

Metodologia_v1 (sob revisão)

Ressalta-se que o objetivo geral do trabalho é avaliar a disponibilização de dados sobre a execução orçamentária dos três Estados mais populosos do Brasil: São Paulo, Minas Gerais e Rio de Janeiro. Portanto é de interesse do grupo analisar como tal disponibilização facilita os processos de **encontrar (A)**, **entender(B)** e **processar(C)** as bases de dados. A seguir estão descritas as métricas empregadas em cada um dos processos.

A. ENCONTRAR os dados - Usabilidade do portal de transparência: neste aspecto é avaliada a interface de pesquisa dos três portais de transparência. Sobre a execução orçamentária são buscadas 2 bases de dados, uma para receita e outra para despesa de 2018, em cada site. A experiência dos usuários será analisada com base nas heurísticas de Nielsen:

1. Visibilidade do status do sistema:

O sistema deve prover feedback aos usuários, mantendo-os informados sobre o que está acontecendo, qual o atual estágio das requisições feitas.

2. Compatibilidade do sistema com o mundo real:

Deve-se avaliar se o sistema utiliza conceitos e linguagem familiares ao usuário em vez de termos orientados ao próprio sistema (jargões técnicos); e se utiliza convenções do mundo real, exibindo informações com uma ordem lógica e natural.

3. Controle do usuário e liberdade:

Dado que os usuários frequentemente escolhem por engano funções do sistema, eles precisam ter claras saídas de emergência para sair do estado indesejado sem a necessidade de percorrer um extenso caminho.

4. Consistência e padrões:

Evitar que seja necessário entender vários padrões e formas de interações diferentes para cada tela. Deve-se partir do pressuposto de que o usuário não pode prever que diferentes palavras, ícones, situações ou ações significam a mesma coisa.

5. Prevenção de erros:

O design deve prevenir que erros que ocorram por desatenção do usuário ou por falta de informação sobre o funcionamento do sistema.

6. Reconhecimento ao invés de lembrança:

O usuário não pode ser forçado a lembrar informações de uma parte do sistema para outra. Ele deve poder fazer suas escolhas sem a necessidade de recordar algum comando específico.

7. Flexibilidade e eficiência de uso:

Atalhos são essenciais para que usuários experientes executem as operações mais rapidamente. Abreviações, teclas de função, duplo clique no mouse são exemplos de facilitadores recomendados.

8. Estética e design minimalista:

O que for irrelevante não deve estar presente na interface. O sistema deve oferecer apenas o que é necessário para o usuário.

9. Ajudar os usuários a reconhecer, diagnosticar e corrigir erros:

As mensagens de erro devem ser claras, concisas e ajudar na correção do problema.

10. Help e documentação:

Se as outras heurísticas falharem, o usuário precisará consultar a ajudado sistema. Portanto a documentação deve conter informações objetivas e estar facilmente acessível.

Para cada um dos itens, pode haver de 0 a n problemas. A cada problema será atribuído um peso, conforme o grau de severidade:

Peso	Severidade	Efeito
1	Sem importância	Não afeta a operação da interface
2	Cosmético	Não há necessidade de solução imediata
3	Simples	Problema de baixa prioridade (<u>pode</u> ser reparado)
4	Grave	Problema de alta prioridade (<u>deve</u> ser reparado)
5	Catastrófico	Muito grave, deve ser reparado com urgência

Portanto, neste processo, quanto **menor** a pontuação do portal, mais bem avaliada é sua usabilidade.

B. ENTENDER os dados - A publicação das bases de dados na Web

O W3C, consórcio internacional de padronização da internet, publicou em janeiro de 2017 um guia de boas práticas para a publicação de dados na Web, *Data on the web Best Practices* (DWBP)¹. Ao realizar download das bases de dados sobre receitas e despesas no contexto da execução orçamentária, é feita uma análise de aderência de boas práticas do DWBP. As 35 boas práticas sugeridas no framework são descritas a seguir:

1. Fornecer metadados

Recomenda-se que a publicação permita aos humanos compreender os metadados, e aos agentes de software serem capazes de processá-los.

2. Fornecer metadados descritivos

Sugere-se que os seres humanos tenham condições de interpretar a natureza do conjunto de dados e suas distribuições, e os agentes de software tenham condições de descobrir automaticamente conjuntos de dados e distribuições.

3. Fornecer metadados estruturais

Diz que os seres humanos devem ser capazes de interpretar o esquema de um conjunto de dados, e os agentes de software capazes de processar automaticamente os dados das distribuições.

4. Fornece informação de licença de dados

As pessoas devem ser capazes de compreender a licença de dados, descrevendo eventuais restrições impostas à utilização de certos dados, agentes de software serão capazes de detectar automaticamente a licença de dados de uma distribuição.

5. Fornecer informação sobre procedência dos dados

Os seres humanos serão capazes de identificar a origem dos conjuntos de dados, e agentes de software serão capazes de processar automaticamente informações de proveniência.

¹ <https://www.w3.org/TR/dwbp/>

6. Fornecer informação sobre qualidade dos dados

Os seres humanos e os agentes de software serão capazes de avaliar a qualidade e, portanto, a adequação de um conjunto de dados para a sua aplicação.

7. Fornecer indicador de versão

Os seres humanos e os agentes de software poderão facilmente determinar qual versão de um conjunto de dados.

8. Fornecer histórico de versão

Os seres humanos e os agentes de software serão capazes de entender como o conjunto de dados muda de versão para versão e como quaisquer duas versões específicas diferem.

9. Usar URIs persistentes como identificadores dos datasets

Os conjuntos de dados ou informações sobre conjuntos de dados poderão ser descobertos e citados ao longo do tempo, independentemente da sua disponibilidade ou do formato dos dados.

10. Usar URIs persistentes como identificadores dentro dos datasets

Os itens de dados serão relacionados em toda a Web criando um espaço global de informação acessível a humanos e máquinas

11. Atribuir URIs para versões e séries dos datasets

Os seres humanos e os agentes de software serão capazes de referenciar versões específicas de um conjunto de dados, séries de conjunto de dados, bem como a versão mais recente de um conjunto de dados

12. Usar formatos de dados processáveis por computador

Máquinas serão capazes de ler e processar dados publicados na Web e os seres humanos serão capazes de usar ferramentas computacionais para manipular os dados.

13. Usar representações de dados com neutralidade de localidade (datas, moedas, números)

Os seres humanos e os agentes de software serão capazes de interpretar o significado do conjunto de caracteres (strings) que representam datas, horas, moedas, números, entre outros, com precisão.

14. Fornecer dados em formatos múltiplos

Tantos usuários quanto possível serão capazes de utilizar os dados sem ter que transformá-los em seu formato preferido.

15. Reusar vocabulários, de preferência padronizados

Interoperabilidade e consenso entre os publicadores e consumidores de dados serão reforçados.

16. Escolher o nível correto de formalização

Os casos de aplicação mais prováveis serão apoiados com não mais complexidade do que o necessário

17. Fornecer download em massa

Transferências de arquivos grandes, ou seja, que exigem mais tempo do que um usuário típico consideraria razoável, serão possíveis por meio de protocolos de transferência de arquivos dedicados.

18. Fornecer subconjuntos para datasets grandes

Os seres humanos e as aplicações serão capazes de acessar subconjuntos de um conjunto de dados, em vez de todo o conjunto. Isso proporcionará aos consumidores o acesso aos dados com uma elevada proporção de dados que são realmente necessários em comparação aos dados desnecessários. Conjuntos de dados estáticos considerados muito grandes poderão ser recuperados em porções menores. APIs podem ser usadas para filtrar os dados disponíveis. A granularidade de acesso aos dados poderá ser definida de acordo com as necessidades do domínio e as demandas de desempenho das aplicações.

19. Usar negociação de conteúdo para servir os dados em múltiplos formatos

Content negotiation permitirá que diferentes recursos ou representações diferentes de um mesmo recurso possam ser servidas de acordo com a requisição feita pelo cliente.

20. Fornecer acesso em tempo real

Aplicações serão capazes de acessar os dados em tempo real ou quase em tempo real; em tempo real significa um intervalo de milissegundos até alguns segundos após a criação de dados.

21. Fornecer dados atualizados

Os dados na Web serão atualizados em tempo hábil para que os dados disponíveis on-line reflitam os dados mais recentes divulgados através de qualquer outro canal. Quando novos dados estiverem disponíveis, logo que possível, serão publicados na Web.

22. Fornecer explicação para os dados que não estiverem disponíveis

Os consumidores saberão que os dados que são referenciados a partir do conjunto de dados não estão disponíveis ou se estão disponíveis sob diferentes condições.

23. Tornar os dados disponíveis por meio de uma API

Os desenvolvedores terão acesso aos dados para uso em seus próprios aplicativos, com dados atualizados e sem a necessidade de esforço por parte dos consumidores. As aplicações serão capazes de obter dados específicos por meio de consultas à API.

24. Usar padrões Web como fundamentos da API

Desenvolvedores que tenham alguma experiência com APIs baseadas em padrões Web, tais como o REST, já deverão ter um conhecimento inicial de como usar a API. Além disso, será mais fácil dar manutenção na API.

25. Fornecer documentação completa para a API

Descrever as funções da API, assinaturas de métodos e quaisquer informações pertinentes a desenvolvedores.

26. Evitar quebras na API

O código do desenvolvedor deve continuar válido após alterações na API. Os desenvolvedores devem ser notificados das melhorias feitas na API e devem ser capazes de fazer uso delas. Quebrar alterações em sua API será raro e, se ocorrer, os desenvolvedores terão tempo e informações suficientes para adaptar o seu código, aumentando a confiança na API. Alterações na API deverão ser anunciadas no site da documentação da API.

27. Preservar os identificadores

A URI de um conjunto de dados sempre levará para o conjunto propriamente dito ou então redirecionará para um recurso com informações sobre ele.

28. Avaliar a cobertura do dataset

Os usuários serão capazes de fazer uso de dados arquivados no futuro

29. Coletar feedback dos consumidores dos dados

Os consumidores de dados serão capazes de fornecer feedback e avaliações sobre conjuntos de dados e distribuições.

30. Tornar o feedback disponível

Os consumidores serão capazes de avaliar os tipos de erros que afetam o conjunto de dados, avaliar experiências de outros usuários e ter a certeza de que o publicador trata os problemas de forma adequada. Os consumidores também serão capazes de determinar se outros usuários já fizeram comentários semelhantes, poupando-lhes a submissão de relatórios desnecessários e poupando os publicadores de terem que lidar com duplicatas.

31. Enriquecer os dados ao gerar novos dados

Os conjuntos de dados com valores nulos poderão ser “corrigidos” a partir do preenchimento de tais valores. Estrutura poderá ser conferida aos dados e sua utilidade poderá ser melhorada se forem adicionadas medidas ou atributos relevantes. Porém, tal adição só deverá ser feita se não alterar os resultados analíticos, o significado ou o poder estatístico dos dados.

32. Fornecer apresentações complementares dos dados

Complementar os conjuntos de dados com possíveis visualizações permitirá que os consumidores humanos tenham uma visão imediata sobre os dados, apresentando-os de forma que possam ser facilmente compreendidos.

33. Fornecer feedback ao publicador original

Uma melhor comunicação entre publicadores e consumidores fará com que seja mais fácil para os publicadores originais determinar como os dados que eles publicam estão sendo usados. Isso ajudará a justificar o investimento na publicação dos dados. Os publicadores também serão informados de medidas que podem ser tomadas para melhorar a qualidade dos seus dados.

34. Seguir os termos de licença

Os publicadores serão capazes de confiar que seu trabalho está sendo reutilizado de acordo com os seus requisitos de licenciamento, tornando-os mais propensos a continuar com a publicação dos dados. Reutilizadores de dados vão ser capazes de licenciar adequadamente os trabalhos derivados a partir de dados previamente publicados.

35. Citar a publicação original

Os consumidores finais serão capazes de avaliar a origem dos dados e os esforços dos publicadores originais serão reconhecidos. A cadeia de proveniência para os dados na Web será rastreável de volta ao seu publicador original.

Para cada um dos três estados deseja-se avaliar se as recomendações são “atendidas”, “atendidas parcialmente” ou “não atendidas”, para as bases de dados sobre receita e despesa no ano corrente (2018).

Peso	Grau
0	Atende a recomendação
1	Atende parcialmente
2	Não atende

A exemplo do processo anterior, quanto **menor** a pontuação da base de dados publicada, maior sua adesão às práticas recomendadas.

C. PROCESSAR os dados - A qualidade dos dados disponibilizados, no contexto da execução orçamentária

Como framework para este aspecto, são utilizados os princípios apresentados no estudo “Assessing Open Government Budgetary Data in Brazil”², de 2013. O estudo apresenta 8 princípios de dados abertos. A seguir, estão descritos como tais princípios serão aplicados na avaliação das 6 bases envolvidas (receita e despesa de cada um dos 3 Estados).

1. **Completos:** quando os dados de receita e despesa são apresentados em uma série temporal. Neste aspecto, para o presente trabalho será empregada a seguinte pontuação:

Pontuação	Grau
0	Dados ausentes
1	Base com alguns dados em branco sem justificativa
2	Base completamente preenchida ou com dados em branco devidamente justificados

2. **Primários:** quando os dados são publicados na forma coletada na fonte, com a mais fina granularidade possível. Neste caso, segue-se a lógica do estudo de 2013 que considerou os níveis de classificação de receita e despesa para gerar as pontuações:

RECEITAS	
Pontuação	Nível de Detalhamento
0	Sem classificação
1	Categoria Econômica
2	Origem
3	Espécie
4	Rubrica
5	Alínea
6	Subalínea

DESPESAS	
Pontuação	Nível de Detalhamento
0	Sem classificação
1	Categoria Econômica
2	Grupo de Natureza de Despesa
3	Modalidade de Aplicação
4	Elemento de despesa
5	Item (subelemento) da despesa

² https://www.researchgate.net/publication/236033125_Assessing_Open_Government_Budgetary_Data_in_Brazil

3. **Atuais:** os dados são disponibilizados com atualizações em tempo real, conforme Art. 48, §1º, inciso II da Lei Complementar 101/2000³. A avaliação aplicada neste princípio será a seguinte:

Pontuação	Grau
0	Não há atualização em tempo real
1	Há atualização em tempo real

4. **Acessíveis:** os dados são publicados para o público mais abrangente possível. Tal princípio já foi avaliado como boa prática no framework do DWBP,
5. **Processáveis por máquina:** os dados são razoavelmente estruturados para possibilitar o seu processamento automatizado. Este princípio não será avaliado nesta etapa, uma vez que já foi contemplado entre as práticas do DWBP.
6. **Acesso não discriminatório:** dados disponibilizados sem necessidade de identificação e registros no portal.

Pontuação	Grau
0	Exige algum registro
1	Portal não exige registro do usuário

7. **Formatos não proprietários:** os dados estão disponíveis em um formato sobre o qual nenhum ente tenha controle exclusivo. Este princípio também não será avaliado nesta etapa, uma vez que já foi contemplado entre as práticas do DWBP.
8. **Livres de licenças:** os dados não estão sujeitos a regulações de direitos autorais, marcas, patentes ou segredo industrial. Também avaliado no DWBP.

Note-se que ao contrário das etapas A e B, nesta avaliação, quanto **maior** a pontuação, mais bem avaliada é a transparência do Estado.

³ http://www.planalto.gov.br/ccivil_03/LEIS/LCP/Lcp101.htm#art48p.