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

Bernadette Farias Lósio  
bfl@cin.ufpe.br  
Center of Informatics - Federal University of Pernambuco  
Recife, Brazil

Caroline Burle  
cburle@nic.br  
Brazilian Network Information Center - NIC.br  
São Paulo, Brazil

Newton Calegari  
newton@nic.br  
Brazilian Network Information Center - NIC.br  
São Paulo, Brazil

ABSTRACT

The Best Practices described on the Data on the Web Best Practices (DWBP) document [3] encourages and enables the continued expansion of the Web as a medium for the exchange of data. In this context, this paper focus on two cases of implementing the DWBP. The first one concerns data published by The Regional Center for Studies on the Development of the Information Society (Cetic.br) of The Brazilian Network Information Center (NIC.br). The second use case shows the experience of the Judiciary Department of Costa Rica (Justicia Abierta) with applying the DWBP Recommendation to publish their data on the Web.

KEYWORDS

Data on the Web, Best Practices, Application in Brazil and Costa Rica, Open Data

ACM Reference Format:

1 INTRODUCTION

One of the main goals of the Data on the Web Best Practices (DWBP) [3] is to facilitate interaction between publishers and consumers of data on the Web. A set of 35 Best Practices was created by W3C as a Web Recommendation 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.

In this context, this paper aims to illustrate the use of the DWBP through real use cases. For this, we present the Cetic.br [1] and the Justicia Abierta [2] use cases for applying the set of 35 Best Practices to publish their data on the Web. These two use cases illustrate some of the main challenges faced when applying the DWBP. On the one hand, their evaluation about the DWBP shows that a large number of best practices are easier to implement, on the other hand, both scenarios found difficulties in applying a subset of the best practices.

The remainder of this paper is organized as follows. Section 2 presents an overview of the DWBP. Section 3 presents the Brazil use case. Section 4 presents the Costa Rica use case. Section 5 presents some conclusions.

2 DATA ON THE WEB BEST PRACTICES

The Best Practices described on the Data on the Web Best Practices document [3] encourages and enables the continued expansion of the Web as a medium for the exchange of data. In broad terms, data publishers aim to share data either openly or with controlled access. Data consumers (who may also be publishers themselves) want to be able to find, use and link to the data, especially if it is accurate, regularly updated and guaranteed to be available at all times. This creates a fundamental need for a common understanding between data publishers and data consumers. Without this agreement, data publishers’ efforts may be incompatible with data consumers’ desires.

In this context, it becomes crucial to provide guidance to publishers that will improve consistency in the way data is managed. Such guidance will promote the reuse of data and foster trust in the data among developers, whatever technology they choose to use, increasing the potential for genuine innovation.

The set of Best Practices proposed in [3] was developed to cover the lack of technical guidance for publishing data on the Web. As described in [3], each Best Practice has an intended outcome, which describes "What it should be possible to do when a data publisher follows the best practice". In general, the intended outcome is an improvement in the way that a data consumer (human or software) can manipulate a dataset published on the Web. In some cases, the intended outcome reflects an improvement in the dataset itself, which will also result in a gain for the data consumer.

In order to encourage data publishers to adopt the DWBP, the recommendation proposes a number of distinct DWBP benefits, including: comprehensibility; processability; discoverability; reuse; trust; linkability; access; and interoperability. The benefits are important to engage data publishers in using the Best Practices because they help them to have a better understanding of "what will be possible" when one or more Best Practices are adopted.

Each benefit is associated to one or more specific Best Practices. For example, the benefit ‘Comprehensibility’ is associated to ten Best Practices, which are related to metadata, data vocabularies, feedback and data enrichment. This means that if a data publisher adopts these Best Practices, the level of comprehensibility will increase, i.e., it will be possible for humans to have a better understanding...
about the data structure, the data meaning, the metadata and the nature of the dataset.

The following list shows the set of best practices linked to the DWBP document:

- 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: Use content negotiation for serving data available in multiple formats
- Best Practice 15: Use Web Standards as the foundation of APIs
- Best Practice 16: Provide complete documentation for your API
- Best Practice 17: Provide bulk download
- Best Practice 18: Provide Subsets for Large Datasets
- Best Practice 19: Use real-time access
- Best Practice 20: Provide data up to date
- Best Practice 21: Provide an explanation for data that is not available
- Best Practice 22: Avoid Breaking Changes to Your API
- Best Practice 23: Preserve identifiers
- Best Practice 24: Assess dataset coverage
- Best Practice 25: Gather feedback from data consumers
- Best Practice 26: Make feedback available
- Best Practice 27: Enrich data by generating new data
- Best Practice 28: Provide Complementary Presentations
- Best Practice 29: Provide Feedback to the Original Publisher
- Best Practice 30: Follow Licensing Terms
- Best Practice 31: Cite the Original Publication

### 3 BRAZIL’S USE CASE

The Center for Studies on the Development of the Information Society (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. Doing this, the Cetic.br allows more people to access and use their data, as well as different devices could recognize and read their data, thus broadening the possibilities of use of the databases published by the Center [1].

![Table 1: Assessing the implementation of DWBP on Cetic.br microdata website](https://docs.google.com/spreadsheets/d/1yXyd3bmvLRLoKvKZ6ubd8_BrB47Y6t5g93eE0HEDw/edit?usp=sharing)

In order to improve the quality of their data publication process, the Center for Studies on the Development of the Information Society (Cetic.br) followed the steps described below to apply the DWBP:

1. Meetings with the teams of Cetic.br, Web Technology Study Center (Ceweb.br), Legal Department and Web Systems Department of NIC.br were held to level the knowledge of these teams on DWBP.

2. A checklist with the 35 Best Practices was made to understand which of them were applicable right away, which should be postponed and for how long, and the ones that were not applicable.

---

[1] https://docs.google.com/spreadsheets/d/1yXyd3bmvLRLoKvKZ6ubd8_BrB47Y6t5g93eE0HEDw/edit?usp=sharing
(3) The DWBP applicable were implemented on Cetic.br Microdata Website

An important step in the process of applying the DWBP concerns the meetings with the teams of Cetic.br, Ceweb.br, Legal Department and Web Systems Department of of NIC.br to level the knowledge of these teams on DWBP. After the meetings, the necessary adaptations - both in the implementation of the BPs and in the processes of the research of Cetic.br - , were implemented. This was also a first step to apply the DWBP and to become part of Cetic.br’s research process and data publication.

Table 1 shows the result of the checklist, meaning which BPs were applicable, which ones were not applicable and which ones should be adapted before being implemented.

The checklist presented in the Table 1 shows that 18 Best Practices could be applied immediately and the BP19 “Use content negotiation for serving data available in multiple formats” should be applied afterwards because the Web System Department of NIC.br needed more time to implement such BP.

Table 1: Result of the checklist

The checklist in Table 1 shows that 18 Best Practices could be applied immediately and the BP19 “Use content negotiation for serving data available in multiple formats” should be applied afterwards because the Web System Department of NIC.br needed more time to implement such BP.

The biggest challenge reported by the Justicia Abierta leader was the lack of knowledge on the part of public officials involved in the choice of those data and also some resistance by some guardians of the data. According to him, these difficulties were overcome by means of conversations and training that took place with the interdisciplinary group.

At the end, after the internal process have been established, of the 35 DWBP, the Justice Department of Costa Rica applied 17 BPs. It is noteworthy that, prior to step 1, only 3 BPs were adopted by the Costa Rica Justice Department.

The Figure 2 shows Justicia Abierta Website open data front page.
<table>
<thead>
<tr>
<th>Dimension</th>
<th>Best Practice</th>
<th>Applies</th>
<th>Applies immediately</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metadata</td>
<td>BP01</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>BP02</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>BP03</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Data Licenses</td>
<td>BP04</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Data Provenance</td>
<td>BP05</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Data Quality</td>
<td>BP06</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Data Versioning</td>
<td>BP07</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>BP08</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Data Identifiers</td>
<td>BP09</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>BP10</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>BP11</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Data Formats</td>
<td>BP12</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>BP13</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>BP14</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Data Vocabularies</td>
<td>BP15</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>BP16</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Data Access</td>
<td>BP17</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>BP18</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>BP19</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>BP20</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>BP21</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>BP22</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Data Access APIs</td>
<td>BP23</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>BP24</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>BP25</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>BP26</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>BP27</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>BP28</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Feedback</td>
<td>BP29</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>BP30</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Data Enrichment</td>
<td>BP31</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>BP32</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Republication</td>
<td>BP33</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>BP34</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>BP35</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

Table 2: Assessing the implementation of DWBP on the Justice Department website

5 DISCUSSION

This paper presented two real use cases of the DWBP application: the Cetic.br [1] and the Justicia Abierta [2] use cases. Both use cases adopted a similar methodology to implement the DWBP, including team meetings as well as a detailed evaluation of the DWBP.

Both Cetic.br and Justicia Abierta use cases demonstrated the importance of gathering all people involved in the process of publishing data. In the case of Cetic.br, even though the data was produced by the Center, holding meetings with the other departments involved such as the Legal Department and Web System Department of NIC.br, as well as with Ceweb.br to help the explanation about the DWBP, was very important to the process. In the case of Open Justice, conversations and trainings that took place with the interdisciplinary group were significant to overcome the difficulties on lack of knowledge and resistance by some guardians of the data.

It is important to note that in both cases some Best Practices could be implemented immediately: 1 to 5 - about Metadata, Data Licenses and Data Provenance; 7 and 8 - about Data Versioning; 12 and 14 - about Data Formats; 15 - about Data Vocabularies; 17 and 21 - on Data Access; 29 - about Feedback; and 32 - about Data Enrichment. This shows that some of the proposed best practices can be easily adopted without major modifications on the data publication process. On the other hand, neither were able to implement the BPs 10 - about Data Identifiers; 20 - on Data Access; 23 to 26 - about Data Access APIs; 31 - about Data Enrichment; and 33 to 35 - about Republication. For implementing these BPs, more technical or advanced knowledge becomes necessary.

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 them without great effort. Furthermore, according to both teams, understand the DWBP as a guide and not as rules to be fulfilled is of great importance. We understand that both Brazil’s and Costa Rica’s use cases may be a reference to other data publishers of how to use the DWBP to facilitate interaction between data publishers and consumers.

REFERENCES