

Uma Proposta de Processo para Implantação de Dados Abertos em Instituições Públicas Brasileiras

Artigo submetido à Edição Especial sobre Sistemas de Informação e os Desafios do Mundo Aberto

Renan de Oliveira Silva¹, Gibeon Soares de Aquino Júnior²

¹Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Norte (IFRN)
Caixa Postal 59550-000 – João Câmara – RN – Brasil

²Departamento de Informática e Matemática Aplicada (DIMAp) – Universidade Federal do Rio Grande do Norte (UFRN)
Caixa Postal 59063-400 – Natal – RN – Brasil

renan.silva@ifrn.edu.br, gibeon@dimap.ufrn.br

***Abstract.** Open Data initiative has been gaining strength in recent times, with increasing participation of public institutions and governments. However, for being a new experience and because there are not many disseminated guides and processes, many difficulties emerge, causing delays in publications and other problems. Therefore, the objective of this work is to define a process that helps brazilian public institutions to open their data, guiding them during the phases and tasks of publication. For this, we carried out a systematic mapping of the literature, in order to discover strategies, best practices and difficulties that exist in the field, as well as other existing processes for study and comparison purposes.*

***Resumo.** A iniciativa de Dados Abertos vem ganhando força nos últimos tempos, com adesão crescente de instituições públicas e governos. Porém, por ser uma experiência nova e por ainda não haver muitos guias e processos bem difundidos, muitas dificuldades acabam surgindo, ocasionando atrasos nas publicações e outros problemas. Assim sendo, o objetivo deste trabalho é definir um processo que auxilie instituições públicas brasileiras a abrirem seus dados, guiando-as durante as fases e tarefas de publicação. Para isso, foi realizado um mapeamento sistemático da literatura, a fim de descobrir estratégias, melhores práticas e dificuldades que existem na área, além de outros processos já existentes para fins de estudo e comparação.*

1. Introdução

As iniciativas de Dados Abertos vêm ganhando força principalmente em instituições governamentais e científicas para, entre outras coisas, conquistar confiança pública, alcançar transparência, estimular inovações e entregar crescimento econômico (OPEN KNOWLEDGE INTERNATIONAL, 2017). Do ponto de vista conceitual, Dados Abertos referem-se aos dados que podem ser livremente usados, reusados e redistribuídos por todos e para qualquer propósito (BARGH, CHOENNI E MEIJER, 2016).

Essa iniciativa possui diversos pontos positivos, como o potencial de gerar produtos e serviços úteis para a sociedade, como fornecer mecanismos de transparência, controle social, cidadania ativa, cooperação, melhorias na administração pública e criação de novas ferramentas, trazendo desenvolvimento e melhoria no bem-estar social (KUCERA et al., 2016). Além disso, permite que a própria sociedade explore os dados disponibilizados de maneira mais efetiva do que seus próprios donos, que geralmente possuem apenas uma visão limitada sobre eles, utilizando-os para atividades simples.

As características mais importantes dos Dados Abertos, segundo a Open Knowledge International (2017), são: **disponibilidade e acesso**, segundo a qual dados devem estar disponíveis como um todo, de forma conveniente e modificável; **possibilidade de reuso e redistribuição**; e **participação universal**, segundo a qual todos devem poder usar, reusar e redistribuir os dados.

Apesar da grande e crescente adoção aos Dados Abertos e dos esforços para tornar a publicação e uso dos dados melhor e mais eficiente, ainda existem diversas dificuldades a serem superadas nesse processo de publicação. Instituições que desejam disponibilizar seus dados ainda enfrentam uma série de desafios e incertezas ao longo desse processo. Um dos desafios é a definição da melhor maneira de disponibilizar esses dados, tendo em vista que hoje existem várias formas e padrões recomendados. Outro ponto relevante é quanto à definição das melhores práticas a serem utilizadas, de forma a guiá-las no processo, dando a segurança de que estão realmente no caminho correto. Organizações como a W3C (World Wide Web Consortium) têm se esforçado para criar padrões relacionados aos Dados Abertos, porém a linguagem utilizada não é facilmente acessível a gestores de instituições públicas, que necessitam de um manual simples que compile informações como essas de maneira mais adequada.

Nota-se que ainda existe uma certa carência de processos que auxiliem as instituições durante a abertura de seus dados (ZUIDERWIJK et al., 2014), de forma semelhante à época em que os processos de desenvolvimento de *softwares* eram escassos, o que ocasionava diversos problemas para a indústria (SOMMERVILLE, 2011). Por causa disso, instituições geralmente enfrentam muitos problemas para disponibilizar seus dados, que ocorrem por diversos motivos, como a falta de expertise na área, a dificuldade de encontrar processos que os auxiliem, insegurança sobre quais dados publicar e até mesmo falta de articulação interna, o que acaba atrasando seus projetos de Dados Abertos.

Assim sendo, o objetivo geral deste trabalho é propor um novo processo para publicação de dados por parte de instituições públicas brasileiras. Para isso, foram definidos como procedimentos metodológicos: a realização de uma Revisão do Estado da Arte em busca de processos, estratégias, boas práticas, desafios e dificuldades existentes na área de publicação de dados institucionais; e a avaliação do processo proposto com base em casos reais de instituições que publicaram seus dados, comparando as atividades realizadas com as propostas, com o intuito de verificar sua utilidade e eficiência.

Este artigo está estruturado como se segue: na Seção 2, é apresentada a Revisão do Estado da Arte realizada, assim como seus resultados; na Seção 3, é apresentado e detalhado o processo criado; na Seção 4, são apresentados dois estudos de casos realizados; por fim, na Seção 5, é feita a conclusão.

2. Revisão do Estado da Arte

Nesta seção, será apresentada uma revisão do estado da arte da literatura baseada na técnica de mapeamento sistemático. O objetivo é obter uma visão geral sobre trabalhos relacionados à disponibilização de dados organizacionais. Além disso, serão comentados os resultados dessa revisão, bem como os principais trabalhos relacionados.

2.1. Planejamento e Execução do Mapeamento Sistemático

De acordo com Petersen, Vakkalanka e Kuzniarz (2015), um mapeamento sistemático é usado para obter a estrutura dos trabalhos publicados e seus resultados, de maneira a possibilitar sua classificação. Com isso, é possível que se tenha uma visão abrangente de uma determinada área. O objetivo desse método de pesquisa é identificar possíveis lacunas no meio científico atual, com o intuito de sugerir novos temas de pesquisa.

Kitchenham e Charters (2007) explicam que mapeamentos sistemáticos são projetados para fornecer uma visão ampla de uma área de pesquisa, para estabelecer se existem evidências de pesquisa sobre um determinado tópico e fornecer uma indicação da quantidade de evidência.

O planejamento e execução da metodologia utilizada foram baseados no guia descrito por Petersen, Vakkalanka e Kuzniarz (2015). A Figura 1 ilustra o processo realizado.



Figura 1. Processo de Revisão do Estado da Arte realizado

O mapeamento sistemático deve responder as seguintes questões de pesquisa:

- Q1: Quais as formas existentes de publicação de dados institucionais na Web?
- Q2: Quais são as estratégias e boas práticas existentes para disponibilização de dados institucionais na Web?
- Q3: Quais desafios/dificuldades são encontrados no processo de abertura de dados de uma instituição?

A *string* de busca definida foi a seguinte:

(“open api” OR “open data”) AND (“process” OR “method” OR “technique”).

Dessa forma, procurou-se principalmente por estudos que se tratassem de processos, ou que possuísem métodos ou técnicas para abertura de dados. Essa *string* de busca¹ foi utilizada diretamente na *engine* de busca do Scopus², tendo sido aplicados, ainda, outros filtros, a saber: subárea de computação, ano de publicação a partir de 2012

¹ O termo “open data” foi incluído na *string* de busca para se referir ao processo de forma geral, enquanto que o termo “open api”, para incluir resultados referentes à publicação por meio de APIs (*Application Programming Interface*), permitindo a comparação das estratégias e cenários de uso dessas duas formas.

² <https://www.scopus.com/>

e busca apenas no título e resumo. O Scopus foi escolhido por consultar diretamente nas bases de pesquisa mais relevantes para este estudo, como, por exemplo: *IEEE Xplore Digital Library*, *ACM Digital Library*, *SpringerLink*, entre outras.

A Tabela 1 ilustra o processo de seleção dos artigos que compõem este estudo, partindo da fase inicial – a busca no Scopus, que encontrou 1213 – até a fase final – da leitura completa dos 31 trabalhos selecionados.

Para que fossem selecionados em cada fase, os artigos em questão deveriam se encaixar em pelo menos um dos seguintes critérios: comentar ou avaliar alguma forma de disponibilização de dados na Web; estabelecer, avaliar ou comentar estratégias e/ou boas práticas existentes para publicação de dados institucionais; comentar sobre experiências e/ou dificuldades no processo de abertura de dados de uma instituição.

Tabela 1. Mapeamento da quantidade de artigos em cada fase

Fase	Quantidade de artigos no início da fase	Descrição da fase	Quantidade de artigos após a fase
0	0	Execução da <i>string</i> de busca na base de dados	1213
1	1213	Leitura e análise de título e resumo	108
2	108	Download de artigos	99
3	99	Leitura e análise de introdução e conclusão	38
4	38	Leitura completa de cada estudo	31

2.1.1. Ameaças à validade

Existem algumas ameaças à validade da revisão do estado da arte deste estudo. Elas estão descritas a seguir, juntamente com as estratégias utilizadas para tentar mitigar cada uma delas:

- **Viés de publicação:** não há como garantir que todos os estudos relevantes foram obtidos. Porém, tentou-se incluí-los ao máximo. Antes da definição da *string* de busca final, foram realizados vários testes para tentar identificar a que retornasse resultados mais significativos, de maneira que apresentasse equilíbrio entre a quantidade de resultados e sua relevância. Além disso, foram observadas as referências dos artigos selecionados, de maneira que as mais relevantes também pudessem contribuir com a pesquisa.
- **Análise de artigos:** durante o processo de análise, os artigos foram classificados baseados no julgamento do avaliador. Para mitigar essa ameaça, quando havia dúvidas sobre a possibilidade de contribuição do estudo em questão, ele era mantido para a fase seguinte, só sendo descartado quando ficava evidente de que fugia ao tema pesquisado.

2.2. Resultados e Discussões

Nesta seção, serão relatados os resultados alcançados com o mapeamento sistemático e que irão auxiliar a responder as questões de pesquisa definidas na Seção 2.1. Os 31

artigos selecionados pelo mapeamento foram listados em uma tabela³ juntamente com as questões de pesquisa que cada um pode ajudar a responder.

Foram encontradas três formas diferentes de abertura de dados: Dados Brutos (DB), Dados Conectados (DC) e API. A primeira forma, DB, é bastante utilizada e consiste em liberar informações “cruas”, como o próprio nome sugere. Normalmente, as informações não são pré-processadas, sendo publicadas da maneira em que são encontradas na sua fonte. Exemplos comuns são bases de dados e planilhas.

A segunda forma de abertura de dados, DC, se refere a um conjunto de melhores práticas para publicação e ligação de dados estruturados na Web. Essas melhores práticas têm sido adotadas por um número crescente de provedores de dados ao longo dos últimos anos, levando à criação de um espaço global de dados que contém vários bilhões de asserções - a nuvem⁴ de Dados Abertos Conectados (JANEV; MIJOVIC; VRANES, 2015). Jovanovik et al. (2015) afirmam que isso permite transformar a Web de uma Web de documentos para uma Web de dados, imaginada na ideia original da Web Semântica, permitindo, através da natureza interligada de seus conjuntos de dados, cenários de casos de uso que antes não eram possíveis, devido a seu isolamento.

O último método reportado, API, se refere a serviços que podem ser utilizados por aplicações na Web para obter dados do fornecedor da API. Suas principais vantagens são a possibilidade de obter dados em tempo real e filtragem de dados (W3C, 2016). Sua desvantagem é que é preciso ter conhecimentos em programação para usá-la.

Tim Berners-Lee (BERNERS-LEE, 2006), criador da *World Wide Web*, propôs uma classificação para avaliar a qualidade de um dado aberto, tendo sido apontada em vários dos estudos analisados. Nela, é estabelecido um padrão de avaliação baseado em estrelas, numa escala que varia de 1 a 5. Para alcançar cada nível, é necessário que o conjunto de dados obedeça aos seguintes critérios:

- **uma estrela:** estar disponível na Web, em qualquer formato, com licença aberta;
- **duas estrelas:** deve estar disponível como dados estruturados que possam ser lidos por máquina, como uma planilha em vez de uma imagem escaneada;
- **três estrelas:** deve atender ao critério de duas estrelas, porém estando em um formato aberto, por exemplo, CSV em vez de Excel;
- **quatro estrelas:** obedecer aos critérios anteriores e usar padrões abertos da W3C, como RDF e SPARQL, estabelecendo um identificador global para identificar coisas, para que as pessoas possam apontar para seus dados.
- **cinco estrelas:** obedecer aos critérios anteriores e ligar seus dados a dados de outras pessoas, com o objetivo de fornecer contexto.

Nota-se que o conceito de Dados Conectados é plenamente atendido quando se atinge a classificação de 5 estrelas. A W3C também sugere DC como a forma mais completa pela qual um dado pode ser publicado (W3C, 2016). Portanto, pode-se concluir que essa seria a forma “plena” para publicação, levando-se em consideração o

³ Disponível em: <https://goo.gl/Auh7LW>

⁴ Disponível em: <http://lod-cloud.net/>

peso e relevância das fontes informadas. Por outro lado, DB seriam uma forma inicial, já que, com ela, só é possível atingir o nível inicial ou, no máximo, um intermediário, segundo a classificação de Berners-Lee (2006). Apesar do uso de DB não permitir atingir os critérios máximos de qualidade, sua importância não pode ser negada, dado que vários países utilizam essa forma de publicação, além de representar um primeiro passo importante, principalmente porque a construção de DC não é trivial e exige bastante esforço, algo que nem todas as instituições têm condições de fazer (BARGH, CHOENI E MEIJER, 2016).

A abertura de dados através de API não foi tão citada pelos artigos analisados, porém não se pode negar sua importância. O uso de APIs é uma prática recomendada pela W3C, pois oferece uma ótima flexibilidade e processabilidade para consumidores de dados, podendo possibilitar uso de dados em tempo real, filtragem na requisição, e a habilidade de trabalhar com os dados em um nível atômico (W3C, 2016). Há situações em que o uso de APIs é ainda mais adequado, como para grandes *datasets*, dados que precisem ser frequentemente atualizados ou disponibilizados em tempo real, ou ainda para dados que sejam muito complexos (W3C, 2016). Pode-se dizer, no entanto, que APIs representam apenas um complemento à publicação na forma de DB ou de DC, e não devem ser utilizadas como a única forma de publicação, pois elas não são acessíveis a todos, tendo em vista que, para utilizá-las, são necessários conhecimentos em programação. Essa é uma possível explicação para o fato de elas não terem sido tão citadas na revisão. Apesar disso, as APIs não perdem sua utilidade, pois podem permitir e facilitar a construção de uma ampla gama de serviços à sociedade e de aplicações.

Dessa forma, para responder à Q1, pode-se chegar às seguintes conclusões:

- **Dados Brutos:** representam o estágio inicial de publicação. Já consiste em um grande passo para instituições e governos. Serão úteis, desde que sejam feitos em formatos legíveis por máquina, como CSV (*Comma-separated Values*). Evitar arquivos em formatos PDF (*Portable Document Format*) ou imagens, pois eles não são facilmente processáveis por máquina.
- **API:** representa um estágio intermediário de publicação. Trará mais uma opção de consulta para programadores de aplicações.
- **Dados Conectados:** estágio mais avançado de publicação de dados. Torna os dados interligados, e mais fáceis de se manipular computacionalmente.

Por ser a forma mais simples de publicação, é recomendável que instituições que não tenham prática ou que estejam começando a publicar seus dados o façam através de DB. Dessa forma, a chance de sucesso provavelmente será maior, assim como a motivação da equipe, por ter mais chances de obter progresso rapidamente nas fases iniciais. Isso seria particularmente adequado em instituições públicas pequenas, que geralmente não possuem pessoal qualificado o suficiente, por exemplo, na área de tecnologia, para já começar com outra forma de publicação mais complexa. Após esse início, caso a instituição tenha os recursos para tal, poderia também fornecer dados através de APIs. Com isso, traria mais uma opção de acesso, facilitando a vida principalmente de programadores, que poderiam criar novas aplicações usando as APIs. Finalmente, se a instituição perceber que a iniciativa está tendo um bom retorno da comunidade e também tiver recursos para isso, seria interessante publicá-los na forma

de DC, que são mais fáceis de se manipular computacionalmente, porém mais difíceis de serem criados. Dessa forma, evita-se o risco de desenvolver algo não trivial e que poderia não ter retorno.

Em relação à resposta para a Q2, uma série de boas práticas, estratégias e processos para publicação de dados foram relatados. Os mais relevantes são descritos na Tabela 2.

Conclui-se que a maior parte das estratégias e boas práticas encontradas na literatura referem-se a Dados Conectados e ao estabelecimento de processos organizacionais, que não abordam aspectos técnicos de publicação de dados. Nota-se a falta de estudos que guiem instituições ao abrirem seus dados quanto a esses aspectos, bem como de estudos genéricos o suficiente que permitam atender a qualquer instituição pública, ou pelo menos a maioria delas, tendo em vista que boa parte dos trabalhos encontrados foram criados para instituições específicas.

Tabela 2. Lista de estratégias, boas práticas e processos mais relevantes encontrados, por artigo

Artigo(s)	Estratégias/Boas Práticas/Processos
Koide et al. (2013); Janev, Mijovic and Vranes (2015); Jovanovik et al. (2015); entre outros.	Conversão de DB para DC; definição de vocabulários (ontologias); interligação de dados; e uso de ferramentas que possam auxiliar no processo de conversão.
Zuiderwijk et al. (2012)	Propõe um guia com várias questões a serem levadas em consideração para decidir se determinados conjuntos de dados judiciais poderiam ser abertos ou não.
Segundo (2013)	Seguir a “Cartilha Técnica para Publicação de Dados Abertos no Brasil” e o “Manual dos Dados Abertos”.
Zuiderwijk et al. (2014)	Estabelece um processo organizacional para auxiliar uma instituição pública holandesa específica a abrir seus dados, estabelecendo papéis e melhorando a comunicação interna.
Agbabiaka e Ojo (2014)	Processo organizacional para avaliação do nível de preparo de uma instituição para disponibilizar dados; envolvimento de cidadãos em decisões.
Kucera et al. (2016)	Utilização de metodologia de publicação com quatro fases (desenvolvimento de plano, preparação, publicação e arquivamento); tarefas executadas em cada fase.
Lobos et al. (2016)	Estabelece um processo de abertura simplificado, no qual a iniciativa não é da instituição e sim do cidadão interessado.

Quanto à Q3, a Tabela 3 ilustra os desafios e dificuldades mais relevantes encontrados:

Os desafios e dificuldades reportados dão a entender que conjuntos de dados que têm sido publicados, principalmente na forma de DB, possuem problemas, tanto quanto à sua interpretação quanto à sua possibilidade de utilização. Isso sugere que as instituições publicadoras talvez não estejam seguindo as melhores práticas. Também são destacadas as dificuldades de se transformar DB em DC. Em alguns casos, devido à ausência de processos difundidos, foi necessário criar processos novos para contornar as dificuldades enfrentadas.

Tabela 3. Lista de desafios/dificuldades mais relevantes encontrados, por artigo

Artigo(s)	Desafios/Dificuldades reportados
Segundo (2013); Jovanovik et al. (2015); entre outros	Conjuntos de dados (DB) difíceis de entender e publicados em formatos proprietários, tornando a leitura por máquina impossível.
Jovanovik et al. (2015); Knoblock et al. (2013); entre outros	Dificuldades em transformar DB em DC, como: definição de ontologia, interligação de informações, volume e diversidade de metadados, e utilização das ferramentas para conversão.
Bergamaschi et al. (2015)	Seleção, uso e integração de dados; construção de métricas para avaliar a qualidade de dados disponíveis.
Koide et al. (2013)	Definição de um processo para publicação de dados; escolha de um sistema de gerenciamento de conteúdo; e configuração de servidor para publicação de dados.
Zuiderwijk et al. (2012) e Zuiderwijk et al. (2014)	Publicação de dados privados ou sensíveis, possibilidade de danos à reputação da organização, problemas de comunicação interna.

Na próxima subseção, será feita uma análise dos processos principais que foram encontrados através do mapeamento sistemático.

2.3. Trabalhos Relacionados

Após a finalização da Revisão do Estado da Arte, notou-se que existem poucos processos relacionados à abertura de dados por instituições públicas que sejam aplicáveis à maioria delas. A seguir, será feita uma análise dos principais trabalhos relacionados que se encaixam nessa descrição.

Kucera et al. (2016) desenvolveram um processo bastante completo para instituições que desejarem publicar seus dados. De certa forma, ele é parecido com o processo de desenvolvimento de *software* conhecido como RUP (*Rational Unified Process*), devido à sua completude e extensividade, sendo um processo pesado, rigoroso e complexo. Por isso mesmo, de acordo com os próprios autores, não envolve instituições públicas pequenas, que, devido à quantidade e complexidade das tarefas, podem não conseguir segui-lo.

Ávila (2015), por sua vez, criou um processo que leva em consideração o nível de maturidade da instituição, ou seja, processos mais simples para instituições com menos recursos. Apesar disso, o processo pode ser considerado complexo e relativamente pesado, principalmente porque ele foi baseado no Modelo de Desenvolvimento de *Software* em Espiral e no PMBoK (*Project Management Body of Knowledge*), possuindo, ainda, um alto número de tarefas. Sendo assim, instituições pequenas provavelmente ainda teriam dificuldades em utilizá-lo e, principalmente, compreendê-lo, dependendo da equipe que possuam. Além disso, o processo em questão se baseou fortemente nas boas práticas sugeridas pela W3C, deixando de lado sugestões espontâneas de outros estudos, que também poderiam ser relevantes. O processo não chegou a ser aplicado em instituições governamentais.

Sendo assim, conclui-se, pelas evidências mostradas na Revisão do Estado da Arte, que os processos atuais na área de publicação de dados governamentais são poucos

e insuficientes. Há a necessidade de um processo que seja adequado para a maioria das instituições governamentais existentes no país e que seja simples, prático e direto, de maneira a facilitar seu uso, pois os que foram encontrados e que poderiam ser utilizados são burocráticos e complexos. Será apresentada nas próximas seções uma proposta de processo que possa suprir essas necessidades. Ele foi desenvolvido a partir da observação e compilação dos conjuntos de boas práticas obtidos com a Revisão do Estado da Arte, o que é mais um diferencial, levando-se em consideração que as sugestões de boas práticas foram obtidas de forma abrangente. O processo recebeu o nome de ProceDA (Processo para Dados Abertos).

3. ProceDA: Um Processo para Dados Abertos em Instituições Públicas Brasileiras

Para a criação do ProceDA, efetuou-se a releitura dos artigos selecionados pelo mapeamento sistemático, os quais foram anteriormente listados. Além disso, verificou-se suas referências mais relevantes. A partir disso, pôde-se realizar a leitura de vários outros estudos importantes, os quais foram listados em uma tabela⁵ à parte. Essa ampliação objetivou obter o maior número de boas práticas e tarefas relacionadas à publicação de dados.

Após a leitura dos materiais em questão, foi possível obter uma lista com as boas práticas e tarefas mais relevantes relacionadas à publicação de dados. Elas foram agrupadas em oito fases, em consonância com suas naturezas. Cada uma dessas fases é composta por tarefas, sendo que cada uma precisa ser executada por um ou mais papéis, conforme será detalhado nas próximas subseções.

3.1 Definição do ProceDA

O processo definido foi baseado no modelo de desenvolvimento de *software* Iterativo e Incremental, baseado em ciclos. Segundo Villazón-Terrazas et al. (2011), isso é importante porque permite a melhoria contínua da publicação de dados, através da execução de várias iterações. A Figura 2 ilustra as fases do processo, bem como sua sequência.



Figura 2. Processo e suas fases

⁵ Disponível em: <https://goo.gl/ksuk17>

As fases deverão ser executadas preferencialmente na ordem definida na Figura 2. Ao concluir a última (Manutenção e Otimização), volta-se para a fase de Seleção de Dados, iniciando um novo ciclo de publicação. No entanto, não é intenção deste processo estabelecer uma sequência rígida de fases. Caso seja necessário, pode-se voltar de uma fase para outra, para corrigir algum problema ou executar alguma tarefa que ficou faltando, por exemplo.

3.2 Especificação do Processo

Nesta seção, será feito o detalhamento completo do processo. Serão explicadas e detalhadas as suas oito fases, bem como as tarefas e papéis associados a cada uma delas.

3.2.1 Estabelecimento de Equipe

Esta é a primeira fase do processo. Nela, deve-se definir os responsáveis por coordenar e por executar o processo de abertura de dados. A liderança envolvida é de extrema importância e sua definição pode impactar diretamente no fracasso ou sucesso da iniciativa.

Os seguintes papéis foram identificados no processo de abertura de dados (GRAY et al., 2011):

- **Órgão Central de Suporte e Controle das Ações (OCSCA):** órgão que ficará responsável por auxiliar e coordenar o processo de abertura de dados, removendo eventuais obstáculos e problemas que surgirem durante o caminho, bem como articulando todas as unidades que irão abrir seus dados. As funções do OCSCA podem ser assumidas por alguma diretoria, coordenação ou secretaria já existente, ou pode ser criada uma nova.
- **Diretor:** responsável pela iniciativa e principal interessado na abertura de dados. Geralmente é um Executivo da alta gestão, que pode ou não estar diretamente associado ao OCSCA. Pode ser o diretor geral da instituição ou algum diretor específico, como um Diretor de TI, por exemplo.
- **Coordenador:** responde pela execução do processo de abertura, articulando entre os responsáveis técnicos e gestores de negócio. Será o coordenador geral, que acompanhará de perto as tarefas executadas, estando abaixo hierarquicamente apenas do Diretor.
- **Gestor Negocial:** tem a capacidade de identificar dados importantes que podem ser abertos, os dados de risco e dados confidenciais. Por exemplo, poderia ser alguém que domine as regras de negócio da instituição em questão.
- **Gestor Técnico:** responde pelas ações técnicas do processo de abertura, como extração e apresentação dos dados. Indivíduo que possua conhecimentos técnicos para elaborar os conjuntos de dados. Poderia ser, por exemplo, alguém com bons conhecimentos em TI.
- **Usuário do Dado:** cliente final do Dado Aberto, sendo pessoa física ou jurídica que tenha interesse de acesso ao dado. Pode ser a sociedade civil, *hackers*, a iniciativa privada ou os próprios gestores públicos.

A instituição que estiver publicando seus dados deverá definir se será necessário

apenas um Coordenador, um Gestor Negocial e um Gestor Técnico, ou se seriam necessários mais de um. Essa definição irá depender do tipo de órgão que esteja disponibilizando seus dados. Por exemplo, para o caso de uma prefeitura, é provável que o mais indicado seja que o Órgão Central de Suporte e Controle das Ações seja a Secretaria de Planejamento, por exemplo. Essa secretaria seria responsável por coordenar as outras secretarias detentoras de dados que serão abertos, como, por exemplo, as Secretarias de Turismo, de Educação, entre outras. Dada a natureza dessas secretarias, é provável que o mais indicado seja haver um Coordenador, um Gestor Negocial e um Gestor Técnico para cada uma delas, visto que uma secretaria costuma trabalhar com dados totalmente diferentes das outras. Por outro lado, levando em consideração o caso de uma Instituição Federal de Ensino Superior (IFES), é provável que um único Coordenador, um único Gestor Negocial e um único Gestor Técnico sejam suficientes para todas as unidades dessa IFES, visto que os dados são muito semelhantes entre essas unidades. Ainda para o caso de uma IFES, um possível Órgão Central de Suporte e Controle das Ações seria a Diretoria de Tecnologia da Informação.

É necessário avaliar os recursos humanos que a instituição possui. A situação ideal é quando ela tem os recursos humanos necessários para todos os papéis, o que normalmente ocorre quando se trata de uma instituição de médio a grande porte. Porém, em casos de instituições menores, como, por exemplo, uma unidade de saúde ou uma escola municipal/estadual, idealmente as responsáveis pela disponibilização dos dados devem ser as secretarias às quais essas instituições estão atreladas. No caso do exemplo dado, as responsáveis deveriam ser, respectivamente, as Secretarias de Saúde e de Educação, considerando que são elas, em grande parte dos casos, as detentoras das informações sobre essas entidades. As secretarias também normalmente possuem mais recursos para permitir a publicação de dados. Em alguns casos, pode acontecer de a instituição possuir apenas alguns dos papéis necessários. Nesses casos, poderia haver uma colaboração técnica entre a instituição em questão e a secretaria ou ministério responsável por ela. Exemplo: consideremos que um hospital estadual deseja publicizar seus dados, porém ele não possui recursos humanos suficientes para isso, pois falta alguém com conhecimentos suficientes para ser o Gestor Técnico. Existem duas abordagens para resolver esse problema. Na primeira delas, a equipe da Secretaria Estadual de Saúde poderia ficar responsável por todo o processo. Na segunda abordagem, a responsabilidade poderia ficar dividida entre a equipe do hospital e da Secretaria Estadual de Saúde, que poderia fornecer alguém para ocupar a função de Gestor Técnico, que estava faltando. Isso só reforça o quanto é importante que a iniciativa de Dados Abertos seja apoiada pela Alta Gestão, incluindo aí principalmente os políticos, sem o apoio e interesse dos quais não é possível que todas as instituições consigam obter sucesso no processo. As particularidades de cada entidade devem, portanto, ser levadas em consideração no momento de atribuição dos papéis.

Após a definição da equipe, é possível passar para a próxima etapa.

3.2.2 Diagnóstico e Planejamento

Nesta fase, os atores envolvidos devem fazer uma análise da situação atual da instituição, verificando uma série de questões.

Conforme destacado em Nations (2013), não é necessário reinventar a roda

quando se trata de desenvolver um projeto de Dados Abertos. Agências governamentais que estejam começando seu processo de publicação devem ser encorajadas a examinar e comparar diferentes abordagens de publicação de dados realizadas por outras instituições, de forma a eliminar dúvidas, que são naturais e que existirão durante o processo. Pode-se, por exemplo, verificar seus portais de catalogação de dados, em busca de várias informações, como, por exemplo, quais dados publicar.

A Tabela 4 lista as tarefas associadas a esta fase. Nas próximas subseções, serão fornecidos maiores detalhes sobre cada uma.

Tabela 4. Tarefas da fase de Diagnóstico e Planejamento e os papéis responsáveis por cumpri-las

Tarefa	Papel(éis)
Realizar estudo de viabilidade inicial	Responsáveis: OCSCA, Diretor, Coordenador. Envolvidos: Gestor Negocial e Gestor Técnico.
Estabelecer Plano de Dados Abertos	Responsáveis: OCSCA, Diretor, Coordenador. Envolvidos: Gestor Negocial e Gestor Técnico.

3.2.2.1. Realizar estudo de viabilidade inicial

No estudo de viabilidade inicial, o órgão deve analisar a extensão de seu comprometimento com a causa, os recursos materiais e humanos que estão disponíveis para executar a iniciativa, a legislação atual, bem como o envolvimento da gestão dentro da instituição e demandas externas (ESPINAR, 2014).

Deve ser feita também uma avaliação do nível de preparo da instituição para disponibilização de dados. É necessário avaliar o órgão quanto aos seguintes aspectos: Pessoas, Processo, Tecnologia e Sistema (AGBABIKA; OJO, 2014). O fator Pessoas é um componente crítico e talvez o mais importante. É necessário avaliar a qualidade e nível de competência da equipe. É preciso, também, avaliar se a equipe tem pleno conhecimento do Processo, pois é necessário que todos estejam apropriados das etapas, tarefas e como executá-las, de modo que tudo flua da forma correta. Quanto ao aspecto Tecnologia, deve ser focada a avaliação das tecnologias e infraestruturas de TI (Tecnologia da Informação), como *softwares*, *hardwares*, rede, banco de dados e portal de catalogação, para verificar se atendem aos requisitos necessários para publicação de dados. Por fim, quanto ao fator Sistema, devem ser verificados aspectos de governança, leis, políticas e práticas atuais.

Esses estudos permitirão que a instituição verifique a viabilidade atual de execução da iniciativa de Dados Abertos. Caso seja concluído que no momento a iniciativa é inviável, o ideal é que se tente encontrar soluções para contornar os problemas detectados para que ela possa, de fato, ser executada.

3.2.2.2 Estabelecer Plano de Dados Abertos

Instituições públicas federais são obrigadas, pelo Decreto Presidencial nº 8.777, de 11 de maio de 2016, a elaborar um Plano de Dados Abertos. Segundo o Ministério do Planejamento, Orçamento e Gestão (2017), esse é o documento que orienta as ações de implementação e promoção de abertura de dados, que deverão obedecer aos padrões

mínimos de qualidade, de forma a facilitar o entendimento e a reutilização de informações. Instituições que não sejam legalmente obrigadas a estabelecer um PDA podem criar um documento simplificado, apenas com questões primordiais, como as que serão comentadas a seguir.

Nesse documento, deve ser feito um cronograma com prazos para publicação, no qual pode ser considerado também como se dará a implantação do portal de catalogação⁶ da instituição. Além disso, devem ser explicitados os critérios para priorização e seleção de dados a serem publicados.

Foram identificadas, na literatura, duas formas para priorização e escolha de conjuntos de dados a serem disponibilizados. Segundo Nations (2013) e Espinar (2014), essas duas formas são: publicações orientadas a demandas e publicações orientadas a ofertas. Na primeira forma de priorização (publicações orientadas a demandas), as instituições precisam encontrar quais conjuntos de dados possuem alta demanda pela sociedade. Isso pode ser feito através da realização de pesquisas ou estudos. Nas publicações orientadas a ofertas, procura-se quais conjuntos de dados são os mais fáceis de publicar. Isso inclui dados que já estejam pré-prontos, em formatos estruturados, ou aqueles que normalmente são publicados em PDF, bastando sua conversão para um formato mais adequado. Publicações pequenas e fáceis podem atuar como catalisadoras para mudanças comportamentais maiores dentro das organizações (NATIONS, 2013).

Pode ser uma boa ideia, ainda, combinar as duas abordagens (NATIONS, 2013). Em cada ciclo do processo, poderá ser aplicada uma estratégia diferente ou, ainda, no mesmo ciclo podem ser disponibilizados dados advindos das duas estratégias.

3.2.3 Seleção de Dados

Nesta fase, devem ser definidos os conjuntos de dados que serão publicados na iteração atual. Para isso, deve-se escolher pelo menos uma das abordagens definidas no Plano de Dados Abertos. Deve-se, então, estabelecer conjuntos de dados prioritários, com base na estratégia de publicação escolhida. Para auxiliar, pode-se identificar benefícios e riscos de abrir cada conjunto de dados, bem como fazer uma estimativa da quantidade de trabalho para abri-los (KUCERA; CHLAPEK, 2014). A Tabela 5 lista as tarefas desta fase.

Tabela 5. Tarefas da fase de Seleção de Dados e os papéis responsáveis por cumpri-las

Tarefa	Papel(éis)
Definir a estratégia de priorização de conjuntos de dados do ciclo atual	Responsáveis: Diretor (opcional), Coordenador. Envolvido: Gestor Negocial.
Definir conjuntos de dados prioritários e escolher os que serão publicados	Responsáveis: Diretor (opcional), Coordenador. Envolvido: Gestor Negocial.

O ideal é que se comece de forma pequena, simples e rápida. Não há necessidade de que todos os conjuntos de dados sejam publicados de uma vez. Várias iniciativas começaram com poucos conjuntos de dados (NATIONS, 2013). Isso permite que a

⁶ Um dos *softwares* mais utilizados para catalogação é o CKAN. Disponível em: <https://ckan.org/>.

instituição adquira experiência e obtenha resultados de forma mais rápida. Como o processo é dividido em várias iterações, novos conjuntos de dados podem ser publicados nos ciclos seguintes, sem maiores problemas.

Os conjuntos de dados selecionados para publicação devem possuir valor agregado para os usuários finais, sob pena de inviabilizar sua reutilização, consequentemente não atingindo plenamente os objetivos dos Dados Abertos (JANSSEN; CHARALABIDIS; ZUIDERWIJK, 2012).

3.2.4 Modelagem de Dados

Na fase de Modelagem de Dados, é realizada a preparação dos conjuntos de dados que foram selecionados para publicação na etapa anterior. A preparação inclui a limpeza e a escolha do melhor formato de publicação, bem como outras considerações. Ao concluir a fase, os dados deverão estar prontos para serem publicados.

O ideal é que os dados selecionados estejam em sua forma mais natural ou bruta, ou seja, na forma em que são encontradas em sua fonte, antes de quaisquer cruzamentos ou agregações com outros dados (STI, 2011). Isso porque quem deve se encarregar de realizar tais cruzamentos é o usuário final, caso necessite.

A Tabela 6 lista as tarefas da fase de Modelagem de Dados e que serão comentadas nos parágrafos seguintes.

Tabela 6. Tarefas da fase de Modelagem de Dados e os papéis responsáveis por cumpri-las

Tarefa	Papel(éis)
Modelar os dados em mais de um formato, caso seja possível	Responsável: Gestor Técnico.
Manter um padrão nos formatos publicados	Responsável: Gestor Técnico.
Definir a forma de publicação: DB, DC e/ou API	Responsáveis: Diretor, Coordenador. Envolvido: Gestor Técnico.
Elaborar metadados	Responsáveis: Gestor Negocial e/ou Técnico.
Limpeza dos dados	Responsável: Gestor Técnico e/ou Técnico.
Tratar informações confidenciais	Responsáveis: Gestor Negocial e/ou Técnico.

Conforme exposto na Tabela 6, é recomendável que se disponibilize os dados em mais de um formato, caso seja possível (STI, 2011). Isso minimiza a possibilidade de introdução de erros no processo de transformação entre formatos pelos usuários, já que eles não terão de fazê-lo, e também facilita seus trabalhos. É interessante, também, que haja um padrão nesses formatos publicados, dando preferência sempre aos mesmos (KUCERA et al., 2016). Formatos comuns estruturados utilizados para publicação de dados são os seguintes: HTML (*Hypertext Markup Language*), XML (*Extensible Markup Language*), CSV, ODS (*Open Document Sheet*), JSON (*JavaScript Object Notation*) e RDF (*Resource Description Framework*). Este último é o formato recomendado para publicação na forma de DC. CSV e JSON são os mais recomendados

para os demais casos (LEE; CYGANIAK; DECKER, 2014).

O tipo de publicação do ciclo atual deve ser escolhido nesta etapa, durante a tarefa “Definir a forma de publicação”. A instituição deve escolher entre DB, API e/ou DC. Para isso, deve levar em consideração as conclusões obtidas através da resposta da Q1 do mapeamento sistemático, descritas na Seção 2.2.

Caso o nível escolhido para publicação seja o de quatro estrelas da classificação de Berners-Lee (2006), deve-se, nesta etapa, escolher os vocabulários a serem utilizados no arquivo RDF, bem como definir as URIs (*Uniform Resource Identifier*) que serão utilizadas para identificar cada conjunto de dados. Caso se escolha pelo nível de 5 estrelas (DC), deve-se conectar os dados a outros, além do exposto anteriormente para 4 estrelas (BAUER; KALTENBÖCK, 2012).

É obrigatório que os dados publicados possuam uma explicação associada, ou seja, possuam metadados (JANSSEN; CHARALABIDIS; ZUIDERWIJK, 2012). Isso pode ser feito através da publicação de um arquivo à parte, explicando todos os dados que são exibidos no arquivo original. Fazendo-se isso, os usuários entenderão os conjuntos de dados mais facilmente, contribuindo para sua reutilização. Essas ações estão previstas na tarefa “Elaborar metadados”.

A tarefa “Limpeza dos dados” indica que eles deverão passar por uma verificação, com o intuito de impedir publicações de informações errôneas, prejudicando a credibilidade da iniciativa e evitando causar impactos negativos na sociedade (BAUER, KALTENBÖCK, 2012).

Por fim, a última tarefa da Tabela 6 informa que é necessário ter cuidado com informações confidenciais. Deve haver uma política para controle de que tipo de informações podem ser publicadas, que irá variar de instituição para instituição (JANSSEN; CHARALABIDIS; ZUIDERWIJK, 2012). Em geral, dados que possam identificar pessoas ou outras informações sensíveis devem ser omitidos, porém o restante das informações deve ser publicado.

3.2.5 Definição de Licença

Dados publicados na Web que não possuem termos de uso ou licença explícita são um problema, pois criam um aspecto de insegurança jurídica para seus reutilizadores, podendo inviabilizar sua utilização. Dessa forma, é indispensável definir os termos de uso dos dados a serem publicados, tornando-os explícitos para os usuários finais. Deve ser escolhida uma licença que permita que os dados sejam abertos, não proprietários, em formatos estáveis e de amplo uso.

Tim Berners-Lee exemplifica duas licenças que poderiam ser utilizadas: Creative Commons CC-BY⁷ e a Licença do Governo Aberto do Reino Unido⁸ (BERNERS-LEE, 2006).

A Tabela 7 ilustra a tarefa de definição de licença e seus papéis associados.

⁷ <https://creativecommons.org/licenses/by/3.0/br/>

⁸ <http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/>

Tabela 7. Tarefa da fase de Definição de Licença e os papéis responsáveis por cumpri-la

Tarefa	Papel(éis)
Especificar uma licença para os dados	Responsáveis: Diretor, Coordenador. Envolvido: Gestor Técnico.

3.2.6 Publicação de Dados

Nesta etapa, será realizada, de fato, a publicação dos dados que foram selecionados e modelados nas fases anteriores. A Tabela 8 identifica as tarefas da fase juntamente aos papéis responsáveis por executá-las.

Tabela 8. Tarefas da fase de Publicação e os papéis responsáveis por cumpri-las

Tarefa	Papel(éis)
Publicar os dados no portal de catalogação	Gestor Técnico
Informar a versão dos conjuntos de dados	Gestor Técnico
Publicar metadados	Gestor Técnico
Elaborar URLs únicas e persistentes para os conjuntos de dados	Gestor Técnico
Explicitar termos de uso no portal de catalogação	Gestor Técnico

De acordo com a primeira tarefa da Tabela 8, os dados devem ser publicados em um catálogo, como o CKAN, que já deve estar previamente configurado e no ar. Ao publicar, é importante informar a versão dos conjuntos de dados, conforme visto na segunda tarefa da Tabela 8. Isso serve para permitir rastreabilidade por parte dos reutilizadores dos dados e pode ser feito inserindo a informação nos metadados, no próprio arquivo, na página do portal ou na própria URL do dado (W3C, 2016).

Ainda de acordo com a Tabela 8, cada conjunto de dados deve possuir um identificador único e persistente, seguindo uma padronização na formação da URL, para que possa ser usado continuamente por aplicativos. Um identificador persistente é aquele que não muda nunca sua URL. Isso é fundamental para evitar que aplicações deixem de funcionar abruptamente.

Por fim, como última tarefa, deve-se explicitar, na página do conjunto de dados no portal de catalogação, seus termos de uso (W3C, 2016), que foram definidos na fase anterior.

3.2.7 Promoção

Após a fase de Publicação, vem a de Promoção. A ideia é que os dados publicados recebam publicidade, em diferentes meios, para que diversos setores da sociedade possam ser alcançados, de modo a incentivar seu reuso.

Para isso, esses conjuntos de dados podem ser publicados em redes sociais oficiais da instituição, bem como em seus sites oficiais, sendo divulgados em seções destacadas desses portais (ÁVILA, 2015). A instituição pode considerar, ainda, outras promoções de reuso, como eventos e comunicações maiores, dependendo do estado atual da publicação e de seu reuso. *Hackathons* ou maratonas de programação também

podem ser organizadas, por exemplo, anualmente, para arrecadar investimentos para *startups* (ESPINAR, 2014).

Por fim, é importante que o catálogo de dados da instituição seja integrado ao catálogo central do governo⁹, pois isso também aumenta sua visibilidade (STI, 2011).

A Tabela 9 mostra as tarefas desta fase, bem como os papéis responsáveis em executá-las.

Tabela 9. Tarefas da fase de Promoção e os papéis responsáveis em cumpri-las

Tarefa	Papel(éis)
Publicar notícias em redes sociais e sites oficiais	Coordenador
Considerar eventos e comunicações maiores	Diretor, Coordenador
Integrar catálogo da instituição ao catálogo central do governo, caso ainda não tenha sido feito	Gestor Técnico

3.2.8 Manutenção e Otimização

A Tabela 10 lista as tarefas da fase de Manutenção e Otimização e os papéis associados a elas.

Tabela 10. Tarefas da fase de Manutenção e Otimização e os papéis responsáveis em cumpri-las

Tarefa	Papel(éis)
Verificar se algum conjunto de dados precisa ser atualizado	Responsáveis: Coordenador, Gestor Negocial, Gestor Técnico
Monitoramento e ajuste da execução do processo	Responsáveis: Diretor, Coordenador. Envolvidos: Gestor Negocial, Gestor Técnico.
Verificar como os dados estão sendo usados pela comunidade	Responsáveis: Coordenador, Gestor Técnico
Facilitar requisições de dados através de portais oficiais	Responsáveis: Coordenador, Gestor Técnico
Facilitar <i>feedback</i> dos usuários sobre conjuntos de dados individuais	Responsáveis: Coordenador, Gestor Técnico

Nesta fase, como primeira tarefa, deve-se dar atenção aos conjuntos de dados que já foram publicados anteriormente. Deve-se verificar se algum conjunto de dados precisa ser atualizado e fazê-lo. Também é preciso fazer o monitoramento e ajuste de toda a execução do processo, tentando identificar possíveis problemas ocorridos e elaborando possíveis estratégias para combatê-los.

Outra tarefa importante é verificar como os dados estão sendo usados pela comunidade e se eles têm atendido à demanda dos usuários finais. Para isso, podem ser

⁹ <http://dados.gov.br/>

elaborados formulários de opinião no catálogo de dados e/ou nos portais oficiais da instituição. Deve-se procurar facilitar requisições de dados através dos portais oficiais, de modo a melhor identificar as necessidades da sociedade, bem como facilitar o *feedback* dos usuários sobre conjuntos de dados individuais, através do próprio portal de catalogação, preferencialmente.

Ao concluir a fase de Manutenção e Otimização, o ciclo do processo também é terminado. A partir disso, a instituição deve decidir se continua as publicações em um novo ciclo ou não.

4. Estudos de Caso

Nesta seção, serão descritos os estudos de casos que foram realizados para avaliar o ProceDA. Esses estudos foram feitos em parceria com duas entidades públicas: Universidade Federal do Rio Grande do Norte (UFRN) e Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Norte (IFRN).

A UFRN concluiu sua política de Dados Abertos em setembro de 2016, através de seu Plano de Dados Abertos (UFRN, 2016). Hoje, a Universidade possui 58 conjuntos de dados publicados em seu portal¹⁰, pertencentes a 11 categorias. O IFRN, por sua vez, concluiu seu Plano de Dados Abertos em março de 2017 (IFRN, 2017). Hoje, o instituto possui 12 conjuntos de dados publicados em seu portal¹¹, pertencentes a oito categorias. A seguir, será feito o detalhamento do estudo de caso.

Os objetivos dos estudos de caso foram:

- verificar se as práticas do ProceDA são condizentes com a realidade, se são úteis e se de fato podem auxiliar entidades públicas a disponibilizarem seus dados;
- avaliar se existem aspectos do ProceDA que devem ser modificados.

Para alcançá-los, as seguintes questões de pesquisa devem ser respondidas:

- **Q1:** O ProceDA é útil para ser utilizado por instituições públicas brasileiras que desejem disponibilizar seus dados na Web?
- **Q2:** Quais são os benefícios, problemas e desafios associados à utilização do ProceDA?

Nas próximas seções, serão fornecidos maiores detalhes sobre os estudos realizados. O planejamento e descrição dos resultados foram baseados nas orientações propostas em Yin (2003), Kitchenham, Pickard e Pfleeger (1995) e Runeson e Höst (2009).

4.1. Execução

Nesta seção, será mostrado como se deu a abertura de dados da UFRN e do IFRN. A Tabela 11 exhibe como essas instituições executaram cada tarefa sugerida pelo ProceDA. Esses dados foram obtidos através da realização de entrevistas com os responsáveis pela iniciativa de Dados Abertos em ambas as instituições.

¹⁰ <http://dados.ufrn.br/>

¹¹ <https://dados.ifrn.edu.br/>

Tabela 11. Lista de fases e tarefas do ProceDA e seu mapeamento com o que foi realizado pelo IFRN e pela UFRN

Estabelecimento de Equipe	
Estabelecimento da equipe responsável	UFRN: a equipe foi composta pelo Comitê Gestor de Tecnologia da Informação (CGTI), Coordenação do Serviço de Informação ao Cidadão, Pró-reitorias e unidades administrativas e Superintendência de Informática. IFRN: a equipe foi composta por dois responsáveis da área de TI, um responsável pelo Planejamento e Desenvolvimento Institucional, um representante da Comunicação Social, um representante da Ouvidoria e o Diretor de Gestão de TI.
Diagnóstico e Planejamento	
Realizar estudo de viabilidade inicial	UFRN: a instituição considerou que possuía as condições necessárias e não realizou essa tarefa de maneira explícita. IFRN: foram avaliados quais dados seriam priorizados e como seria a metodologia de implantação e construção. Não foi necessário lidar com o pessoal envolvido, no sentido de remover dúvidas e obstáculos.
Estabelecer Plano de Dados Abertos	UFRN e IFRN: Os Planos de Dados Abertos das instituições estão disponíveis ao público em seus portais oficiais.
Seleção de Dados	
Definir a estratégia de priorização de conjuntos de dados	UFRN: foram priorizados os dados mais solicitados por meio do Serviço de Informação ao Cidadão, dados já armazenados nos sistemas de informação, informações já publicadas no Portal da Transparência da instituição e, por fim, dados relevantes contidos nas páginas mais acessadas nos portais da UFRN. Em resumo, foram seguidas as publicações orientadas a demandas e a ofertas. IFRN: Foram avaliados os dados mais solicitados pela Ouvidoria e Acesso à Informação, bem como quais dados estariam disponíveis através do sistema estruturante do IFRN, o SUAP (Sistema Unificado de Administração Pública). Em outras palavras, foram seguidas as publicações orientadas a demandas e a ofertas.
Definir conjuntos de dados prioritários e escolher os que serão publicados	UFRN e IFRN: em cada ciclo de publicação, foram definidos os conjuntos de dados de acordo com as estratégias de priorização.
Modelagem de Dados	
Modelar os dados em mais de um formato	UFRN: por questão de agilidade, os dados são publicados apenas em CSV. IFRN: as publicações são feitas em CSV e JSON.
Manter um padrão nos formatos publicados	UFRN: o formato de publicação é sempre o mesmo: CSV. IFRN: os arquivos são sempre publicados em CSV e JSON.
Definir a forma de publicação	UFRN: foram feitas publicações na forma de DB e API. IFRN: foram realizadas publicações na forma de DB.
Elaborar metadados	UFRN e IFRN: metadados são sempre elaborados.
Limpeza dos dados	UFRN: os responsáveis pelos dados devem verificá-los, homologando-os antes de serem publicados. IFRN: a instituição não realizou esta tarefa.

Tratar informações confidenciais	UFRN e IFRN: informações confidenciais ou sensíveis não são publicadas.
Definição de Licença	
Especificar uma licença para os dados	UFRN: foi escolhida a Licença Aberta. IFRN: foi escolhida a <i>Creative Commons Attribution</i> .
Publicação de Dados	
Publicar os dados no portal de catalogação	UFRN e IFRN: os dados encontram-se publicados nos portais de catalogação oficiais das instituições.
Informar a versão dos conjuntos de dados	UFRN e IFRN: o número da versão e a última data de atualização são exibidos nas páginas de cada conjunto de dados.
Publicar metadados	UFRN e IFRN: foram publicados para todos os conjuntos de dados.
Elaborar URLs únicas e persistentes para os conjuntos de dados	UFRN e IFRN: as URLs de cada conjunto de dados são únicas, não se alteram com o tempo e possuem um padrão de formação.
Explicitar termos de uso no portal de catalogação	UFRN e IFRN: na página de cada conjunto de dados, é exibida sua respectiva licença.
Promoção	
Publicar em redes sociais e sites oficiais sobre os conjuntos de dados	UFRN: foram publicadas notícias quando o portal de catalogação foi disponibilizado e para promoção do I <i>hackathon</i> da instituição. IFRN: foram publicadas notícias quando o portal de catalogação foi disponibilizado.
Considerar eventos e comunicações maiores	UFRN: foi realizado um <i>hackathon</i> . IFRN: encontra-se em planejamento.
Integrar catálogo da instituição ao catálogo central do governo	UFRN e IFRN: a integração foi realizada.
Manutenção e Otimização	
Verificar se algum conjunto de dados precisa ser atualizado	UFRN e IFRN: cada conjunto de dados possui seu tempo de atualização pré-definido.
Monitoramento e ajuste da execução do processo	UFRN: está previsto no PDA e sendo executada. IFRN: está sendo realizada. Uma proposta é a criação de ferramenta para extração e atualização automática de dados.
Verificar como os dados estão sendo usados pela comunidade	UFRN: está sendo estudada uma forma de realizar esta tarefa. IFRN: está prevista a realização de pesquisas de opinião e de melhoria dos conjuntos de dados.
Facilitar requisições de dados através de portais oficiais	UFRN e IFRN: é possível realizar requisições através do Serviço de Informação ao Cidadão.
Facilitar <i>feedback</i> dos usuários sobre conjuntos de dados individuais	UFRN: Ainda não é possível. IFRN: é disponibilizado o e-mail da Diretoria de TI.

Os responsáveis pelas iniciativas de Dados Abertos de ambas as instituições afirmaram, ainda através de entrevistas, que não consideram nenhuma tarefa do ProceDA como não importante ou desnecessária. Além disso, nenhuma das instituições

precisou realizar tarefas que não tivessem sido sugeridas pelo processo. Foi dito, ainda, que um processo como o ProceDA poderia auxiliar outras instituições a disponibilizarem seus dados, aumentando sua produtividade.

4.2. Ameaças à validade

A captura dos dados deste estudo de caso foi realizada por meio de entrevistas com as equipes, o que não garante que todos os aspectos importantes foram obtidos. Procurou-se, no entanto, conseguir o máximo de informações relevantes possíveis. Além disso, a execução dos processos de disponibilização de dados foi guiada pelas próprias instituições, o que não garante que o ProceDA foi plenamente seguido. Porém, conforme exposto na Tabela 11, houve uma boa correspondência.

4.3. Respostas às questões de pesquisa e Discussões

Percebe-se, ao analisar a Tabela 11, que grande parte das sugestões de tarefas do ProceDA foram realizadas pela UFRN e pelo IFRN. Isso evidencia que a proposta possui sugestões reais e que realmente podem ajudar a orientar e guiar instituições pelo caminho correto.

Quanto à fase de estabelecimento de equipe executada pelo IFRN, pode-se fazer uma analogia entre os papéis propostos pelo ProceDA e os que foram utilizados pela instituição. Os técnicos em TI utilizados pelo IFRN podem ser considerados como os Gestores Técnicos propostos pelo ProceDA; o responsável pelo Planejamento e Desenvolvimento Institucional e o responsável pela Ouvidoria podem ser considerados como Gestores Negociais do ProceDA, tendo em vista que eles detêm o conhecimento das regras de negócio e dados da instituição; o Diretor de Gestão de TI do IFRN exerceu acumuladamente os papéis de Diretor e de Coordenador sugeridos pelo ProceDA; por fim, o responsável pela Comunicação Social do IFRN exerceu a publicação de notícias em portais oficiais da instituição, o que no ProceDA eram tarefas previstas para o Coordenador. O Órgão Central de Suporte e Controle das Ações não foi instituído formalmente pela entidade, porém observou-se que ele foi desempenhado informalmente pela Diretoria de Gestão de TI. Portanto, conclui-se que o único papel que o IFRN utilizou que não possui correspondência direta com os propostos pelo ProceDA foi o de Comunicação Social. Os demais foram utilizados, embora não com a mesma nomenclatura.

Pode-se fazer, também, uma comparação entre os papéis propostos pelo ProceDA e os que foram utilizados pela UFRN. Pode-se dizer que o CGTI exerceu, de forma simultânea, as funções do OCSCA, de Diretor e de Coordenador do ProceDA, pois esse comitê foi quem realizou a maior parte das tarefas propostas para esses três papéis. A Coordenação do Serviço de Informação ao Cidadão e as pró-reitorias e unidades administrativas podem ser consideradas como equivalentes ao Gestor Negocial do ProceDA, tendo em vista que elas detêm o maior conhecimento sobre os dados a serem publicados e suas relevâncias. Por fim, a Superintendência de Informática exerceu o papel de Gestor Técnico. Portanto, conclui-se que houve correspondência total entre os papéis sugeridos pelo ProceDA e os que foram utilizados pela UFRN, apesar de alguns deles terem sido acumulados, algo que não é proibido pelo processo em questão.

Quanto às demais fases e tarefas, houve uma boa correspondência entre o que foi

previsto e o que foi executado pelas instituições.

Assim sendo, para responder a Q1, conclui-se que este estudo de caso serve como um indício de que o ProceDA possui tarefas factíveis e que podem auxiliar outras instituições quanto à publicação de seus dados. A UFRN e o IFRN, que são casos de sucesso, executaram grande parte do que é proposto pelo ProceDA e não realizaram nenhuma atividade que não está presente no processo. Isso evidencia que ele é simplificado e completo ao mesmo tempo.

Quanto à Q2, pode-se apontar como benefícios do ProceDA o fato de ele ser um processo reproduzível e que possui apenas tarefas essenciais à publicação de dados, além de representar um guia com um bom potencial de aumentar a produtividade das instituições que o utilizarem. Neste estudo de caso, não foram identificados problemas ou desafios associados à utilização do ProceDA.

5. Conclusão

Conforme visto no mapeamento sistemático realizado na Seção 2, existem poucos processos aptos para serem utilizados e seguidos por todos os tipos de instituições públicas brasileiras. Por isso, como resultado deste trabalho, propôs-se um novo processo para publicação de dados por essas instituições. Com o ProceDA, elas possuirão um guia de fácil entendimento para executarem, de maneira ágil, sucinta e controlada, seus projetos de Dados Abertos, seguindo as melhores práticas da área, que foram traduzidas em suas tarefas. O processo em questão é adequado para todos os tipos de instituições públicas brasileiras, sejam elas grandes ou pequenas, pois se trata de um processo leve, de fácil entendimento e com poucos papéis envolvidos, o que representa um diferencial com relação a outros já existentes. Sem um processo como este, as equipes dessas instituições poderiam ter dificuldades, gerando um alto grau de insegurança em grande parte do tempo, o que poderia ocasionar vários problemas, como atrasos, publicações mal feitas e a consequente não reutilização dos dados por parte dos usuários, impedindo que se atinja os benefícios dos Dados Abertos.

Os estudos de caso realizados demonstraram que o ProceDA é capaz de guiar as instituições pelo caminho correto. Ambos foram de sucesso, demonstrando que as sugestões do processo são realizáveis e correspondem à realidade. Devido a isso, considera-se que o ProceDA está apto a auxiliar outras instituições.

5.1. Contribuições

As contribuições do trabalho realizado podem ser definidas como as seguintes: Revisão do Estado da Arte dos processos de publicação de dados existentes; listagem das boas práticas, estratégias e dificuldades relacionadas à publicação de dados; definição de um processo, denominado ProceDA, para publicação de dados governamentais por instituições públicas brasileiras; realização de estudos de caso que verificaram a utilidade e a executabilidade do ProceDA.

5.2. Limitações

A principal limitação deste trabalho está relacionada ao fato de ele não ter sido, ainda, colocado em prática em um grande número de instituições públicas com naturezas diferentes. Por isso, não é possível precisar como se daria sua implantação e se os

benefícios esperados realmente se concretizariam em todos os casos. Contudo, como atenuante, durante a fase de Manutenção e Otimização do processo, na tarefa “Monitoramento e ajuste da execução do processo”, permite-se a realização de ajustes ao mesmo, de modo a sanar os principais problemas e dificuldades encontrados. Assim, acredita-se e espera-se que eventuais problemas não venham a prejudicar sobremaneira o andamento dos demais ciclos nem inviabilizar o processo como um todo.

5.3. Trabalhos Futuros

Espera-se poder colocar o processo em prática em um maior número de instituições públicas brasileiras e avaliar seu desempenho. É interessante, ainda, desenvolver um modelo de maturidade para avaliar o grau de aderência das organizações com relação às práticas do processo, estabelecendo um plano para que possam chegar a um ponto de excelência.

Por fim, pelo que se observou nos estudos de casos e na revisão do estado da arte, seria interessante o desenvolvimento de ferramentas genéricas para resolução de problemas comumente enfrentados. São exemplos: 1) extração de dados presentes em vários tipos de bancos de dados existentes e sua conversão em formatos pré-definidos, como CSV; 2) atualização automática de dados que já estejam publicados em portais de catalogação, eliminando a necessidade de intervenção humana.

Referências

- AGBABIKA, O.; OJO, A. Framework for assessing institutional readiness of government organisations to deliver open, collaborative and participatory services. In: Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance. [S.l.]: Association for Computing Machinery, 2014. p. 186–189.
- ÁVILA, T. J. T. Uma proposta de modelo de processo para publicação de Dados Abertos Conectados Governamentais. Tese (Projeto de Diplomação) — Instituto de Computação, UFAL, Maceió, AL, Brasil, nov. 2015.
- BARGH, M. S.; CHOENNI, S.; MEIJER, R. Meeting open data halfway: On semi-open data paradigm. In: Proceedings of the 9th International Conference on Theory and Practice of Electronic Governance. [S.l.]: Association for Computing Machinery, 2016. p. 199–206.
- BAUER, F.; KALTENBÖCK, M. Linked Open Data: The Essentials. 1. ed. Vienna: edition mono/monochrom, 2012.
- BERGAMASCHI, S. et al. Open data for improving youth policies. In: Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management. [S.l.]: SciTePress, 2015. p. 118–129.
- BERNERS-LEE, T. Linked Data. jul. 2006. Jul., 2006. Disponível em: <<https://www.w3.org/DesignIssues/LinkedData.html>>. Acesso em Novembro 20, 2016.
- ESPINAR, M. Álvarez. Government Data Openness and Re-Use. 1. ed. Catalunya: Autonomous Body for the Official Gazette, 2014.

GRAY, J. et al. Manual dos Dados Abertos: Governo. 1. ed. São Paulo: Comitê Gestor da Internet no Brasil, 2011. Cooperação técnica científica entre Laboratório Brasileiro de Cultura Digital e o Núcleo de Informação e Coordenação do Ponto BR (NIC.br).

IFRN. Plano de Dados Abertos. 2017. Disponível em: <http://portal.ifrn.edu.br/conselhos/consup/resolucoes/2017/resolucao-no-21-2017/at_download/file>. Acesso em 16 Jan., 2018.

INTERNATIONAL, O. K. OPEN DATA HANDBOOK. out. 2017. Disponível em: <<http://opendatahandbook.org>>. Acesso em 13 Out., 2017.

JANEV, V.; MIJOVIC, V.; VRANES, S. Linked data approach to the psi directive implementation: Supporting tools and lessons learned. In: Proceedings of the 23rd Telecommunications Forum. [S.l.]: Institute of Electrical and Electronics Engineers Inc, 2015. p. 823–826.

JANSSEN, M.; CHARALABIDIS, Y.; ZUIDERWIJK, A. Benefits, adoption barriers and myths of open data and open government. Information Systems Management (ISM), v. 29, n. 4, p. 258–268, 2012.

JOVANOVIK, M. et al. Linked open data for medical institutions and drug availability lists in macedonia. In: Proceedings of the 18th East European Conference on Advances in Databases and Information Systems and Associated Satellite Events. [S.l.]: Springer Verlag, 2015. p. 245–256.

KITCHENHAM, B.; CHARTERS, S. Guidelines for performing Systematic Literature Reviews in Software Engineering. [S.l.], 2007.

KITCHENHAM, B.; PICKARD, L.; PFLEEGER, S. L. Case studies for method and tool evaluation. IEEE Softw, v. 12, n. 4, p. 52–62, 1995.

KOIDE, S. et al. An lod practice lessons learned from open data meti. In: Proceedings of the Joint International Workshop on 2013 Linked Data in Practice Workshop. [S.l.]: CEUR-WS, 2013. p. 12–26.

KNOBLOCK, C. A. et al. Publishing data from the smithsonian american art museum as linked open data. In: Proceedings of the 12th International Semantic Web Conference. [S.l.]: CEUR-WS, 2013. p. 129–132.

KUCERA, J.; CHLAPEK, D. Comparison of approaches to publication of open government data in two czech public sector bodies. JeDEM, v. 6, n. 1, p. 106–111, 2014.

KUCERA, J. et al. Methodologies and best practices for open data publication. In: Proceedings of the Dateso 215. [S.l.]: CEUR-WS, 2016. p. 52–64.

LEE, D.; CYGANIAK, R.; DECKER, S. Open Data Ireland: Best Practice Handbook. 1. ed. Ireland: Insight, 2014.

LOBOS, V. E. et al. Proposal of a unified bpm model for open data publication. In: Proceedings of the Third International Conference on eDemocracy and eGovernment (ICEDEG). [S.l.]: IEEE, 2016. p. 93–98.

Ministério do Planejamento Orçamento e Gestão. Manual para a Elaboração de Plano de

- Dados Abertos. Disponível em: http://www.planejamento.gov.br/secretarias/upload/arquivo/governo-aberto/copy_of_manual_elaboracao_plano_dados_abertos.pdf/@@download/file/manual_elaboracao_plano_dados_abertos.pdf>. Acesso em: 22 nov. 2017.
- NATIONS, U. Guidelines on Open Government Data for Citizen Engagement. 1. ed. New York: United Nations, 2013.
- OPEN KNOWLEDGE INTERNATIONAL. The Open Data Handbook. Disponível em: <http://opendatahandbook.org/guide/en/>>. Acesso em: 18 abr. 2017.
- PETERSEN, Kai; VAKKALANKA, Sairam; KUZNIARZ, Ludwik. Guidelines for conducting systematic mapping studies in software engineering: An Update. [s. L.], p. 1-18. ago. 2015
- RUNESON, P.; HöST, M. Guidelines for conducting and reporting case study research in software engineering. Empirical Software Engineering, v. 14, n. 2, p. 131–164, 2009.
- SEGUNDO, J. E. S. Tecnologias de informação e comunicação para disponibilização de dados abertos em formato semântico. Ibersid, v. 7, n. 1, p. 33–40, 2013.
- SOMMERVILLE, I. Engenharia de Software. 9. ed. São Paulo: Pearson, 2011.
- STI. Cartilha Técnica para Publicação de Dados Abertos no Brasil. 2011. Disponível em: <http://dados.gov.br/cartilha-publicacao-dados-abertos/>>. Acesso em 22 Nov., 2016.
- UFRN. Plano de Dados Abertos. 2016. Disponível em: <http://www.ufrn.br/resources/documentos/planodedadosabertos/Plano-PDA-7out2016.pdf>>. Acesso em 24 Jan., 2018.
- VILLAZÓN-TERRAZAS, B. et al. Methodological guidelines for publishing government linked data. In: Wood D. (eds) Linking Government Data. [S.l.]: Springer, 2011. p. 27–50.
- W3C. Data on the Web Best Practices. ago. 2016. Disponível em: <https://www.w3.org/TR/dwbp/>>. Acesso em Novembro 23, 2016.
- YIN, R. K. Case Study Research: Design and Methods. 1. ed. S. l.: SAGE Publications, 2003.
- ZUIDERWIJK, A. et al. Design principles for improving the process of publishing open data. In: Transforming Government: People, Process and Policy. [S.l.]: Emerald Group Publishing Ltd, 2014. p. 185–204.
- ZUIDERWIJK, A. et al. Issues and guiding principles for opening governmental judicial research data. In: Proceedings of the 11th IFIP WG 8.5 International Conference on Electronic Government. [S.l.]: Springer Verlag, 2012. p. 90–101.