Implementation of the Best Practices for Data on the Web in Brazil and Costa Rica

Bernadette Farias Lóscio, Caroline Burle and Newton Calegari

Universidade Federal de Pernambuco

Motivation

The Data on the Web Best Practices (DWBP) provide guidance to different challenges related to data on the Web publishing and consumption.

Data on the Web Best Practices

W3C Recommendation 31 January 2017

This version:
https://www.w3.org/TR/2017/REC-dwbp-20170131/

Latest published version:
https://www.w3.org/TR/dwbp/

Latest editor's draft:
http://w3c.github.io/dwbp/bp.html

Implementation report:
http://w3c.github.io/dwbp/dwbp-implementation-report.html

Previous version:
https://www.w3.org/TR/2016/PR-dwbp-20161215/

Editors:
Bernadette Farias Lóscio, CIn - UFPE, Brazil
Caroline Burle, NIC.br, Brazil
Newton Calegari, NIC.br, Brazil

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 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

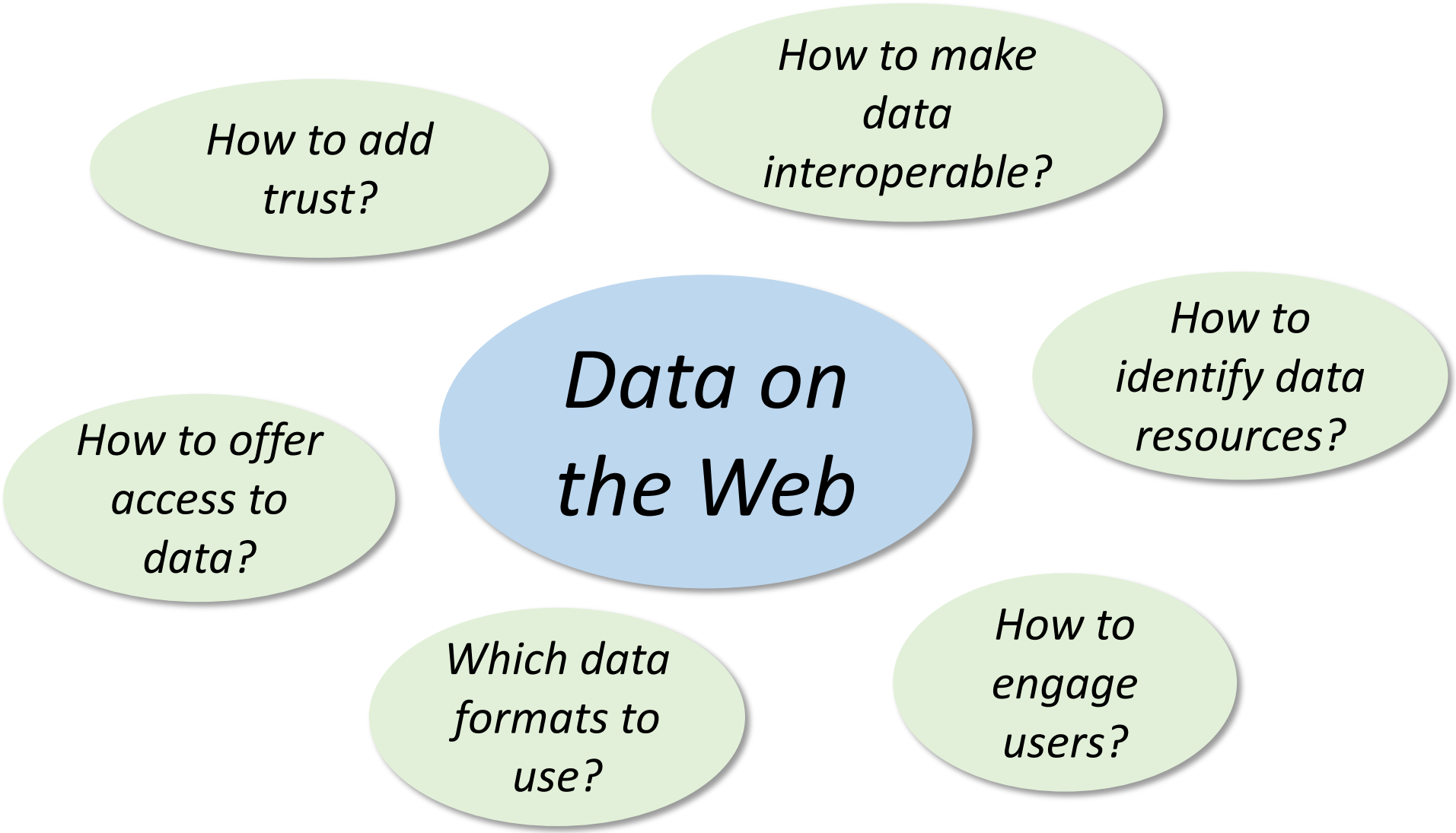
Best Practice 31: Enrich data by generating new data

Best Practice 32: Provide Complementary Presentations

Best Practice 33: Provide Feedback to the Original Publisher

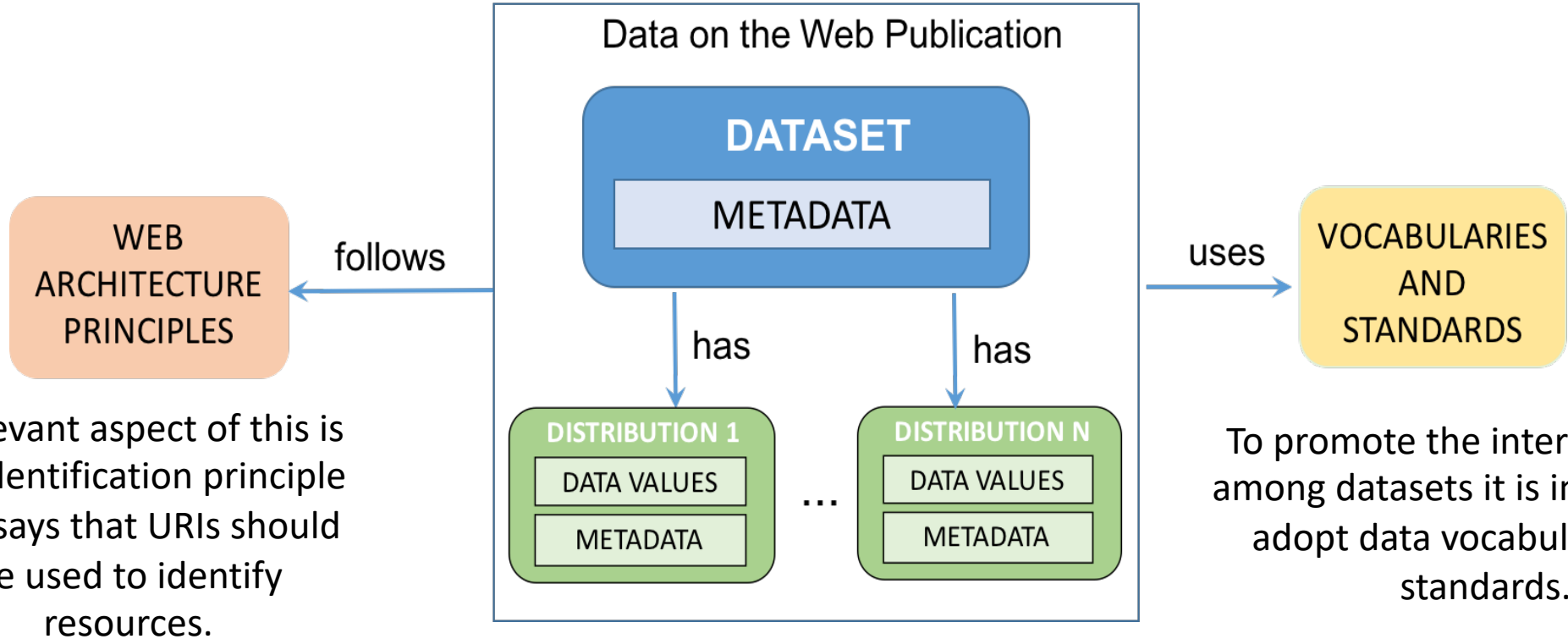
Best Practice 34: Follow Licensing Terms

Best Practice 35: Cite the Original Publication



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

Motivation – DWBP context



A relevant aspect of this is the identification principle that says that URIs should be used to identify resources.

Data is published in different distributions, which are specific physical form of a dataset

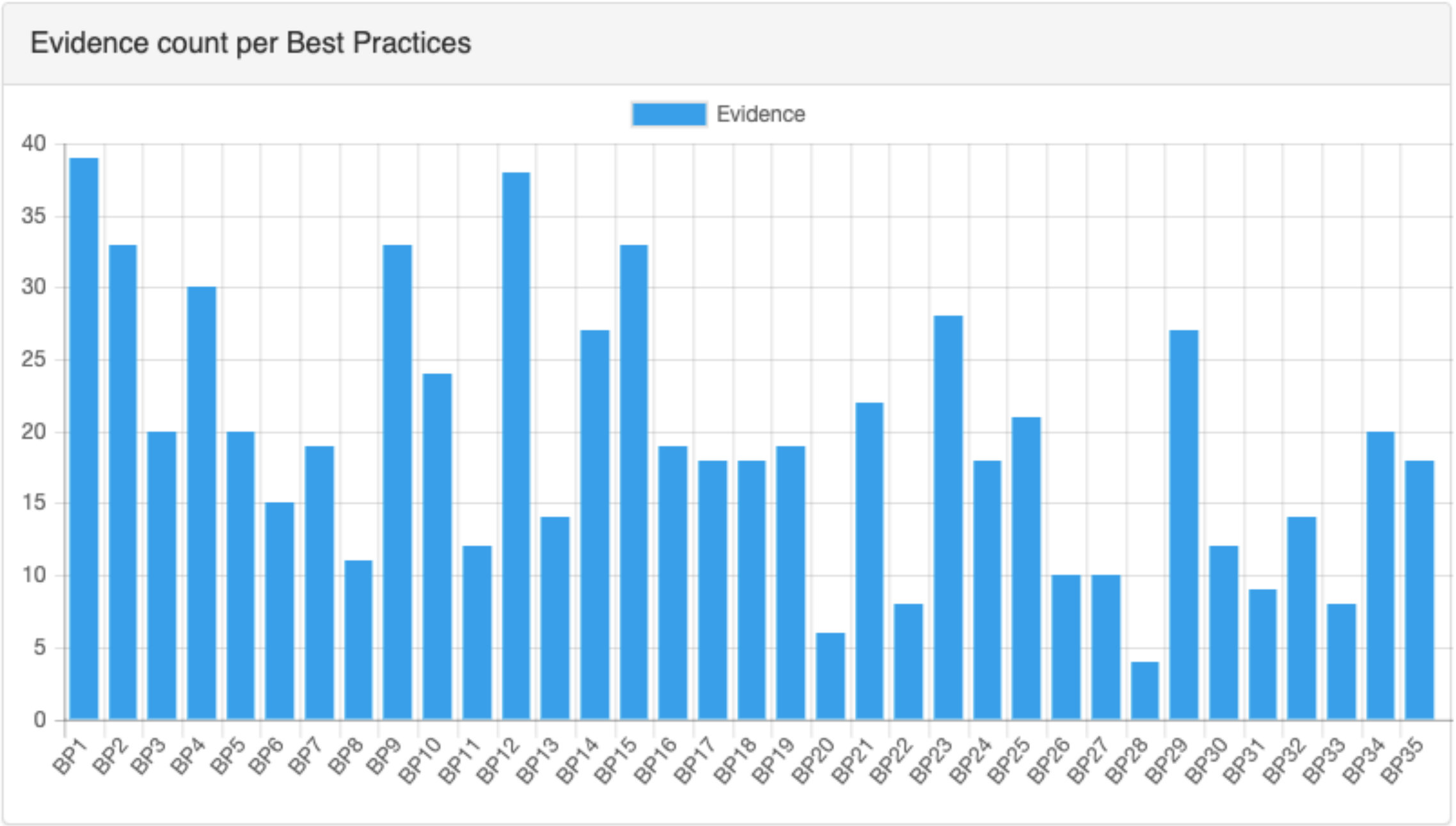
To promote the interoperability among datasets it is important to adopt data vocabularies and standards.

LÓSCIO; BURLE; CALEGARI, 2017 - DWBP Recommendation

Motivation – DWBP Implementation Report

To show that the DWBP are implementable as well broadly adopted and referenced by well-known organizations, we collected evidence in the form of datasets, data portals, documents, references and guidelines.

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


Evidence count per Best Practices
Evidence
40
35
30
25
20
15
10
5
0
BP1
BP2
BP3
BP4
BP5
BP6
BP7
BP8
BP9
BP10
BP11
BP12
BP13
BP14
BP15
BP16
BP17
BP18
BP19
BP20
BP21
BP22
BP23
BP24
BP25
BP26
BP27
BP28
BP29
BP30
BP31
BP32
BP33
BP34
BP35

Use Cases

- The Regional Center for Studies on the Development of the Information Society (**Cetic.br**) of The Brazilian Network Information Center (**NIC.br**).
- The Judiciary Department of Costa Rica (**Justicia Abierta**).

applied the DWBP Recommendation

to publish their data on the Web

Brazil's Use Case

- **Cetic.br** produces indicators and statistics on the use of information and communication technologies in Brazil since 2005.
- In 2017, it started the process of providing microdata databases of its sample surveys.

Figure 1: Cetic.br microdata website

https://cetic.br/pesquisa/domicilios/microdados

Costa Rica's Use Case

- In 2017, the Judiciary Department of Costa Rica was supported by the The Trust for the Americas Foundation to create the Justicia Abierta (Open Justice) and open its data.

Figure 2: Costa Rica Justicia Abierta website

http://datosabiertospj.eastus.cloudapp.azure.com/dataset

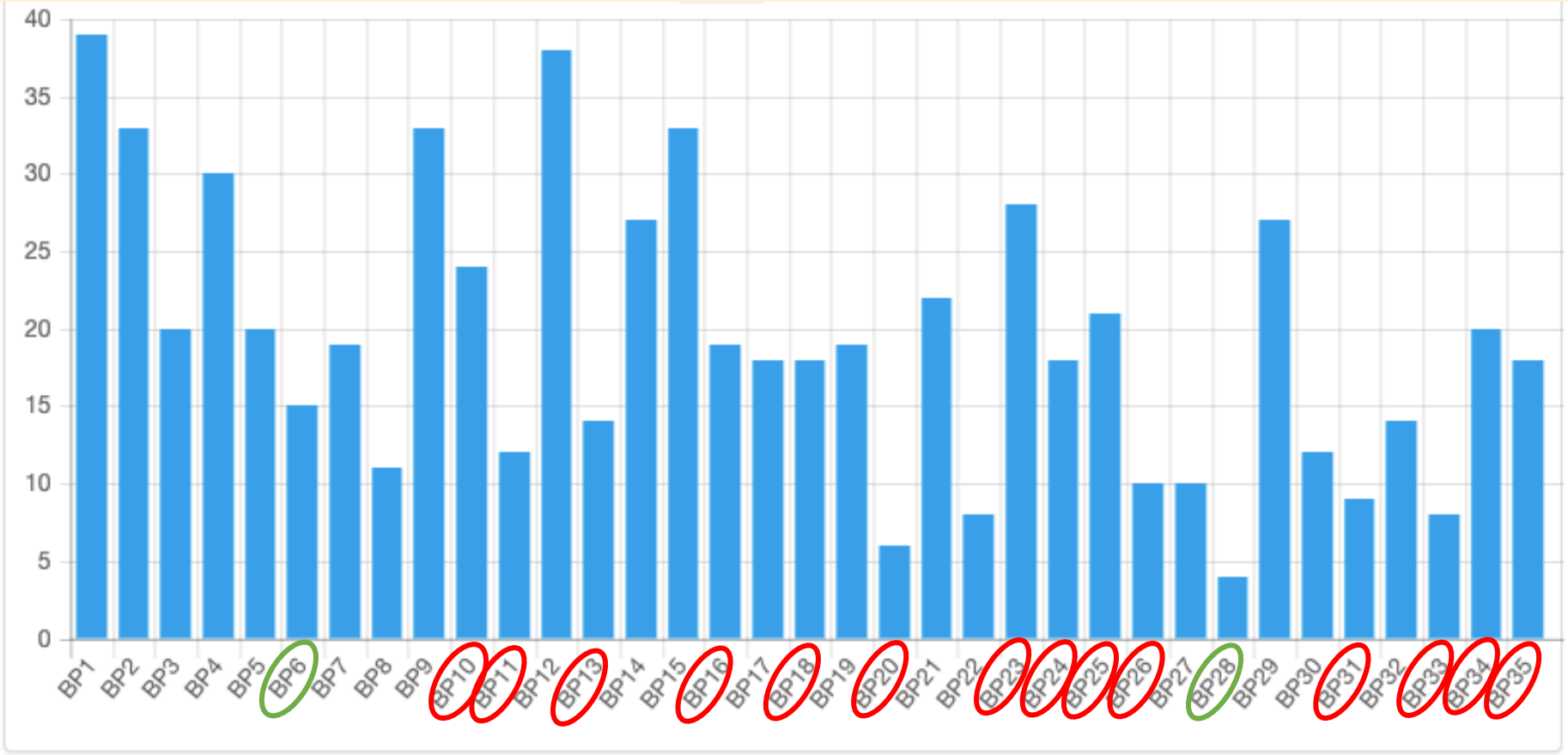
Discussion

- Brazil’s use case
  - 18 Best Practices could be applied immediately;
  - They needed more time to implement the BP19 "Use content negotiation for serving data available in multiple formats”;
  - 16 BPs not applicable

- Costa Rica’s use case
  - 16 Best Practices could be applied immediately;
  - Technical limitations postponed the implementation of 06 BPs;
  - 13 BPs not applicable.

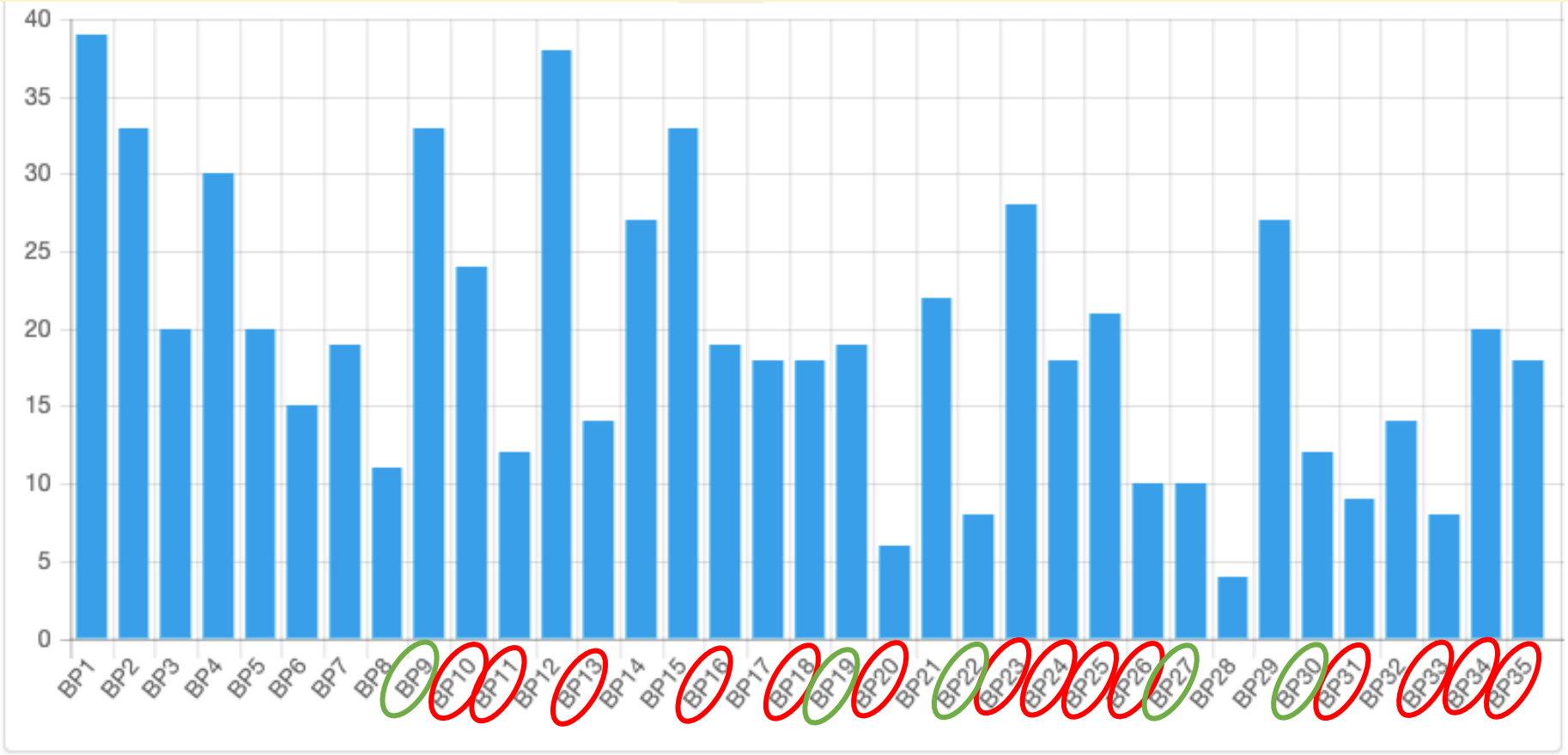
Discussion – Cetic.br

Cetic.br: **BP06** (Data Quality); **BP10**, **BP11** (Data Identifiers); **BP13** (Data Formats); **BP16** (Data Vocabularies); **BP18**, **BP20** (Data Access); **BP23, BP24, BP25, BP26** (Data Access APIs); **BP28** (Data Preservation); **BP31** (Data Enrichment); **BP33, BP34, BP35** (Republication)



Discussion – Justicia Aberta

Costa Rica: **BP09, BP10, BP11** (Data Identifiers); **BP13** (Data Formats); **BP16** (Data Vocabularies); **BP18, BP19**, **BP20**, **BP22** (Data Access); **BP23, BP24, BP25, BP26** (Data Access APIs); **BP27** (Data Preservation); **BP30** (Feedback); **BP31** (Data Enrichment); **BP33, BP34, BP35** (Republication)



Discussion

- Both Cetic.br and Justicia Abierta use cases demonstrated the importance of gathering the whole team involved in the process of publishing data.

- Although neither use cases implemented all the Data on the Web Best Practices, both use cases showed that it is possible to implement some of the DWBP without great effort.

Obrigada!

Bernadette Farias Lóscio - bfl@cin.ufpe.br

Caroline Burle - cburle@nic.br

Newton Calegari - newtoncalegari@gmail.com