



# Boas Práticas para Dados na Web: Desafios e Benefícios

Bernadette Lóscio, Caroline Burle and Newton Calegari



*Oficina Dados abertos, ciência aberta: questões legais para pesquisa  
12 de abril de 2017*



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

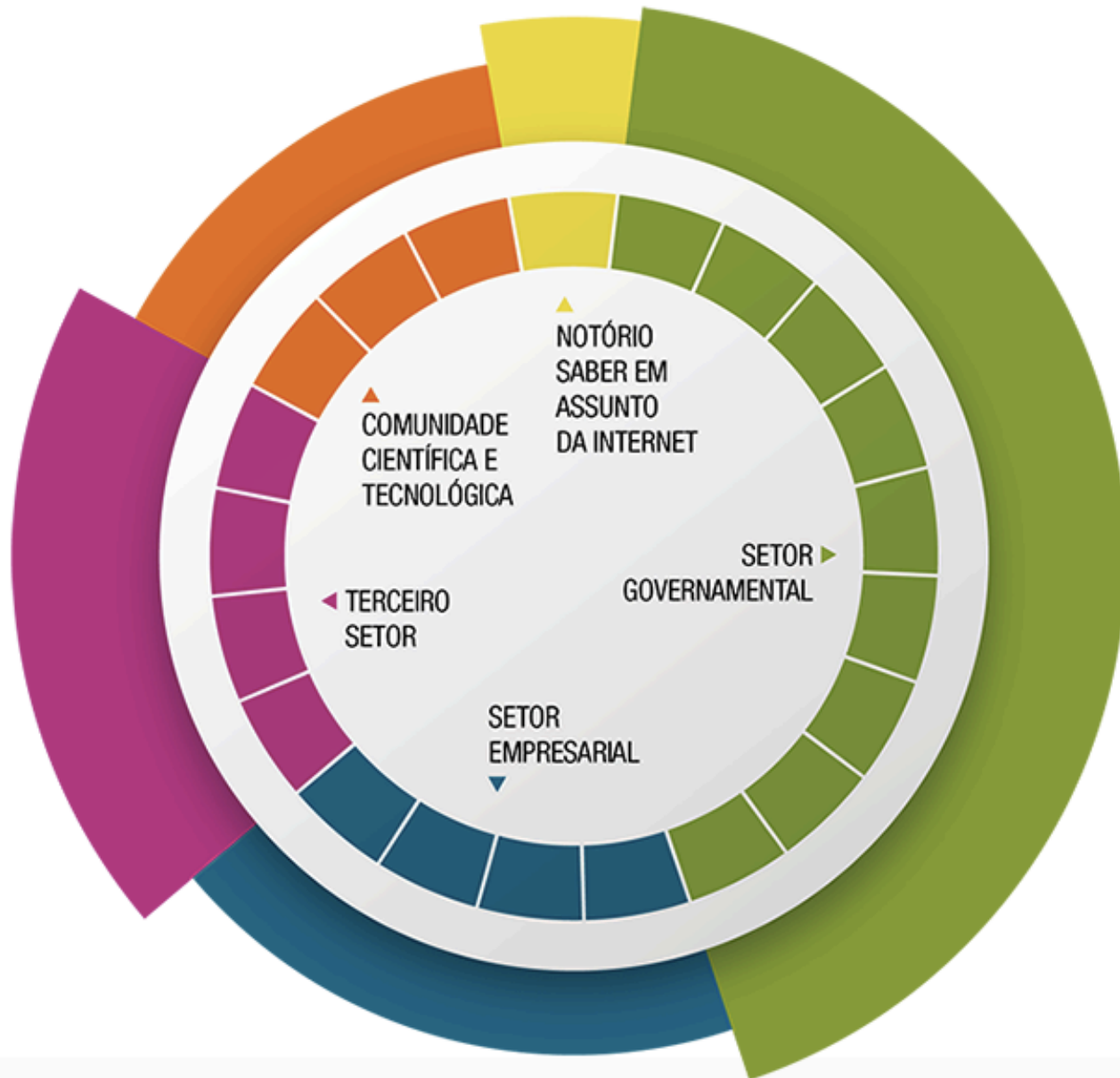
Tecnologias Web

ix.br

Troca de Tráfego

W3C  
Brasil

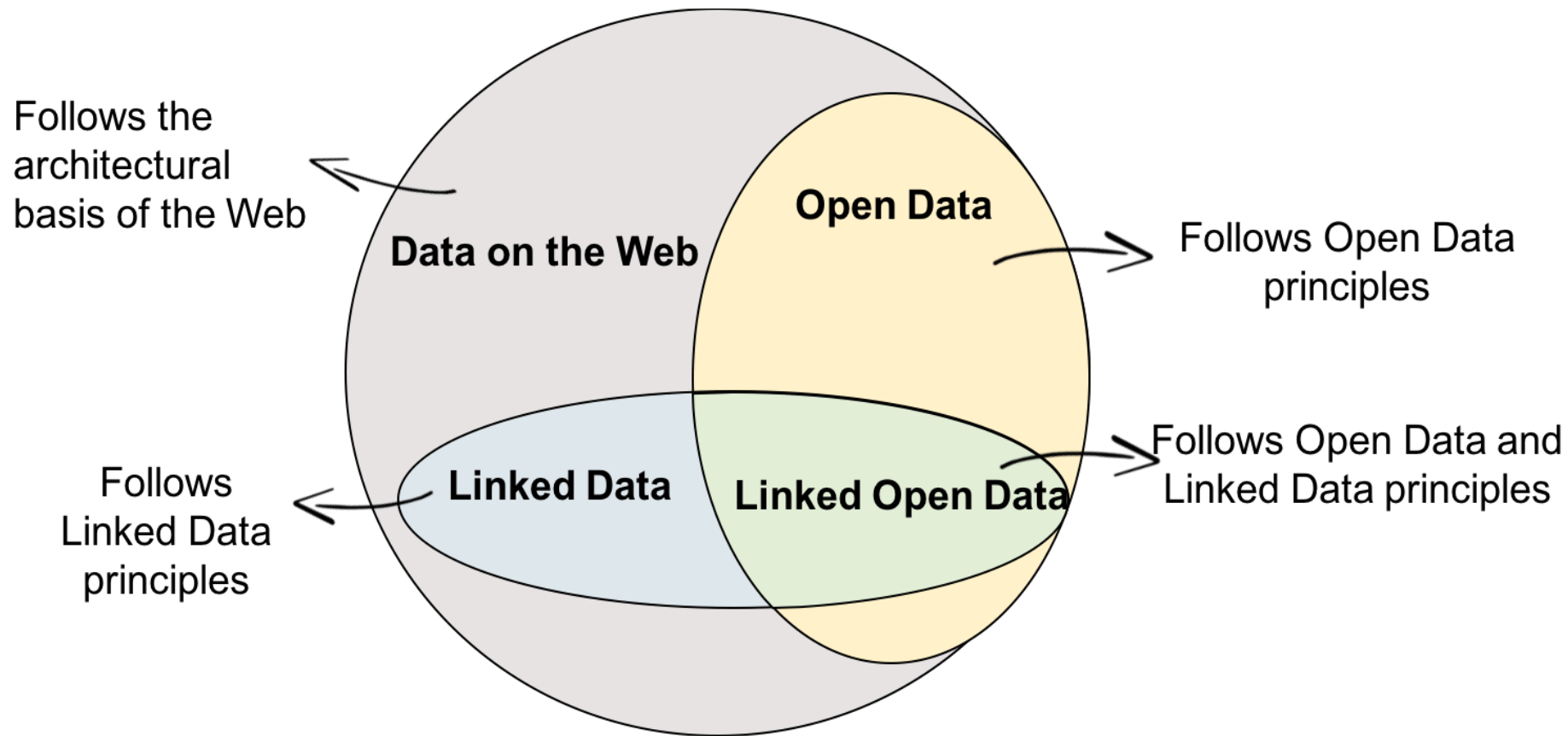
Padrões Web



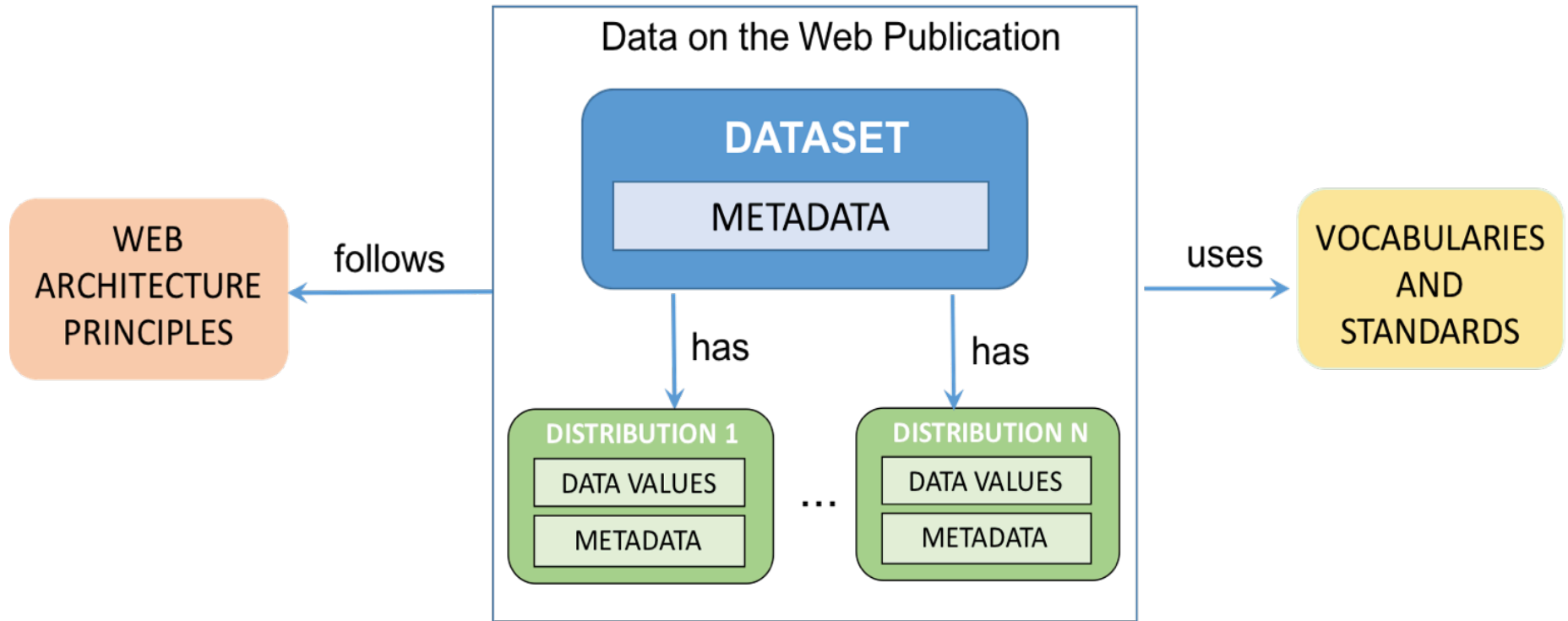
# Tópicos a serem discutidos

- Contexto da Web de Dados
- Casos de Uso de Dados na Web
- 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



# Contexto de Dados na Web

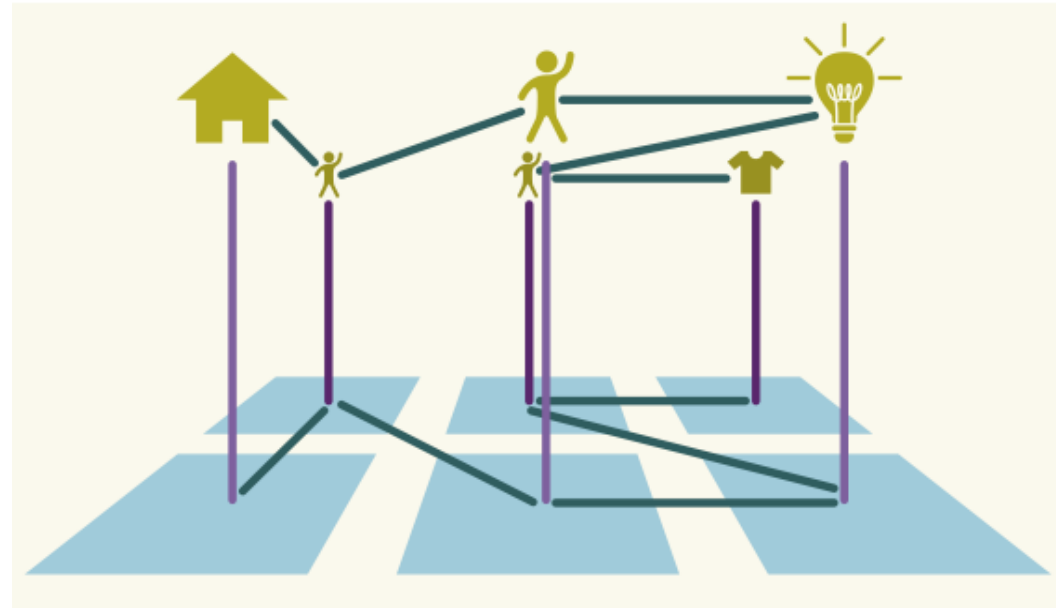


# Atores do ecossistema de Dados na Web

*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?

*Um 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



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):

# 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](mailto:Insight@NUIG), Ireland)

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

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

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

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

# Publicação de dados na Web

*Como disponibilizar dados?*

*Quais dados publicar?*

*Como tornar os dados interoperáveis?*

*Quais são as fontes de dados?*

*Como identificar recursos de dados?*

*Quais formatos de dados utilizar?*

*Como obter feedback?*

*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

# 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*

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

## Evidence

Relevant requirements: [R-ProvAvailable](#), [R-MetadataAvailable](#)

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

## Intended Outcome

Humans will know the origin or history of the dataset and software agents will be able to automatically process provenance information.

[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 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



# Benefícios DWBP

*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

## Linkability

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

## 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

## 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

## Best Practice 1: Provide metadata

Metadata must be provided for both human users and computer applications

### Why

Providing metadata is a fundamental requirement for publishers and data consumers may be unaware of the metadata that helps human users and computer applications understand aspects that describes a dataset or a distribution.

### Intended Outcome

Human-readable metadata will enable human users to understand metadata will enable computer applications to process data.

### Possible Approach to Implementation

Possible approaches to provide *human-readable* metadata:

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

Possible approaches to provide *machine-readable* metadata:

- machine readable metadata may be provided in multiple formats. If it can be embedded in the HTML page or published separately, they should be provided in multiple formats. The presence of multiple formats is best achieved by providing a single source of the metadata.
- when defining machine readable metadata, standard vocabularies are strongly recommended. For example, Dublin Core metadata (DCMI) terms [[DC-TERMS](#)] and Data Catalog Vocabulary [[VOCAB-DCAT](#)] should be used to provide descriptive metadata.

## Benefícios das Boas Práticas

- **Compreensão:** pessoas capazes de melhor entender a estrutura dos dados, do significado dos dados, bem como os metadados e a natureza do conjunto de dados.
- **Processabilidade:** máquinas serão capazes de processar automaticamente e manipular os dados de um conjunto de dados.
- **Descoberta:** máquinas serão capazes de automaticamente descobrir um conjunto de dados ou os dados nele contidos.
- **Reuso:** as chances de reuso de dados ou conjuntos de dados por diferentes grupos irão aumentar.

Datasets must be identified by a persistent URI.

## Why

Adopting a common identification system by any stakeholder in a reliable way. The and reuse.

Developers may build URIs into their code to provide a dereference to the same resource over time.

## Intended Outcome

Datasets or information about datasets with known status, availability or format of the data.

## Possible Approach to Implementation

To be persistent, URIs must be designed to be stable. For creating a Web site designed for human consumption on a topic, see, for example, the European Commission's approach to many other resources.

Where a data publisher is unable or unwilling to use a native approach is to use a redirection service. These provide persistent URIs that can be used to identify ephemeral. The [software behind such services](#) is freely available so that it can be installed and managed locally if required.

Digital Object Identifiers ([DOIs](#)) offer a similar alternative. These identifiers are defined independently of any Web technology but can be appended to a 'URI stub.' DOIs are an important part of the digital infrastructure for research data and libraries.

# Benefícios das Boas Práticas

- **Vinculação de links:** será possível criar links entre recursos de dados (conjuntos de dados e itens de dados).
- **Interoperabilidade:** será mais fácil obter consenso entre publicadores e consumidores de dados.
- **Confiança:** melhorará a confiança que consumidores possuem em relação aos conjuntos de dados.
- **Acesso:** pessoas e máquinas serão capazes de acessar dados atuais e em diferentes formatos.

# 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.

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

#### 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**

# Spatial Data on the Web Best Practices

## TABLE OF CONTENTS

1. Introduction
2. Audience
3. Scope
  - 3.1 Spatial data
  - 3.2 Best practice criteria
  - 3.3 Privacy considerations
4. Best Practices Summary
5. Spatial Things, Features and Geometry
6. Coverages: describing properties that vary with location (and time)
7. Spatial relations
8. Coordinate Reference Systems (CRS)
9. Linked Data
10. Why are traditional Spatial Data Infra-structures not enough?
11. How to use these best practices

## Spatial Data on the Web Best Practices

W3C Working Group Note 30 March 2017

### This version:

<https://www.w3.org/TR/2017/NOTE-sdw-bp-20170330/>

### Latest published version:

<https://www.w3.org/TR/sdw-bp/>

### Latest editor's draft:

<https://w3c.github.io/sdw/bp/>

### Previous version:

<https://www.w3.org/TR/2017/NOTE-sdw-bp-20170216/>

### Editors:

Jeremy Tandy, [Met Office](#)

Linda van den Brink, [Geonovum](#)

Payam Barnaghi, [University of Surrey](#)

### OGC Document Number:

OGC 15-107

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



Fonte: <https://www.w3.org/TR/sdw-bp/>

# Caso Prático com Linked Open Data

## SOBRE O PROJETO

Este projeto busca melhorar o Ambiente de Negócios por Meio da Transparência no Estado de São Paulo. O projeto é uma parceria entre a Embaixada Britânica no Brasil, o Governo do Estado de São Paulo, por meio da Unidade de Inovação em Governo (iGovSP) da Subsecretaria de Parcerias e Inovação da Secretaria de Governo, e o Comitê Gestor da Internet (CGI.br) / Núcleo de Informação e Coordenação do Ponto BR (NIC.br), por meio o Ceweb.br (Centro de Estudos sobre Tecnologias Web) como agente implementador. Ele busca reforçar a política de uso de dados abertos, baseando-se na experiência do Reino Unido.

### Objetivos

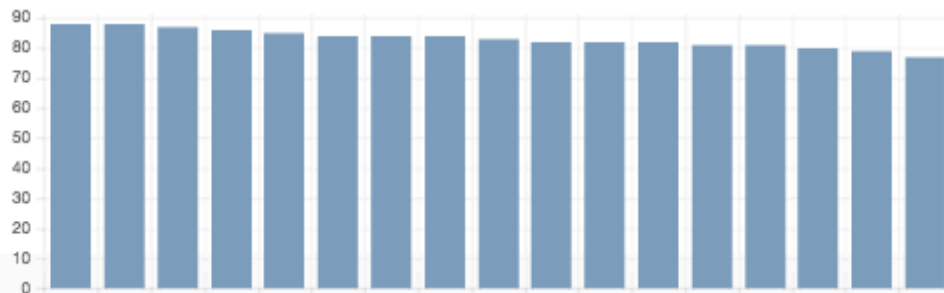
Os objetivos são: o Governo do Estado de São Paulo aumentar em pelo menos 70% o número de bases de dados em formato aberto, plenamente acessíveis e disponibilizadas em linguagem compreensível para qualquer cidadão; e ter um mínimo de 3% das bases abertas com dados estruturados em conformidade com os preceitos da Web Semântica, no site [Governo Aberto SP](#).

### Ontologia

Foi criada uma Ontologia que descreve os conceitos relacionados à descrição de dados e indicadores da área de Saúde. Na versão atual da ontologia foi contemplado o indicador que diz respeito à proporção do número de casos de pacientes curados até 31 de dezembro do ano de avaliação em relação ao número de casos diagnosticados de determinada doença identificada a partir do código CID10. Os indicadores são avaliados anualmente de acordo com espaço geográfico municipal do estado de São Paulo, dando a possibilidade de recuperação dos dados por Departamento Regional de Saúde (DRS) e Redes Regionais de Atenção à Saúde (RRAS), abrangendo informações demográficas municipais.

[Visualizar Ontologia](#)

### % MÉDIA DE CURADOS POR DRS



### TOP 10 CIDADES

Ranking das 10 cidades com maior número de pessoas curadas

- 1º - São Paulo com 17975 pessoas curadas
- 2º - Guarulhos com 1498 pessoas curadas
- 3º - São Vicente com 1411 pessoas curadas
- 4º - Santos com 1054 pessoas curadas
- 5º - Campinas com 982 pessoas curadas
- 6º - Guarujá com 938 pessoas curadas

# Materiais de Referência



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

# 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