

A Lei de Acesso à Informação brasileira e os desafios tecnológicos dos dados abertos governamentais

Louise Pedroso, Asterio Tanaka, Claudia Cappelli

Programa de Pós-Graduação em Informática
Universidade Federal do Estado do Rio de Janeiro (UNIRIO)
Rio de Janeiro, RJ – Brazil

{louise.pedroso, tanaka, claudia.cappelli}@uniriotec.br

Abstract. *The Brazilian Law of Information Access determines the spontaneous publicity (active transparency) of diverse information by public institutions. The law has filled a social political, historic gap in the presence of international initiatives and agreements, but it exposes the lack of technical maturity in open government data. In fact, the execution of the law is being done amid very recent R&D efforts around themes from open formats and interoperability standards to linked data in the context of the Semantic Web. This position paper presents research perspectives and technological challenges in the area of open government data and transparency.*

Resumo. *A Lei de Acesso à Informação (LAI) determina a divulgação espontânea (transparência ativa) de informações diversas por instituições públicas. A LAI veio a preencher uma histórica lacuna político-social, mediante iniciativas e acordos internacionais, porém, ela evidencia a falta de maturidade técnica em dados abertos governamentais. De fato, a implantação da LAI vem ocorrendo em meio a esforços muito recentes de P&D sobre temas que vão de formatos abertos e padrões de interoperabilidade a dados ligados no contexto da Web Semântica. Este artigo de posicionamento apresenta perspectivas de pesquisa envolvendo a LAI e os desafios tecnológicos na área de dados abertos governamentais e transparência.*

1. Introdução

Nos últimos anos, seguindo o movimento de controle social e transparência pública que cresce em todo o mundo, o Brasil tem investido em políticas de acesso à informação. Não por acaso, o país integra a Parceria para Governo Aberto, firmada entre governos que se comprometem a manter uma agenda em questões como disponibilidade de informações, participação do cidadão, integridade profissional na administração e acesso a novas tecnologias para abertura e responsabilidade final, a qual implicou na criação da Infraestrutura Nacional de Dados Abertos.

É neste contexto que a Lei nº 12.527, conhecida como Lei de Acesso à Informação – LAI, sancionada em 18/11/2011, entrou em vigor em 16/5/2012. Com a LAI, o Brasil deu mais um passo na direção da transparência pública e da consolidação da democracia, além de iniciativas anteriores de controle social, como o Portal da Transparência do Governo Federal – PTGF.

De acordo com a LAI e seu regulamento (Decreto nº 7.724, de 16/5/2012), “é dever dos órgãos e entidades públicas promover, independentemente de requerimentos, a divulgação em local de fácil acesso, no âmbito de suas competências, de informações de interesse coletivo ou geral por eles produzidas ou custodiadas”. Um conjunto mínimo

de informações deve ser divulgado de forma espontânea e proativa por órgãos e entidades públicas, no que é denominado de transparência ativa. Estes instrumentos legais obrigam a divulgação dessas informações através de publicação em sítios oficiais na Internet e, quanto ao formato, estabelecem os seguintes requisitos:

- possibilitar a gravação de relatórios em diversos formatos eletrônicos, inclusive abertos e não proprietários, tais como planilhas e texto, de modo a facilitar a análise das informações;
- possibilitar o acesso automatizado por sistemas externos em formatos abertos, estruturados e legíveis por máquina;
- divulgar em detalhes os formatos utilizados para estruturação da informação.

O Guia para Criação da Seção de Acesso à Informação nos Sítios Eletrônicos dos Órgãos e Entidades Federais [CGU, 2012] define o termo “formato aberto” como sendo uma “especificação publicada para armazenar dados digitais, livre de limitações legais no uso”, ou seja, restringindo-se ao aspecto não proprietário do formato e suas implicações legais. O Guia recomenda que as informações estejam disponíveis em formato modificável e que o acesso a elas não dependa da aquisição de um software proprietário; como exemplo, recomenda a preferência à publicação de dados nos formatos CSV, XML e JSON. Entretanto, o Guia aponta para aspectos mais amplos sobre formatos abertos, ao mencionar o Manual de Dados Abertos [W3C, 2011] e o [e-PING – Padrões de Interoperabilidade de Governo Eletrônico](#).

Este artigo explora os aspectos tecnológicos dos dados abertos governamentais, e suas implicações técnicas na implantação da LAI, e está estruturado da seguinte forma: a seção 2 apresenta as motivações para as pesquisas decorrentes, bem como trabalhos relacionados; a seção 3 enumera os problemas de pesquisa identificados; a seção 4 aponta propostas de solução e possíveis métodos de pesquisa; e a seção 5 apresenta a conclusão.

2. Motivação e trabalhos relacionados

A definição de dados abertos, adotada em [W3C, 2011], vem da [Open Definition](#): “um dado é aberto se qualquer pessoa for livre para usá-lo, reusá-lo e redistribuí-lo”. Esta definição implica em algumas propriedades essenciais: disponibilidade e acesso, reuso e redistribuição, e participação universal.

Em reforço e em complemento a essas propriedades, os dados governamentais são considerados abertos quando publicados de acordo com oito princípios, elencados como atributos dos dados abertos governamentais [W3C, 2011]: completos, primários, atuais, acessíveis, compreensíveis por máquina, não discriminatórios, não proprietários e livres de licenças. Um aspecto essencial que se busca em quaisquer definições de dados abertos é a interoperabilidade, isto é, a capacidade de sistemas e organizações interoperarem com diferentes conjuntos de dados.

A abertura de dados, como decorrência da LAI, não é uma atividade trivial e, tampouco, uma decorrência natural da implantação dos dispositivos legais. O fato da LAI ter tido uma data de entrada em vigor cumprida estritamente, sob a condução de um órgão de controle rigoroso e abrangente como a CGU, propiciou a implantação bem sucedida no primeiro momento da data de vigência, no âmbito do Poder Executivo da Administração Pública Federal. Entretanto, uma rápida visita às Páginas de Acesso a

Informação dos diversos órgãos e entidades federais, alguns meses após a entrada em vigor da LAI, permite verificar a inobservância de vários requisitos de dados abertos governamentais, como mencionado. São gritantes as inadequações de formatos, que impedem a interoperabilidade, assim como é evidente a existência de dados incompletos, desatualizados, incompreensíveis por máquina e dependentes de licenças proprietárias, entre outras inconformidades com os preceitos e padrões adotados.

Tais deficiências refletem a ausência de processos organizacionais e da adoção de padrões existentes, assim como a falta de pessoas capacitadas a darem conta das atividades necessárias ao planejamento e implantação dos processos. As organizações, de um modo geral, e os órgãos públicos, em particular, encontram-se em diferentes estágios de maturidade quando se trata de transparência organizacional e dados abertos.

Neste contexto, modelos de maturidade têm sido propostos para classificar e qualificar as organizações quanto ao estágio em que se encontram e que devem almejar.

O esquema de cinco estrelas proposto por Tim Berners-Lee (Tabela 1), para dados abertos ligados [Berners-Lee, 2009], no contexto da Web Semântica, vem sendo proposto como base para um modelo de maturidade em dados abertos.

Tabela 1: Esquema progressivo de 5 Estrelas de Tim Berners-Lee.

☆	Dado disponível na web (em qualquer formato) com licença aberta – OL
☆☆	Dado disponível em formato estruturado legível por máquina (ex: Excel ao invés de imagem digitalizada de uma tabela) – OL+RE
☆☆☆	Formatos não proprietários devem ser utilizados (ex: CSV ao invés de Excel) - OL+RE+OF
☆☆☆☆	Usar padrões abertos do W3C (RDF e SPARQL) para identificar coisas, de modo que pessoas possam apontá-las - OL+RE+OF+URI
☆☆☆☆☆	Ligar os dados a dados de outras pessoas para prover contexto - OL+RE+OF+URI+LD

A atual situação das Páginas de Acesso à Informação sugere que, neste primeiro momento, a estratégia do governo seja mudar a cultura governamental de sigilo das informações para publicidade das mesmas, e efetivamente divulgar as informações, ainda que incompletas ou com alguma outra inconformidade. É o primeiro passo para avançar no esquema de Berners-Lee e na própria abertura de dados governamentais.

Em 2010, Dan Morgan se baseou no esquema de Berners-Lee e nos princípios de dados abertos para lançar uma proposta inicial de um Modelo de Maturidade de Dados Abertos de quatro níveis (Emergente, Praticante, Habilitado, Condutor) apoiado sobre três dimensões: Estratégia e Políticas, Disponibilidade, e Práticas de Descrição e Documentação. Uma característica do modelo é que ele pressupõe que os dados sejam divulgados, de modo que uma instituição que não disponibilize seus dados não se enquadraria no modelo, sendo classificada num nível zero adicional [Morgan, 2010].

Recentemente, o Núcleo de Pesquisa em CiberDemocracia da UNIRIO iniciou o desenvolvimento de um Modelo de Maturidade de Transparência Organizacional [CiberDem, 2013]. A intenção deste trabalho é o uso de um modelo de níveis que defina práticas para implementação das características de transparência e regras de avaliação sobre a implementação destas características. A primeira possibilitará às organizações

definirem o que deve ser inserido em seus procedimentos e políticas para aderência às características de transparência. A segunda irá fornecer mecanismos para que cada organização possa identificar seu nível de maturidade quanto à transparência.

A transparência organizacional envolve tanto processos como informações, portanto é abrangente de forma a indicar práticas que permitam às organizações estarem aderentes aos princípios de dados abertos.

3. Problemas de pesquisa

Além do aspecto legal das falhas no cumprimento da LAI, muitas das questões apontadas na seção anterior não são propriamente problemas de pesquisa, mas de gestão, alguns dos quais de nível estratégico gerencial e outros de nível operacional. No primeiro caso, estão a questão cultural de “propriedade” das informações, e a falta de atenção à gestão dos processos organizacionais e à capacitação dos recursos humanos, que denunciam a imaturidade das instituições públicas em relação à transparência organizacional. Em geral, são de nível operacional os problemas de inconformidade em relação às propriedades desejadas e aos princípios dos dados abertos.

Um dos pontos mais abordados na LAI e no guia da CGU é a orientação para o uso de formatos abertos. Ainda assim, a maioria das páginas de Acesso à Informação apresenta as informações em formato PDF, o que impossibilita a interpretação por máquina. Elas atendem tão somente ao requisito de licença aberta (no caso, do software leitor de PDF), estando na escala inicial no esquema de [Berners-Lee, 2009]. Também encontram-se no nível emergente do modelo de [Morgan, 2010], em que os dados são disponíveis para leitura humana, com contexto e significado dependentes de relatórios.

Um olhar apurado sobre as informações divulgadas até o momento no âmbito da transparência ativa revela que, no caso de convênios, por exemplo, estas ainda não são suficientes para que sociedade, em geral, e o cidadão, em particular, possam efetuar um controle mínimo sobre esses instrumentos. A lista de dados que devem ser divulgados obrigatoriamente, segundo o Guia da CGU, não o é nem mesmo pelo PTGF.

Determinar o que deve ser divulgado sobre um determinado assunto não é algo simples, já que requer a definição de critérios e, por vezes, o conhecimento do domínio da informação. Ainda assim, esses critérios certamente devem incluir os princípios dos dados abertos, principalmente a completude e a primariedade.

No caso dos convênios, foi observado que nem todos os partícipes do convênio são exibidos, impossibilitando a descoberta de todos os convênios em que uma determinada instituição atua. Isso fere, diretamente, o princípio da completude.

A falta de definição dos critérios para que as informações divulgadas sejam consideradas suficientes pode ser explicada, em parte, pela pesquisa de [Nascimento, 2011], que realizou entrevistas com as equipes envolvidas na criação de portais de transparência, inclusive o PTGF. De acordo com os relatos, dado o prazo reduzido imposto pela CGU para o seu desenvolvimento, os cidadãos não foram envolvidos durante a concepção e construção do PTGF, negando-se-lhes a oportunidade de opinar sobre os dados e informações e sobre a maneira mais adequada de disponibilizá-los. Ainda de acordo com o pesquisador, o portal se baseou na estrutura do SIAFI, o que é refletido nos dados brutos publicados.

Na área da interoperabilidade, a divulgação de dados em PDF ou em arquivos de imagem digitalizados ainda é o maior entrave. A ausência de uma solução apoiada em consulta a banco de dados, com divulgação em formato legível por máquinas, também acarreta um retrabalho para órgãos e entidades, que são obrigados a recriar documentos em PDF e imagens digitalizadas e publicá-los com a frequência exigida por lei.

Entre aquelas informações já disponíveis em formato interpretável por máquina, como CSV e planilhas, o desafio é interoperar informações em formatos distintos e correlacionar informações de mesmo significado. A falta de dicionário de dados e layout de relatórios são indicações de que ainda não passariam do nível emergente no modelo de [Morgan, 201], enquanto estariam com duas ou três estrelas na escala de [Berners-Lee, 2009], dependendo do formato publicado ser proprietário ou aberto.

É perceptível a necessidade de mais investimentos neste tema, começando pelo próprio Guia da CGU que, ao detalhar os dados acerca de convênios, despesas, licitações e contratos, que devem estar disponíveis no âmbito da transparência ativa, não expõe o seu significado. Como consequência, órgãos e entidades, que muitas vezes trabalham com jargão próprio, deduzem essas informações. A sociedade e os cidadãos, por sua vez, tentam interpretar o significado dos conceitos e trabalham com informações com o mesmo rótulo mas que podem não significar a mesma coisa. A comparação entre essas informações, que é um dos usos mais promissores dos dados abertos, pode ficar comprometida e levar a conclusões equivocadas.

4. Abordagens de Solução

Alguns dos problemas descritos na seção anterior podem ser atacados com abordagens de pesquisa. A falta de maturidade tende a atenuar com a padronização e a adoção “*de facto*” de um modelo de maturidade que permita medir o nível em que a organização se encontra e apontar os passos necessários para evoluir a níveis mais maduros de transparência organizacional e, em consequência, de abertura de dados. Em geral, as organizações se preocupam com aspectos de visibilidade externa quando estes possam ser medidos, segundo a lógica de que “não se pode controlar o que não se pode medir”.

Outra boa parte dos problemas pode ser minimizada a partir da sensibilização e capacitação de pessoal, inclusive gestores, para trabalhar com as informações. Porém também é preciso pensar nas informações do ponto de vista da sociedade e estruturá-las de modo intragovernamental. Isto requer o aperfeiçoamento de sistemas estruturantes da Administração Pública, como o SICONV, o SIAFI, o SIAPE, o SCDP, o SIORG, e o ComprasNet, entre outros. Requer ainda o envolvimento da sociedade nos projetos de transparência, através de consultas públicas.

Também é necessário pesquisar a melhor forma de divulgar informações textuais, ou com estruturas variadas, e ainda se elas devem ser apresentadas no mesmo nível das demais informações. No exemplo dos convênios, a prestação de contas pode possuir pareceres, datas e valores. Cada uma dessas informações pode ter uma estrutura característica, e que pode mudar de instituição para instituição.

A solução para algumas das questões de inconformidade aos princípios de dados abertos, como completos, primários e atuais, passa pela investigação das causas desses problemas. Uma parte deles pode ser tratada no campo da qualidade dos dados; técnicas de pré-processamento de textos, largamente utilizadas em processos de recuperação de informação, podem ser úteis na resolução de conflitos usuais como sinonímia e

homonímia. É possível ainda que o esforço para que os dados sejam publicados na fonte, também contribua para a redução dessas inconformidades. A Web Semântica tem linhas de pesquisa para estes tipos de problemas.

As questões relacionadas ao jargão específico de cada órgão ou entidade podem ser minimizadas com a criação de ontologias ou com um trabalho de realinhamento de ontologias. Outra opção é a unificação de termos utilizados pelos diversos órgãos e entidades, criando uma ontologia única a ser adotada por toda a administração pública.

As áreas de Mineração de Informações e Gestão de Conhecimento podem em muito contribuir com soluções para identificação de grupos de informações relevantes para cada contexto e o uso destas para inibição de corrupção e fraudes.

No âmbito dos convênios firmados por órgãos e entidades da administração pública, propõe-se um estudo de caso exploratório para avaliar se os dados disponibilizados atendem aos princípios de dados abertos e se são suficientes, claros e úteis para os cidadãos e para a sociedade. Durante o estudo de caso, através de técnicas de questionário e entrevistas, deverão ser coletados e analisados dados como inconformidade, dificuldades encontradas, usos potenciais dos dados e influências que possam advir da divulgação e uso dos dados.

5. Conclusão

Fato é que a implantação da LAI e de conceitos como Dados Abertos e Interoperabilidade necessitam ainda de muitos estudos e pesquisa. Muitos dos problemas que as organizações estão tendo que enfrentar ainda não têm solução sistematizada. Apesar dos esforços que vêm ocorrendo tanto nas organizações quanto no meio acadêmico, ainda há muito que ser desbravado. O grande objetivo deste artigo foi apresentar frentes aos problemas identificados, possibilidades de linhas de trabalho que possam efetivamente construir soluções aplicáveis ao universo das organizações que, de fato, buscam a transparência organizacional.

Referências

- Berners-Lee, Tim. (2009) “Linked Data”. Disponível em <http://www.w3.org/DesignIssues/LinkedData.html>. Acesso em 20/01/2013.
- CGU – Controladoria Geral da União. (2012) “Guia para Criação da Seção de Acesso à Informação nos Sítios Eletrônicos dos Órgãos e Entidades Federais”.
- CiberDem – Núcleo de Pesquisa em Ciberdemocracia. (2013). Disponível em <http://www2.unirio.br/unirio/ccet/unirio-cgu>. Acesso em 22/01/2013.
- Morgan, Dan. (2010) “Toward an Open Data Maturity Model”. Disponível em <http://www.phaseonecg.com/blog/archives/319>. Acesso em 21/01/2013.
- Nascimento, L.P. do (2011). A Transparência dos Portais Brasileiros de Transparência Pública: Um Estudo de Três Casos. Dissertação. UFRJ/COPPE.
- OGP - Open Government Partnership. (2012) “Open Government Declaration”. Disponível em: <http://www.opengovpartnership.org/open-government-declaration>. Acesso em 07/06/2012.
- W3C – World Wide Web Consortium – Escritório Brasil. (2011) “Manual dos Dados Abertos: Governo”.