Environment Grid \(em inglês\)](#) publica uma listagem de referências de URIs para elementos de escalas temporais geológicas da *International Commission on Stratigraphy's Chronostratigraphic Chart* tais como:
<https://resource.geosciml.org/classifier/ics/ischart/Paleozoic> (em inglês) para a Era Paleozóica:

```
:dataset-005 a dcat:Dataset ;  
    dct:temporal <http://resource.geosciml.org/classifier/ics/ischart/Paleozoic
```

5. O Google mantém a especificação [General Transit Feed Specification](#) que define um formato para a publicação de dados de transporte público. Este formato baseia-se em um conjunto de campos tais como `nome_atalho_rota` (do inglês `route_short_name`) ou `tipo_rota` (do inglês `route_type`), que são cuidadosamente definidos e expostos a constantes *feedbacks* da comunidade para facilitar o consenso. As definições incluem especificações de valores codificados tais como os utilizados em `tipo_rota`:

```
0 - Bondes, veículos leves sobre trilhos. Qualquer sistema de veículo leve sobre trilhos.  
1 - Metrô, transporte subterrâneo. Qualquer tipo de sistema de transporte por trilhos.  
2 - Trilhos. Utilizados para transporte intermunicipal e viagem de longa distância.
```

Observe que em uma modalidade que não seja de Dados Conectados, estes campos e códigos não possuem identificadores Web individuais, nem tampouco uma semântica legível por máquinas. Portanto, explorá-los demanda que os implementadores analisem a documentação e codifiquem interpretações em cada aplicação individual que consome os dados.

Como Testar

Utilize repositórios de vocabulários como o repositório [Linked Open Vocabularies \(em inglês\)](#) ou listagens de serviços mencionados em documentos de Boas Práticas específicas de tecnologia, tais como o [Best Practices for Publishing Linked Data \[LD-BP\]](#) ou o [Core Initial Context for RDFa and JSON-LD \(em inglês\)](#); certifique-se de que as classes, propriedades, termos, elementos ou atributos utilizados para representar um conjunto de dados não repliquem aqueles definidos por vocabulários utilizados para outros conjuntos de dados.

Verifique se os termos ou códigos no vocabulário a serem utilizados estão definidos em uma organização de desenvolvimento de padrões tal como IETF, OGC e W3C etc., ou tenham sido publicados por uma autoridade adequada, tais como agências governamentais.

Evidência

Requisitos Relevantes: [R-MetadataStandardize \(em inglês\)](#), [R-MetadataDocum \(em inglês\)](#), [R-QualityComparable \(em inglês\)](#), [R-VocabOpen \(em inglês\)](#), [R-VocabReference \(em inglês\)](#)

Benefícios



Reúso
(Reuse)



Facilidade de Processamento
(Processability)



Compreensão
(Comprehension)



Confiança
(Trust)



Interoperabilidade
(Interoperability)



Boa Prática 23: Disponibilizar dados por meio de uma API

Disponibilizar uma API para servir os dados caso você tenha recursos para tanto.

Porque

Uma API oferece aos consumidores de seus dados maior flexibilidade e facilidade de processamento. Ela pode habilitar o uso de dados em tempo real, realizar filtragens a partir de solicitações e permite trabalhar com os dados em um nível atômico. Caso o seu conjunto de dados seja grande, frequentemente atualizado ou altamente complexo, é provável que uma API seja a melhor opção para publicar seus dados.

Resultado Pretendido

Os desenvolvedores terão acesso programático aos dados para usar em suas próprias aplicações, com os dados atualizados e sem exigir esforço por parte dos consumidores. As aplicações Web terão a capacidade de obter dados específicos consultando uma interface programática.

Possível Abordagem para implementação

Criar uma API é um pouco mais complexo do que publicar dados para *download*. Demanda algum conhecimento de como construir uma aplicação Web. No entanto, não é necessário construir uma a partir do zero. Caso uma plataforma de gerenciamento de dados seja utilizada, tal como a CKAN, você poderá habilitar uma API já existente. Muitas estruturas de desenvolvimento Web incluem suporte para APIs e também disponibilizam estruturas projetadas especificamente para a construção de APIs personalizadas.

Rails (Ruby), Django (Python) e Express (NodeJS) são alguns exemplos de estruturas de desenvolvimento Web que oferecem suporte para criação de APIs. Exemplos de estruturas API incluem Swagger, Apigility, Restify e Restlet.

EXEMPLO 23

Além de fornecer *downloads* de dados em massa sobre transporte público, Adrian decide oferecer um mecanismo mais flexível de acesso aos dados. Para tanto ele desenvolve uma API para oferecer acesso aos pontos de ônibus, às rotas de ônibus e às informações em tempo real sobre os pontos de ônibus. Consulte os [exemplos deste uso \(em inglês\)](#).

DWBP API Documentation

API URL: `http://data.mycity.example.com/transport/api/`

List all bus routes

GET `http://data.mycity.example.com/transport/api/bus/routes`

Response

```
{
  "routes": [
    {"id": "1", "uri": "http://data.mycity.example.com/transport/bus/routes/1"},
    {"id": "3", "uri": "http://data.mycity.example.com/transport/bus/routes/3"},
    {"id": "5", "uri": "http://data.mycity.example.com/transport/bus/routes/5"},
    {"id": "8", "uri": "http://data.mycity.example.com/transport/bus/routes/8"},
    {"id": "10", "uri": "http://data.mycity.example.com/transport/bus/routes/10"},
    {"id": "12a", "uri": "http://data.mycity.example.com/transport/bus/routes/12"}
  ]
}
```

List all bus stops

GET `http://data.mycity.example.com/transport/api/bus/stops`

Response using fields defined in [dwbp-example.html](#)

Como Testar

Verifique se um cliente de teste pode simular chamadas e se a API responde de acordo com o previsto.

Evidência

Requisitos Relevantes: R-AccessRealTime (em inglês), R-AccessUpToDate (em inglês)

Benefícios



Reúso
(Reuse)



Facilidade de Processamento
(Processability)



Interoperabilidade
(Interoperability)



Facilidade de Acesso
(Access)

Boa Prática 29: Coletar feedback de consumidores de dados

Fornecer meios fáceis de se encontrar para que os consumidores deem feedback.

Porque

Receber *feedbacks* ajuda os publicadores a compreenderem as necessidades de seus consumidores de dados, além de auxiliá-los a melhorar a qualidade dos dados publicados. Também aumenta a confiança na medida em que demonstra aos consumidores que o publicador se importa e se preocupa em atender às suas necessidades. Especificar claramente um mecanismo de *feedback* remove a inconveniência para o consumidor de dados de ter que procurar uma maneira de fornecer *feedbacks*.

Resultado Pretendido

Os consumidores de dados poderão fornecer *feedbacks* e avaliações sobre os conjuntos de dados e suas distribuições.

Possível Abordagem para implementação

Forneça aos consumidores de dados um ou mais mecanismos para o envio de *feedbacks*, incluindo (mas não limitando a) um formulário de contato, botões para ranquear a qualidade de dados ou uma caixa de comentários. Para aproveitar ao máximo os *feedbacks* recebidos de consumidores, uma boa ideia é coletá-lo por meio de um sistema de rastreamento que captura cada item em uma base de dados, assim permitindo quantificação e análise. Outra boa ideia é a captura por tipo de item de *feedbacks* por exemplo, sua motivação (edição, classificação [avaliação], comentário ou questionamento), de maneira que cada item possa ser expresso utilizando o vocabulário *Dataset Usage Vocabulary* [[VOCAB-DUV](#)].

EXEMPLO 29

Exemplo de formulário de *feedback* (em inglês)

Como Testar

Verifique se pelo menos um mecanismo de *feedback* foi fornecido e se este pode ser facilmente encontrado por consumidores de dados

Evidência

Requisitos Relevantes: [R-UsageFeedback \(em inglês\)](#), [R-QualityOpinions \(em inglês\)](#)

Benefícios



Reúso
(*Reuse*)



Compreensão
(*Comprehension*)



Confiança
(*Trust*)

Boa Prática 30: Compartilhar o feedback disponível

Disponibilizar publicamente mecanismos de feedback de consumidor sobre conjuntos de dados e distribuições.

Porque

Ao compartilhar *feedback* com consumidores, os publicadores demonstram aos usuários que suas contribuições estão sendo levadas em conta, e podem evitar o envio de relatórios de erros duplicados. Compartilhar *feedback* também ajuda os consumidores a compreender quaisquer questões que possam afetar sua capacidade de utilizar dados, assim como estimula um sentimento de comunidade entre eles.

Resultado Pretendido

Os consumidores serão capazes de avaliar os tipos de erros que afetam o conjunto de dados, analisar as experiências de outros usuários com o mesmo, e assegurar-se de que o publicador está abordando ativamente os problemas conforme o necessário. Também poderão determinar se outros usuários já forneceram *feedbacks* semelhantes, poupando-os do trabalho de enviar relatórios desnecessários e evitando que os responsáveis tenham que lidar com duplicidade.

Possível Abordagem para implementação

O *feedback* pode ficar disponível como parte de uma página Web em HTML, mas também pode ser fornecido em um formato legível por máquinas utilizando o vocabulário *Dataset Usage Vocabulary* [[VOCAB-DUV](#)].

EXAMPLE 30

```
:stops-2015-05-05
  a dcat:Dataset ;
  dct:title "Bus stops of MyCity" ;
  dcat:keyword "transport","mobility","bus" ;
  dct:issued "2015-05-05"^^xsd:date ;
  dcat:contactPoint <http://data.mycity.example.com/transport/contact> ;
  dct:temporal <http://reference.data.gov.uk/id/year/2015> ;
  dct:spatial <http://sws.geonames.org/3399415> ;
  dct:publisher :transport-agency-mycity ;
  dct:accrualPeriodicity <http://purl.org/linked-data/sdmx/2009/code#freq-A>
  dcat:theme :mobility ;
  dcat:distribution :stops-2015-05-05.csv
  .

:stops-2015-05-05.csv
  a dcat:Distribution ;
  dct:title "CSV distribution of stops-2015-05-05 dataset" ;
  dct:description "CSV distribution of the bus stops dataset of MyCity" ;
  dcat:mediaType "text/csv;charset=UTF-8"
  .

:comment1Content
  a oa:TextualBody ;
  rdf:value "This dataset is missing stop 3"
  .
```

```
:comment1
```

```
  a oa:Annotation ;  
  a duv:UserFeedback ;  
  oa:hasBody :comment1Content ;  
  oa:hasTarget :stops-2015-05-05 ;  
  dct:creator :localresident ;  
  oa:motivatedBy oa:assessing
```

```
  .
```

```
:comment2Content
```

```
  a oa:TextualBody ;  
  rdf:value "Are tab delimited formats also available?"
```

```
  .
```

```
:comment2
```

```
  a oa:Annotation ;  
  a duv:UserFeedback ;  
  oa:hasBody :comment2Content ;  
  oa:hasTarget :stops-2015-05-05.csv ;  
  dct:creator :localresident ;  
  oa:motivatedBy oa:assessing
```

```
  .
```

```
:localresident
```

```
  a foaf:Person ;  
  foaf:name "Alan Law"
```

```
  .
```

Como Testar

Verifique se qualquer *feedback* fornecido por consumidores de dados para um determinado conjunto de dados ou distribuição esteja publicamente disponível.

Evidência

Requisitos Relevantes: [R-UsageFeedback \(em inglês\)](#), [R-QualityOpinions \(em inglês\)](#)

Benefícios



Reúso
(Reuse)



Confiança
(Trust)

Enriquecimento de dados

Boa Prática 31: Enriquecer dados por meio da geração de novos dados

Enriqueça seus dados gerando novos dados, pois ao fazê-lo você estará aumentando o valor dos mesmos.

Boa Prática 32: Fornecer visualizações complementares

Enriqueça seus dados apresentando-os em formas complementares e diretamente informativas, tais como visualizações, tabelas, aplicações Web ou resumos.

Republicação de dados

Boa Prática 33: Fornecer feedback para o publicador original

Informar ao publicador original quando você está promovendo o reúso de seus dados. Informe-o caso encontre um erro, ou tenha sugestões ou elogios a tecer.

Boa Prática 34: Obedecer os termos de licença

Encontrar e seguir os requisitos de licença indicada pelo publicador original do conjunto de dados.

Boa Prática 35: Citar a publicação original do conjunto de dados

Reconhecer a fonte de seus dados nos metadados. Caso forneça uma interface de usuário, inclua a citação claramente na interface.

Dataset Exchange Working Group Charter

The **mission** of the [Dataset Exchange Working Group](#) is to:

- Maintain and revise the Data Catalog Vocabulary, DCAT, taking into account feature requests from the DCAT user community.
- Define and publish guidance on the specification and use of application profiles when requesting and serving data on the Web.

[Join the Dataset Exchange Working Group.](#)

Start date	05 February 2020
End date	31 January 2022
Charter Extension	See Change History .
Chairs	Peter Winstanley (Invited Expert), Caroline Burle (NIC.br - Brazilian Network Information Center)
Team Contacts	Philippe Le Hegaret (0.05 FTE)
Meeting Schedule	Teleconferences: 1-hour calls will be held weekly Face-to-face: twice per year, expected to include the W3C's annual Technical Plenary week.

1. Goals

Sharing data among researchers, governments and citizens, whether openly or not, requires the provision of metadata. Different communities use different metadata standards to describe their datasets, some of which are highly specialized. At a general level W3C's Data Catalog Vocabulary, [DCAT](#), is in widespread use, but so too are [CKAN](#)'s native schema, schema.org's [dataset description](#) vocabulary, [ISO 19115](#), [DDI](#), [SDMX](#), [CERIF](#), [VoID](#), [INSPIRE](#) and, in the healthcare and life sciences domain, the [Dataset Description vocabulary](#) and [DATS \(ref\)](#) among others. This variety is a clear indication that no single vocabulary offers a complete and universally accepted solution.

TABLE OF CONTENTS

1. **Introduction**
2. **Motivation for change**
3. **Namespaces**
 - 3.1 Normative namespaces
 - 3.2 Non-normative namespaces
4. **Conformance**
5. **Vocabulary overview**
 - 5.1 DCAT scope
 - 5.2 RDF considerations
 - 5.3 Basic example
 - 5.4 Classifying datasets thematically
 - 5.5 Classifying dataset types
 - 5.6 Describing catalog records metadata
 - 5.7 Dataset available only behind some Web page
 - 5.8 A dataset available as a download and behind some Web page
 - 5.9 A dataset available through a service
6. **Vocabulary specification**
 - 6.1 RDF representation

Data Catalog Vocabulary (DCAT) - Version 2

W3C Recommendation 04 February 2020

**This version:**

<https://www.w3.org/TR/2020/REC-vocab-dcat-2-20200204/>

Latest published version:

<https://www.w3.org/TR/vocab-dcat-2/>

Latest editor's draft:

<https://w3c.github.io/dxwg/dcat/>

Implementation report:

<https://w3c.github.io/dxwg/dcat-implementation-report/>

Previous version:

<https://www.w3.org/TR/2019/PR-vocab-dcat-2-20191119/>

Previous Recommendation:


<https://www.w3.org/TR/2014/REC-vocab-dcat-20140116/>

Editors:

[Riccardo Albertoni](#)  (CNR - Consiglio Nazionale delle Ricerche, Italy)

[David Browning](#) (Refinitiv)

[Simon Cox](#)  (CSIRO)

[Alejandra Gonzalez Beltran](#)  (Scientific Computing Department, Science and Technology Facilities Council, UK) (Previously at the University of Oxford)

[Andrea Peregó](#)  (European Commission, Joint Research Centre)

[Peter Winstanley](#) (Scottish Government)

Publicações do Ceweb.br



Acesse:
<https://acervo.ceweb.br/>

<https://acervo.ceweb.br/acervos/itens/0f760d5c-01ee-49cd-b973-6e8128b81ac8>

Publicações do Ceweb.br



Acesse:
<https://acervo.ceweb.br/>

<https://acervo.ceweb.br/acervos/itens/0f760d5c-01ee-49cd-b973-6e8128b81ac8>

ceweb.br nic.br cgi.br

OBRIGADA



Caroline Burle

E-mail: cburle@nic.br

Twitter: [@carolburle](https://twitter.com/carolburle)

www.ceweb.br

Design Beatriz Rocha