

Web de Dados - abertura de bases e web semântica

Vagner Diniz, Centro de Estudos sobre Tecnologias Web (CeWeb.br),
Vagner.diniz@gmail.com, telefone (11) 5509-3511.

Caroline Burle dos Santos Guimarães - Centro de Estudos sobre Tecnologias Web
(CeWeb.br), carolburlesg@gmail.com, telefone (11) 5509-3511.

Abril, 2015

Resumo

A Web de Dados se aplica ao contexto das discussões sobre governo eletrônico, dados abertos e web semântica. É padronizada no W3C, por meio do Data Activity "*Building the Web of Data*"¹. Lidar com dados na Web exige-se que se considere o ciclo de vida dos dados e também o gerenciamento do ciclo de dados abertos.

Dados abertos governamentais são dados produzidos pelo governo e disponibilizados em formato aberto para que sejam lidos por qualquer pessoa e por máquinas, permitindo o cruzamento com outros dados de diferentes fontes.

Em 2009, Tim Berners-Lee definiu o termo "Linked Data" para explicar a Web Semântica: usar a Web para conectar dados que não estavam previamente ligados. Para se considerar "Linked Data", Tim criou um modelo progressivo de maturidade baseado em cinco estrelas: 1. Os dados devem estar disponíveis na Web; 2. Além disso, devem estar estruturados; 3. Disponíveis, estruturados e também em formato não-proprietário; 4. Adiciona-se que os dados devem usar padrões estabelecidos pelo W3C e possuir URL própria; 5. Todas as regras anteriores devem ser seguidas e também vincular os dados a outros dados de outras fontes, para estabelecer um contexto.

Alcançar cinco estrelas não é trivial e exige esforço técnico e político. Nessa perspectiva, o Centro de Estudos sobre Tecnologias Web (CeWeb.br) coordena o desenvolvimento dos Guias de Dados Abertos e de Web Semântica, contribuindo com a evolução das formas de disponibilização dos dados governamentais.

¹ Site do W3C Data Activity: <<http://www.w3.org/2013/data/>>.

Introdução

Desde que Tim Berners-Lee criou a Web, em 1989, a partir de tecnologias básicas e simples (HTTP², URL³ e HTML⁴), um número incontável de aplicações e plataformas desenvolvidas no topo dessas tecnologias modificaram o modo de uso e o número de usuários da Internet em todo o mundo. Aplicações como Skype, Facebook, LinkedIn, Instagram dependem da Web parcial ou totalmente para funcionar. Até mesmo as aplicações de troca de mensagens, como Whatsapp, Telegram e e-mails conhecidos como webmails já estão também usando a Web como plataforma para conectar pessoas.

Os indicadores mais recentes apontam que mais de um terço da população mundial já tem acesso à Internet, de acordo com a pesquisa "TIC domicílios e usuários 2013", do CETIC (2013). Esses números indicam o avanço exponencial do número de pessoas conectadas à rede mundial de computadores. Esse avanço só foi possível por causa de dois vetores: a invenção da Web há 26 anos – uma apresentação ao mundo de uma “aplicação matadora”⁵ – que possibilitou o consumo de conteúdos digitais independentemente do local onde esteja armazenado; e a popularização do uso dos telefones celulares “inteligentes” com acesso à Internet.

De fato, a invenção da Web criou um mundo de documentos ou recursos digitais conectados, pois somente com a URI ou endereço único de um documento digital (por exemplo, um endereço tipo www.documento.com.br) é possível ter acesso ao seu conteúdo na Web. Por isso a Web atual ainda é chamada de Web de documentos. Somente com o desenvolvimento das tecnologias Web e da infraestrutura da Internet foi possível também referenciar fotos, vídeos, códigos e outras aplicações, também conhecidos pelo nome genérico de recursos digitais.

² HTTP é a sigla para Hyper Text Transmission Protocol, nome dado ao protocolo que permite a transmissão de textos entre diferentes computadores conectados à Internet e que disponibilizam e consomem conteúdos.

³ URL é a sigla para Uniform Resource Locator que é a maneira de oferecer um endereço para localizar um determinado recurso de maneira unívoca. Atualmente, é mais comum encontrar a sigla URI para Uniform Resource Identifier.

⁴ HTML é a sigla para Hyper Text Markup Language, nome dado à linguagem de marcação que instrui os computadores, por meio de navegadores, como e o que será exibido nas telas quando se acessa um recurso na Web, por exemplo, documentos, imagens e vídeos.

⁵ Tradução para a expressão em inglês “killer application” que indica uma solução tecnológica que eleva para um patamar superior o desenvolvimento de uma determinada tecnologia,

A possibilidade de referenciar tantos recursos digitais se realiza em diversas aplicações como blogs, sites, jogos, comércio eletrônico, internet banking, pagamentos, governo eletrônico e redes sociais, que fazem a alegria de quem navega pela Web. Acontece que não são simples recursos digitais que trafegam pela Internet, mas conteúdos derivados e acompanhados de dados com significados. São dados que podem ser visualizados como preço e descrição de uma mercadoria, origem e destino de uma viagem, formulários com endereço, CPF e outros dados pessoais. São também dados não perceptíveis numa página Web como o dia, hora e quais sites foram acessados em um navegador, ou tipo de produto mais pesquisado por uma pessoa, metadados sobre um recurso digital, como data, hora, tipo de câmera que registrou uma foto e texto explicativo da foto.

Essa quantidade gigantesca de dados coletados a cada segundo sobre cada transação na Web, armazenadas em diferentes repositórios de grandes a pequenos atores do ecossistema da Internet, é um acervo precioso pela possibilidade técnica de conectá-los entre si para oferecer uma experiência de navegação muito mais rica.

A conexão entre dados permite aplicações como, ao escrever um texto em um blog sobre um destino turístico para um determinado período do ano, poder oferecer automaticamente no próprio texto as ofertas de diferentes sites sobre os melhores voos, melhores hotéis e melhor custo de aluguel de carro. Ou, ao buscar por um indicador social do País, receber automaticamente também uma lista do mesmo indicador para outros países para efeito de comparação.

A necessidade de conectar dados vai crescer com a emergência da Internet das Coisas e Web das Coisas.⁶ Os dispositivos conectados, como sensores, atuadores, carros, Smart TV, aparelhos domésticos, câmeras fotográficas, coletarão e transmitirão dados. A combinação desses dados tem potencial para elevar a Internet a mais um novo patamar tecnológico. A Web de documentos torna-se, portanto, a Web de dados.

⁶ Internet das Coisas é o nome dado à possibilidade de conectar à Internet diferentes dispositivos, além dos próprios computadores de mesa e notebooks. Hoje, já é bastante comum conectar sensores e atuadores instalados em casa, carros, iluminação pública. Assim, esses dispositivos conectados recebem um endereço na Internet e coletam e transmitem dados. Web das Coisas é o nome dado para a possibilidade utilizar interfaces Web para visualização e interação humana com os dispositivos.

Objetivos

O objetivo geral desse trabalho é apresentar a importância da construção de um ambiente favorável na administração pública para a abertura de dados e exposição destes com características semânticas, no contexto do crescimento exponencial da coleta e transmissão de dados na Web, a partir do relato de um caso real.

São objetivos específicos:

- Mostrar que o processo de abertura de dados, com o objetivo de permitir conexões entre eles, não é trivial e exige esforço técnico e político dos atores envolvidos.
- Relatar como está sendo feito o esforço de abertura de dados pelo Governo do Estado de São Paulo nos anos de 2014 e 2015, tendo como referência a experiência do Governo Britânico.
- Apresentar a construção de Guias de Dados Abertos e de Web Semântica, como um ponto de partida necessário para o processo e como contribuição com a evolução das formas de disponibilização dos dados governamentais.

Metodologia

Esse artigo é um relato de uma experiência viva em um projeto de abertura de dados ainda não concluído. Os autores são também atores do ecossistema de dados abertos que vem sendo construído no Governo do Estado de São Paulo. A participação ativa e constante em todas as etapas do projeto, desde a sua concepção e planejamento até a sua execução tem permitido aos autores uma visão privilegiada de observação.

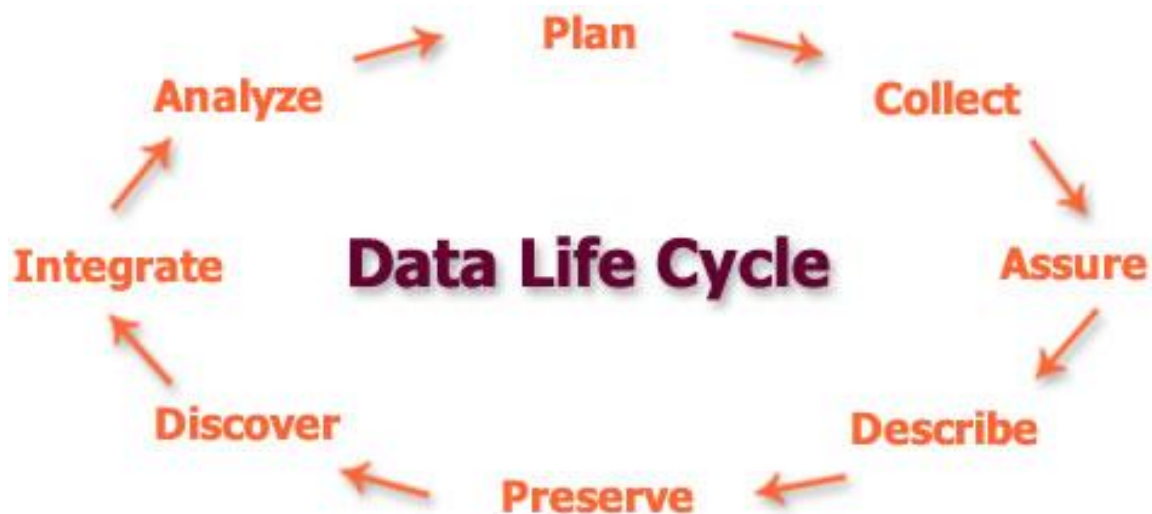
Os autores dos Guias de Dados Abertos e Web Semântica tiveram a possibilidade conversar com gestores do projeto, especialistas britânicos que são referência no tema de dados abertos e órgãos públicos que estão engajados nesse projeto para publicar algumas de suas bases de dados em formato aberto. Os autores também administram a aplicação de recursos financeiros e profissionais necessários ao projeto.

O relato aqui feito buscou informações e referências na documentação gerada pelo projeto com metodologia ágil, na qual a documentação é criada em momentos diferentes. No modelo cascata de desenvolvimento de software a documentação é criada antes do desenvolvimento do projeto (SOMMERVILLE, 2007). Já no modelo ágil, a documentação é criada e desenvolvida ao longo das etapas do projeto, com pouca chance de ficar desatualizada. O projeto em foco, além de documentar cada etapa, cada encontro, cada “sprint”, tem uma plataforma digital para documentação, colaborativa com acesso por todos os participantes.

Processo de abertura de dados

Disponibilizar dados em formato aberto exige elaborar a publicação dos dados, que podem estar previamente estruturados ou não. Há um ciclo de vida dos dados que deve ser considerado. De acordo com as boas práticas do (DATA, 2015) há oito passos a serem seguidos nesse ciclo: planejamento, coleta, segurança, descrição, preservação, descoberta, integração e análise, ilustrados na figura abaixo.

Figura 1 – Ciclo de vida dos dados



Fonte: DATA, 2015.

A figura acima mostra que, idealmente, o ciclo tem início com o planejamento dos dados que serão disponibilizados. É preciso ter uma política de backup e um

sistema de armazenamento de dados. Também deve-se definir quais tipos de dados serão produzidos e em quais formatos e identificar os dados sensíveis, de modo que não exista o risco de ferir a privacidade. Ainda na fase de planejamento, deve-se considerar que o plano de gerenciamento de dados pode ser aprimorado durante o ciclo de vida dos dados.

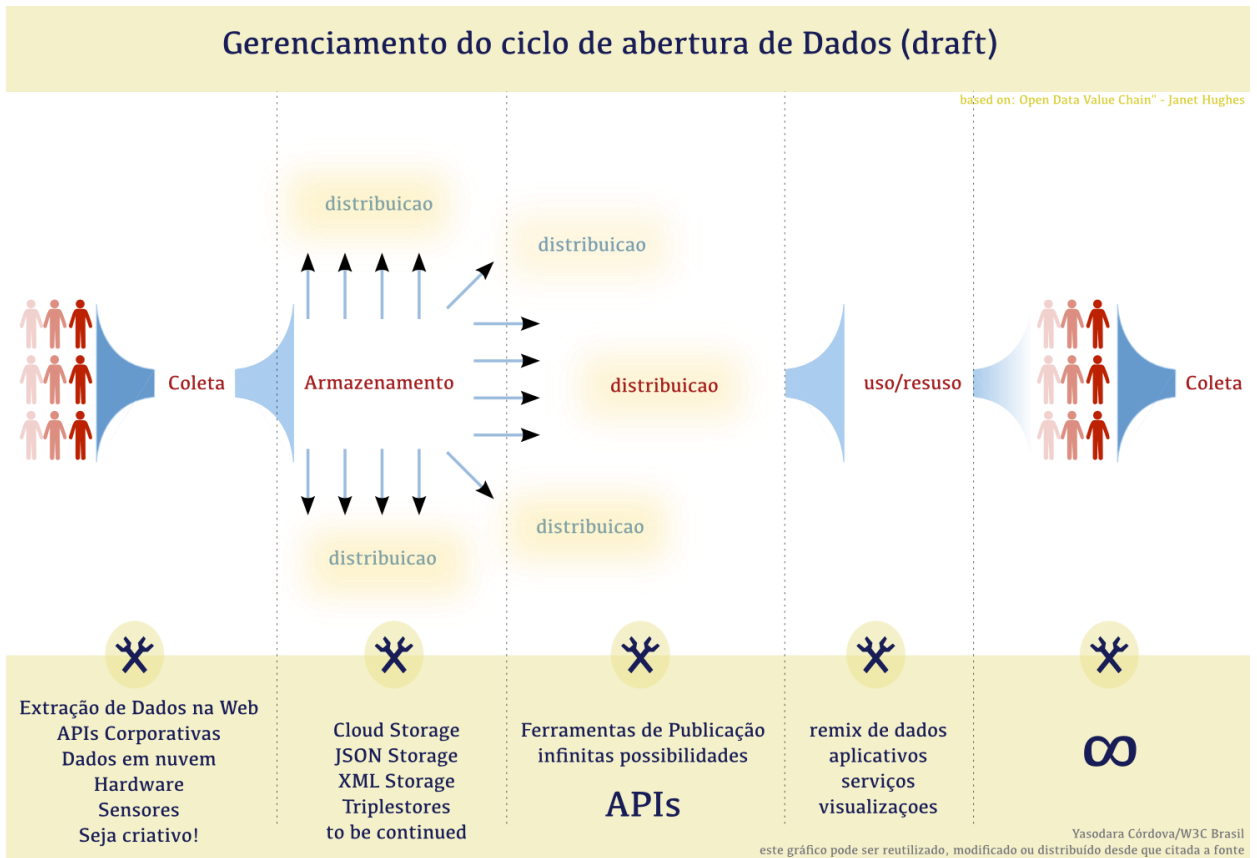
Após cumpridas as etapas de planejamento, esses dados devem ser coletados de maneira segura, com a devida descrição do seu significado. É importante que os dados sejam consistentes e que tenham um controle de qualidade. O conjunto de dados (*datasets*) precisam ser integrados, compatíveis e comparáveis.

Em seguida é preciso pensar como esses dados serão preservados, de modo a manter sua integridade. Deve-se ainda cuidar para que os dados sejam encontrados facilmente, sendo que os nomes dos arquivos precisam refletir o seu conteúdo. Esses dados devem estar acessíveis na Web, considerando os padrões de Acessibilidade na Web (W3C, 2015).

Além disso, os dados devem ter a possibilidade de serem integrados, para isso a identificação e o uso de padrões de metadados são importantes de modo a garantir que os dados sejam utilizados e possam ser reproduzidos. Finalmente, para analisar esses dados, é necessário descrever as etapas do processo de derivação de um dado e o método utilizado.

A figura abaixo, mostra de maneira simplificada como seria o ciclo de gerenciamento de abertura de dados. Verifica-se que há uma semelhança entre os dois ciclos, porém, na abertura de dados existe uma preocupação com as ferramentas e os padrões a serem utilizados.

Figura 2- Gerenciamento do ciclo de abertura de dados



Fonte: CORDOVA, 2013.

De acordo com o gráfico, o ciclo que um determinado dado percorre pode começar a partir da fase de COLETA. Em tempos de uma sociedade informatizada e com dispositivos de hardware à disposição por preços relativamente acessíveis, é hora de aproveitar a abundância de possibilidades de fontes de informação para ser criativo na hora de acumular dados que possam servir para beneficiar as pessoas em um futuro próximo. (CORDOVA, 2013).

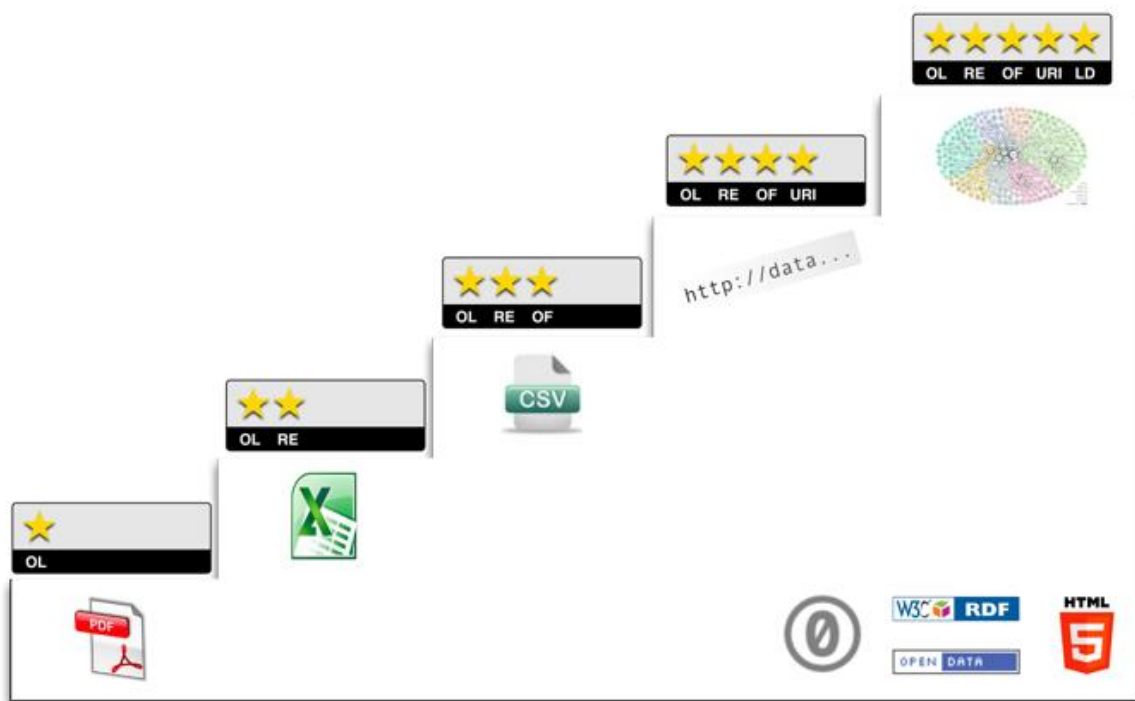
O processo de abertura de dados envolve considerar o ciclo de vida dos dados, mas também o gerenciamento do ciclo de abertura de dados. Este último mostra a importância de pensar em como os dados podem ser disponibilizados na Web em formato aberto. Enfatiza-se que dados abertos podem ser lidos por qualquer pessoa e por máquinas, permitindo o cruzamento com outros dados de diferentes fontes, para serem livremente reutilizados pela sociedade. Nesse sentido, dados abertos

governamentais são dados produzidos pelos governos, que devem ser colocados à disposição de qualquer cidadão e para qualquer fim. (W3C BRASIL, 2011, p.4). As três regras dos dados abertos governamentais foram estabelecidas por Eaves (2009):

1. Se o dado não pode ser encontrado e indexado na web, ele não existe.
2. Se não estiver aberto e disponível em formato compreensível por máquina, ele não pode ser reaproveitado.
3. Se algum dispositivo legal não permitir sua replicação, ele não é útil.

É preciso, logo, considerar os formatos que os dados são disponibilizados. Um caminho é seguir o modelo progressivo de maturidade baseado em cinco estrelas, criado por Tim Berners-Lee (2009) para definir o termo "Linked Data" e explicar a Web Semântica, de maneira a usar a Web para conectar dados que não estavam previamente ligados. Conforme a figura abaixo:

Figura 3 – As cinco estrelas dos dados abertos



Fonte: Berners-Lee, Tim. Linked Data. 2009.

Essa figura mostra a classificação para atingir as cinco estrelas dos dados abertos e disponibilizá-los em “Linked Data”: 1. Os dados devem estar disponíveis na Web; 2. Além disso, devem estar estruturados; 3. Disponíveis, estruturados e também em formato não-proprietário; 4. Adiciona-se que os dados devem usar padrões estabelecidos pelo W3C e possuir URL própria; 5. Todas as regras anteriores devem ser seguidas e também vincular os dados a outros dados de outras fontes, para estabelecer um contexto.

Entende-se, portanto, que é necessário verificar e fazer uso tanto do ciclo de vida dos dados como do ciclo de abertura de dados para disponibilizar dados em formato aberto. O processo de abertura de dados é tido como base para o desenvolvimento do projeto SPUK – Melhoria do ambiente de negócios por meio da Transparência no Estado de São Paulo, conforme verificar-se-á em seguida.

Projeto SPUK

O projeto SPUK – Melhoria do ambiente de negócios por meio da Transparência no Estado de São Paulo⁷ foi concebido pelo Governo do Estado de São Paulo, em parceria com o Governo do Reino Unido, em meados de 2013. O Governo de São Paulo realizou, em seguida, uma parceria com o Centro de Estudos sobre Tecnologias Web (Ceweb.br), por meio do Comitê Gestor da Internet (CGI.br) e do o Núcleo de Informação e Coordenação do Ponto BR (NIC.br), com o objetivo melhorar o ambiente de negócios no Estado de São Paulo, por meio da disponibilização e reuso de dados abertos governamentais.

A governança do projeto ocorre por meio de um Comitê Gestor, composto por entidades⁸ do Governo do Estado de São Paulo, pela Embaixada e Consulado Britânicos e pelo Centro de Estudos sobre Tecnologias Web (Ceweb.br). Esse comitê reúne-se⁹ a cada dois ou três meses, com o intuito de fazer um balanço das atividades

⁷ O site do projeto SPUK está disponível em: <<http://igovsp.net/spuk/>>.

⁸ Fazem parte do Comitê Gestor a da Unidade de Inovação da Subsecretaria de Parcerias e Inovação da Secretaria de Governo do Estado de São Paulo, a Assessoria Especial para Assuntos Internacionais, a FUNDAP, a Fundação Sead e a Controladoria-Geral da Administração.

⁹ As atas das reuniões estão disponíveis em: <<http://igovsp.net/spuk/reunioes/>>.

que estão sendo ou já foram realizadas e daquelas que ainda ocorrerão, dentro do escopo do projeto.

Os principais objetivos do projeto são aumentar em 70% o número de bases de dados em formato aberto, plenamente acessíveis e disponibilizadas em linguagem compreensível para qualquer cidadão e ter um mínimo de 3% das bases abertas com dados estruturados em conformidade com os preceitos da Web Semântica. Para cumpri-los, o Ceweb.br coordena o desenvolvimento e publicação de guias que darão as diretrizes para: Abertura de Dados, Web Semântica, Participação Cidadã e Fomento a Negócios Sociais¹⁰.

Mapeamento da experiência britânica

O consultor Marco Túlio Pires, visitou sete instituições britânicas, entre os dias 13 e 17 de outubro de 2014, com o objetivo de mapear a experiência britânica relativa à abertura de dados e implementação da web semântica. Túlio reuniu-se com representantes do governo britânico, de empresas e da sociedade civil para compreender a percepção desses atores sobre o ecossistema de dados abertos. O relatório desse mapeamento, anexo a este artigo, resume as entrevistas de modo a produzir sugestões que possam ser aproveitadas pelo Governo do Estado de São Paulo.

Embora Pires (2014) conclua que “não há qualquer modelo no governo britânico que deva ser seguido à risca”, o relatório ilumina caminhos e mostra possíveis cenários a serem considerados ao pensar na abertura de dados e na implementação da web semântica. Enfatiza-se a ênfase dada por quase todos os entrevistados à importância do Primeiro Ministro Britânico ter encabeçado a abertura de dados no país, de modo que a política de dados abertos fosse adotada por todos os gestores públicos.

Outro tópico evidenciado pelos entrevistados é de que a centralização da política de transparência, que inclui abertura de dados e dados linkados, deve ser centralizada e com a anuência de um órgão que confira autoridade e autonomia para que essa

¹⁰ Informações disponíveis em: <<http://ceweb.br/projetos/acordos-cooperacao/>>.

política seja articulada em todas as secretarias e órgãos. Isso não quer dizer que apenas um grupo de pessoas ficará responsável pela quantidade homérica de trabalho necessário para abrir as bases de dados do governo, mas que essa equipe terá um apoio "de cima" e de parceiros por todo o governo. (PIRES, 2014)

Os entrevistados também citaram a importância de formar parcerias com organizações interessadas em trabalhar com o setor público nessa empreitada. As parcerias podem gerar benefícios para o governo e também para os cidadãos, de modo a aproximar a sociedade desse tema, podendo contribuir com o uso dos dados.

Modelo ágil de desenvolvimento

O desenvolvimento dos Guias foi feito, conforme mencionado no item sobre a metodologia deste artigo, com base na metodologia ágil. Assim, os autores compreenderam a demanda apresentada pelo Governo do Estado de São Paulo para, então, iniciar a pesquisa do estado da arte com atores envolvidos com os temas de dados abertos e web semântica no Reino Unido. Realizou-se um mapeamento *in loco* com esses atores, que gerou uma vasta documentação baseada em entrevistas, cuja contribuição foi fundamental para o desenvolvimento dos Guias de Dados Abertos e Web Semântica.

A aplicação da metodologia ágil foi de grande importância para o aprimoramento do conteúdo desenvolvido. Reuniões semanais foram realizadas entre os autores dos respectivos guias e os coordenadores do projeto por parte do Governo do Estado de São Paulo e do Centro de Estudos sobre Tecnologias Web (Ceweb.br). A colaboração entre os autores e os coordenadores, com diálogo constante, facilitou que o desenvolvimento se desse no tempo previsto, com os guias finalizados no final de abril de 2015, e deverão ser publicados em junho do mesmo ano.

Para o acompanhamento das atividades, utilizou-se o método *Scrum*, por meio da ferramenta Trello¹¹. Essa ferramenta online proporcionou fácil interação entre os envolvidos relacionando as tarefas a fazer, tarefas que estavam sendo executadas e

¹¹ A ferramenta está disponível em: <<https://trello.com/>>.

aquelas que foram concluídas. Dessa maneira, os coordenadores e os autores tiveram ampla visão das atividades a serem realizadas e do status de cada tarefa.

Com uso da metodologia ágil de desenvolvimento, os Guias de Dados Abertos e Web Semântica foram desenvolvidos com o intuito de proporcionar “o caminho das pedras” aos projetos pilotos¹² que o Governo do Estado de São Paulo desenvolverá para chegar ao objetivo de ter um mínimo de 3% das bases abertas com dados estruturados em conformidade com os preceitos da Web Semântica.

Guias de Dados Abertos e Web Semântica

O guia de dados abertos foi desenvolvido com o objetivo de pautar os gestores públicos sobre o assunto, explicando a importância de publicar dados em formato aberto. Descreveu-se, portanto, o que são dados abertos, seus benefícios, decorreu-se sobre o artigo 8º da Lei de Acesso à Informação, sobre as cinco estrelas dos dados abertos, sobre os formatos de bases de dados, viabilizaram-se exemplos de uso de dados abertos e iniciou-se a conversa sobre quais licenças podem ser utilizadas.

Já o Guia de Web Semântica tem é focado para os técnicos que colocarão em prática a abertura dos dados governamentais, de modo a contemplar o “Linked Data”. Dissertou-se sobre a Web e a sua evolução desde a Web de Documentos até a Web de Dados. Considerou-se também o ecossistema de Dados na Web, o que é Web Semântica e o que são dados conectados, para compreender a importância dos vocabulários e das ontologias.

Considerações finais

A Web de Dados é uma realidade sem volta. A necessidade de conectar dados crescerá, especialmente com a chegada da Internet das Coisas e Web das Coisas. Os dispositivos estarão cada vez mais conectados e os dados serão a essência da vida dos usuários, sejam eles visíveis ou não.

¹² O Governo do Estado de São Paulo terá ao menos três projetos pilotos nas Secretarias da Saúde, do Planejamento e também no Metrô.

A importância dos dados governamentais estarem disponíveis em formato aberto foi compreendida pelos governos e assimilada no artigo 8º da Lei de Acesso à Informação. Porém, ainda há obstáculos políticos e técnicos a serem ultrapassados para que os dados sejam de fato disponibilizados conforme os preceitos da Web Semântica.

O projeto SPUK – Melhoria do ambiente de negócios por meio da Transparência no Estado de São Paulo foi concebido com o intuito de aprimorar a disponibilização dos dados abertos gerenciados pelo Governo do Estado, e publicar no mínimo 3% dos seus dados em formato estruturado com base na Web Semântica. Ao designar projetos pilotos para essa empreitada, verificou-se a necessidade de produzir Guias, de Dados Abertos e de Web Semântica, que proporcionassem o caminho a ser seguido.

Referências

BERNERS-LEE, T.. “Linked Data”. 2009. Disponível em: <<http://www.w3.org/DesignIssues/LinkedData.html>>. Acesso em 14 maio 2015.

CEWEB.br, Relatório Sintético do Workshop de 24 de março de 2015 do Projeto SPUK. No prelo.

CETIC, Br, 2013. TIC Domicílios e Usuários 2011 - Total Brasil. Disponível em: <http://www.cetic.br/media/docs/publicacoes/2/TIC_DOM_EMP_2013_livro_eletronico.pdf>. Acesso em 14 abr. 2015.

CORDOVA, Yasodara. Gerenciamento do ciclo de abertura de dados, 2013. Disponível em: <<http://platform.od4d.org/article?locale=pt&uri=http%3A%2F%2Fplatform.od4d.org%2Fposts%2F77>>. Acesso em 14 de maio de 2015.

DATA, One. Best Practices - Data Life Cycle, 2015. Disponível em <<https://www.dataone.org/best-practices>>. Acesso em 10 maio 2015.

DWBP, W3C Data on the Web Best Practices Working Group. Disponível em: <<https://www.w3.org/2013/dwbp/>>. Acesso em 14 abr. 2015.

PIRES, Marco Túlio. Guia de Abertura de Dados. São Paulo, 2015. No prelo.

_____. Mapeamento da experiência britânica. Relatório de visita ao Reino Unido, 2014. Anexo.

LAUFER, Carlos César. Guia de Web Semântica. São Paulo, 2015. No prelo.

SOMMERVILLE, I. Engenharia de software. São Paulo: Pearson, 2007.

W3C. Web Content Accessibility Guidelines (WCAG) 2.0. Tradução Autorizada em Português do Brasil. 2014. Disponível em: <<http://www.w3.org/Translations/WCAG20-pt-br/>>. Acesso em 14 de maio de 2015.

Sites:

Centro de Estudos sobre Tecnologias Web (Ceweb.br) - <http://ceweb.br/>

SPUK – Melhoria do ambiente de negócios por meio da Transparência no Estado de São Paulo - <http://igovsp.net/spuk>

Trello - <https://trello.com/>

ANEXO – Relatório do mapeamento britânico

Melhorando o ambiente de negócios no Estado de São Paulo por meio da transparência

Mapeamento da experiência britânica
Relatório de visita ao Reino Unido

Marco Túlio Pires

São Paulo
Outubro 2014

Índice

[Apresentação](#)

[Visitas](#)

[Entrevistados](#)

[Alan Smith \(Office for National Statistics - ONS\)](#)

[Antonio Acuna \(data.gov.uk\)](#)

[Mark Braggings \(HampshireHUB\)](#)

[Amanda Smith \(Open Data Institute - ODI\)](#)

[Emma Prest \(DataKind UK\)](#)

[Rufus Pollock \(Open Knowledge\)](#)

[Tracie Gyateng \(New Philanthropy Capital - NPC\)](#)

Conclusões

1. Apresentação

O projeto "Melhorando o ambiente de negócios em São Paulo por meio da transparência", liderado pela Secretaria de Planejamento do Estado de São Paulo, em parceria com o escritório do W3C no Brasil e a embaixada britânica no Brasil, tem dentro do seu escopo de objetivos a produção de *guidelines* para fortalecer o ecossistema de transparência em todo o estado paulista. Na atual fase do projeto, dois tópicos estão recebendo foco, a saber, "abertura de bases" e "dados conectados".

Sendo o Reino Unido um dos países à frente dos esforços globais em promover a transparência e o acesso à informação, a presente equipe organizou uma missão à Grã Bretanha para extrair lições úteis de diferentes atores envolvidos com o mundo dos dados públicos. O Reino Unido possui uma política de acesso à informação que vem sendo consolidada ao longo das últimas décadas. Seu portal de dados abertos possui 20 mil bases de dados disponíveis para download — em comparação, o portal de dados abertos do governo federal brasileiro possui 327 bases de dados —, dispostos em uma plataforma que oferece recursos legíveis por máquina e interface de programação de aplicações (API). A publicação das bases é feita de forma periódica por meio do engajamento de diversos setores do serviço público e do envolvimento da sociedade civil.

Além da questão meramente numérica, há um esforço em todas as esferas governamentais, do primeiro ministro às autoridades locais, para publicar dados e promover o acesso à informação. Os problemas também existem, e é essa mistura de soluções e lições aprendidas na experiência britânica que merecem atenção de qualquer governo que almeja ampliar seus esforços de transparência e acesso à informação.

Entre os dias 13 e 17 de outubro de 2014, representantes do governo britânico, de empresas e da sociedade civil foram ouvidos pelo consultor Marco Túlio Pires, para que pudessem dar sua percepção frente ao ecossistema de dados abertos e conectados do Reino Unido. Diante das informações recolhidas durante as entrevistas, que representam um pequeno, mas valioso retrato da experiência britânica com dados abertos e conectados, o presente relatório tem como objetivo compilar as entrevistas de forma a apresentar os resultados mais úteis, reunir recursos que possam ser reaproveitados pela equipe do governo do estado de São Paulo e fazer recomendações sobre a produção das *guidelines* sobre "abertura de bases" e "dados conectados".

2. Visitas

Entre os dias 13 e 17 de outubro a missão deste projeto visitou o Reino Unido para entrevistar atores do setor público, privado e da sociedade civil sobre a experiência britânica com dados abertos e linkados. Os contatos foram feitos e as entrevistas marcadas a partir de São Paulo com base na rede de relacionamentos da equipe e do W3C Brasil e na disponibilidade das fontes. Ao todo, foram sete entrevistas, representando uma gama diversificada do ambiente britânico de transparência. O objetivo foi ouvir não apenas vozes do setor público, mas também do setor privado e da sociedade civil. Todas as entrevistas foram gravadas e o áudio está à disposição para fins de consulta.

2.1. Entrevistados

Setor	Nome	Organização	Cargo	Contato
Governo	Alan Smith	ONS	Chefe de Conteúdo Digital	alan.smith@ons.gsi.gov.uk
Governo	Antonio Acuna	data.gov.uk	Coordenador	antonio.acuna@cabinetoffice.gov.uk
Governo	Mark Braggings	HampshireHUB	Coordenador	mark.braggings@hants.gov.uk
Setor Privado	Amanda Smith	ODI	Gerente de Comunidades	amanda.smith@theodi.org
Sociedade Civil	Emma Prest	DataKind UK	Coordenadora	emma@datakind.org.uk
Sociedade Civil	Rufus Pollock	Open Knowledge	Fundador	rufus.pollock@okfn.org
Sociedade Civil	Tracie Gyateng	NPC	Especialista de Dados	tracey.gyateng@thinknpc.org

2.2. Alan Smith (Office for National Statistics - ONS)

2.2.1. Sobre o ONS

O Escritório de Estatísticas Nacionais (ONS, na sigla em inglês) é o maior produtor independente de estatísticas do Reino Unido e é o instituto de estatísticas reconhecido do país. Ele funciona como o Instituto Brasileiro de Geografia e Estatística, porém, a parte de geografia na Grã Bretanha é coordenada por outro órgão. O ONS é responsável por coletar e publicar estatísticas relacionadas à economia, população e sociedade nos níveis nacionais, regionais e locais. É ele também que conduz o censo da Inglaterra e do País de Gales a cada 10 anos. O escritório tem um papel fundamental nas boas práticas nacionais e internacionais na produção de estatísticas oficiais.

2.2.2. Resumo da entrevista

Alan Smith é designer e cartógrafo, chefe de conteúdo digital no ONS. Trabalha na produção de conteúdo do ONS. Sua equipe de 15 programadores, jornalistas e designers é responsável por vasculhar o amplo espectro de bases de dados do ONS em busca de informações que possam ser de interesse público. Smith trabalha como se fosse um editor-chefe de uma redação, aprovando e publicando matérias de caráter jornalístico no site de interatividade do ONS.

O objetivo dessa equipe é fazer um tipo diferente de trabalho de comunicação, com grande engajamento de pessoas na internet e trabalho próximo de redações de jornais, como a do The Guardian e da BBC. O trabalho da equipe não começou com essa característica. Inicialmente, o trabalho era concentrado na construção de APIs que pudessem facilitar o trabalho de diferentes equipes dentro do ONS. Depois que essas ferramentas estavam funcionando, elas foram apresentadas para a administração de modo que elas pudessem ser abertas para todo o público. Depois disso, houve um interesse crescente em divulgar informações interessantes encontradas dentro do ONS e a equipe nasceu.

Em relação ao ONS propriamente dito, Smith relata que o portal de dados abertos do Reino Unido não é visto como a plataforma de longo prazo para dados estatísticos. Atualmente, o site da web do ONS é precário e estão refazendo toda a estrutura. Existem três milhões de planilhas eletrônicas no formato Excel, todas geradas a partir de áreas independentes dentro da organização. A ideia agora é construir um raspador que varra todas essas planilhas e produza arquivos CSV que serão hospedados no portal principal.

2.2.3. Recomendações

- Realizar diagnóstico na SEADE para saber se há interesse e recursos locais para a produção de conteúdo digital especializado.
- O diagnóstico também poderá investigar a viabilidade da realização de pilotos com a imprensa paulista, de modo que o órgão abra suas bases para as equipes de interatividade dos jornais e meios de comunicação para a produção de matérias jornalísticas de interesse.
- Mapear o fluxo de trabalho dentro da SEADE para encontrar possíveis gargalos e equipes mais aptas a trabalharem com desenvolvimento de APIs e abertura de bases.

2.2.4. Links importantes

- Site principal: <http://www.ons.gov.uk/ons/index.html>
- Conteúdo interativo: <http://www.ons.gov.uk/ons/interactive/index.html>
- Dados geoespaciais (com dados linkados):
<https://geoportal.statistics.gov.uk/geoportal/catalog/main/home.page>
- Estatísticas de bairros: <http://neighbourhood.statistics.gov.uk>

2.3. Antonio Acuna (data.gov.uk)

2.3.1. Sobre o data.gov.uk

O portal de dados abertos do Reino Unido é subordinado à Casa Civil. Possui uma equipe de oito pessoas, sendo a maior parte delas programadores. A plataforma utilizada é o CKAN, armazém de dados desenvolvido pela organização da sociedade civil Open Knowledge, que possui um capítulo no Brasil. O governo federal brasileiro utiliza a mesma plataforma para abrir seus dados. A subordinação do portal de dados abertos do Reino Unido à Casa Civil não é coincidência, trata-se de uma posição altamente estratégica. A equipe possui autonomia e autoridade para supervisionar os esforços de abertura de dados em todos os ministérios e secretarias.

2.3.2. Resumo da entrevista

A administração do portal de dados abertos do Reino Unido é feito por uma equipe enxuta, com 5 programadores (dois especialistas em Drupal, dois especialistas em CKAN e outra posição que é mantida aberta para trazer técnicos de diferentes habilidades), um gerente técnico de projetos, um funcionário que lida com os contatos que chegam e saem do time, e o próprio Antonio Acuna, que coordena a equipe. O governo do Reino Unido pode manter uma equipe reduzida assim, pois o país possui múltiplos mecanismos de articulação da política de abertura de dados e transparência. A equipe é reduzida, porém possuem parceiros em todos os órgãos governamentais e autoridade para cobrar e coordenar esses parceiros. Há uma equipe de transparência na Casa Civil que auxilia a equipe do data.gov.uk nos contatos com outras agências. A equipe do data.gov.uk também depende do trabalho de "Representantes de Transparência" alocados em todas as secretarias e órgãos. Essas pessoas desempenham um papel fundamental para resolver a problemática da abertura de dados nos bastidores do governo.

Não há uma fórmula pronta para acelerar o processo de abertura de bases e dados linkados. Houve uma onda intensa de trabalho em todas as esferas para que o assunto da transparência realmente ficasse arraigado no governo britânico. O entrevistado relata que há uma mistura de intimidação, utilizando a autoridade conferida pelo primeiro ministro, e uma abordagem mais envolvente, com a criação de eventos para estimular o contato de servidores detentores de dados entre múltiplas secretarias e reuniões mensais de avaliação do trabalho. Essas reuniões ocorrem de forma permanente de modo a cuidar dos servidores que são substituídos e manter o assunto sempre na agenda do órgão.

Em relação ao envolvimento com outras partes do governo, a equipe do data.gov.uk desistiu de treinar os servidores para publicar dados em formatos abertos. O caminho que eles optaram foi de pedir que todos enviem no formato mais conveniente para eles (planilha eletrônica, no mínimo) e o data.gov.uk faz a conversão dos arquivos para formatos mais abertos. O esforço é concentrado em fazer acordos em relação aos campos presentes nessas planilhas e mantê-los na forma concordada permanentemente. Assim, há garantias de que as bases serão atualizadas de forma coesa.

Na opinião do coordenador do data.gov.uk, hackdays não servem a qualquer propósito, a não ser ao de criar laços entre pessoas. O mais importante não seria o produto do hackday em si, mas que os servidores possam ter a oportunidade de participar e perceber como a abertura de dados pode facilitar a vida deles e aumentar sua satisfação profissional. Os hackdays devem ser inclusivos, um momento em que os servidores participam ativamente do evento, em vez de meros observadores. Também recomenda a contratação de um

mediador para organizar os eventos, de modo a colocar todos no mesmo nível e evitar conflitos de egos.

Para os órgãos que possuem mais dificuldade em aderir à cultura de abertura de dados, ele sugere a busca por casos onde essa prática resolva um ou dois problemas para o servidor e para a sociedade. Essa abordagem daria um senso de propósito mais tangível do que simplesmente convidar as pessoas à abertura de dados porque é algo nobre. Um dos exemplos citados foi o de uma startup que reuniu dados de receitas médicas: onde elas estavam sendo emitidas, por quem e qual remédio. Descobriram, dentre outras coisas, que o remédio mais prescrito era a variante mais cara do mercado. O segundo mais barato, e igualmente eficiente, era metade do preço. Se os médicos da rede pública estivessem receitando o segundo, a economia chegaria a 200 milhões de libras por ano.

Acuna também mencionou esforços sobre Linked Data e seus desdobramentos. As iniciativas ainda são incipientes, mas eles já possuem uma estrutura de dados linkados e uma série de recomendações para que qualquer um possa publicar, em algumas áreas. Outras, como educação, ainda estão identificando quais dados existem e quais fazem sentido publicar, sob qual nomenclatura etc. Mais informações nos links abaixo, que incluem exemplos de trabalhos nas áreas de Meio Ambiente, Fazenda, Poder Legislativo, Geolocalização e Transporte.

Entre os desafios atuais estão melhor qualidade nos metadados e melhor entendimento de como os dados se relacionam uns com os outros. Nos últimos três anos, ele tem percebido que não há uma demanda grande por serviços ou APIs, mas por dados mais consistentes, que não tenham sua estrutura alterada de tempos em tempos, mantendo suas colunas e seus valores base inalterados, com metadados contextualizados e atualizados periodicamente.

2.3.3. Recomendações

- Na oportunidade do governo trabalhar com organizações interessadas em dados abertos, exigir uma contrapartida de que elas ajudem no trabalho de divulgação e evangelização da cultura de dados abertos no setor público.
- Envolvimento do governador para enviar uma mensagem clara a todo o governo de que a liderança apoia a iniciativa de dados abertos.

- Evento com o Secretário de Transparência do Reino Unido e o governador para estreitar os laços entre os países e fortalecer a política de transparência.
- Trabalhar para que as pessoas possam publicar os dados na plataforma escolhida, independentemente de as bases estarem limpas. Depois, trabalhar com as mesmas pessoas para melhorar a publicação desses dados.
- Concentrar os esforços em manter as bases consistentes, firmando acordo de nomenclaturas, dicionários, e metadatos.

2.3.4. Links importantes

- Discurso do primeiro ministro sobre transparência (vídeo disponível no data.gov.uk): https://www.youtube.com/watch?v=0stXV_fWWtU
- Prescribing Analytics (startup que reuniu dados de receitas médicas): <http://www.prescribinganalytics.com/>
- Linked data, quem está fazendo o quê?: <http://data.gov.uk/linked-data/who-is-doing-what>
- Pilotos de dados linkados: <http://data.gov.uk/blog/linked-data-pilot-implementations-update>
- Panorama geral sobre dados linkados no Reino Unido: <http://data.gov.uk/linked-data/overview-of-linked-data>
- Desenvolvendo URIs para o setor público no Reino Unido: <https://www.gov.uk/government/publications/designing-uri-sets-for-the-uk-public-sector>

Guia para ontologia de pagamentos: <http://data.gov.uk/resources/payments>

-

2.4. Mark Braggings (HampshireHUB)

2.4.1. Sobre o HampshireHUB

O HampshireHUB é uma iniciativa isolada do governo local de Hampshire para organizar um portal de dados abertos. É a experiência mais próxima do contexto do estado de São Paulo encontrada durante a visita. No governo de Hampshire, três pessoas ligadas à secretaria de planejamento local se reuniram para começar, do zero, uma iniciativa de dados abertos que tivesse foco nas informações locais. Os trabalhos começaram em 2010 na forma de um blog e foram progredindo lentamente até incluir recursos mais complexos, como dados linkados. Hoje, a equipe faz a manutenção do portal e a parte de relacionamento com outros órgãos. Além disso, contrataram uma empresa privada, chamada Swirrl, para fazer a estruturação de dados linkados.

2.4.2. Resumo da entrevista

A equipe vem construindo gradativamente o que eles chamam de Hampshire (Open Data) HUB. Começou como um blog onde eles diziam o que eles queriam fazer e o projeto foi crescendo e ganhando recursos do governo. Trata-se de um observatório/repositório/divulgador de dados abertos e linkados. Eles estão prestes a lançar um portal de dados abertos e linkados próprio, independente do data.gov.uk.

A experiência no HampshireHUB também não encontrou padrões estabelecidos para fazer com que outras agências e órgãos publicassem suas bases. As perguntas que ainda cercam as atividades de dados abertos giram em torno do quanto de dinheiro vai se gastar. Se for muito, pouco provável que passe. Mas a pressão necessária pra fazer com que algo que não é aberto se abra é bem menor que no Brasil.

A equipe foi escolhendo pessoas de diferentes áreas que tinham mais afinidade com dados e, principalmente, as pessoas que são as "detentoras" das bases de dados. O trabalho de convencimento era focado nessas pessoas. Se elas fossem convencidas, o restante do trabalho, que inclui a identificação dessas bases, a conformização de nomes, treinamento para utilização de plataforma etc, fica facilitado.

Para popular o repositório de dados abertos e linkados numa plataforma contratada e desenvolvida por uma empresa inglesa chamada Swirrl, eles

contrataram uma segunda empresa que baixa os dados das secretarias, preenche na plataforma os metadados de acordo com definições já estabelecidas (seguindo orientações do W3C) e repete o processo periodicamente, num contrato inicial de dois anos. Depois disso, eles esperam que o trabalho seja o de apenas atualizar as bases, não os seus metadados.

2.4.3. Recomendações

- Refletir sobre o interesse de terceirizar, inicialmente, o preenchimento de metadados e conformização de bases.
- Mapear servidores interessados em dados, mapear os detentores das bases de interesse em todos os órgãos do governo.
- Concentrar o trabalho de treinamento e aproximação nesses servidores.

2.4.4. Links importantes

- Dados linkados com mapas: <http://linkeddata.hants.gov.uk/>
- HampshireHUB: <http://protohub.net/>
- Swirrl (empresa de dados linkados): <http://www.swirrl.com>

2.5. Amanda Smith (Open Data Institute - ODI)

2.5.1. Sobre o ODI

O ODI foi fundado por Tim Berners-Lee e Nigel Shadbolt, como uma empresa sem fins lucrativos e apartidária. O instituto recebeu 10 milhões de libras do governo britânico ao longo de cinco anos (por meio da agência de inovação) e 750 mil libras da Omidyar Network. A empresa precisa equiparar o total investido com produtos e serviços oferecidos à medida que o seu modelo de negócios vai sendo implementado. Dentre as estratégias adotadas para captar recursos estão a formação de servidores públicos no setor de dados abertos, treinamento para empresas, como jornais, em jornalismo de dados, uma incubadora de startups, um centro de pesquisa, desafios temáticos, capítulos ao redor do mundo e cursos de curta e média duração.

2.5.2. Resumo da entrevista

A primeira rodada de encubação do ODI começou em 2013, com os temas meio ambiente e finanças. As empresas que conseguiram contratos acima de 250 mil libras foram consideradas "graduadas" e agora podem seguir modelos mais sustentáveis, fora do ODI. Dentre as startups do primeiro grupo está a Open Corporates, que analisa dados e relações entre empresas e seus contratos, usando dados públicos. A partir de junho de 2014, 12 startups começaram uma nova rodada de encubação.

O ODI possui um envolvimento estreito com o movimento de dados abertos no Reino Unido. Um dos focos do instituto é na pesquisa, com três grandes projetos. Os maiores se chamam Data and Platform as a Service, ou DaPaaS, e o Open Data Monitor. O primeiro envolve a participação de acadêmicos e empresas para construir ferramentas que facilitem a construção de bases de dados linkadas. O segundo funciona como um monitor de plataformas de dados abertos em nível nacional e local, principalmente na Europa, informando quais dados foram publicados, por quem, quando.

Na opinião da Amanda, o movimento de dados abertos no Reino Unido se beneficiou por duas frentes: mobilização da sociedade civil, por meio de ativistas respeitados, como Tim Berners-Lee, e vozes fortes dentro do governo. Ela também acredita que houve muito apoio à transparência e prestação de contas governamental, sem que isso trouxesse à tona a ideia de que os dados poderiam ser utilizados comercialmente.

Outro fator importante para ajudar a descrever o cenário de transparência no Reino Unido seria o envolvimento direto da Casa Civil em todos os processos. Exemplo disso são duas equipes fundamentais para oxigenar o ecossistema de transparência no governo, o time de transparência e a equipe que administra o dados.gov.uk; ambos subordinados à Casa Civil e com autonomia e autoridade para coordenar e cobrar resultados dos demais órgãos do governo.

O ODI percebe o processo de engajamento de empresas privadas na questão dos dados abertos em duas etapas. A primeira consiste em levar uma personalidade que demonstra alto envolvimento e conhecimento técnico em relação a abertura de dados. Essa pessoa faria o primeiro contato com a empresa por meio de oficinas ou palestras, de modo a sensibilizar a administração e seus funcionários de que dados abertos e públicos podem ser um caminho interessante a ser seguido. A segunda fase consiste em identificar quais são os órgãos governamentais que possuem as bases de interesse do setor privado e organizar fins de semana, hackatonas, em que servidores públicos e funcionários das empresas possam, juntos, trabalhar nessas bases de dados e encontrar formas úteis de utilizá-las. O principal instrumento do ODI

que atua nessa frente se chama Open Data Challenge Series. Mais informações na lista de links abaixo.

2.5.3. Recomendações

- Identificação de profissionais que possam ser os porta-vozes da política de dados abertos do governo de São Paulo para apresentação de casos nas empresas
- Formulação de desafios temáticos, para que o governo possa concentrar recursos em áreas específicas com resultados de alto impacto
- Trabalho de sensibilização com organizações do setor privado para entender quais são os dados que elas precisam
- Envolvimento das secretarias e órgãos relacionados, por meio de hackatonas que misturam servidores públicos e funcionários de empresas interessadas

2.5.4. Links recomendados

- Open Policy Making: <https://openpolicy.blog.gov.uk/>
- Mastodon C (startup de Big Data): <http://www.mastodonc.com/>
- Open Data Challenge Series:
<http://opendatainstitute.org/opendatachallengeseries>
- Guias da ODI: <http://opendatainstitute.org/guides>
- Cursos da ODI: <http://opendatainstitute.org/international-courses> &
<http://opendatainstitute.org/courses>
- Programa de incubadoras de startups de dados abertos:
<http://opendatainstitute.org/start-ups>

2.6. Emma Prest (DataKind UK)

2.6.1. Sobre o DataKind UK

A DataKind é uma ONG global que procura unir organizações que possuem algum problema de dados (visualização, análise, coleta etc) com cientistas de dados para aumentar a qualidade do acesso e o entendimento dos dados no setor social. Fundada em 2011, a DataKind tem escritórios em Nova Iorque, Washington DC, Bangalore, Dublin, São Francisco, Singapura e no Reino Unido. O capítulo na Grã Bretanha é liderado pela Emma Prest. Ela tem contato com cerca de 60 ONGs interessadas em abrir seus dados.

2.6.2. Resumo da entrevista

O trabalho do DataKind não está necessariamente incluído no ecossistema de abertura de dados do governo, mas a experiência da ONG pode dar *insights* interessantes sobre o contexto no Reino Unido. Na opinião da Emma, apesar de o Reino Unido possuir uma estrutura de dados abertos madura, esses dados são utilizados por uma parcela ínfima, quase desprezível, da sociedade.

Isso ocorreria porque a sociedade civil não está familiarizada com o trabalho que o governo faz para abrir os dados. Além disso, falta mão de obra especializada nessas ONGs, profissionais que sejam mais sensíveis ao mundo dos dados e possam apontar caminhos de modo que essas organizações possam tirar mais proveito do que os dados podem oferecer a elas. Há também uma sensação de que o governo não oferece ajuda suficiente para envolver as empresas em um processo de crescimento mútuo dentro do panorama de abertura de dados e transparência.

Dentre as principais queixas, estaria a falta de uma ação coordenada pelo governo para que as organizações possam conhecer e entender quais bases estão públicas e o que é possível fazer com elas. O nível técnico das organizações em relação aos dados força o trabalho da DataKind, por exemplo, a ser ainda bastante basal, no sentido de ter que realizar um processo de formação das organizações, que passa pela conscientização, convencimento e envolvimento.

2.6.3. Recomendações

- Parceria do governo com organizações e redes que possam ajudar no trabalho de divulgação e treinamento.
- Participação ativa das secretarias em eventos/encontros para divulgar bases abertas e linkadas.

Estímulo à criação de ambientes onde a sociedade civil e servidores públicos possam interagir como expedição de dados (Escola de Dados/Open Knowledge Brasil), Hackatonas (Transparência Hacker), spinoffs do InovaDay (Secretaria de Planejamento).

-

2.7. Rufus Pollock (Open Knowledge)

2.7.1. Sobre a Open Knowledge

A Open Knowledge é uma rede sem fins lucrativos que promove a abertura de conhecimento em todas as instâncias, principalmente no governo e por meio de ferramentas digitais. Dentre seus grupos de trabalho encontram-se ativistas ligados às áreas de Ciência Aberta, Recursos Educacionais Abertos, Código Aberto e Dados Abertos. Fundada em 2004 na Universidade Cambridge, na Inglaterra, o grupo tem representação em mais de 40 países, inclusive no Brasil. Ao longo dos anos seus projetos globais vêm ganhando ímpeto, os mais notáveis sendo a Escola de Dados (com orçamento de um milhão e meio de reais) e a Parceria pelos Dados Abertos (POD, na sigla em inglês), entre o Banco Mundial e o Open Development Institute, com dois milhões e meio de reais.

2.7.2. Resumo da entrevista

De acordo com Rufus Pollock, uma política de dados abertos precisa ter uma narrativa de curto, médio e longo prazo: onde estamos? Onde queremos chegar? Quais são os beneficiados diretos pela abertura de dados? Cidadãos? Jornalistas? Servidores públicos? Empresas? Cientistas? Infomediários? Como chegar até eles? Essas são algumas das perguntas que precisam de respostas claras e objetivas dentro da organização e do planejamento de qualquer política de transparência e abertura de dados.

O governo britânico já está colhendo bons frutos por ter optado abrir seus dados. Um dos casos mais emblemáticos foi a descoberta, por causa de dados abertos, que diversos setores de informática do governo estavam comprando o mesmo serviço de consultoria da empresa Gartner. Ao analisar os dados, a administração percebeu que poderia economizar seis milhões de libras diminuindo o número de horas de consultoria contratada. A cifra é o equivalente a todo recurso necessário para pagar o programa de transparência do governo britânico. Com efeito, a questão da transparência está tão arraigada no Reino Unido que já ocorreram casos de um ministério usar a lei de acesso à informação para conseguir dados de outro.

A Open Knowledge desenvolveu um conceito chamado "[Frictionless Data](#)" ou algo como "Dados Sem Atrito", uma forma de analisar a problemática de dados abertos sob o ponto de vista da simplificação de processos e do pragmatismo. O conceito descreve "Dados Sem Atrito" aqueles que são abertos, ou seja, legíveis

por máquina e que possuam licença livre; estruturados, de modo que haja um padrão acordado entre os atores sobre a forma de consolidá-los; e públicos, em que o maior número de pessoas possível consegue encontrá-lo de forma facilitada.

Para a produção dos *guidelines*, Rufus faz uma série de recomendações, como exigir que todos os dados tenham licença aberta por padrão; fazer uma lista (pequena) de exceções, de documentos que não podem ser publicados; elencar quais bases de dados são as mais importantes, desejadas dentro e fora do governo e priorizar os esforços nessas bases. O formato mínimo recomendado é o CSV, e, na pior das hipóteses, uma planilha do Excel. Além disso, recomenda também seguir o exemplo do governo mexicano, que estabeleceu um "esquadrão de dados", uma força-tarefa que ia de órgão em órgão dentro do governo ajudando a iniciar os processos de abertura de bases. Essa equipe poderia ter um período de validade, seis meses, por exemplo, e depois seria reavaliada a necessidade dela. Também sugeriu a organização de encontros na hora do almoço em cada uma das secretarias, para identificar quem são as pessoas que estão motivadas em cada departamento. Não é possível dizer se essa última sugestão seria viável no Brasil, por causa de diferenças de costumes no horário do almoço entre os dois países. O brasileiro costuma almoçar em restaurantes, o britânico leva o almoço para o escritório.

Sobre o Portal de Dados abertos, Rufus pensa que trata-se de uma plataforma e não um produto final. É preciso refletir sobre que tipo de informação entra e quais tipos de produtos podem sair de um Portal de Dados Abertos, de modo que as integrações de bastidores (a formulação de scripts para automatizar o processo periódico de publicação de bases, por exemplo) façam parte do fluxo de trabalho dos servidores, sem atrapalhar a rotina deles. O Portal é também uma forma de o governo rastrear quais dados abertos estão sendo publicados e mantê-los organizados e atualizados.

Dentre os produtos que podem depender de um Portal de Dados Abertos bem administrado estão os *dashboards*, ou painéis dinâmicos. Essas ferramentas poderiam mostrar a performance da cidade ou do estado a partir de diversos indicadores construídos em cima das bases de dados disponíveis no Portal. Essa estratégia seria interessante pois os dados utilizados precisam ser abertos e atualizados para que o *dashboard* funcione, criando um ciclo virtuoso. Os painéis podem ser criados de acordo com a demanda do governador, dos secretários e dos cidadãos, áreas como "criminalidade", "finanças", "saúde", "meio ambiente" ou "transporte". A escolha poderia ser feita a partir de uma tempestade de ideias com servidores, gestores e a sociedade civil ou por meio de uma consulta pública. Essas informações consistiriam no argumento para o desenvolvimento, que poderia ser feito por meio de competições. O governo colocaria as bases e as APIs à disposição dos programadores e empresas, e

eles fariam o protótipo. A empresa vencedora poderia ganhar um contrato de um ano para ampliar e melhorar o protótipo vencedor.

Um dos fatores mais importantes, segundo Rufus, é a contratação, por parte do governo, de um gerente de comunidades. Alguém que possa "sair por aí falando sobre os dados e as possibilidades de se trabalhar com os dados abertos do governo". Na opinião dele, gasta-se demais na parte técnica, mas muito pouco na divulgação dos resultados conquistados. Essa pessoa poderia frequentar encontros de startups, visitar empresas, redações, organizações da sociedade civil, eventos, e até outras agências dentro do mesmo governo. Essa contratação pode ocorrer por parceria com ONGs ou agências que realizam trabalho pro-bono.

2.7.3. Recomendações

- Adoção de uma política de "Dados Sem Atrito", nos moldes recomendados pela Open Knowledge.
- Criação de uma força-tarefa por um período de seis meses, capaz de iniciar o processo de abertura de dados em diversas secretarias.
- Utilizar a plataforma CKAN como o armazém de dados do Estado de São Paulo.
- Fortalecer o Portal de Dados Abertos com a criação de painéis dinâmicos que mostrarão informações de interesse do executivo e da sociedade.
- Painéis dinâmicos podem ser criados por áreas temáticas por meio de competições onde o prêmio seria a assinatura de contrato para melhoria do produto durante um ano.
- Contratação de um gerente de comunidades, ou por meio de parceria, para que o trabalho de dados abertos ganhe visibilidade dentro e fora do governo.

2.7.4. Links recomendados

- "Frictionless Data": <http://data.okfn.org/>
- Open Data Handbook: <http://opendatahandbook.org/>
- Open Government Toolkit (Banco Mundial):
<http://data.worldbank.org/open-government-data-toolkit>

2.8. Tracie Gyateng (New Philanthropy Capital - NPC)

2.8.1. Sobre o NPC

A NPC é uma organização não-governamental que faz a ponte entre organizações e indivíduos que querem doar dinheiro para causas diversas e ONGs que trabalham por essas causas. O objetivo é maximizar o impacto do trabalho realizado pelas ONGs. O funcionamento é parecido com o IDIS (Instituto para o Desenvolvimento do Investimento Social), com sede em São Paulo.

2.8.2. Resumo da entrevista

A Tracey Gyateng foi responsável por vender e organizar uma idéia promissora no Ministério da Justiça britânico e que agora está ganhando tração para atingir outras áreas, como previdência social, trabalho, saúde e educação. O projeto se chama [Data Labs](#) e tenta melhorar o trabalho de organizações que complementam serviços do governo por meio da abertura de dados.

A NPC se aproximou do Ministério da Justiça, que possui informações sobre o sistema carcerário (quem cometeu o crime, se há reincidência etc), para realizar uma parceria de abertura de dados. Juntos, desenvolveram um processo que permite a uma série de organizações que realizam trabalhos com presos, egressos do sistema prisional e jovens infratores, receber um relatório de performance sobre o trabalho que realizam com essas pessoas. Eles enviam os dados pessoais das pessoas que atenderam para o Ministério da Justiça, que cruza esses dados com as bases de dados protegidas. Eles então devolvem um relatório geral sobre a atuação da organização e esse relatório permanece público.

Um exemplo citado foi o de uma organização que trabalha com egressos do sistema prisional e ela quer saber a taxa de pessoas atendidas nos últimos cinco anos que acabaram voltando para a prisão. Apenas o Ministério tem essa informação e ela é confidencial. Contudo, a forma como os relatórios são gerados (nos últimos anos, 40% das pessoas atendidas seriam reincidentes, por exemplo) permite que a ONG tenha um retorno sobre o trabalho que realiza e isso se encaixa na agenda de transparência do ministério, devido a determinação do primeiro ministro a aumentar o volume de dados abertos publicados. A ideia do Data Labs está sendo levada para outros ministérios com o mesmo intuito.

Outro aspecto digno de nota que surgiu durante a entrevista é que o portal de dados do Reino Unido não é bem visto pelas organizações da sociedade civil, apesar de que a utilidade do portal para o usuário final é contestada por outro entrevistado, Rufus Pollock, da Open Knowledge.

2.8.3. Recomendações

- Apresentar a ideia do Data Labs para o IDIS e medir o interesse de organizações da sociedade civil que queiram trabalhar com dados abertos.
- Se o interesse for identificado, aproximar-se dos órgãos de governo que detém esses dados e avaliar o interesse da administração em replicar uma experiência parecida com a do Reino Unido.
- Caso o interesse seja confirmado, marcar reuniões de mentoria com a NPC para adaptação do modelo Data Labs para o contexto brasileiro

3. Conclusões

A visita ao Reino Unido não foi uma tentativa de mapear de forma exaustiva a experiência britânica em relação aos dados abertos e linkados. Contudo, a amostra contém vozes significativas dentro do governo britânico e dão um panorama amplo e, por vezes, detalhado de iniciativas que vem dando certo e de algumas que deram errado no Reino Unido. Alguns temas são recorrentes em todas as entrevistas. O mais evidente é que a política de transparência na Grã Bretanha tem sido exitosa devido ao envolvimento e comprometimento dos mais altos escalões do governo, incluindo, e especialmente, o Primeiro Ministro. Sem esse apoio, a resposta é unânime: seria praticamente impossível fazer com que a política de transparência fosse adotada por todos.

Outro tópico evidenciado pelos entrevistados é de que a centralização da política de transparência, que inclui abertura de dados e dados linkados, deve ser centralizada e com a anuência de um órgão que confira autoridade e autonomia para que essa política seja articulada em todas as secretarias e órgãos. Isso não quer dizer que apenas um grupo de pessoas ficará responsável pela quantidade homérica de trabalho necessário para abrir as bases de dados do governo, mas que essa equipe terá um apoio "de cima" e de parceiros por todo o governo.

Igualmente importante é a preocupação que o governo deve depositar na formação de parcerias com organizações que estejam interessadas em trabalhar conjuntamente com o setor público para a abertura de bases. Sozinho, será demasiadamente complicado para o governo executar seus planos para fundamentar uma política de transparência que dê certo. Essas parcerias podem gerar muitos benefícios tanto para a sociedade civil quanto para o setor público, na formação de equipes que podem acelerar a abertura de dados em diferentes órgãos e na contratação de profissionais que vão ajudar a divulgar o trabalho do governo em diferentes cenários.

Por fim, não há qualquer modelo no governo britânico que deva ser seguido à risca. Todas as recomendações contidas neste relatório devem ser lidas à luz do contexto paulista de governo, suas dificuldades, seus limites, suas características únicas e suas qualidades.