



# Boas Práticas para Dados na Web: Desafios e Benefícios

Caroline Burle, Newton Calegari e Bernadette Lóscio



*30 de maio de 2017, CONIP*



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

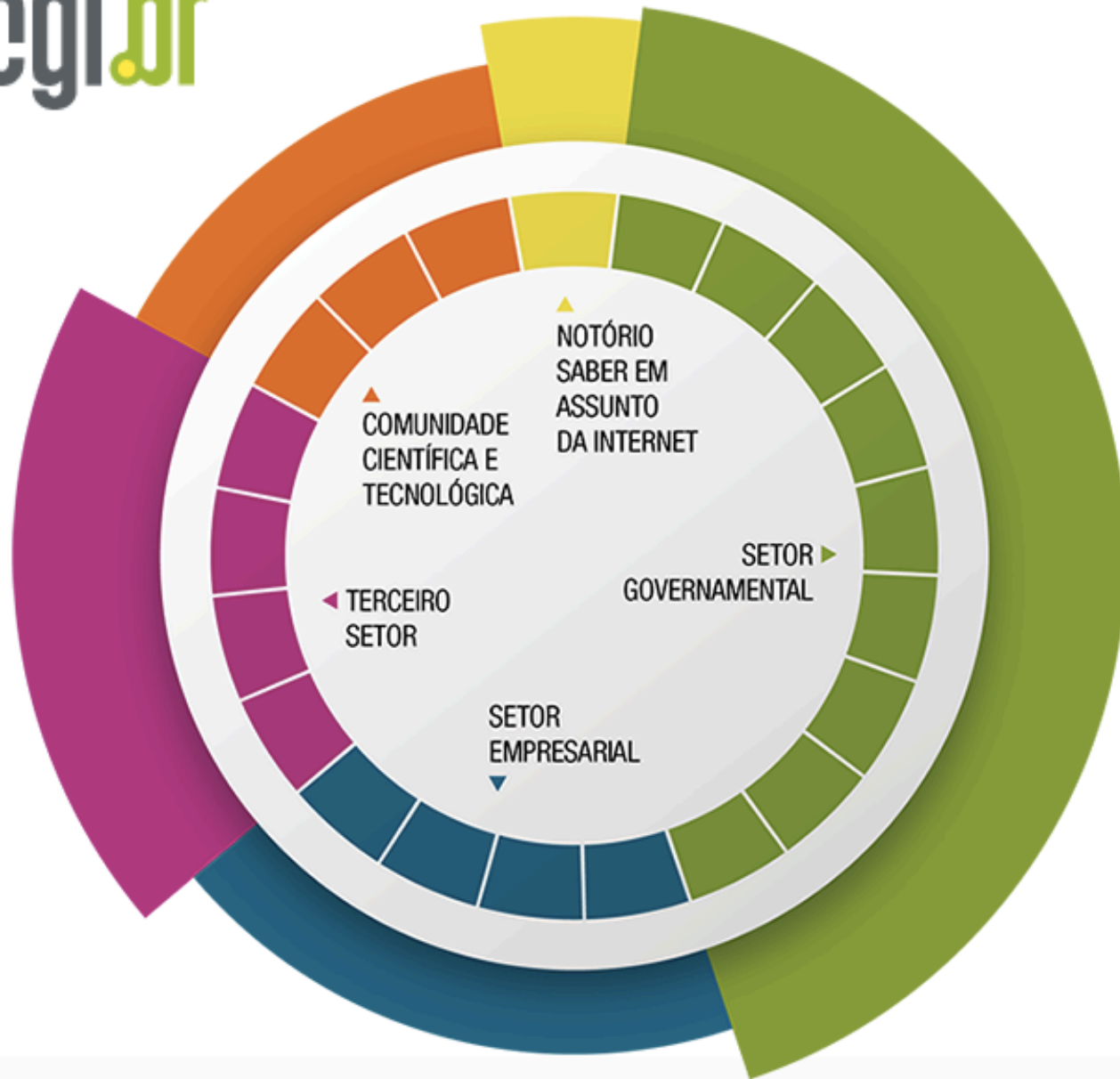
Tecnologias Web

ix.br

Troca de Tráfego

W3C  
Brasil

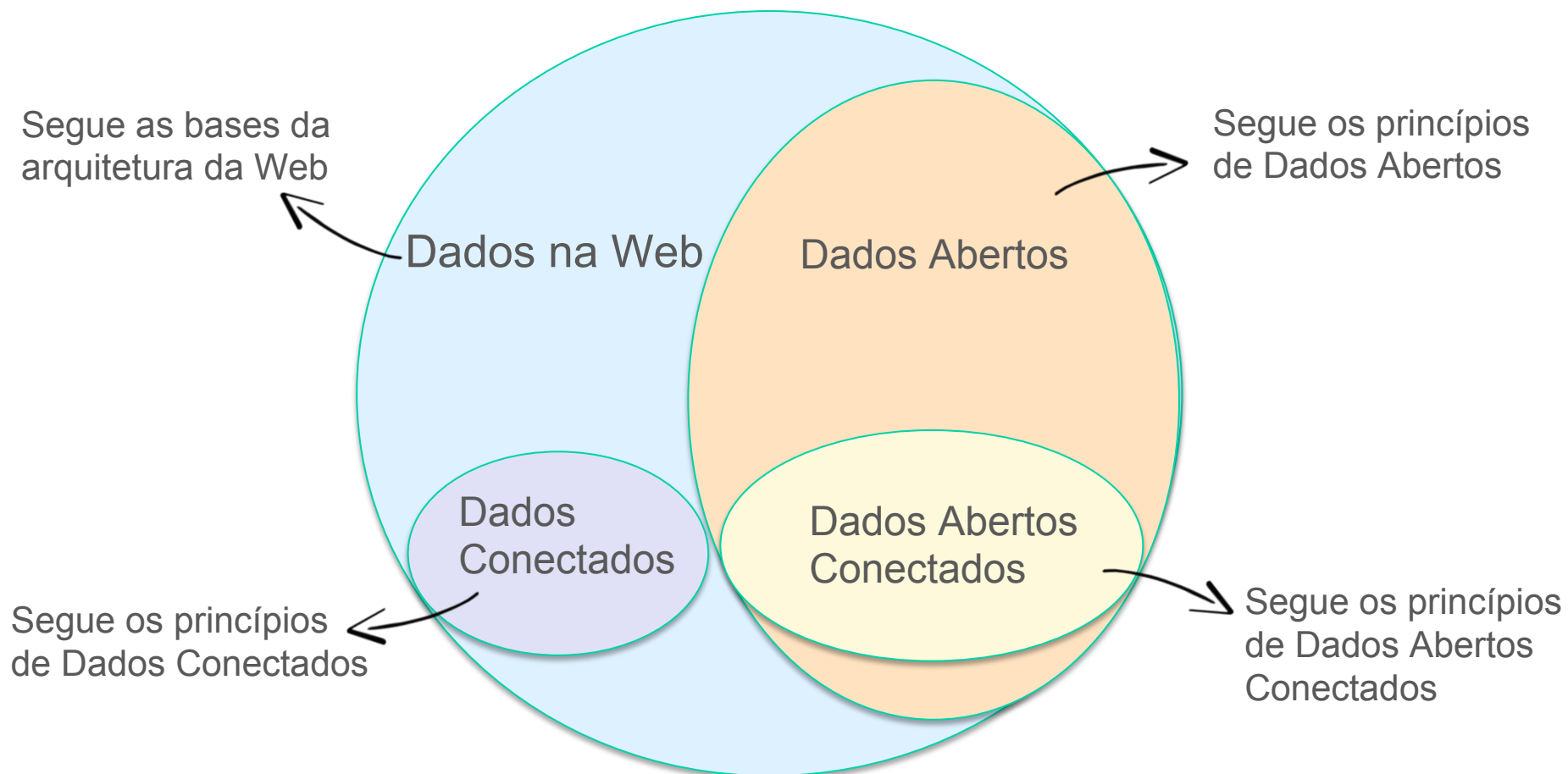
Padrões Web



# Tópicos a serem discutidos

- Dados na Web x Dados Abertos x Dados Conectados
- Desafios e Requisitos de Dados na Web
- Boas Práticas de Dados na Web
- Benefícios das Boas Práticas de Dados na Web

# Dados na Web x Dados Abertos x Dados Conectados

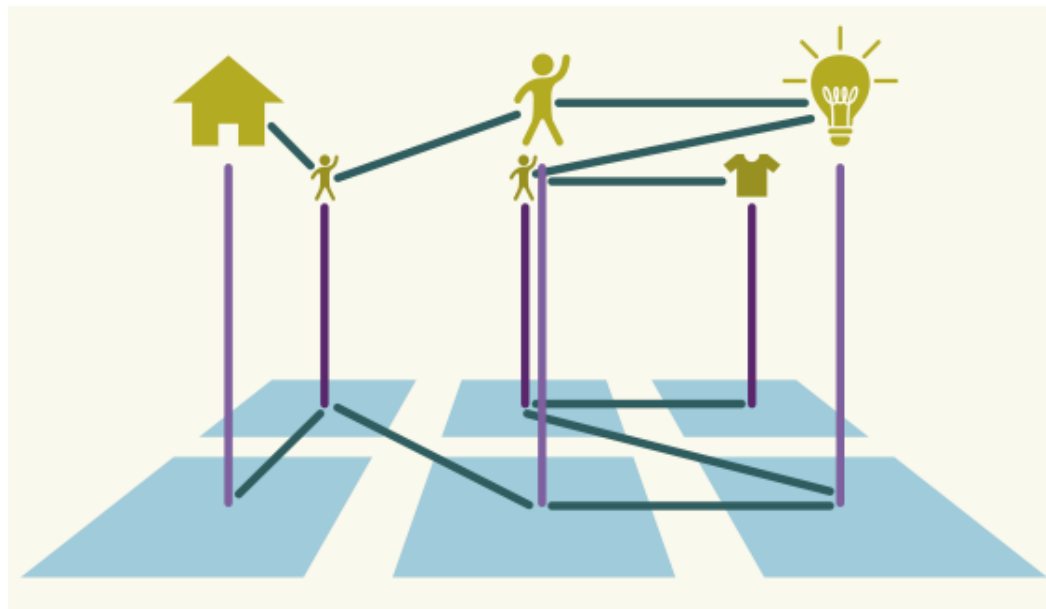


# Atores do ecossistema de Dados na Web

*Diversos tipos de recursos de dados (sistemas transacionais, sensores, dispositivos móveis, documentos...)*

Publicadores de dados:  
*publicam e compartilham dados*

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



Source: <http://ceweb.br/livros/dados-abertos-conectados/capitulo-1/>

*Como possibilitar o reuso dos dados?*

# Como possibilitar o reuso dos dados?

*O entendimento comum entre os publicadores e consumidores de dados é fundamental.*

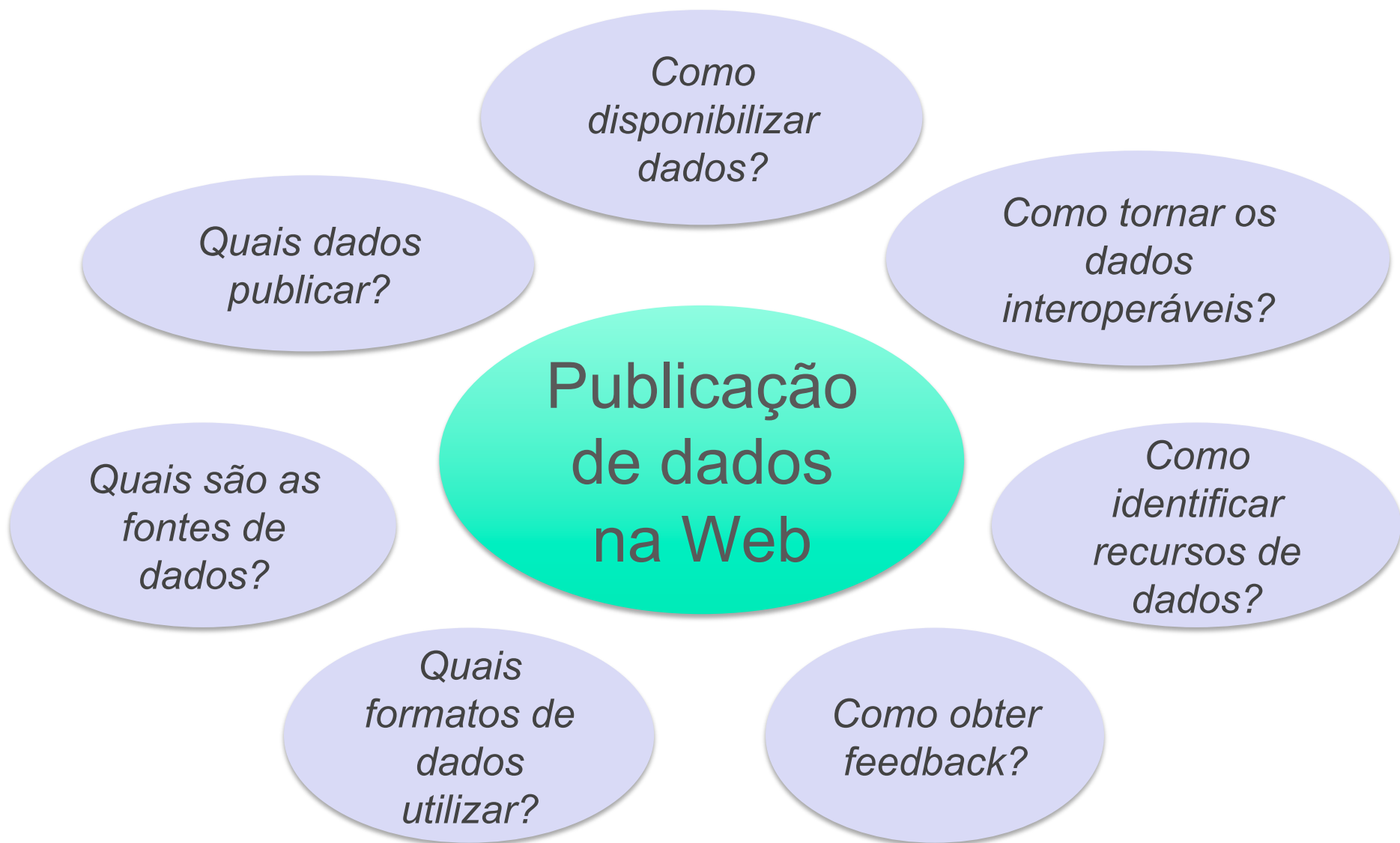
*Sem esse entendimento o esforço dos publicadores pode ser incompatível com o desejo dos consumidores.*



Consome dados



Publica dados



*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 conjuntos de dados 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

# Data on the Web Best Practices

W3C Candidate Recommendation 30 August 2016



## This version:

<https://www.w3.org/TR/2016/CR-dwbp-20160830/>

## Latest published version:

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

## Latest editor's draft:

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

## Implementation report:

[https://www.w3.org/2013/dwbp/wiki/BP\\_Implementation\\_Report](https://www.w3.org/2013/dwbp/wiki/BP_Implementation_Report)

## Previous version:

<http://www.w3.org/TR/2016/WD-dwbp-20160519/>

## Editors:

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

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

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

## Contributors:

Annette Greiner

Antoine Isaac

Carlos Iglesias

Carlos Laufer

Christophe Guéret

Deirdre Lee

Eric G. Stephan

Eric Kauz

Ghislain A. Ateazing

Hadley Beeman

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

*Audiência:*

*As BPs foram criadas para atender as necessidades de profissionais que trabalham com gerenciamento de informações, desenvolvedores, e grupos como cientistas interessados em compartilhar e reutilizar dados na Web*

# 35 Boas Práticas para Dados na Web

[Best Practice 1](#): Provide metadata

[Best Practice 2](#): Provide descriptive metadata

[Best Practice 3](#): Provide structural metadata

[Best Practice 4](#): Provide data license information

[Best Practice 5](#): Provide data provenance information

[Best Practice 6](#): Provide data quality information

[Best Practice 7](#): Provide a version indicator

[Best Practice 8](#): Provide version history

[Best Practice 9](#): Use persistent URIs as identifiers of datasets

[Best Practice 10](#): Use persistent URIs as identifiers within datasets

[Best Practice 11](#): Assign URIs to dataset versions and series

[Best Practice 12](#): Use machine-readable standardized data formats

[Best Practice 13](#): Use locale-neutral data representations

[Best Practice 14](#): Provide data in multiple formats

[Best Practice 15](#): Reuse vocabularies, preferably standardized ones

[Best Practice 16](#): Choose the right formalization level

[Best Practice 17](#): Provide bulk download

[Best Practice 18](#): Provide Subsets for Large Datasets

[Best Practice 19](#): Use content negotiation for serving data available in multiple formats

[Best Practice 20](#): Provide real-time access

[Best Practice 21](#): Provide data up to date

[Best Practice 22](#): Provide an explanation for data that is not available

[Best Practice 23](#): Make data available through an API

[Best Practice 24](#): Use Web Standards as the foundation of APIs

[Best Practice 25](#): Provide complete documentation for your API

[Best Practice 26](#): Avoid Breaking Changes to Your API

[Best Practice 27](#): Preserve identifiers

[Best Practice 28](#): Assess dataset coverage

[Best Practice 29](#): Gather feedback from data consumers

[Best Practice 30](#): Make feedback available

[Best Practice 31](#): Enrich data by generating new data

[Best Practice 32](#): Provide Complementary Presentations

[Best Practice 33](#): Provide Feedback to the Original Publisher

[Best Practice 34](#): Follow Licensing Terms

[Best Practice 35](#): Cite the Original Publication

## **Best Practice 1: Provide metadata**

*Provide metadata for both human users and computer applications.*

---

### **Why**

Providing metadata is a fundamental requirement when publishing data on the Web because data publishers and data consumers may be unknown to each other. Then, it is essential to provide information that helps human users and computer applications to understand the data as well as other important aspects that describes a dataset or a distribution.

### **Intended Outcome**

Humans will be able to understand the metadata and computer applications, notably user agents, will be able to process it.

### **Possible Approach to Implementation**

Possible approaches to provide *human-readable metadata*:

- to provide metadata as part of an HTML Web page
- to provide metadata as a separate text file

## Possible approaches to provide *machine-readable metadata*:

- machine-readable metadata may be provided in a serialization format such as Turtle and JSON, or it can be embedded in the HTML page using [[HTML-RDFA](#)] or [[JSON-LD](#)]. If multiple formats are published separately, they should be served from the same URL using [content negotiation](#) and made available under separate URIs, distinguished by filename extension. Maintenance of multiple formats is best achieved by generating each available format on the fly based on a single source of the metadata.
- when defining machine-readable metadata, reusing existing standard terms and popular vocabularies are strongly recommended. For example, Dublin Core Metadata (DCMI) terms [[DCTERMS](#)] and Data Catalog Vocabulary [[VOCAB-DCAT](#)] can be used to provide descriptive metadata. Such vocabularies are designed to be very flexible so it is often helpful to use a specific *profile* of a vocabulary such as the European Commission's [DCAT-AP](#)

### EXAMPLE 1

#### **Human-readable**

[Example page](#) with a human-readable description of an available dataset.

#### **Machine-readable**

[Example file](#) with a machine-readable description of an available dataset.

## How to Test

Check if human-readable metadata is available.

Check if the metadata is available in a valid machine-readable format and without syntax error.

## Evidence

Relevant requirements: [R-MetadadataAvailable](#), [R-MetadadataDocum](#), [R-MetadadataMachineRead](#)

## Benefits



Reuse



Comprehension



Discoverability



Processability

# Benefícios DWBP

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



*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.



*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.

# W3C® 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, umentando o potencial para inovação.



Fonte: [https://www.w3.org/2013/dwbp/wiki/Main\\_Page](https://www.w3.org/2013/dwbp/wiki/Main_Page)

# Como participar agora?



## DWBP Implementation Report

W3C Document 29 January 2017

### Editors:

[Bernadette Farias Lôscio, CIN - UFPE, Brazil](#)

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

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

Copyright © 2017 W3C® (MIT, ERCIM, Keio, Beihang). W3C liability, trademark and document use rules apply.

### Abstract

This document reports on evidence and implementations of the Data on the Web Best Practices [Candidate Recommendation](#). In particular, it demonstrates that the DWBP are already in use and are also implementable.

### Status of This Document

This document is merely a W3C-internal document. It has no official standing of any kind and does not represent consensus of the W3C Membership.

### 1. Introduction

One of the main goals of the Data on the Web Best Practices ([DWBP](#)) is to facilitate interaction between publishers and consumers of data on the Web. A set of 35 Best Practices were created to cover different [challenges](#) related to data publishing and consumption, such as Metadata, Data licenses, Data provenance, Data quality, Data versioning, Data identification, Data formats, Data vocabularies, Data access and APIs, Data preservation, Feedback, Data enrichment and Data republication.

### TABLE OF CONTENTS

1. **Introduction**
  - 1.1 Methodology
  - 1.2 Meeting the exit criteria
2. **DWBP Evidence**
  - 2.1 Datasets, Data portals and Vocabularies
  - 2.2 Documents and References
  - 2.3 Guidelines
3. **General analysis**
4. **DWBP and Data Catalogs**
5. **Set of Best Practices**
6. **Acknowledgements**

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

# Materiais de Referência

Data on the Web Best Practices

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

Livro Dados Abertos Conectados



<http://ceweb.br/publicacoes/indice/>

Guias sobre Dados Abertos



# Cursos online e gratuitos



## Publicação de dados em formato aberto (gratuito)

REALIZAÇÃO:

[ceweb.br](http://ceweb.br) [nic.br](http://nic.br) [cgj.br](http://cgj.br)

Oferece capacitação sobre dados abertos governamentais, explicando, de forma contextualizada, o potencial gerado pela abertura das informações. Apresenta ainda o papel dos dados abertos dentro da cultura de transparência de informações do governo federal.

\* Ver ementa

✓ Fazer inscrição



## Dados abertos conectados (gratuito)

REALIZAÇÃO:

[ceweb.br](http://ceweb.br) [nic.br](http://nic.br) [cgj.br](http://cgj.br)

Qualifica profissionais em métodos e ferramentas para modelar e estruturar os dados de maneira adequada, para que eles possam ser utilizados e reutilizados por programas de computador. Além disso, pretende-se apresentar técnicas de desenvolvimento de software utilizando tecnologias avançadas da área de Web Semântica e Linked Data, para uso efetivo dos dados publicados na Web.

\* Ver ementa

✓ Fazer inscrição

<http://ceweb.brcursos>

# Obrigada(o)!

[www.ceweb.br](http://www.ceweb.br) - [www.cin.ufpe.br](http://www.cin.ufpe.br)

© cburle@nic.br

© @carolburle

© bfl@cin.ufpe.br

© @bernafarias

© newton@nic.br

© @newtonclegari