

Abertura e Disponibilização de Dados Abertos Governamentais: Estudos de Caso

Andrés M. R. Martano¹, Gisele S. Craveiro¹

¹ Escola de Artes, Ciências e Humanidades - Universidade de São Paulo (USP)
Av. Arlindo Bértio, 1000 – 03828-000 – São Paulo – SP – Brazil

{andres.martano, giselesc}@usp.br

Abstract. *The opening of government data is still an underexplored process. This study aims to monitor and participate in one of these processes throughout their development, documenting it in detail and helping it to be built collaboratively with diverse groups of society. Finally, there will be an evaluation of results and possible impacts of the whole process. We intend to achieve this goal through interviews, public events and the development of an open source tool for data publishing according to the requirements raised with stakeholders.*

Resumo. *A abertura de dados governamentais é um processo ainda pouco explorado. Este estudo se propõe a acompanhar e participar de um desses processos ao longo de seu desenvolvimento, documentando-o em detalhe, auxiliando para que seja construído de forma colaborativa com os diversos grupos da sociedade interessados e, por fim, avaliando os resultados e possíveis impactos da publicação dos dados. Pretende-se atingir tal fim através de entrevistas, eventos abertos à população e o desenvolvimento de uma aplicação aberta para disponibilização dos dados segundo os requisitos levantados junto com os grupos interessados.*

1. Introdução

A abertura de dados governamentais é um tema que, de alguma forma, tem alcançado governos pelo mundo todo nos últimos anos. Porém, segundo [Denis and Goëta 2014], o esforço dispendido dentro da máquina pública para disponibilizá-los é ainda uma face pouco pesquisada da mesma questão. Entrevistando gestores públicos na França eles defendem que, antes dos dados poderem ser disponibilizados, é necessário que haja uma série de processos internos no órgão público para: descobrir quais bases ele possui (exploração), extrair essas bases dos sistemas internos (extração) e, por fim, tratá-las de forma a se tornarem genéricas o bastante para serem usáveis por pessoas externas ao órgão (fase a qual deram o nome de ‘rawification’).

O artigo deles, entretanto, não chega a abordar o que poderíamos chamar de um quarto passo: a disponibilização dos dados em si. Sendo que a forma pela qual os dados são disponibilizados é fruto de uma série de escolhas não triviais e, muitas vezes, políticas. Membros de movimentos sociais, gestores públicos, jornalistas, acadêmicos e *hackers*¹ são usuários frequentes de dados governamentais, apesar disso, eles costumam

¹Neste contexto usa-se o termo hacker para designar um indivíduo que tenha disposição e o domínio da computação necessário para desenvolver aplicativos ou visualizações que usem dados abertos governamentais.

ter necessidades distintas quanto à disponibilização dos mesmos. Uma cópia de uma base extraída em *SQL* pode ser perfeita para *hackers*, mas totalmente inútil para o cidadão comum. O contrário pode ocorrer com a disponibilização de dados em formato PDF.

No trabalho de [Lislaine Krupek Braz de Oliveira 2013] pode ser encontrado um exemplo de tal problema. Entrevistas com gestores, responsáveis por um portal de transparência, e cidadãos, potenciais usuários do mesmo portal, mostraram que apesar dos gestores considerarem o portal acessível, cerca da metade dos cidadãos entrevistados possuíam opinião contrária.

De maneira semelhante, [Davies and Frank 2013] encontram problemas ao tentar reutilizar dados disponibilizados pelo governo como abertos, e se utilizaram da situação para explorar o processo que levou a aquela disponibilização, incompleta ao uso que queriam dar.

Um outro tipo de desafio enfrentado pela abertura de dados não diz respeito aos dados em si, mas à vontade política de abri-los. Entrevistando gestores, [Alfano 2011] discute sobre os desafios organizacionais internos aos órgãos públicos, quando os esforços de abertura se chocam com a cultura burocrática vigente.

Por outro lado, fazendo uma análise crítica, [Janssen et al. 2012] atenta para uma série de mitos sobre a abertura de dados. Destaca que a disponibilização de dados geralmente não é interessante como um fim em si, mas sim como um meio de se produzir algum impacto social positivo. Fato que dificilmente irá ocorrer se a sociedade não conseguir ou não possuir interesse em reutilizar os dados disponibilizados.

Tendo isso em vista, esse projeto pretende, em conjunto com gestores do executivo municipal de São Paulo, desenvolver e analisar, ao longo do processo de abertura, uma metodologia para realização do mesmo que permita uma maior participação da sociedade. Tal escolha se baseia na hipótese de que, se a sociedade estiver mais envolvida no processo, maiores são as chances de que suas necessidades de disponibilização dos dados sejam atendidas, de que esses dados sejam reutilizados de alguma forma, e de que algum impacto positivo seja produzido.

A participação no processo de abertura também permitirá um acompanhamento mais detalhado do mesmo, documentando-se as atividades realizadas e entrevistando-se os diversos grupos afetados por ele.

2. Objetivos da Pesquisa

Analisando a bibliografia levantada, observou-se que os estudos foram realizados depois que a disponibilização dos dados já havia ocorrido. Observou-se também que nenhum trabalho detalhou os tipos de desafios técnicos atrelados à disponibilização dos diferentes tipos de bases de dados. Este último fato talvez relacionado ao primeiro já que, quando o estudo é feito a posteriori, informações mais detalhadas do processo podem ter se perdido.

Além disso, a maioria dos estudos se focou na opinião dos gestores públicos, e pouco nos outros grupos interessados nos dados. Por fim, na maioria dos processos de abertura de dados públicos, a sociedade é apenas notificada ao fim de que os dados estão disponíveis em um site, em um determinado formato.

Esse projeto, por sua vez, busca inovar acompanhando e participando de um pro-

cesso de abertura ainda em desenvolvimento, atentando e incluindo os diversos grupos interessados.

O acompanhamento terá como objetivo identificar: desafios internos ao órgão público encontrados no processo de abertura, grupos e interesses afetados pela abertura dos dados e possíveis impactos da abertura desses dados. E será feito através de entrevistas aos diversos atores envolvidos no processo e documentação de todas as atividades necessárias para a conclusão do mesmo, desde eventos públicos a transformações pelas quais os dados precisarem ser submetidos antes de disponibilizados.

A participação, por sua vez, permitirá: construir um processo de abertura de dados governamentais colaborativo, em que os diversos setores da sociedade possam definir requisitos para a disponibilização dos dados. E por fim, será desenvolvida, de forma interativa, uma ferramenta de código aberto que permita a disponibilização dos dados segundo os requisitos levantados. Ao final, pretende-se avaliar, novamente através de entrevistas, os resultados do processo sob a ótica dos diversos grupos envolvidos e afetados.

3. Contribuições Esperadas

Espera-se que a participação nos processos internos da abertura, ainda enquanto ela se desenvolve, permitirá produzir uma documentação mais detalhada do mesmo. Isso possibilitaria uma melhor análise tanto por esse estudo, como por terceiros.

A documentação consistirá de uma descrição das atividades técnicas necessárias para disponibilizar os dados, dos grupos envolvidos no processo, os interesses dos mesmos, os requisitos levantados pelos grupos interessados, formas de contato com esses grupos, contribuições feitas por eles e os demais pontos que se julgarem relevantes de serem documentados.

Pretende-se também contribuir com a experiência e metodologia de um processo de abertura de dados mais aberto do que o convencional, em que funcionários públicos, acadêmicos e sociedade civil possa participar da definição de quais e como serão disponibilizados os dados.

4. Resultados já Alcançados

Uma parceria foi firmada com a Controladoria Geral do Município (CGM) de São Paulo e o processo inicial de acompanhamento da abertura das bases está sendo feito através de reuniões regulares com funcionários internos desse órgão.

Está sendo definido também um calendário de eventos públicos onde a sociedade poderá debater e manifestar seus anseios quanto a disponibilização das bases de dados. Além disso, a própria ferramenta que fará a disponibilização dos dados deverá possuir um canal para coleta de comentários dos usuários. A partir desses eventos e retornos, serão definidos os outros grupos a serem entrevistados.

A CGM fez um levantamento interno de quais bases ela possuía para serem disponibilizadas de forma rápida. Foram constatadas as seguintes bases: arquivos contendo todos os textos e imagens publicados no Diário Oficial Municipal desde 2003, arquivos em papel sobre todos os imóveis pertencentes à prefeitura e dados sobre os imóveis atualmente cedidos a terceiros. Além disso, a prefeitura pretende desenvolver um sistema

para monitoramento de convênios firmados com outras entidades. Decidiu-se então desenvolvê-lo de forma que os dados fiquem abertos, configurando mais uma base.

Pode-se verificar que as bases possuem características bastante distintas, permitindo que sejam estudados diversos dos desafios enfrentados na disponibilização de dados públicos. A base do Diário Oficial Municipal configura o primeiro estudo de caso, seu grande tamanho torna não trivial tanto o download integral da base, como a busca dentro dela. O segundo estudo de caso, base com todos os imóveis, está em papel, necessitando um esforço considerável para a digitalização. Já no terceiro estudo de caso, da base apenas com os imóveis cedidos, apesar de estar em formato digital, ele não é processável facilmente por máquina, requerendo uma conversão manual de formato. Por fim, no quarto estudo de caso, relacionado aos convênios, a base de dados ainda não existe, necessitando ainda uma norma para definir como os dados serão coletados.

Para se discutir melhor cada um desses estudos de caso, será utilizada uma estrutura de análise sugerida por [Tim Davies 2013] que consiste em utilizar as seguintes perspectivas:

1. Contexto: conjuntura política, organizacional, legal, técnica, social e econômica relacionada a dados abertos.
2. Suprimento de Dados Abertos: determinações legais, licenças e atores envolvidos na coleta de dados.
3. Padrões e Técnicas: formatos, *APIs* e catálogos utilizados na disponibilização dos dados.
4. Contexto Específico: questões de governança específicas ao estudo de caso.
5. Intermediários: como e por quem os dados foram disponibilizados.
6. Ações e Impactos: experiências de reuso dos dados e suas consequências.

Os estudos de caso serão agora detalhados, de acordo com as informações preliminares levantadas, aplicando-se a estrutura de análise descrita acima.

4.1. Primeiro Estudo de Caso

O Diário Oficial publica diariamente os atos oficiais no âmbito do governo municipal, tais como leis, decretos, portarias, editais de licitação, nomeações e exonerações dentre outros; e a Secretaria Municipal de Planejamento, Orçamento e Gestão (SEMPLA) é a responsável por ele ². A SEMPLA, por sua vez, contrata a Imprensa Oficial (a mesma empresa responsável pela publicação do Diário Estadual) para realizar a publicação do Diário Municipal.

A Imprensa Oficial publica os diários em formato digital na Internet, porém, não permite o download integral da base. Além disso, o formulário de busca do site não realiza busca por categoria, e as páginas do Diário são disponibilizadas no site apenas em formato PDF.

No Diário Oficial constam informações que afetam o funcionamento da prefeitura, de empresas e os cidadãos em geral, logo existem diversos grupos interessados em possuir essa base completa ou simplesmente em poderem ter mais liberdade de busca nela.

Tendo isso em vista, a CGM requisitou à SEMPLA os arquivos relativos à essa base, e obteve todos os artigos publicados, desde 2003, e em formato TXT. Além disso,

²<http://www.docidadesp.imprensaoficial.com.br>

cada artigo publicado está em um arquivo e possui, no nome, alguns códigos tabelados que permitirão classificações.

Planeja-se agora disponibilizar essa base de forma a melhor atender às necessidades do grupos consumidores desses dados. Uma das formas previstas é possibilitando o download integral da mesma, para aqueles que desejarem executar seus próprios algoritmos ou análises sobre ela. Em segundo permitindo busca textual online em toda a base, com filtros como categoria e data, para aqueles usuários que não têm interesse na base como um todo, mas apenas em um artigo ou outro. E, por fim, disponibilizar uma versão em RDF da base, permitindo que os dados sejam ligados a outros e que análises mais sofisticadas sejam realizadas.

4.2. Segundo e Terceiro Estudos de Caso

A prefeitura possui um arquivo em papel com informações sobre os imóveis de sua propriedade. Tal base poderia ser usada, por exemplo, para definição de imóveis a serem usados como creches, moradia para a população, postos de saúde, entre outros. Porém, estando ela em papel, fica difícil sua consulta tanto pela população como pela própria prefeitura. Portanto, pretende-se digitalizar pelo menos parte dessa base.

A linguagem usada nos documentos é específica da área, requerendo um profissional especializado para obter as informações desejadas. E, antes de iniciar a digitalização, será necessário elencar quais informações se deseja obter dos imóveis. Para tal, deverá se consultar a comunidade em um evento público, onde os cidadãos possam indicar quais informações consideram relevantes.

Enquanto a digitalização da base completa de imóveis não ocorre, pretende-se disponibilizar um subconjunto da mesma. Este possui apenas os imóveis de propriedade da prefeitura que estejam atualmente cedidos a terceiros. Esta base, ao contrário da anterior, já encontra-se em formato digital, em um conjunto de arquivos DOC. Como os dados não estão organizados de forma que possam ser facilmente extraídos automaticamente, eles estão sendo extraídos manualmente e colocados em uma planilha.

Esse subconjunto dos dados possui uma característica específica: a decisão de para quem ceder um imóvel e sob quais condições, é muitas vezes vista como arbitrária. A disponibilização de tal base possibilitaria submeter tais dados ao escrutínio público.

Com o intuito aumentar as possibilidades posteriores de uso desses dados, pretende-se realizar a geolocalização dos mesmos, acrescentando-se a longitude e latitude de cada imóvel à base. Como a base é muito numerosa, pretende-se realizar um processo automático e corrigir possíveis falhas manualmente.

4.3. Quarto Estudo de Caso

A prefeitura constantemente firma convênios com outras entidades, sejam elas empresas ou Organizações Não Governamentais. Esses convênios podem visar as mais diversas finalidades, como a manutenção de uma creche ou o atendimento à população em situação de rua, e geralmente envolvem repasse de dinheiro público. Porém, não existe um sistema interno na prefeitura que unifique o controle desses convênios e, além disso, não há uma norma que padronize como esse controle deve ser feito nas diversas secretarias municipais.

Visando aumentar o controle sobre esse tipo de convênio, a gestão atual da prefeitura está debatendo a criação de um decreto que normatize e centralize essa prestação de contas. Por outro lado, há uma preocupação por parte das entidades conveniadas de que o controle seja muito rígido, engessando-as e impossibilitando a prestação de serviço.

Após esse processo de definição do decreto, será implementado um sistema para coletar os dados e viabilizar o controle, tanto por parte da prefeitura como pela população. A ideia nesse caso é fazer com que o próprio sistema que coletará os dados primários seja responsável pela transparência dos mesmos, disponibilizando-os na web. Estando a preocupação com a transparência inclusa no próprio sistema usado internamente pela prefeitura, acredita-se que os dados finais disponibilizados serão mais fidedignos e atuais.

5. Conclusão

Apesar de muitos governos estarem abrindo seus dados, as atividades necessárias para tal e o impacto produzido ainda são ainda pouco conhecidos. Além disso, a maioria dos processos de abertura de dados são conduzidos de forma fechada, sem levar em consideração as necessidades da sociedade, o que pode levar ao não reuso dos dados disponibilizados.

Visando, possivelmente, uma abertura mais eficiente que atente para as necessidades da sociedade e possibilite o reuso dos dados, esse trabalho pretende desenvolver e documentar um processo de abertura de dados mais colaborativo. Este seria construído em conjunto com membros da academia, da sociedade civil e funcionários públicos interessados na abertura dos dados.

Atualmente, já foram definidas algumas bases para serem abertas e estão sendo realizadas reuniões com o executivo municipal para definir os próximos passos. Serão realizados então eventos públicos para debater os requisitos de disponibilização, desenvolvimento de uma ferramenta aberta que atenda aos requisitos e lançamento da bases disponibilizadas. Todo o processo será documentado por entrevistas e descrições das atividades realizadas.

Referências

- Alfano, G. (2011). Adapting bureaucracy to the Internet. The case of Venice Local Government. *Information Polity*, 16(1):5–22.
- Davies, T. and Frank, M. (2013). 'There's no such thing as raw data': exploring the socio-technical life of a government dataset. In *Proceedings of the 5th Annual ACM Web Science Conference*, pages 75–78. ACM.
- Denis, J. and Goëta, S. (2014). Exploration, Extraction and 'Rawification'. The Shaping of Transparency in the Back Rooms of Open Data. In *After The Reveal. Open Questions on Closed Systems - Neil Postman Graduate Conference*, New York, USA.
- Janssen, M., Charalabidis, Y., and Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Information Systems Management*, 29(4):258–268.
- Lislaine Krupek Braz de Oliveira, C. M. (2013). Anais do WTRANS13 - Workshop de Transparência em Sistemas, Brasília, Brasil, Setembro 29, 2013.
- Tim Davies, Fernando Perini, J. M. A. (2013). *Researching the emerging impacts of open data*.