

O ecossistema de dados abertos do governo federal: um estudo sobre a composição e desafios

José Antônio de Carvalho Freitas

Mestrado profissional em Gestão do Conhecimento e Tecnologia da Informação pela Universidade Católica de Brasília (UCB) - Brasília, DF - Brasil. Analista de Finanças e Controle da Controladoria-Geral da União (CGU) - Brasília, DF - Brasil. Professor do Instituto de Educação Superior de Brasília (IESB) - Brasília, DF - Brasil. <http://lattes.cnpq.br/9863103348924299>

E-mail: jacfreitas@gmail.com

Remis Balaniuk

Pós-Doutorado pelo Institut National Pour La Recherche En Informatique Et Automatique (INRIA) - França. Pós-Doutorado pela Stanford University (STANFORD) - Estados Unidos. Doutorado em Doctorat en Informatique et Automatique pelo Institut National Polytechnique de Grenoble (INPG) - França. Professor da Universidade Católica de Brasília (UCB) - Brasília, DF - Brasil. Analista de Finanças e Controle Externo do Tribunal de Contas da União (TCU) - Brasil.

<http://lattes.cnpq.br/8456124569258786>

E-mail: remis_balaniuk@yahoo.com

Ana Paula Bernardi da Silva

Doutora em Engenharia Elétrica pela Universidade de Brasília (UnB) - Brasília, DF - Brasil. Professora da Universidade Católica de Brasília (UCB) - Brasília, DF - Brasil. <http://lattes.cnpq.br/2644093939475294>

E-mail: anapaula@ucb.br

Vitória Santiago da Silveira

Graduanda em Engenharia Civil pela Universidade Católica de Brasília (UCB) - Brasília, DF - Brasil.

<http://lattes.cnpq.br/6000001984618351>

E-mail: vitoriasant_go@hotmail.com

Submtido em : 21/08/2017. Aprovado em : 18/06/2018 . Publicao em : 26/09/2018.

RESUMO

As tecnologias de Internet, associadas aos dados abertos disponíveis nos portais de governo, possuem forte potencial para promover a transparência e o engajamento social. A disponibilidade de grande volume de informações de governo faz nascer um verdadeiro ecossistema de dados abertos, em que diversos atores e instâncias se revezam em seus componentes e papéis. Este trabalho, de vertente exploratória, procurou apresentar uma descrição dos principais componentes do ecossistema de dados abertos no âmbito do governo federal. Para isso, foram levantados dados com especialistas, mediante consultas a grupos de discussão e aos próprios repositórios oficiais de dados abertos. Ao final, foi possível, por meio dos componentes identificados, analisar criticamente alguns pontos que podem contribuir para o entendimento desse campo emergente e promissor, destacando que o ecossistema de dados abertos do governo federal encontra-se em desenvolvimento e enfrenta desafios estruturantes.

Palavras-chave: Aplicativo cívico. Participação cívica. Ecossistema de dados. Dados abertos.

The open data ecosystem of federal government: compositions and challenges

ABSTRACT

Internet technologies, associated with open data available on the portals of government, have great potential to promote transparency and social engagement. The availability of large volumes of information from government formed a true ecosystem of open data, in which different actors and instances take turns in its components and roles. This work, with an exploratory approach, sought to present a description of the main components of the data ecosystem open within the Federal Government. For this, data were collected to experts, through consultations with virtual groups and official repositories of open data. At the end, it was possible by means of identified components, critically analyze some points that can contribute to the understanding of this emerging and promising field, noting that the Federal Government's open data ecosystem is developing and faces structural challenges.

Keywords: Civic applications. Data ecosystem. Open data.

El ecosistema de datos abiertos del gobierno federal: composición y desafíos

RESUMEN

Las tecnologías de Internet, asociadas a los datos abiertos disponibles en los portales de gobierno, poseen gran potencial para promover la transparencia y el compromiso social. La disponibilidad de gran volumen de información de gobierno hace nacer un verdadero ecosistema de datos abiertos, en el que diversos actores e instancias se turnan en sus componentes y papeles. Este trabajo, de vertiente exploratoria, trató de presentar una descripción de los principales componentes del ecosistema de datos abiertos en el ámbito del Gobierno Federal. Para ello, se levantaron datos con expertos, a través de consultas a grupos de discusión ya los propios repositorios oficiales de datos abiertos. Al final, fue posible, por medio de los componentes identificados, analizar críticamente algunos puntos que pueden contribuir al entendimiento de ese campo emergente y prometedor, destacando que el ecosistema de datos abiertos del Gobierno Federal se encuentra en desarrollo y enfrenta desafíos estructurantes.

Palabras clave: Aplicaciones cívicas. Ecosistema de datos. Datos abiertos.

INTRODUÇÃO

Com a concretização da democracia e dos valores republicanos na sociedade civil, as contribuições dos avanços nas tecnologias da informação e comunicação (TICs) não estão restritas ao setor privado. Muitas dessas inovações podem e devem fazer parte de uma agenda positiva de inclusão, aumento da transparência das decisões de governo, maior participação popular nas formulações de políticas e incremento real na cidadania e no controle social. Conceitos como governo aberto, dados governamentais abertos e participação cidadã por meio de tecnologias cívicas são explorados e apresentam-se interligados em uma teia de recursos, metodologias e tecnologias, com amplo nível de interdependência.

Um tema que surge nesse contexto é o movimento pela transparência, que tem adquirido força com as iniciativas de Governo Aberto e por força da Lei de Acesso à Informação (LAI), (2011). Esse movimento é representado pela disponibilização de dados governamentais por meio de tecnologias envolvendo a Internet, a exemplo do Portal da Transparência, mantido pelo Ministério da Transparência e Controladoria-Geral da União (CGU), e pelo Portal Brasileiro de Dados Abertos, mantido pelo Ministério do Planejamento, Orçamento e Gestão (MPOG).

Nesse contexto, o presente trabalho exploratório aborda a problemática de descrever a situação atual do ecossistema de dados abertos no Brasil, objetivando mapear seus elementos e apresentar contribuições para impulsionar o papel desse ambiente na criação de condições para o incremento de tecnologias cívicas, usando plataformas digitais que já fazem parte do cotidiano das pessoas, aproveitando os dados abertos e seu potencial para mudar a sociedade e o próprio governo.

Para atingir esses objetivos, dados foram coletados com (i) especialistas que atuam no contexto de dados abertos, com (ii) três grupos de discussão de comunidades de infomediários (dois via Google Groups e dois via Telegram) e (iii) por meio da análise dos repositórios de dados e conteúdo normativo, acessíveis via Internet.

REFERENCIAL TEÓRICO

GOVERNO ABERTO

O movimento de Governo Aberto (GA) iniciou-se com as medidas do governo americano relacionadas ao aumento dos mecanismos de transparência (VELJKOVIĆ; BOGDANOVIĆ-DINIĆ; STOIMENOV, 2014). Sua consolidação veio com a declaração oficial do Open Government Partnership (OGP), apresentando a ideia de que as pessoas estão “reivindicando maior participação cívica nos assuntos públicos e buscando meios de tornar seus governos mais transparentes, receptivos, responsáveis e eficientes” (OGP, 2011).

Juntamente com o Brasil, outros sete países compõem a OGP desde 2011, assumindo como compromissos o aumento da disponibilidade de informações sobre as atividades governamentais, o apoio à participação cívica, a melhoria nos padrões de integridade profissional dos governos e o acesso a novas tecnologias para fins de abertura e prestação de contas (OGP, 2011). No caso nacional, esses compromissos já obtiveram diversos frutos, com destaque para a LAI que, desde 2011, promove o acesso a informações públicas como regra, tendo as hipóteses de sigilo como exceção. Assim, um dos pilares do Governo Aberto é promover a transparência por meio de divulgação de dados governamentais, que, nos dias atuais, é feita por meio de elementos de TICs.

A concretização do GA passa pela definição de padrões, procedimentos e tecnologias para orientar as diversas esferas de governo na execução de suas iniciativas. Desses padrões, podem ser destacados os Padrões de Interoperabilidade de Governo Eletrônico (ePing) e a Infraestrutura Nacional de Dados Abertos (INDA).

O ePing é uma estrutura básica de padrões e políticas para a implementação de soluções de governo eletrônico (E-GOV) para o governo federal, sendo sugerido para os demais poderes e esferas de governo. Como políticas gerais têm-se a adoção preferencial de padrões abertos, uso de software público e/ou software livre, transparência, segurança e existência de suporte ao mercado.

Com mais ênfase ao escopo desse trabalho, destaca-se a dimensão semântica das políticas, que recomenda o desenvolvimento, manutenção, ontologias e adoção de disseminação de dados e informações (BRASIL, 2016).

A INDA, por sua vez, especifica dentro da arquitetura do ePing “um conjunto de padrões, tecnologias, procedimentos e mecanismos de controle necessários para atender às condições de disseminação e compartilhamento de dados e informações públicas no modelo de Dados Abertos” (BRASIL, 2012).

DADOS ABERTOS

Segundo o Open Knowledge Foundation, dados abertos “são dados que podem ser livremente usados, reutilizados e redistribuídos por qualquer pessoa - sujeitos, no máximo, à exigência de atribuição da fonte e compartilhamento pelas mesmas regras”. O adjetivo aberto implica disponibilidade e acesso, reutilização e redistribuição e participação universal. Por disponibilidade e acesso entende-se que os dados “devem estar disponíveis como um todo e sob custo não maior que um custo razoável de reprodução”. Em reutilização e redistribuição, compreende-se que “os dados devem ser fornecidos sob termos que permitam a reutilização e a redistribuição, inclusive a combinação com outros conjuntos de dados”. Já a participação universal quer dizer que “todos devem ser capazes de usar, reutilizar e redistribuir - não deve haver discriminação contra áreas de atuação ou contra pessoas ou grupos”. Esse conceito não apresenta a simples abertura como essência, na verdade é preciso assegurar a interoperabilidade dos dados para que se possa integrar as diversas bases e assim permitir o uso combinado de múltiplos sistemas (OKF, 2015).

A avaliação de iniciativas de dados abertos governamentais é feita com uso de diversas metodologias. Dessas, destacam-se os oito princípios da OGD e o modelo de cinco estrelas de Tim Berners Lee.

Os oito princípios da OGD, abordando aspectos sobre a natureza do dado e sobre a forma de acesso, são: completeza, primariedade, temporalidade, acessibilidade, compreensível por máquina, acesso irrestrito e livre de licenças (TAUBERER, 2007). O modelo de cinco estrelas de Tim Berners Lee, por sua vez, aborda a dimensão tecnológica na forma de publicação dos dados abertos com um esquema evolutivo em cinco estágios cumulativos, partindo da mais simples disponibilização na Internet até o momento em que é possível ter esses dados interligados.

A situação brasileira, com base nessas metodologias de avaliação, indica necessidade de ajustes na implementação de dados abertos. Com base nos oito princípios da OGD, é preciso incrementar a quantidade de dados e disponibilizá-los diretamente da fonte primária, sem tratamento posterior. No caso do método das cinco estrelas, o país alcança três somente, sendo necessário incrementar o uso de URIs e interconectividade desses dados, para atingir a quarta e quinta estrela respectivamente (MATHEUS, RIBEIRO, VAZ, 2012; OGD, 2010).

APLICATIVOS BASEADOS EM DADOS ABERTOS

A disponibilização de dados abertos governamentais possibilita a criação de aplicações específicas para reutilizar informações na promoção da cidadania e de serviços públicos. Esse fato transforma a estrutura de governo em uma plataforma na qual o tripé transparência, participação e colaboração ganha espaço ao associar-se com tecnologias da Internet (O'REILLY, 2011).

O surgimento de aplicativos cívicos¹ possibilita a criação de uma camada de valor agregado aos dados abertos governamentais, viabilizando o emprego de uma abordagem de modelo de negócio que incrementa ainda mais as possibilidades de interação com a sociedade (ALBANO, 2014; GERMANO, 2013).

¹ Essa designação alcança todas as soluções tecnológicas criadas no contexto de dados abertos, quer seja via web ou dispositivos móveis. Observa-se maior referência a esse termo nos trabalhos realizados após a abertura de dados do governo americano em 2009, segundo Eyller-Werve e Carlson (2012).

A maior parte das iniciativas desses aplicativos tem sido produzida por meio de competições promovidas por entidades públicas ou filantrópicas.

Como visto em Eyler-Werve e Carlson (2012), essas competições permitem a criação de uma arena de engajamento de organizações sem fins lucrativos, empresas e agências governamentais, criando redes de relacionamento para a solução de problemas comuns com o uso de tecnologias. Esses autores ressaltam que os ganhos abarcam desde os princípios de responsabilização, até a melhoria da eficiência dos governos, passando pelo incentivo à inovação e a criação de comunidades de práticas.

Como exemplos dessas competições podem ser citados a I Maratona Hacker Câmara dos Deputados (2013), o Concurso Inovapps do Ministério das Comunicações (2014), e mais recentemente o INOVApps (2015) e o Controle Social Digital – TCU (2016). No contexto dessas competições, cabe destaque a participação do Laboratório de Dados Abertos Brasil (vide dadosabertosbr.com), iniciativa da Universidade Federal de Pernambuco (UFPE). Cabe mencionar também o sítio aplicativos.gov.br, que contém um “catálogo de aplicativos para dispositivos móveis criados por órgãos do poder executivo federal” (BRASIL, 2016). A experiência de desenvolvimento dessas aplicações tem mostrado que, para assegurar que elas consigam de fato agregar valor público, a infraestrutura de dados abertos brasileira carece de melhorias em alguns pontos. Dentre eles, podem ser citados o baixo número de bases disponibilizadas (datasets), a necessidade de manter os dados atualizados, a remoção da exigência de registro ou obtenção de licença, a melhoria da centralização e padronização dos dados e a criação de canais de retorno, capturando a interação com usuário nas aplicações (BRITO et al., 2015).

ECOSSISTEMA DE DADOS ABERTOS

Para o Banco Mundial, o ecossistema de dados abertos é uma abordagem para tratar a questão de dados abertos com enfoque não somente nos dados em si, mas também numa percepção mais ampla que envolve diversos atores (desenvolvedores, universidades e setor privado) e dimensões (liderança, político/legal) (THE WORLD BANK, 2013).

Essa noção advém do conceito de ecossistema de software proposto por Jansen, Finkelstein e Brinkkemper (2009): um conjunto de empresas e produtos funcionando na mesma plataforma e compartilhando o mercado de software e serviços entre si. Os autores citam como exemplo o ecossistema de softwares do iPhone, representado pela AppStore, e da Microsoft e sua rede de parceiros e plataformas (JANSEN; FINKELSTEIN; BRINKKEMPER, 2009).

Para auxiliar na definição do que vem a ser um ecossistema de dados abertos, algumas pesquisas apresentam modelos no formato de propriedades, relações e componentes, que caracterizam de modo peculiar esse tipo de arranjo de elementos:

- i. Zuiderwijk, Janssen, Davis (2014) trazem sete propriedades elementares e necessárias para um ecossistema: 1) liberação e publicação de dados abertos na internet; 2) criação de mecanismos de procura, descoberta, avaliação e visualização de dados e suas licenças relacionadas; 3) limpeza, análise, melhoria, combinação e interligação de dados; 4) interpretação, discussão e fornecimento de feedback para o provedor dos dados e outras partes interessadas. 5) formas de sinalizar ao usuário como os dados abertos podem ser usados; 6) sistema de gestão da qualidade e 7) diferentes tipos de metadados capazes de conectar os elementos que compõe os dados;

Para compor a descrição do ecossistema brasileiro de dados abertos no âmbito federal, serão utilizadas as três abordagens citadas anteriormente, com destaque para o modelo de Dawes, Vidiasova e Parkhimovich (2016), por representar uma compilação ampla de outros trabalhos e uma construção empírica, tendo como base pesquisa nos ecossistemas de Nova Iorque e São Petersburgo. Com base nesse modelo, são listados a seguir as dimensões e os itens que guiarão a análise:

- Provedores de dados abertos – agências governamentais:
 - d. regulação/influência nacional e global;
 - e. políticas e estratégias;
 - f. dados publicados.
- Infomediários - usuários dos dados abertos – comunidade de tecnologias cívicas e defensores de transparência pública:
 - a. características da comunidade;
 - b. uso de dados e produtos;
 - c. feedback e comunicação.
- Consumidores de produtos e serviços ligados a dados abertos.
 - a. benefícios;
 - b. características da comunidade.

PROCEDIMENTOS METODOLÓGICOS

Para se atingir o objetivo de descrever o ecossistema de tecnologias cívicas no governo federal foi utilizado o modelo de Dawes, Vidiasova e Parkhimovich (2016) como referência para a determinação de quais componentes seriam analisados. Para o aproveitamento do modelo, foram utilizados dados coletados em questionários e entrevistas com especialistas em dados abertos e consultas aos repositórios de informações on-line do governo federal e de normativos afetos ao tema.

Como o objeto de pesquisa deste trabalho é o ecossistema de tecnologias cívicas baseados em dados abertos brasileiros, nada mais natural do que estudá-lo por meio do processo de desenvolvimento de um tipo de componente chave desse ecossistema, ou seja, um aplicativo cívico. Foram percorridos todos os caminhos típicos para esse produto, do desenvolvimento até sua disponibilização na loja de apps (Google Play2), e, a partir da exploração desse processo, junto com um mapeamento e descrição do ecossistema, foi feito um diagnóstico e identificaram-se lacunas que podiam ser melhoradas para facilitar o florescimento dessas tecnologias.

Para isso, realizaram-se coletas de dados em três momentos distintos: (i) durante o processo de descrição do ecossistema e desenvolvimento da solução, a partir da experiência e do contato com as ferramentas e dificuldades encontradas; (ii) durante a disponibilização ao público, utilizando um grupo previamente orientado a instalar a solução e (iii) por último, dados de usuários aleatórios que baixaram a ferramenta, sem estímulo direto. Esse posicionamento marca esta pesquisa como de concepção filosófica pragmática, ou seja, buscou-se atingir os objetivos por meio de uma aplicação prática, que, de forma indutiva, procura apresentar contribuições para o contexto maior dessas soluções (CRESWELL, 2007).

Quanto ao delineamento para coleta de dados, este trabalho buscou atingir seus objetivos por meio de um experimento, também caracterizável como um estudo de caso, envolvendo o desenvolvimento de uma aplicação de software (CRESWELL, 2007), tendo como base um conjunto de dados abertos disponíveis no contexto da Agência de Vigilância Sanitária (ANVISA), especificamente os dados da Câmara de Regulação do Mercado de Medicamentos (CMED)³.

² Loja de aplicativos do Google, disponível em https://play.google.com/store?hl=pt_BR. Acesso em 13/11/2016.

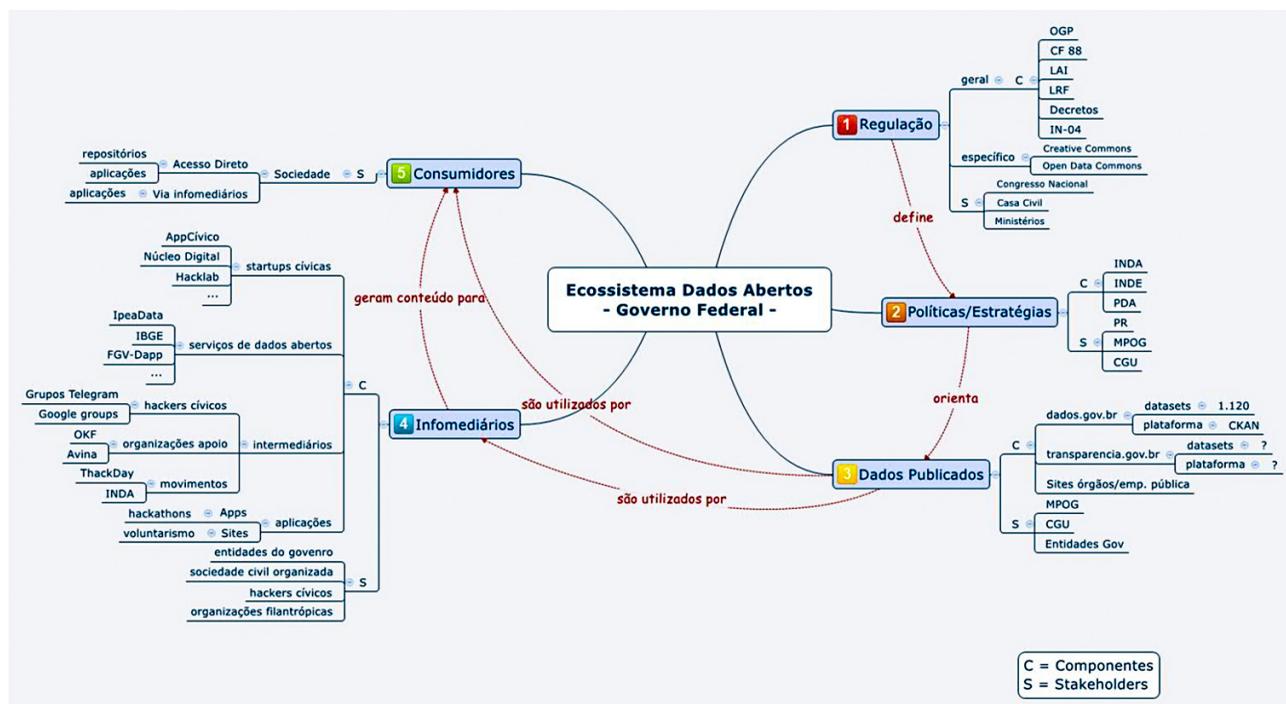
³ Entidade interministerial responsável pela condução da política de preços de medicamentos. Mais detalhes em <http://portal.anvisa.gov.br/cmcd>. Acesso em 12/11/2016.

Também foram feitas coletas de dados nos repositórios institucionais do governo federal e em grupos de discussão de aplicativos móveis (Telegram⁴ e Google Groups⁵), por meio de observação e aplicação de questionário.

RESULTADOS

Com os dados coletados foi possível esboçar o mapa mental da figura 2, que serve de guia e apresenta, inicialmente, no formato de elemento gráfico, uma descrição geral do ecossistema de dados abertos no âmbito do governo federal. Nessa figura, destacam-se as dimensões e seus principais componentes e *stakeholders*, bem como possíveis linhas de relacionamento, mostrando-se eventual interdependência, aspectos essenciais para descrição de um ecossistema desse tipo, alinhando-se com o que é dito em Dawes, Vidiasova, Parkhimovich (2016); Heimstädt, Saunderson, Heath (2014) e Zuiderwijk, Janssen, Davis, (2014).

Figura 2 – Ecossistema de dados abertos – governo federal



Fonte: Elaboração própria com base no modelo de Dawes, Vidiasova e Parkhimovich (2016).

⁴ Aplicativo utilizado para troca de mensagens. Disponível em <https://telegram.org/>. Acesso em 12/11/2016.

⁵ Ferramenta disponibilizada pela empresa Google para criação de grupos de discussão on-line. Disponível em <https://groups.google.com/forum/#!homeredir>. Acesso em 12/11/2016.

Enriquecendo a descrição inicial mostrada na figura anterior, passa-se a discorrer sobre aspectos relevantes de cada dimensão do ecossistema.

DIMENSÃO DE REGULAÇÃO

Essa dimensão trata de normatizar o pilar básico do ecossistema, que é a disponibilização dos dados abertos pelos órgãos de governo. Com base no sítio oficial governoaberto.cgu.gov.br, foram encontrados vários componentes que formam um contínuo de normatização para o ecossistema ao partir da referência internacional e da própria Constituição Federal, passando por diversas leis e chegando a decretos, portarias e instruções normativas, de natureza infralegal. Essa dimensão pode ser analisada com base em dois pontos de vista: (i) um abordando a legislação mais geral, que define a transparência como regra e direito do cidadão e, a partir disso, sedimenta a necessidade de abertura de dados por parte dos órgãos de governo; e (ii) outro abordando a questão do licenciamento do dado aberto em si.

Sobre o primeiro aspecto, por ordem de abrangência e formando um contínuo de regulação, são apresentados a seguir os principais normativos aplicáveis ao ecossistema no contexto do governo federal:

1. Parceria para Governo Aberto (OGP), envolvendo 69 países⁶ trazendo princípios e políticas. A OGP é um componente de destaque desta dimensão pois, enquanto iniciativa internacional, apresenta princípios que balizam a noção de abertura de dados como política pública, definindo-se o que se chama de Governo Aberto: Transparência; Prestação de Contas e Responsabilização; Participação Cidadã; e Tecnologia e Inovação. Esses princípios são refletidos nos planos de ação nacionais, em que cada país membro assume diversos compromissos envolvendo melhoria de serviços públicos, aumento de integridade e gestão de recursos;

2. a Constituição Federal - CF de 1988, por meio do inciso XXXIII do art. 5º, do inciso II do § 3º do art. 37 e do § 2º do art. 216, também apresenta orientação para a transparência, alinhando-se naturalmente com os princípios da OGP e assumindo também posição de destaque na dimensão, pois, como norma maior, cria o direito de acesso à informação, refletido na regulação complementar e no comportamento dos diversos agentes públicos;
3. desdobrando-se da CF, destacam-se a Lei Complementar nº 101, de 4 de maio de 2000 - LRF, que define a necessidade de transparência de dados e transações envolvendo finanças públicas; e a Lei nº 12.527, de 18 de novembro de 2011 - LAI, que generaliza o instituto da transparência como regra e o sigilo como exceção, exigindo do poder público medidas para publicação de dados e atendimento a pedidos de informação;
4. em nível mais operacional e regulamentando as leis mencionadas anteriormente, destacam-se o Decreto nº 7.724, que regulamenta a LAI; o Decreto nº 8.777, que institui a Política de Dados Abertos do Poder Executivo federal; e o Decreto nº 6.666, que detalha a Infraestrutura Nacional de Dados Espaciais (INDE). Complementando-se essa lista, tem-se e a Instrução Normativa nº 4, de 12 de abril de 2012, que cria a Infraestrutura Nacional de Dados Abertos (INDA), elemento que especifica os detalhes técnicos do espaço de dados abertos dos diversos entes do governo federal.

Esse contínuo de regulação apresentado já é responsável por quantidade considerável de dados abertos e divulgados para a sociedade, como será mostrado adiante na dimensão de dados publicados, o que faz com que a transparência já possa ser considerada uma realidade em alguns órgãos federais.

⁶ Lista de países participantes disponível em <https://www.opengovpartnership.org/participants>. Acesso em: 24 abr.2018.

Acredita-se que a discussão avança no sentido de aumento das informações divulgadas, e na necessidade de se entender ainda mais o que fazer com esse volume considerável de informação, levantando-se até a questão se a transparência seria um fim em si mesma, ou uma ferramenta viabilizadora de participação cidadã que transforma a sociedade e resolve problemas. Como *stakeholders* destacam-se o Congresso Nacional (CN) e a Presidência da República (PR), inter-relacionados no processo legislativo e refletindo os anseios da sociedade civil por meio de seus representantes, além dos próprios órgãos públicos que implementam esses mandamentos como planos de ação para abertura dos seus dados.

Em relação à segunda perspectiva, da regulação do uso dos dados, observa-se o aspecto da forma de licenciamento dos dados abertos disponibilizados nos portais, sendo esse um dos itens fundamentais para a própria conceituação do dado como aberto, permitindo livre uso e modificação. Nessa linha, a INDA orienta que os dados abertos estejam sobre licença aberta, representando um “acordo de fornecimento de dados que conceda amplo acesso para que qualquer pessoa os utilize, os reutilize, e os redistribua, estando sujeito a, no máximo, a exigência de creditar a sua autoria e compartilhar pela mesma licença”(BRASIL, 2012). Mas deixa claro que é preciso estabelecer um modelo de licenciamento próprio para o contexto brasileiro, objetivo descrito no Plano de Ação Nacional sobre Governo Aberto, considerado finalizado, mas ainda não publicado e portando não disponível para utilização. Nesse plano, esse objetivo tinha como meta temporal agosto de 2014. Em análise, como já se passaram mais de dois anos sem um modelo de licenciamento próprio, entende-se que essa questão ainda está em aberto. Na verdade, a necessidade de um licenciamento mais claro já vem sendo discutida em outras pesquisas, a exemplo de Brito et al. (2015) e Matheus, Ribeiro e Vaz (2012) nas suas pesquisas de uso de dados abertos em aplicações.

Em consulta ao repositório dados.gov.br, tem-se que o licenciamento dos dados ali disponibilizados é do tipo Creative Commons (CC), um tipo genérico de licença de dados que permite que eles possam ser compartilhados e adequados a outros usos, incluindo comercial. Isso já facilita sobremaneira a criação de soluções diversas sobre os dados abertos, gerando segurança jurídica quanto às possibilidades de questionamentos futuros sobre os usos das informações tornadas públicas. Em consulta ao repositório portaldatransparencia.gov.br, não foi possível identificar alguma licença de uso dos dados, o que também foi notado em bases espalhadas por sites específicos de órgãos públicos, como a base de medicamentos disponível no site da Anvisa.

Reforçando a questão do licenciamento dos dados abertos, pode-se citar a pesquisa Global Open Data Index⁷, que avalia a abertura de dados em 13 categorias, colocando o Brasil em 12º lugar em uma lista de 122 países avaliados. As categorias são avaliadas com base em nove critérios, por exemplo da existência de licença, da legibilidade por máquina e da atualização. Nessa pesquisa, observa-se que o critério da ausência de licença específica para abertura dos dados é o mais recorrente nas 13 categorias, estando presente em 89% das ocorrências identificadas.

DIMENSÃO DE POLÍTICAS E ESTRATÉGIAS

Para a dimensão de Políticas e Estratégias, os componentes observados decorrem de desdobramentos naturais do arcabouço da dimensão de Regulação. Destacam-se nesse ponto o Plano de Ação Nacional sobre Governo Aberto, a Política de Dados Abertos do Poder Executivo federal e a INDA, executados pelo MPOG, com o intuito de consolidar nos órgãos do governo federal o compromisso institucional de promover a divulgação dos dados de governo.

⁷ O Global Open Data Index é uma pesquisa internacional realizada pela instituição sem fins lucrativos Open Knowledge International, agregando 94 localidades e avaliando o nível de publicação de dados abertos.

O Plano de Ação Nacional sobre Governo Aberto, já na sua segunda versão, foi desenvolvido pelo governo brasileiro e por integrantes da sociedade civil por meio de um grupo de trabalho. Segundo o plano, a sociedade poderia participar de debates virtuais realizados via plataforma on-line, o fórum E-democracia. Além disso, foi realizado encontro presencial entre organizações da sociedade civil e o governo. Foram encaminhadas 32 propostas priorizadas pela sociedade (BRASIL, 2015, p.3).

O Plano de Ação Nacional é subdividido em ações em torno de cinco eixos: Aumento da integridade pública, Melhoria dos serviços públicos, Aumento da responsabilidade corporativa, Criação de comunidades mais seguras e Gestão mais efetiva dos recursos públicos.

Para implementar os objetivos planejados, o governo federal, por meio do MPOG, publicou a Instrução Normativa nº 4, de 12 de abril de 2012, que instituiu a INDA, responsável pela ordenação da política de dados abertos, estabelecendo seu modelo de funcionamento, os padrões técnicos a serem adotados, promovendo compartilhamento de recursos de TICs entre os órgãos e entidades do governo federal. A INDA também estabelece os conceitos básicos, a exemplo dos seguintes termos – dado: sequência de símbolos ou valores, representados em algum meio, produzidos como resultado de um processo natural ou artificial; informação: conjunto de dados organizados de tal forma que tenham valor ou significado em algum contexto;

- i. dado público: qualquer dado gerado ou sob a guarda governamental que não tenha o seu acesso restrito por legislação específica;
- ii. formato aberto: formato de arquivo não proprietário, cuja especificação esteja documentada publicamente e seja de livre conhecimento e implementação, livre de patentes ou qualquer outra restrição legal quanto à sua utilização;

- iii. licença aberta: acordo de fornecimento de dados que conceda amplo acesso para que qualquer pessoa os utilize, os reutilize e os redistribua, estando sujeito, no máximo, à exigência de creditar a sua autoria e compartilhar pela mesma licença;
- iv. dados abertos: dados públicos representados em meio digital, estruturados em formato aberto, processáveis por máquina, referenciados na rede mundial de computadores e disponibilizados sob licença aberta que permita sua livre utilização, consumo ou cruzamento; e
- v. metadado: informação que descreve características de determinado dado, explicando-o em certo contexto de uso (BRASIL, 2012).

Para disciplinar a utilização da INDA em cada órgão, o governo faz uso da Política de Dados Abertos do Poder Executivo federal, que orienta a criação do Plano de Dados Abertos (PDA) para cada unidade e assim promove mecanismos para publicação automática de bases de dados que não estejam enquadradas nas opções legais de sigilo. Os objetivos dessa política são os trazidos a seguir:

- i. promover a publicação de dados contidos em bases de dados de órgãos e entidades da administração pública federal direta, autárquica e fundacional sob a forma de dados abertos;
- ii. aprimorar a cultura de transparência pública;
- iii. franquear aos cidadãos o acesso, de forma aberta, aos dados produzidos ou acumulados pelo Poder Executivo federal, sobre os quais não recaia vedação expressa de acesso;
- iv. facilitar o intercâmbio de dados entre órgãos e entidades da administração pública federal e as diferentes esferas da federação;
- v. fomentar o controle social e o desenvolvimento de novas tecnologias destinadas à construção de ambiente de gestão pública participativa e democrática e à melhor oferta de serviços públicos para o cidadão;

- vi. fomentar a pesquisa científica de base empírica sobre a gestão pública;
- vii. promover o desenvolvimento tecnológico e a inovação nos setores público e privado e fomentar novos negócios;
- viii. promover o compartilhamento de recursos de tecnologia da informação, de maneira a evitar a duplicidade de ações e o desperdício de recursos na disseminação de dados e informações; e
- ix. promover a oferta de serviços públicos digitais de forma integrada. (DECRETO Nº 8.777, DE 11 DE MAIO DE 2016).

Para ajudar na implementação desse plano, o MPOG disponibiliza o Kit de dados abertos. O kit consiste em documentos que descrevem o processo, métodos e técnicas para a implementação de uma política de dados abertos, sendo direcionado para órgãos e entidades da administração pública federal, focando na criação e implementação de seus respectivos Planos de Dados Abertos (PDAs).

Em consulta ao repositório de discussões da INDA, percebe-se que ainda é baixa a implementação do PDA pelos órgãos federais. Como mostra a tabela 1 a seguir, apenas cerca de 7% da meta de 210 órgãos foi atingida, sendo que conforme o Decreto 8.777/2016, o prazo para implementação seria até 11/07/2016.

Tabela 1 – Acompanhamento da implementação de PDAs

Segmento	Meta	Realizado
Administração Direta (ministérios)	24	9
Autarquias e Fundações	85	6
Universidades Federais	63	0
Institutos Federais de Ensino	38	0
Total	210	15

Fonte: INDA - <http://wiki.gtinda.ibge.gov.br/Plano-de-Dados-Abertos.ashx?HL=pda>. Acesso em 10/08/2016.

Essa situação compromete o processo de abertura de dados e enfraquece a própria natureza do ecossistema de dados abertos do governo federal, pois não existe ecossistema sem uma quantidade razoável e diversificada de dados abertos. Ressalta-se que é possível haver dados abertos pelo órgão mesmo sem um PDA formalizado, como é o caso da Agência Nacional de Vigilância Sanitária (Anvisa) e do próprio Ministério da Transparência e Controladoria Geral da União (CGU), que veio liberar seu PDA em 03/08/2016 e já divulga dados abertos desde 2009.

Os *stakeholders* da dimensão de Políticas e Estratégias compreendem a Presidência da República, o MPOG e, principalmente, os órgãos públicos que implementam as estratégias, além da própria sociedade civil, diretamente participando em grupos de trabalhos.

DIMENSÃO DOS DADOS PUBLICADOS

A partir da vinculação normativa e da orientação técnica das dimensões anteriores surge a dimensão de Dados publicados, que, por sua vez, agrega os mecanismos que efetivam a abertura dos dados por parte das diversas entidades do governo federal. Como uma dimensão que depende fortemente de tecnologias (TICs), sua implementação envolve desafios sobre a escolha de modelos, padrões e plataformas de software, bem como investimentos em infraestrutura específica (ATTARD et al., 2015). Para a descrição dessa dimensão, pode-se complementar o modelo de Dawes, Vidasova e Parkhimovich (2016) com o modelo de cinco estrelas de Berners-Lee, que aborda a disponibilização dos dados abertos em um esquema evolutivo de cinco estágios (HAUSENBLAS, 2012). Sua aplicação direta no caso brasileiro pode ser visto em Matheus, Ribeiro e Vaz (2012), no qual é mostrado que o atual estágio de dados abertos no Brasil encontra-se na posição três do modelo, ou seja, os dados, apesar de abertos e disponíveis em diversos formatos, ainda não são passíveis de serem interligados, ou correlacionados uns com os outros, o que permitiria melhor aproveitamento e entendimento por parte das aplicações feitas com essas informações.

Dessa leitura combinada, foram extraídos os seguintes pontos de análise: infraestrutura de disponibilização, formato dos dados e quantidade de conjuntos disponibilizados.

Sobre a infraestrutura de distribuição, no âmbito do governo federal, os dados estão disponíveis publicamente na Internet por meio de diversos repositórios, e os principais componentes desse ponto são mostrados na tabela 2:

Tabela 2 – Listagem dos principais repositórios de dados abertos do ecossistema do governo federal – consulta realizada em 02/08/2016

Endereço	Descrição	Bases disponíveis	API disponível?	Usa web service?	Órgão Mantenedor
dados.gov.br	O Portal Brasileiro de Dados Abertos armazena os metadados e links para os dados que são nele catalogados pelos órgãos e entidades do executivo federal.	1.122	Sim	Não	MPOG
portaldatransparencia.gov.br	O Portal da Transparência do Governo Federal é uma iniciativa da Controladoria-Geral da União (CGU), lançada em novembro de 2004, para assegurar a boa e correta aplicação dos recursos públicos. O objetivo é aumentar a transparência da gestão pública, permitindo que o cidadão acompanhe como o dinheiro público está sendo utilizado e ajude a fiscalizar.	Despesas, Receitas, Convênios, Sanções, Servidores, CEPIM, Imóveis funcionais e Grandes eventos.	Não	Não	MTFC/CGU
www12.senado.leg.br/orcamento/sigabrasil	O SIGA Brasil é um sistema de informações sobre orçamento público, que permite acesso amplo e facilitado ao SIAFI e a outras bases de dados sobre planos e orçamentos públicos, por meio de uma única ferramenta de consulta. Use os links à direita para acessar o sistema mediante acesso livre ou especialista.	Orçamento Público	Não	Não	Senado Federal
Sites específicos de entidades públicas	Dados abertos disponibilizados diretamente pela entidade, relativos a seus processos de negócios	-	-	-	Órgãos públicos, Empresas Públicas, Autarquias, etc.

Fonte:Elaborada pelo autor (2016).

Cabe destacar que o repositório dados.gov.br é tido como repositório oficial de dados abertos do governo federal, disponibilizando os dados seguindo as orientações da INDA e conforme a Política de Dados Abertos do Poder Executivo federal. Isso não impede que outros repositórios existam, como o é o caso dos portais de transparência, que possuem como base normativa a Lei de Responsabilidade Fiscal. Em especial, o dados.gov.br utiliza a plataforma CKAN (Comprehensive Knowledge Archive Network), acompanhando a tendência de outras nações em disponibilizar seus dados seguindo uma estrutura tecnológica já desenvolvida para esse objetivo, conforme visto em Attard et al., (2015), Qanbari, Rekabsaz e Dustdar (2015) e Maksimović, Veljković e Stoimenov (2011).

Sobre o formato dos dados, vale lembrar que sua importância consta na própria estrutura do modelo de Berners-Lee que, na verdade, marca o avanço dos estágios pela sucessão de diferentes formatos, partindo-se da abordagem que emprega qualquer formato (estágio 1) até a que utiliza formatos de total interligação (estágio 5), permitindo que tanto sistemas como usuários consigam tirar proveitos dos dados e seus relacionamentos com outros dados.

Com base na infraestrutura dos repositórios, a disponibilização dos dados pode se concretizar de diferentes formas:

(1) diretamente no portal acessando as bases e escolhendo um dos formatos (arquivo texto, xml, json, etc) e fazendo o *download* dos dados diretamente. No caso do dados.gov.br, os formatos mais comuns são CSV (arquivo texto separado por vírgulas), HTML (arquivo para ser exibido em navegadores de páginas internet), gejson e KML (ambos formatos para uso em aplicações de georeferenciamento). Outro exemplo é a possibilidade de consulta direta da informação de forma granular, sem baixar um conjunto de dados completo, como acontece com as opções do Portal da Transparência;

(2) de forma automatizada, por meio de outra aplicação que utiliza os dados do portal, disponibilizados por meio de sua API (Application Program Interface) específica. Uma técnica que auxilia nesse sentido é o *mashup*, que representa uma tecnologia que agrega diversas fontes de dados em uma solução única, quer seja uma página de Internet ou uma aplicação para celular, utilizando também o recurso de API. Uma API significa que o portal exterioriza seu acesso a dados para que outras aplicações possam utilizar, provendo de forma mais automatizada os dados para outra aplicação. Nessa forma o provedor cria pontos de consulta para que outras soluções de software possam ser desenvolvidas usando os dados abertos, promovendo assim o reuso da informação pública. Essa forma de acesso automatizado aos dados representa o potencial para criação de novas aplicações a partir dos dados existentes. Nota-se que, com base na tabela com a listagem dos principais repositórios (tabela 2), apenas o dados.gov.br possui API nativa.

Outro recurso que facilita o desenvolvimento de novas soluções a partir dos dados existentes é a utilização de web services, que na verdade é um tipo específico de API, baseada fortemente em tecnologias de Internet. Seu uso torna o processo de manutenção das aplicações mais fácil, incluindo a tarefa de atualização das bases. Nos repositórios levantados percebe-se que nenhum deles disponibiliza o recurso de *web service* para acesso aos seus dados.

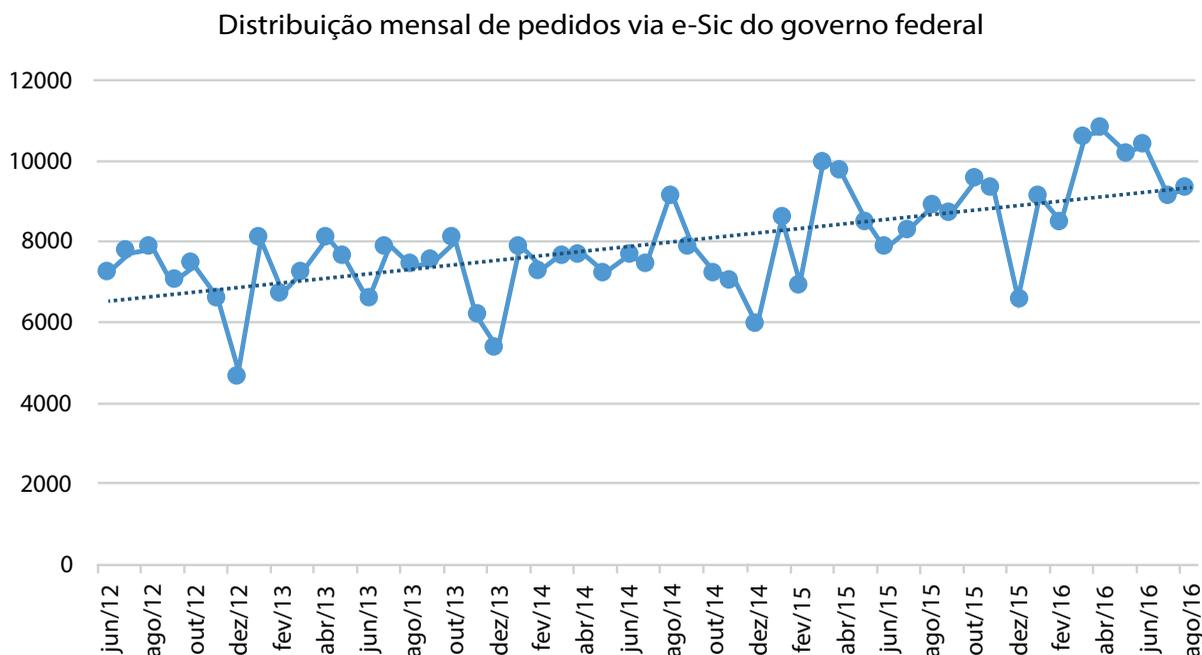
Percebe-se que essa infraestrutura de disponibilização de dados pela Internet possibilita também dois tipos básicos de consumo:

- i. consumo intermediário (automatizado – APIs, web services) – como é o caso típico do portal dados.gov.br, que usa plataforma específica e disponibiliza como padrão o conjunto inteiro de dados para reutilização, exigindo do usuário conhecimentos mais técnicos como tipos de formatos, tabelas e dicionários de dados. Convém destacar nessa forma a experiência inovadora do TCU com a implantação da Nuvem Cívica, que contém dados abertos tratados e disponibilizados para uso automatizado via web services;
- ii. consumo final (direto) – caso geral do Portal da Transparência e outros sites, em que a informação está em formato mais amigável, sendo exibida diretamente na tela de consulta, guiada por parâmetros selecionáveis pelo usuário. Outra fonte direta também são os vários Sistemas Eletrônicos do Serviço de Informação ao Cidadão (e-SICs) espalhados pelos diversos órgãos da administração federal, onde o usuário pode solicitar informação específica de um contexto do órgão. Essas solicitações chegaram a 405.849 pedidos no período de 2012-2016, média mensal de 7.804.

Cabe destacar que o consumo via e-SIC implica que o dado não está disponível naturalmente, sendo necessário provocar o órgão gestor para disponibilizar a informação desejada, que é respondida diretamente ao cidadão que a solicitou. Logo, quando se tem os dados disponíveis, não se precisa necessariamente provocar o gestor para obter alguma informação, pois ela já foi disponibilizada, sendo que essas solicitações diretas servem como indicativo de quão abertas estão as bases governamentais.

Assim, como mostrado no gráfico 1, a seguir, nota-se tendência de crescimento nesses pedidos, o que indica, naturalmente, que a abertura dos dados ainda não supre totalmente as necessidades, sem embargo do aumento natural de interessados, o que é normal a partir do conhecimento de que essas informações existem:

Gráfico 1 – Distribuição mensal de pedidos de e-SIC – governo federal



Fonte:Elaborado pelo autor (2016), com base nos dados retirados de http://esic.cgu.gov.br/sistema/site/relatorios_estatisticos.html em 11/09/2016.

Ainda sobre a quantidade de dados disponíveis, ressalta-se que, nos diversos repositórios de órgãos federais observados, não é possível precisar uma quantidade objetiva de conjuntos de dados disponibilizados. Apenas no dados.gov.br há diretamente a referência à quantidade de conjuntos de dados encontrados, chegando, no momento da escrita deste trabalho, a 1.122 conjuntos. Na tabela 2, mostrada anteriormente, apresenta-se em uma das colunas resumo desses quantitativos.

Ressalta-se ainda que os dados abertos existem ainda de forma dispersa em diferentes repositórios públicos na Internet, embora haja o repositório oficial. Essa dispersão de fontes também se replica nos formatos e na dificuldade de se mensurar a real magnitude do montante de informações tornadas abertas.

DIMENSÃO DOS INFOMEDIÁRIOS

Por sua vez, a dimensão dos Infomediários, caracterizada como elemento intermediador do acesso aos dados abertos, posicionando-se entre o fornecedor do dado bruto e o consumidor final, assume papel de elevada importância na produção de tecnologias que utilizam os dados disponíveis, seguindo a linha do que é discutido em Janssen e Zuiderwijk (2014).

Para facilitar a análise desta dimensão, complementa-se o modelo teórico com as ideias de Magalhães, Roseira e Strover (2013), que apresentam elementos para uma terminologia aplicada aos intermediários que atuam com dados abertos. Nesse trabalho, foram identificadas três categorias de intermediários: *startups* cívicas, serviços de dados abertos e infomediários.

Como *startup* cívica, têm-se organizações, geralmente em estágio inicial, que atuam criando soluções para uso da comunidade, seguindo modelos de negócios próprios do mercado de TIC. No contexto brasileiro não foram identificados nos mecanismos de busca da Internet muitos exemplares dessa categoria, destaques para as empresas AppCívico, Núcleo Digital e Hacklab.

A AppCívico atua disponibilizando serviço de nuvem (*cloudservices*) para aplicações cívicas, sua atuação está fortemente baseada no contexto da cidade de São Paulo, com os projetos Observa Sampa, Eu Voto, De Olho nas Metas e Programa Cidades Sustentáveis, não sendo observados projetos específicos com dados do governo federal. Esse enfoque na cidade de São Paulo também foi verificado nos projetos tocados pelo Núcleo Digital e pelo Hacklab, como os projetos Gestão Urbana SP e Catraca Livre, respectivamente.

Como serviços de dados abertos enquadram-se as soluções criadas por organizações públicas ou privadas, geralmente já consolidadas no mercado, utilizando dados abertos e abertas ao acesso público. Como exemplo na esfera privada, observou-se que a Fundação Getúlio Vargas e o projeto Transparência Orçamentária apresentam de forma gratuita e inovadora os dados abertos do Orçamento Federal, com destaque para as funcionalidades Mosaico Orçamentário e Geologia Orçamentária. Ambas usam fortemente visualizações gráficas para mostrar a evolução do gasto público por função ou órgãos. Como exemplo na esfera pública pode-se citar os serviços de informação do site do Instituto de Geografia e Estatística (IBGE) e do Instituto de Pesquisa Econômica e Aplicada (Ipea).

Como infomediários, detalhando mais a categorização já mostrada, podem ser identificadas as seguintes subcategorias: jornalistas investigativos, hackers cívicos, organizações de apoio e movimentos cívicos. Como exemplos desses grupos têm-se a Open Knowledge Brasil (OKF), o grupo Transparência Hacker - Thackday, o grupo da Infraestrutura Nacional de Dados Aberto (INDA) e a Fundação Avina. Os três primeiros grupos comunicam-se por meio de ferramentas gratuitas de colaboração, como o *googlegroups*, e a ferramenta de mensagens instantâneas Telegram, sendo observados dois grupos com mais de 300 usuários cadastrados, discutindo temas sobre dados abertos, suas ferramentas e abordagens, além de troca de material.

A Fundação Avina, em especial, enquanto instituição filantrópica que atua na América Latina desde 1994, tem promovido a colaboração para a sustentabilidade, focando também em aplicações cívicas. Um projeto dessa fundação que se destaca é o Tecnologia para Mudança Social, que promove o uso de tecnologias cívicas, apoiando a agenda de dados abertos governamentais, a exemplo do portal focado na cidade de São Paulo. Destaca-se que não foram detectados projetos específicos para dados do governo federal. Outro exemplo de serviço é a biblioteca de aplicativos cívicos. Nesse serviço é possível navegar por aplicativos de diversos países latino-americanos, sendo que no caso do Brasil, há mais aplicações de ecossistemas locais, como da cidade do Rio de Janeiro e São Paulo.

Outro exemplo de produção, no contexto do governo federal, envolvendo hackers cívicos, é o projeto Serenata de Amor, que possui como foco divulgar dados sobre a verba indenizatória (Câmara dos Deputados - CD), trazendo a tecnologia para ajudar no combate à corrupção. Em mais detalhes, o projeto facilita o acesso às informações que a CD disponibiliza abertamente, mas em formato não muito acessível pelo usuário comum, pois tratam-se de arquivos grandes (mais de 300MB) e difíceis de serem abertos por programas comuns de edição de texto. Os arquivos grandes são compilados em versões menores (um décimo do tamanho original) e mais fáceis de serem consultados por qualquer interessado em acompanhar esses gastos. Os autores chamam a atenção para um detalhe importante no processo de abertura de dados, o fato de que disponível é diferente de acessível, pois o formato e o tamanho dos arquivos podem dificultar o consumo em virtude das limitações da computação pessoal. Cabe destacar também que esse projeto conta com modelo de negócio baseado em doações on-line (via Cartase), aceitando também Bitcoins.

Após a descrição dos principais componentes de infomediários, convém comentar como se dá o processo de comunicação entre essa dimensão e a dimensão anterior de dados publicados.

Na dimensão de dados publicados, as entidades públicas disponibilizam seus dados para o consumidor acessar diretamente ou para os intermediários da dimensão de infomediários utilizarem para suas soluções. Nesse processo, destaca-se a promoção de desafios ou concursos de ideias para estimular a criação de soluções sobre os dados disponibilizados. Um dos tipos mais comuns desses concursos é o hackathon⁸, que corresponde a uma designação genérica para eventos de tecnologia em que vários profissionais colaboram na produção de soluções com foco definido, com fins sociais ou educacionais.

No presente contexto, os *hackathons* são geralmente promovidos pelo próprio governo, representando uma fonte clássica para o incentivo à comunidade de desenvolvedores e ativistas ou hackers cívicos na busca de soluções para determinados problemas ou desafios a partir de dados abertos. A orientação do governo é que o PDA já proponha uma ação de comunicação junto à sociedade, como mostra o seguinte trecho extraído do repositório sobre Governo Aberto:

De todo modo, o Plano de Dados Abertos recomenda uma ação de comunicação conjunta sempre que uma organização publica um conjunto de dados. As principais sugestões são encontros abertos reunindo possíveis interessados e a realização de concursos de aplicativos (“Hackathons”). Os concursos de aplicativos são fundamentalmente ações promocionais, mas que muitas vezes criam soluções inovadoras que podem ser utilizadas pelos próprios gestores para melhor visualização dos dados de suas políticas públicas. Algumas organizações planejam inclusive incorporar alguns aplicativos criados em concursos.

Em levantamento feito para este trabalho, foi possível observar que nove *hackathons* no período de 2013 a 2016 foram promovidos por entidades do governo federal, como pode ser visto na tabela 3, a seguir:

⁸ O termo hackathon refere-se a um tipo de competição que pode durar horas ou dias e que possui como objetivo a construção de soluções computacionais de forma colaborativa. Disponível em: <<https://en.oxforddictionaries.com/definition/hackathon>>. Acesso em: 20 abr. 2018.

Tabela 3 – Lista de Hackathons – governo federal – 2013 - 2016

Hackathon	Ano	Entidade Promotora	Quantidade de Projetos/ Participantes
Hackathon de Transparência legislativa e participação popular	2013	Câmara dos Deputados - LABHacker	24
Concurso Nacional de Dados Abertos	2013	MPOG	15
I hackathon de dados educacionais	2013	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – INEP	30
II hackathon de dados educacionais	2014	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – INEP	40
Hackathon de Gênero e Cidadania	2014	Câmara dos Deputados - LABHacker	75
INOVAApps	2014	Ministério das Comunicações	865
INOVAApps	2015	Ministério das Comunicações	527
Hackathon sobre Participação no Combate à Corrupção	2016	Ministério da Justiça - LapPI	49
Controle Social Digital 2016	2016	Tribunal de Contas da União - TCU	Em andamento...
Total			1.625

Fonte:Elaborado pelo autor (2016).

Nota-se na tabela 3 que esses eventos têm tido nível de participação considerável, tendo em vista a quantidade de projetos submetidos (1.625 em três anos), contando com vários tipos de premiações, que vão do apoio logístico (*hackathon* de dados educacionais) até a liberação de recursos financeiros para o desenvolvimento de projetos (INOVAApps), sendo os projetos desenvolvidos ou iniciados no momento do próprio concurso ou posteriormente. Como se trata de um concurso, ao final do evento são escolhidos os projetos que receberão atenção para seu desenvolvimento, sendo que, teoricamente, a ideia é que seu potencial de agregar valor aos dados abertos aconteça não somente no momento da competição.

Dos nove concursos apresentados, verificou-se que, dos 1.625 projetos, foram selecionados 225 como finalistas e aptos a desenvolvimento final. Excluindo-se os vencedores do *Hackathon* sobre Participação no Combate à Corrupção, que ainda está em fase de finalização, chega-se a 186 soluções baseadas em dados abertos e patrocinadas por entidades do governo federal, via *hackathons*.

Das 186 soluções, observou-se que 49% (90) não estão mais disponíveis para acesso, conforme consulta aos repositórios oficiais das plataformas Android, iTunes, para aplicações, e Google, para os casos de páginas web. Das 95 disponíveis para acesso (51%), 32 (17%) não estavam atualizadas, exibindo ainda os dados originais da época do concurso.

Assim, restam 33% das soluções produzidas disponíveis e atualizadas, ou seja, apenas um terço dessas aplicações sobrevivem aos primeiros anos pós-*hackathon*. Em outra perspectiva, pode-se dizer que o índice de mortalidade de um app cívico é de 67%, ou de cada três, dois não conseguem sobreviver aos primeiros anos.

DIMENSÃO DOS CONSUMIDORES

A dimensão de Consumidores representa a visão da utilização das soluções de dados abertos pela sociedade como um todo, indo do simples acesso até a solução de problemas da comunidade e benefícios advindos disso (DAWES; VIDIASOVA; PARKHIMOVICH, 2016). Para descrição dessa dimensão, faz-se necessário diferenciar três tipos de consumidores: (i) os consumidores diretos dos dados, acessando as informações diretamente da fonte; (ii) os consumidores infomediários, tratados na dimensão anterior, que acessam os dados para produzir novas aplicações e, a partir delas, disponibilizar novamente os dados para o consumidor direto e (iii) os consumidores diretos dos dados via aplicações geradas pelos infomediários. Dos três tipos, derivam as seguintes relações de consumo: (i) consumidor final e dados abertos; (ii) consumidor infomediário e dados abertos; e (iii) o consumidor final e aplicações geradas pelo consumidor infomediário.

Para analisar as duas primeiras relações, foram tomados dois conjuntos de estatísticas de acesso dos dois principais repositórios de dados abertos do governo federal: o dados.gov.br e portaldatransparencia.gov.br, exibidos na tabela 4, a seguir:

Tabela 4 – Quantidade de acessos mensais por repositório

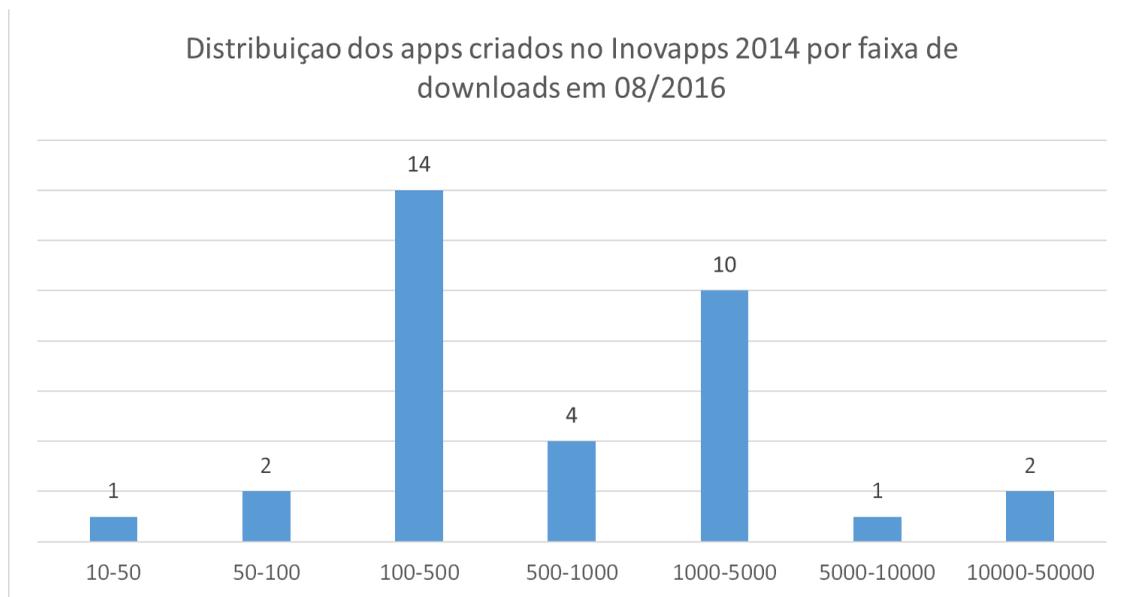
Mês/Ano	Portal da Transparência	dados.gov.br
mai/15	1.439.877	74.613
jun/15	1.374.952	74.837
jul/15	1.362.421	70.482
ago/15	1.427.392	79.670
set/15	1.376.790	79.035
out/15	1.374.348	77.726
nov/15	1.343.182	68.924
dez/15	1.382.915	51.336
jan/16	1.280.412	62.422
fev/16	1.246.294	67.598
mar/16	1.537.466	118.672
abr/16	1.465.138	108.782
Média	1.384.266	77.841
Total	17.995.453	1.011.938

Fonte: Elaborado pelo autor (2016), posição 08/2016.

A superioridade da quantidade de acessos do Portal da Transparência sobre o portal dados.gov.br pode ser atribuída à natureza diferenciada que esses dois conjuntos de dados possuem. O primeiro apresenta interface mais amigável, destinando-se ao usuário final, além de trazer consultas diretas a temas sensíveis, como é o caso da consulta dos vencimentos de servidores federais diretamente pelo nome e da consulta de recursos transferidos a municípios. Essa é uma solução tipicamente voltada para o consumidor direto, que busca a informação pronta direto na fonte.

Já o outro portal possui formatação mais voltada para infomediários, o segundo tipo de consumo, para quem deseja desenvolver soluções a partir dos dados abertos, sendo a interface menos amigável ao usuário final, como mostra o uso de termos genéricos, como conjuntos de dados e recursos e a disponibilização em formatos que não fazem parte do senso comum, como json, kml e xml.

Gráfico 2 – Distribuição de aplicativos por faixa de downloads



Fonte:Elaborado pelo autor (2016).

Nota-se que esses dois enfoques não podem ser tidos como rivais, mas sim complementares, representando a diversidade que caracteriza naturalmente qualquer ecossistema de dados abertos.

Em outro levantamento para essa dimensão, focando-se agora na terceira relação (consumidor final e aplicações geradas por infomediários) em que os consumidores diretos acessam as aplicações produzidas por infomediários, foram catalogadas as quantidades de downloads dos aplicativos vencedores do *hackathon* Inovapps 2014, conforme mostrado no gráfico 2. Como pode ser notado, 50% desses aplicativos estão abaixo de 500 instalações (*downloads*), o que representa uma parcela ínfima perto do espaço de usuários de 53 milhões de pessoas. O fato mostra que a utilização de aplicações de dados abertos é um ainda um desafio a ser enfrentado pelos desenvolvedores desse tipo de solução.

CONSIDERAÇÕES FINAIS E CONCLUSÕES

Para uma apropriação valorativa do que foi apresentado até então, resgatam-se as já comentadas propriedades básicas de ecossistemas de dados abertos identificadas por Heimstädt, Saunderson e Heath (2014), consideradas como critérios de maturidade: capacidade de retroalimentação; sustentabilidade, fornecimento de dados baseado por demanda da sociedade e interdependência entres os diversos atores (gestores de dados, infomediários e consumidores). Passa-se então a considerar como cada critério pode ser percebido com base nas dimensões e componentes identificados:

- capacidade de retroalimentação: para esse critério, entende-se que os dados abertos ainda estão fluindo em sentido único, ou seja, da Dimensão de Dados Publicados para a Dimensão de Consumidores, não sendo percebida nenhuma iniciativa que produzisse informações no sentido inverso, retroalimentando as instâncias de governo e caracterizando o aspecto cíclico da informação.

A Dimensão de Consumidores apropria-se dos dados abertos ou das soluções de infomediários em um modo de leitura em que as informações são apenas exibidas sem produção ou agregação de novos dados para posterior reenvio. Nota-se que a Dimensão de Infomediários, por meio de suas interações em competições de aplicativos ou mesmo via demandas de e-Sic ou ouvidoria, pode sensibilizar a Dimensão de Dados Publicados, caracterizando algum nível de retroalimentação. Assim, entende-se que esse critério não está totalmente atendido;

- b. sustentabilidade: aqui o governo é ainda a espinha dorsal do ecossistema, tanto no aspecto de manutenção do repositório de dados abertos, quanto na capacidade de movimentar a Dimensão de Infomediários, por meio dos desafios de aplicativos e pagamento de premiações. Aqui também se insere a sustentabilidade no sentido mais estrito, ligada à capacidade de sobrevivência das soluções geradas pela Dimensão de Infomediários. Nota-se que o alto índice de mortalidade de aplicativos é prova de que não há sustentabilidade dessas aplicações. Entende-se que não se atende ao critério de sustentabilidade mais geral também;
- c. fornecimento de dados baseado em demandas da sociedade: a partir da Dimensão de Políticas e Estratégias, percebe-se que um texto padrão da maioria dos PDAs publicados é que deverá ser usado como um critério para divulgação de dados “o grau de relevância para o cidadão, observando-se as demandas encaminhadas via eSIC, bem como os setores e serviços mais procurados nos sítios eletrônicos da entidade”. Ressalvas para o PDA do Ministério da Saúde, que não traz explicitamente essa diretriz. Entende-se que, pelo menos no nível de compromisso firmado nos planos, há direcionamento para abertura de dados seguindo as demandas da sociedade (refletidas nos pedidos via e-SIC). Contudo, como a formalização desses planos está baixa (7%), acredita-se que ainda não se pode afirmar

que os dados publicados são orientados pelo consumidor final das informações. A ideia desse critério é que as informações abertas seguissem a demanda dos agentes do ecossistema e não somente fossem “empurradas” pelos órgãos gestores. Entende-se assim que esse critério resta parcialmente atendido;

- d. interdependência entre os diversos atores: aqui percebe-se que há certo nível de interdependência entre os agentes das diversas dimensões: órgãos gestores, infomediários e consumidores finais. Prova disso são os diversos momentos em que órgãos e infomediários se apresentam nos desafios de aplicativos cívicos e quando consumidores finais acessam diretamente seus dados ou aplicações. Destaca-se que essa relação é em sentido único, carecendo de retroalimentação, como já mostrado em critério específico.

A partir das descrições das dimensões e comparação com os critérios de maturidade, percebe-se que o ecossistema de dados abertos do governo federal está em estágio de desenvolvimento, sua existência é fato, mas seu nível de maturidade ainda é iniciante, carecendo de melhorias e tendo que lidar com desafios estruturais. Em especial, como ter assegurada a sustentabilidade das soluções criadas, e do próprio ecossistema, sem depender eternamente das iniciativas governamentais, embora sejam necessárias em um primeiro momento para mostrar o valor social e o potencial para modelos de negócio.

Esses desafios também são refletidos nas dimensões analisadas. Assim, entende-se que mesmo com um aspecto regulatório definido, sua adoção é lenta e gradual, principalmente quando se pensa num país como o Brasil, em que o pacto federativo é um desafio e a União, representada pelo governo federal, assume papel de destaque na criação e financiamento de políticas públicas.

Convém ressaltar que grande parte das pesquisas acadêmicas sobre ecossistemas de dados abertos foca em cidades, a exemplo de Londres (HEIMSTÄDT; SAUNDERSON; HEATH, 2014), de Boston (MAHONEY, 2013), de Chicago (O'NEIL, 2013), de Nova Iorque e de São Petersburgo (DAWES; VIDIASOVA; PARKHIMOVICH, 2016). No Brasil, segundo a bibliografia pesquisada, assumem destaque os ecossistemas de poucas cidades, como é o caso de São Paulo, Rio de Janeiro e Recife (BRITO et al., 2014a).

O enfoque do ecossistema em localidades pode ser atribuído à maior proximidade das administrações locais com a sociedade, permitindo o enfrentamento direto de problemas do dia a dia, na linha da proposta do movimento americano Code for America e seu sistema de brigadas, espalhando por cidades a filosofia de construção de soluções tecnológicas por meio de dados abertos. Percebe-se que a organização do próprio Estado impacta na forma de concepção e análise de ecossistemas de dados abertos, sendo um dos desafios do governo federal brasileiro fazer chegar ao nível local as experiências bem-sucedidas de dados abertos, alcançando diretamente a população e permitindo visualizar de forma mais clara a correlação da utilização dos dados abertos e seus benefícios para as pessoas.

Em especial, com relação à dimensão de dados publicados, percebe-se o problema do reduzido número de conjuntos de dados, enquanto o repositório data.gov.uk (Reino Unido) possui 38.487 conjuntos de dados e o data.gov (EUA) 185.698, o dados.gov.br (Brasil) possui 1.122 (acesso em 20/08/2016), o que indica lenta e gradual abertura de dados por parte dos órgãos públicos federais, sendo que mais esforços são necessários para deixá-los no formato comum para permanecer no repositório, atendendo às normas da INDA. Destaca-se que essa questão do número reduzido de dados abertos disponíveis já vem sendo apontada por outras pesquisas no contexto brasileiro, a exemplo de Brito et al. (2015) e Matheus, Ribeiro e Vaz (2012).

Outro problema detectado é a baixa disponibilização de APIs e *web services*, o que eleva sobremaneira o trabalho de manutenção dos dados para reutilização, demandando esforço manual para baixa, limpeza e carga nas aplicações desenvolvidas, o que poderia ser evitado se os repositórios disponibilizassem formas mais automatizadas de acesso aos dados, tanto na totalidade como em nível mais granular da informação.

Na dimensão de infomediários, a questão não poderia ser diferente, e apesar da existência dos *hackathons*, que promovem o primeiro contato e comunicação direta com a comunidade, ainda se percebe pouco envolvimento e incipiente produção de tecnologias consumindo dados abertos; as poucas comunidades que se formaram ainda discutem a abertura de dados; e suas formas de divulgação, como mostra a observação direta nos grupos de discussão pelas ferramentas Telegram e Google Groups. Observou-se que a temática ainda não venceu o estágio inicial da preocupação com a abertura dos dados, o que aparenta ser natural, tendo em vista o já discutido reduzido número de conjuntos de dados disponíveis no repositório oficial. Como já apontado por Dawes, Vidasova e Parkhimovich (2016), a ação dos infomediários se mostra crítica para o desenvolvimento e sustentabilidade de um ecossistema de dados abertos.

Em resumo, sugere-se a novas pesquisas sobre o ecossistema de tecnologias cívicas procurar detalhar as diversas dimensões e aspectos apresentados, promovendo debate sobre novas estratégias para as soluções cívicas, incluindo *gamefication* e Internet das coisas, sem esquecer de testar novas abordagens de compreensão da dimensão dos infomediários, que se mostra como peça chave no ecossistema, capaz de disparar o real potencial dos dados abertos no concurso de soluções de problemas sociais.

Ratificam-se as contribuições no sentido de promover também o debate quanto ao papel institucional do governo federal não só no provimento puro de dados abertos, mas sim em todos os aspectos do ecossistema, incluindo o apoio ao desenvolvimento das soluções cívicas, a melhoria dos dados e a comunicação com os consumidores finais. Na verdade, acredita-se em uma convergência da noção de Governo Eletrônico e de Governo Aberto, focando na prestação de serviços melhorados por meio de tecnologias desenvolvidas por infomediários e acessadas pela sociedade, com uso intensivo das tecnologias já tornadas comuns atualmente.

Esse caminho de digitalização do Estado é sem volta e acompanha a própria tendência da sociedade que se faz cada vez mais social em redes digitais. A transparência, o aumento da participação social e do consumo de serviços digitais públicos são favorecidos por tecnologias da informação e representam uma evolução natural desse processo de digitalização.

REFERÊNCIAS

ALBANO, C. S. *Dados governamentais abertos*: proposta de um modelo de produção e utilização de informações sob a ótica conceitual da cadeia de valor. 2014. 217 f. Tese (Doutorado em Administração) - Universidade de São Paulo, 2014.

BRASIL. Portal Brasileiro de Dados Abertos. *Instrução Normativa da INDA*. Disponível em: <<http://dados.gov.br/instrucao-normativa-da-inda/>> Acesso em: 17 jul. 2016.

BRASIL. *2º Plano de Ação Brasileiro*. Disponível em: <<http://www.governoaberto.cgu.gov.br/no-brasil/planos-de-acao-1/2o-plano-de-acao-brasileiro/2o-plano-de-acao-brasileiro>>. Acesso em: 17 jul. 2016.

BRASIL. *Guia de Aplicativos do Governo Federal*. Disponível em: <<http://www.aplicativos.gov.br/>>. Acesso em: 25 ago. 2015a.

BRASIL. Ministério do Planejamento, Orçamento e Gestão. Secretaria de Tecnologia da Informação. Departamento de Governo Digital. *Padrões de Interoperabilidade de Governo Eletrônico*. Brasília, 2017. Disponível em: <<http://eping.governoeletronico.gov.br/>>. Acesso em: 11 set. 2016.

BRASIL, *Programa de Governo Eletrônico Brasileiro* - Sítio Oficial. Disponível em: < <http://www.governoeletronico.gov.br/>>. Acesso em: 13 out. 2015b.

BRITO, K. S. et al. Is Brazilian Open Government Data Actually Open Data?: An Analysis of the Current Scenario. *International Journal of E-Planning Research*, v. 4, n. 2, p. 57–73, 2015.

DAWES, S. S.; VIDIASOVA, L.; PARKHIMOVICH, O. Planning and designing open government data programs: An ecosystem approach. *Government Information Quarterly*, v. 33, n. 1, p. 15–27, 2016.

EYLER-WERVE, K.; CARLSON, V. *Civic Apps Competition Handbook*. [S.l]: O'Reilly Media, Inc., 2012.

GERMANO, E. C. *Modelos de negócios adotados para o uso de dados governamentais abertos*: estudo exploratório de prestadores de serviços na cadeia de valor dos dados governamentais abertos. 2013. 169 f. Dissertação (Mestrado em Administração) - Universidade de São Paulo, 2013.

JANSEN, S.; FINKELSTEIN, A.; BRINKKEMPER, S. A sense of community: A research agenda for software ecosystems. In: INTERNATIONAL CONFERENCE ON SOFTWARE Engineering – ICSE, 31., 2009, Vancouver. *Proceedings...Canada*: IEEE, 2009. Disponível em: <http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5070978>. Acesso em: 2 jul. 2015.

LEE, D. *Building an open data ecosystem*: an Irish experience. In: INTERNATIONAL CONFERENCE ON THEORY AND PRACTICE OF ELECTRONIC GOVERNANCE, 8., 2014, Portugal. *Proceedings... New York*: ACM, 2014. Disponível em: <<http://dl.acm.org/citation.cfm?id=2691258>>. Acesso em: 29 nov. 2015

OPEN GOVERNMENT PARTNERSHIP - OGP. *Declaração de Governo Aberto*. 2011. Disponível em: <<http://www.governoaberto.cgu.gov.br/central-de-conteudo/documentos/arquivos/declaracao-governo-aberto.pdf>>. Acesso em: 13 out. 2015.

OPEN KNOWLEDGE INTERNATIONAL - OKF. *The Open Data Handbook*. Disponível em: <<http://opendatahandbook.org/>>. Acesso em: 13 out. 2015.

O'REILLY, T. *Government as a Platform*. *Innovations: Technology, Governance, Globalization*, v. 6, n. 1, p. 13–40. 2011.

THE WORLD BANK. *Open Government Data Toolkit*. Disponível em: <<http://opendatatoolkit.worldbank.org/en/index.html>>. Acesso em: 9 dez. 2015.

VELJKOVIĆ, N.; BOGDANOVIĆ-DINIĆ, S.; STOIMENOV, L. Benchmarking open government: An open data perspective. *Government Information Quarterly*, v. 31, n. 2, p. 278-290, 2014.

ZUIDERWIJK, A.; JANSSEN, M.; DAVIS, C. Innovation with open data: Essential elements of open data ecosystems. *Information Polity*, v. 19, n. 1-2, p. 17-33, 2014.