Data on the Web Best Practices: Challenges and Benefits

Bernadette Lóscio, Caroline Burle and Newton Calegari

Topics to be discussed

- Data on the Web Context
  - Data on the Web x Open Data x Linked Data
- Data on the Web use cases
  - Data on the Web Challenges and Requirements
- Data on the Web Best Practices
  - Data on the Web Best Practices Benefits

Open Data Charter principles

- Open by Default
- Timely and Comprehensive
- Accessible and Usable
- Comparable and Interoperable
- For Improved Governance and Citizen Engagement
- For Inclusive Development and Innovation

Open Data must be on the Web

How to enable the data reuse?

A common understanding between data publishers and data consumers becomes fundamental.
Without this agreement, data publishers' efforts may be incompatible with data consumers' desires.

Best Practices

Consumes data

Publishes data

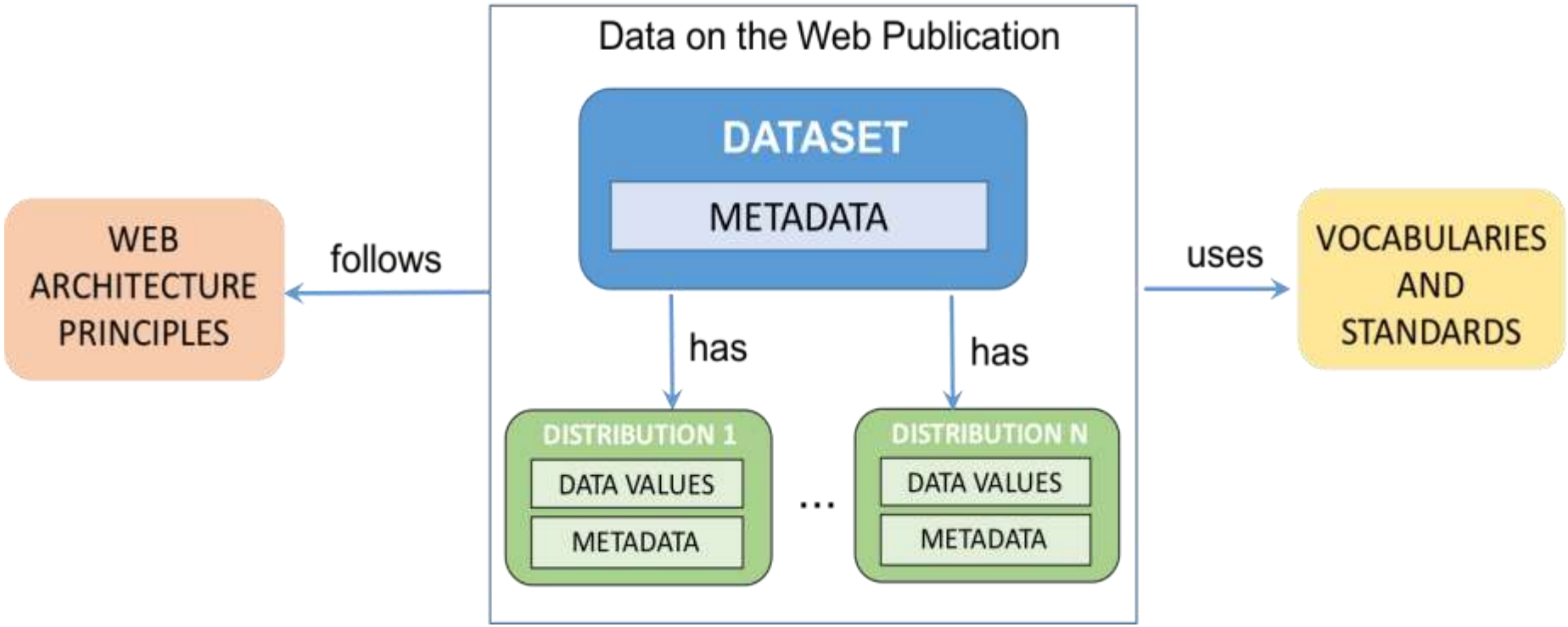
Data on the Web Best Practices Working Group

The **Mission** of the Data on the Web Best Practices Working Group, part of the Data Activity, is:

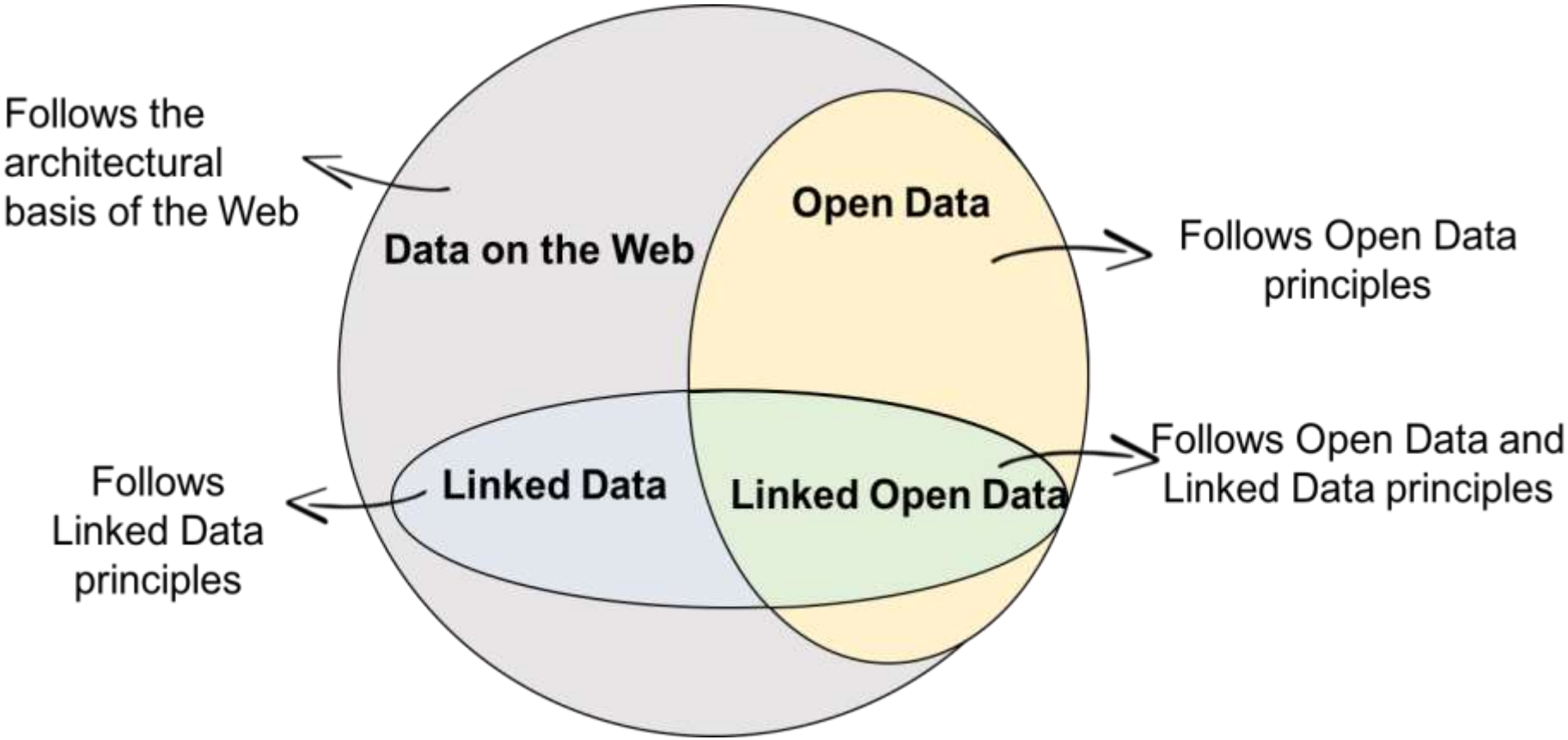
1. to develop the **open data ecosystem**, facilitating better communication between developers and publishers;
2. to provide **guidance to publishers** that will improve consistency in the way data is managed, thus promoting the re-use of data;
3. to **foster trust in the data** among developers, whatever technology they choose to use, increasing the potential for genuine innovation.

Source: https://www.w3.org/2013/dwbp/wiki/Main_Page

Data on the Web Context



Data on the Web x Open Data x Linked Data



Data on the Web use cases

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

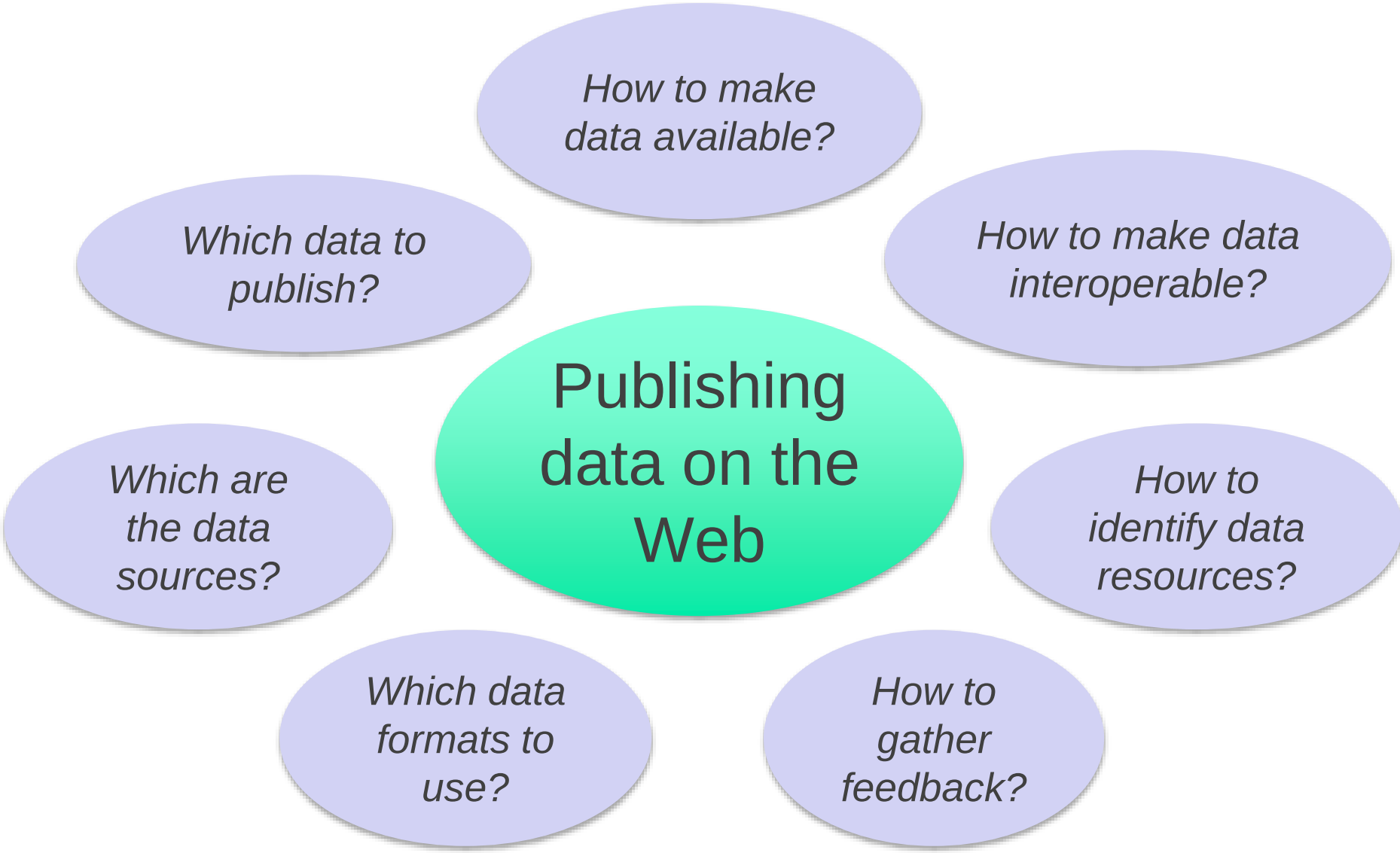
https://www.w3.org/TR/dwbp-ucr/

Data on the Web use cases

12 challenges and
42 requirements

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



Publishing data on the Web is more than just publishing data!

Data on the Web Challenges

- Metadata *(for humans & machines)*
- Data Licenses *(how to permit & restrict access?)*
- Data Provenance & Quality *(how to add trust?)*
- Data Versioning (*tracking dataset versions)*
- Data Identification *(identifying datasets and distributions)*
- Data Formats *(which data formats to use?)*

Data on the Web Challenges

- Data Vocabularies *(how to promote interoperability?)*
- Data Access *(access options)*
- Data Preservation
- Feedback *(how to engage users?)*
- Data Enrichment *(adding value to data)*
- Data Republication *(reuse data responsibly)*

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

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. Atemezing
Hadley Beeman

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

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

DWBP Benefits

Each benefit represents an improvement in the way how datasets are available on the 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

Processibility

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 re
lishers and data consumers may be unk
that helps human users and computer a
aspects that describes a dataset or a di

Intended Outcome

Human-readable metadata will enable h
metadata will enable computer applicati

Possible Approach to Implementatio

Possible approaches to provide *human*

- to provide metadata as part of an H
- to provide metadata as a separate

Possible approaches to provide *machin*

- machine readable metadata may b
  it can be embedded in the HTML pa
  published separately, they should b
  nance of multiple formats is best ac
  a single source of the metadata.
- when defining machine readable metadata, reusing existing standard terms and popular vocabular-
  ies 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.

BP Benefits

- **Comprehension**: humans will have a better understanding about the data structure, the data meaning, the metadata and the nature of the dataset.
- **Processability**: machines will be able to automatically process and manipulate the data within a dataset.
- **Discoverability:** machines will be able to automatically discover a dataset or data within a dataset.
- **Reuse**: the chances of dataset reuse by different groups of data consumers will increase.

DWBP Call for Implementations

DWBP Evidences Form

Thank you for your help to collect implementation evidence of the Data on the Web Best Practices (DWBP), a document to be released as a W3C Recommendation in 2016.

There are 35 forms, each one of them is referring to one Best Practice and there is the possibility to create more evidences for each BP.

The Data on the Web Best Practices document is available at https://www.w3.org/TR/dwbp.

The Organisation name will be mentioned as one of the organisations that tested the **Data on the Web Best Practices** in order to become a W3C Recommendation.

You may want to start filling your name and email and then the name of the organisation that published the resource or dataset. After that please start filling the respective forms for the best practices that were implemented.

If you have some questions, feel free to send us a message (public-dwbp-comments@w3.org).

Contact information

Name

Email

Publisher's information *(Organisation that published the resource or dataset)*

Publisher

Site

Save info

http://w3c.br/form-dwbp/

Obrigada!

www.ceweb.br - www.cin.ufpe.br

 cburle@nic.br

 @carolburle

 bfl@cin.ufpe.br

 @bernafarias

 newton@nic.br

 @newtoncalegari

Madrid, Spain
October 5, 2016