



## Dados abertos governamentais (Open Government Data): instrumento para exercício de cidadania pela sociedade

*Claudio Jose Silva Ribeiro, Reinaldo Figueiredo de Almeida*

### Resumo

O maior acesso à informação gerada e disseminada pelos governos tem sido amplamente debatido pela sociedade e tem como principal catalisador o uso do ambiente Web. O tema Dados Abertos Governamentais (Open Government Data) tem sido explorado e começa a ser tratado com prioridade, pois o volume de informações geradas e publicadas cresce vertiginosamente, mas com possibilidades restritas de tratamento automatizado por programas e agentes de software, que podem ser desenvolvidos pela sociedade civil organizada. Este relato apresenta o desenvolvimento do projeto que entrega para a sociedade um conjunto de dados, que está sendo desenvolvido segundo os princípios para uso de Dados Abertos de Governo. São apresentadas as ações que estão sendo encaminhadas pela Dataprev para estruturar o arcabouço tecnológico que sustentará as atividades para gerar este conjunto de dados.

**Palavras-chave:** Dados Abertos. Governo Eletrônico. LOD. Dataprev

### 1. INTRODUÇÃO

Atualmente, quando falamos de governo eletrônico dois marcos políticos podem ser trazidos como sendo determinantes para as definições das opções tecnológicas a serem conduzidas neste campo.

O primeiro marco é a publicação do decreto nº 6.932, de 11 de agosto de 2009, assinado pelo presidente da República, o qual “dispõe sobre a simplificação do atendimento público ao cidadão, ratifica a dispensa do reconhecimento de firma em documentos produzidos no Brasil, institui a ‘Carta de Serviços ao Cidadão’ e dá outras providências”. Para tanto, este decreto traz no seu art. 1º, incisos II e VI, que o Poder Executivo nas relações com cidadão deverá compartilhar informações e aplicar soluções tecnológicas, que além de simplificar processos e procedimentos, propiciem melhores condições para tornar disponível a informação para o cidadão.

Além disto, o decreto prevê no seu art. 4º, que as bases de dados oficiais geridas pelo Poder Executivo, sejam colocadas à disposição de órgãos e entidades públicas interessadas no acesso às informações constantes nas mesmas.

O segundo marco determinante foi a publicação pelo presidente dos EUA, Barack Obama, do documento *Memorandum on Transparency and Open Government* (OBAMA, 2009), o qual define as diretrizes do governo dos EUA para a sua gestão, no que tange ao acesso dos dados do governo americano, indo desde a questão da gestão interna da informação até a publicação dos dados do governo para o público em geral. O fato é que após



a publicação deste documento, o que se verificou foram iniciativas nos EUA em torno do chamado *Open Government*, fortalecendo as ações relacionadas ao tema e implicando na disponibilização de bases de dados oficiais para a comunidade em geral. Este marco também fomentou outras discussões relacionadas, principalmente sobre padrões tecnológicos referentes ao tema, tais como, RDF, OWL e SKOS, além de definições sobre o padrão de licença a ser atribuída aos dados publicados.

Deste modo, de acordo com o citado anteriormente, a atual concepção de governo eletrônico passa pelo investimento em *Open Government Data*, o que permitirá a ampliação do acesso dos cidadãos aos serviços públicos, além de permitir que os mesmos possam interagir com bases governamentais, gerando novas informações e novos conhecimentos.

A idéia básica para *Open Government Data* é de tornar disponível um grande número de bases dados oficiais, cobrindo todos os aspectos da atividade governamental, permitindo que além do acesso aos dados, o público em geral possa compartilhá-los, fazendo o uso de processamento automatizado em computadores, conforme suas necessidades de informação.

Para implementar estas idéias, a plataforma tecnológica básica que se coloca é a Web Semântica, mas especificamente, conceito fundamental formulado por Tim Berners-Lee de *Linked Data* ou *Linked Open Data* (LOD), segundo o qual, deve ser uma prática expor e compartilhar, tanto os elementos de ligação de dados quanto às informações e conhecimento. Esta plataforma tem como base o uso de URIs (*Universal-Uniform Resource Identifier*) para identificação e endereçamento de unidades de informação, além de RDF (*Resource Description Framework*) para representá-las. Assim, a estruturação do espaço Web não se fará apenas por meio de links e conexões entre documentos, mas sim com uma nova abordagem onde documentos e dados passam a ser ligados, fazendo uso de um conjunto de melhores práticas para publicar e tornar disponível o acesso a estes dados.

Em síntese, Bizer, Heath e Berners-Lee (2011) observam que a noção de Dados Ligados (*Linked Data*) consiste em fazer uso da Web para criar ligações entre dados de diferentes fontes. Estas fontes podem ser tanto bancos de dados mantidos por outras organizações, quanto um conjunto de sistemas heterogêneos residentes em diferentes ambientes de computação. Tecnicamente, dados ligados se referem a dados publicados na Web que são estruturados de forma que sejam lidos por máquinas e com o seu significado definido explicitamente.

Bizer, Heath e Berners-Lee continuam e esclarecem que:

O projeto Linked Open Data (LOD) é a aplicação mais visível dos princípios de Linked Data, pois tem o objetivo de promover a noção de Web of Data



por meio da identificação do conjunto de dados abertos que estão disponíveis sob uma versão livre de licenciamento, representados em formato RDF e publicados na Web. (BIZER, HEATH, BERNERS-LEE, 2011, p.5)<sup>1</sup>

A aplicação de *Linked Data* nos leva a um esforço para construção de um esquema de nomenclatura global, a fim de permitir a troca de dados. Em outras palavras, isto quer dizer que é necessário a elaboração de dicionários, com vocabulários que possam ser controlados e publicados, mas que possam ser expressos em modelos de alto valor, fazendo uso de ontologias e permitindo o incremento da semântica para estes dados ligados.

No Brasil, o desenvolvimento de projetos para suportar as ações de governo eletrônico e intrumentalizar a adoção da “Carta de Serviços ao Cidadão” têm exemplos em diferentes segmentos onde dados abertos estão agregando valor aos serviços. Ações ligadas ao Portal da Transparência ([www.portaltransparencia.gov.br/](http://www.portaltransparencia.gov.br/)) e sítio Comprasnet ([www.comprasnet.gov.br/](http://www.comprasnet.gov.br/)), são esforços brasileiros que fomentam o maior controle das ações do governo pela sociedade. Ademais, estes esforços podem ser estruturados em acordo com outras abordagens automatizadas, de forma a possibilitar uma maior participação popular propondo inovações em produtos e serviços, além de auxiliar na melhoria da eficiência dos serviços governamentais por meio da medição do impacto das políticas públicas implementadas (MANUAL, 2011).

Convalidando estes esforços, também é possível identificar em nível mundial, diversos projetos que materializam as ações e princípios citados anteriormente. Como intuito de enriquecer este relato, seguem alguns exemplos sobre projetos para tornar visíveis gastos de governos: Tax Tree ([www.mindtrek.org/2009/node/127](http://www.mindtrek.org/2009/node/127)), da Finlândia; e Where Does My Money Go ([wheredoesmymoneygo.org](http://wheredoesmymoneygo.org) – para onde vai o meu dinheiro?) da Grã-Bretanha, que mostram como os recursos dos impostos estão sendo gastos pelo governo. No Canadá, a abertura de dados permitiu que se economizassem 3.2 bilhões de dólares canadenses em fraudes fiscais de caridade. Outros sítios, como o Folketsting ([folketsting.dk](http://folketsting.dk)), da Dinamarca, acompanham as atividades parlamentares e o processo legislativo, permitindo o acompanhamento de projetos no âmbito do parlamento (MANUAL, 2011).

Também em nível mundial, no âmbito da melhoria de serviços e da participação popular, é possível citar projetos como o Find Toilet ([www.findtoilet.dk](http://www.findtoilet.dk)) que mostra em um mapa todos os banheiros públicos da Dinamarca. Em uma primeira análise este projeto pode parecer de pouca utilidade, contudo, as informações disponibilizadas possibilitam que pessoas

---

<sup>1</sup> Tradução do autor para “The original and ongoing aim of the project is to bootstrap the Web of Data by identifying existing data sets that are available under open licenses, converting these to RDF according to the Linked Data principles, and publishing them on the Web.”



com diferentes problemas de saúde possam sair de casa com maior frequência. Na Holanda, o sítio Vervuilings Alarm ([www.vervuilingsalarm.nl](http://www.vervuilingsalarm.nl)) avisa com uma mensagem se a qualidade do ar da vizinhança está perto de atingir um nível prejudicial. Há projetos como o Mapumental ([www.mapumental.channel4.com](http://www.mapumental.channel4.com)), no Reino Unido, e o Mapnificent ([www.mapnificent.net](http://www.mapnificent.net)), na Alemanha, que permitem encontrar locais para morar de acordo com características indicadas, como duração do trajeto até o local de trabalho, além dos preços das casas e outros atrativos das regiões. Todos são todos exemplos de serviços que utilizam dados abertos. (MANUAL, 2011)

Dentro deste contexto é que surge a proposta de um projeto para usar *Linked Open Data* na publicação dos dados do Anuário Estatístico da Previdência Social, permitindo não só o acesso aos dados contidos no Anuário, mas o uso destes dados em operações de troca com outras bases, trazendo-os para um espaço de informações comum e permitindo a exploração de interligações ainda não pensadas para as informações do Anuário.

## 2. O ARCABOUÇO CONCEITUAL DO PROJETO

### 2.1 *Open Government Data* ou Dados Abertos Governamentais

Dados Abertos Governamentais são entendidos como o esforço para a publicação e disseminação das informações do setor público na Web, permitindo a re-utilização e a integração destes dados. Além de propiciar a análise destes dados segundo vários pontos de vista, esta disseminação permite que a sociedade construa uma visão mais clara sobre o desempenho do governo frente às metas estabelecidas, bem como sobre o desenvolvimento de políticas públicas. Estas informações devem estar compartilhadas em formato bruto e aberto<sup>2</sup>, além de serem compreensíveis logicamente, pois estas características permitirão sua reutilização em aplicações informatizadas desenvolvidas pela sociedade (BERNES-LEE, 2009).

Segundo David Eaves, pesquisador canadense, a trilha para implementar *Open Government Data*, dentre outros fundamentos, deverá obedecer as três leis<sup>3</sup> para promover o uso de Dados Abertos:

1. Se o dado não pode ser encontrado e indexado na Web, ele não existe;

---

<sup>2</sup> O formato bruto corresponde à forma primária em que o dado foi coletado pelo sistema informatizado disponível no governo, respeitando-se a segurança e privacidade de cada cidadão. O formato aberto diz respeito à independência de licenças comerciais e/ou softwares específicos para seu uso.

<sup>3</sup> Traduzido de <http://eaves.ca/2009/09/30/three-law-of-open-government-data/>



2. Se o dado não estiver aberto e disponível em formato compreensível por máquina, ele não pode ser reaproveitado;
3. Se algum dispositivo legal não permitir sua reaplicação, ele não é útil.

A estas leis, acrescentam-se os pressupostos sugeridos pela International Aid Transparency Initiative (IATI), contidos no Manual dos Dados Abertos<sup>4</sup>, e materializados em oito princípios<sup>5</sup>:

1. Dados precisam ser completos: Todos os dados públicos estão disponíveis. Dado público é o dado que não está sujeito a limitações válidas de privacidade, segurança ou controle de acesso;
2. Dados precisam ser primários: Os dados são apresentados como os coletados na fonte, com o maior nível possível de granularidade e sem agregação ou modificação;
3. Dados precisam ser atuais: Os dados são disponibilizados tão rapidamente quanto necessário à preservação do seu valor;
4. Dados precisam ser acessíveis: Os dados são disponibilizados para o maior alcance possível de usuários e para o maior conjunto possível de finalidades;
5. Dados precisam estar em formato compreensível por máquinas: Os dados são razoavelmente estruturados de modo a possibilitar processamento automatizado;
6. Dados precisam ser não discriminatórios: Os dados são disponíveis para todos, sem exigência de requerimento ou cadastro;
7. Dados precisam estar disponíveis em formato não proprietário: Os dados são disponíveis em formato sobre o qual nenhuma entidade detenha controle exclusivo;
8. Dados precisam estar livres de licenciamento: Os dados não estão sujeitos a nenhuma restrição de direito autoral, patente, propriedade intelectual ou segredo industrial. Restrições sensatas relacionadas à privacidade, segurança e privilégios de acesso são permitidas.

Especificamente em relação a Dados Abertos Governamentais, junto aos oito princípios devem ser observados quatro aspectos fundamentais: disponibilização dos dados, publicação de informações sobre os dados disponíveis a fim de facilitar o processo de

---

<sup>4</sup> Este Manual foi lançado no 12º. FISL (Fórum Internacional de Software Livre). Disponível em <http://softwarelivre.org/fisl12> (MANUAL, 2011)

<sup>5</sup> Traduzido de <http://www.opengovdata.org/home/8principles> e também disponível em (MANUAL, 2011)



intercambiamento, definição de plataformas tecnológicas capazes de dispor os dados de modo facilmente acessível e intercambiável (inclusive de modo automático), e licenças ou autorizações públicas a fim de permitir não só o acesso aos dados disponibilizados, mas o pleno uso dos seus conteúdos. (AUER, 2007)

O primeiro aspecto a ser observado se refere ao processo de disponibilização das bases de dados oficiais por parte dos Governos, que além de observar os pressupostos listados acima, devem observar e garantir: (AUER et. al., 2007; GRAY et. al., 2009)

- a) O sigilo e a inviolabilidade de dados pertencentes a pessoas físicas e jurídicas;
- b) A não rastreabilidade dos acessos praticados junto aos dados disponíveis;
- c) A promoção da previsibilidade legal e da segurança.

Um segundo aspecto fundamental é o que se refere à publicação de informações sobre os dados disponíveis a fim de facilitar o processo de intercambiamento. Segundo Gray et. al. (GRAY et. al., 2009), as bases de dados oficiais, uma vez tornadas acessíveis, deverão estar associadas a dicionários que detalhem no seu conteúdo informações referentes aos dados publicados, com semântica adequada ao público em geral.

Em “*An ontology for e-government public services*”, Vassilakis e Lepouras (VASSILAKIS & LEPOURAS, 2006), chama a atenção para o terceiro aspecto fundamental que deve ser observado: as plataformas tecnológicas a serem empregadas. Visando atender a proposta de Dados Abertos Governamentais, estas plataformas deverão permitir que sejam atendidos os princípios citados acima e seguir, preferencialmente, a padrões propostos por organismos internacionais multilaterais como o W3C e o IATI, objetivando garantir o acesso às bases de dados publicadas de todos os cidadãos interessados.

Finalmente, um aspecto crítico observado nas discussões conduzidas sobre Dados Abertos, especialmente quando se aborda o conjunto de dados oriundos do âmbito governamental, está relacionado às questões de licença e autorização para acesso aos dados e aos seus conteúdos, bem como também ao uso pleno dos mesmos (AUER et. al., 2007; GRAY et. al., 2009). Neste aspecto crítico observa-se, inclusive, a diferença legal existente entre os conteúdos individuais de uma base de dados (o conteúdo de cada um dos campos de dados) e a coleção de conteúdos de uma base de dados, ou seja, todos os campos de dados de uma base e os seus respectivos conteúdos. (DAVIES, 2010)

De qualquer modo, é importante que os entes governamentais adotem políticas que eliminem direitos de uso e de autoria para as bases constantes dos programas de Dados Abertos Governamentais, adotando e institucionalizando o pelo uso de protocolos como o *Protocol for Implementing Open Access Data*. Este protocolo já vem sendo adotado pelos





EUA e pela União Européia, institucionalizando o assim chamado *Public Domain* ou Política do Direito Público. (GRAY et. al., 2009)

Portanto, Dados Abertos Governamentais deve ser uma política capaz de dispor dados oficiais de fácil acesso e plenamente compreensível a pessoas não técnicas, em formato legível, preferivelmente, para homens e máquinas, com previsibilidade legal, segurança, e devidamente atualizados. (DAVIES, 2010)

## 2.2 Dados Abertos Governamentais no contexto brasileiro

Como observado na introdução deste relato, muitas ações estão sendo desenvolvidas na direção de trabalhar com dados abertos, tanto em nível mundial quanto no Brasil. Um dos exemplos mais contundentes sobre esta percepção foi a realização de evento recente em nosso país para tratar deste tema. O IV CONSEGI (Congresso Internacional – Software Livre e Governo Eletrônico<sup>6</sup>) teve como temática central o uso de Dados Abertos para a Democracia na Era Digital e trouxe diferentes pesquisadores para debater ações de implementação de dados abertos.

Além disto, o conjunto de experiências de outros países tem sido utilizado como referência para os esforços de implementação que estão sendo desenvolvidos pelo governo brasileiro. Nesta direção, os sítios do governo americano ([www.data.gov](http://www.data.gov)), do Reino Unido ([data.gov.uk](http://data.gov.uk)), da Alemanha ([www.offenedatten.de](http://www.offenedatten.de)) e do Canadá ([www.datadotgc.ca](http://www.datadotgc.ca)), apresentam diferentes visões em relação à estruturação e organização para o conjunto de dados que será alvo de publicação (DADOS, 2011).

Por outro lado, existem no âmbito do próprio governo brasileiro ações para melhorar a qualidade das informações e criar instrumentos para facilitar o intercâmbio de dados. As iniciativas em curso podem ser constatadas no relato de Franzosi et. al., que pode ser considerado como a gênese da problemática para o tratamento da informação nas bases de dados do Governo:

A imensa quantidade de código dos Sistemas Estruturantes Corporativos de Governo, somado à complexidade do negócio que apóiam e as inúmeras diferentes plataformas computacionais que os suportam remetem a diversas tentativas de evolução e integração desses sistemas. Esta dificuldade natural em implementar melhorias apontadas em ambientes complexos incentivou, ao longo de anos, a criação de um Cinturão de Dados, CD para apoio ao processo decisório de governo. [...] Os dados dos SGAs [Sistemas de Gestão Administrativa], muitas vezes extraídos e consolidados em planilhas, são acrescidos com outras informações declarativas, outros documentos, notícias, etc, contituindo uma base importante para tomada de decisão no nível estratégico do Governo. (DADOS, 2011, p. 25)

<sup>6</sup> Disponível em [www.consegi.gov.br](http://www.consegi.gov.br)



É dentro deste contexto que também está sendo desenvolvido o projeto DadosGov<sup>7</sup>. Este projeto objetiva criar um Catálogo de Informações Aberto para aprimorar a gestão pública e facilitar o acompanhamento pela sociedade. Desenvolvido pelo Comitê de Organização de Informações da Presidência da República (COI-PR), este catálogo foi construído a partir de alguns pressupostos para disponibilização de informações (DADOS, 2011).

- Apresentação das informações com organização em árvores temáticas e séries históricas;
- Uso de grupos de informação para facilitar a aquisição e atribuição da responsabilidade pela qualidade da estrutura de dados e seu conteúdo;
- Especificar e estruturar os requisitos do nível geral para os níveis mais específicos;
- Organizar o armazenamento de dados em formato padrão de catalogação.

Os trabalhos de modelagem deste catálogo também foram desenvolvidos com o apoio do Modelo de Dados Global, onde, por meio do mapeamento das Bases de Dados Eletrônicas dos sistemas do Governo Federal, buscou-se integrar a estrutura, a semântica e os processos envolvidos na atualização dos respectivos repositórios de dados.

Some-se a isto o conjunto de esforços para organização deste tema, que estão sendo desenvolvidos pela Secretaria de Logística em Tecnologia de Informação (SLTI) do Ministério do Planejamento. Em recente debate também realizado no IV CONSEGUI, a Sra. Miriam Chaves – Diretora da Sec. Logística e Tecnologia da Informação do Ministério do Planejamento – apresentou em linhas gerais a proposta para integração e uso da infraestrutura de LOD no governo (Figura 1).

---

<sup>7</sup> Disponível em <http://i3gov.planejamento.gov.br/dadosgov/>



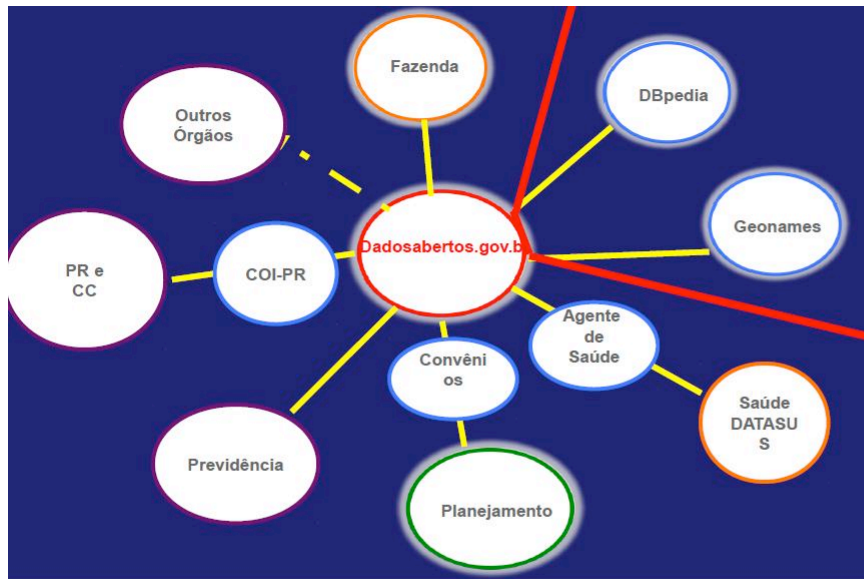


Figura 1: Estrutura LOD e Dados do governo (CHAVES, 2011)

Os esforços da SLTI culminaram com a parceria entre o governo e a sociedade, representada por organismos de padronização, universidades e Organizações não governamentais, no sentido de desenvolver a INDA<sup>8</sup> (Infraestrutura Nacional de Dados Abertos). Esta infraestrutura segue o modelo que foi praticado para geração da INDE<sup>9</sup> (Infraestrutura Nacional de Dados Espaciais) e é composta por 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, em conformidade com o disposto no projeto e-Ping<sup>10</sup> (Projeto de Interoperabilidade no Governo).

É dentro deste cenário repleto de alternativas e exemplos de estruturação segundo pressupostos para dados abertos, que se insere o projeto em desenvolvimento pela Dataprev (Empresa de Tecnologia e Informações da Previdência Social).

### **3. VENCENDO O DESAFIO: ESTRUTURANDO E IMPLEMENTANDO O PROJETO DE DADOS ABERTOS NA PREVIDÊNCIA SOCIAL**

Em sintonia com a evolução dos ambientes tecnológicos e para atender a estruturação proposta pela SLTI/Min. Planejamento, a Dataprev investiu na organização de um projeto para inserir a Previdência no contexto dos Dados Abertos.

<sup>8</sup> Disponível em <http://wiki.gtinda.ibge.gov.br/MainPage.ashx>

<sup>9</sup> Disponível em <http://www.inde.gov.br/>

<sup>10</sup> Disponível em <http://www.governoeletronico.gov.br/acoes-e-projetos/e-ping-padroes-de-interoperabilidade>



Nesta direção, destaca-se, inicialmente, o papel da Dataprev, que administra grandes cadastros sociais do governo federal brasileiro. As bases de dados do sistema de Previdência Social pública, o Cadastro Nacional de Informações Sociais (que contém os registros referentes a todos os vínculos de contribuições previdenciárias e trabalhistas da população economicamente ativa brasileira) e o Cadastro Nacional de Óbitos (que concentra o registro de todos os óbitos registrados em cartórios do Brasil), são exemplos do conjunto de dados que estão sob a guarda da Dataprev.

Portanto, para dar conta deste desafio, foi necessário compreender o domínio que seria alvo de publicação, compondo a fase de planejamento do trabalho (AUTOR, 2008). Esta fase foi encaminhada por meio de reuniões para estruturação do escopo, que foram conduzidas pelo Departamento de Gestão de Informações em parceria com a Presidência da Dataprev. Nestas reuniões foram encaminhados debates internos sobre o conjunto de necessidades de informação para a sociedade, principalmente as demandas provocadas pela comunidade de pesquisadores que já fazem solicitações de conjunto de dados para a Dataprev. Estes debates foram conduzidos no sentido de avaliação dos técnicos da Dataprev sobre a relação das demandas, bem como a relevância destas para a comunidade.

O resultado destas reuniões foi a decisão de publicação dos dados de acidente de trabalho presentes no Anuário Estatístico da Previdência Social, pois apesar destes dados já serem publicados por meio de ferramenta que permite a extração de consultas pela internet<sup>11</sup>, esta publicação não segue os princípios para uso de dados abertos que foram apresentados na seção anterior, dificultando o reuso destas informações pela sociedade em geral.

O Anuário Estatístico da Previdência Social – AEPS juntamente com o Boletim Estatístico da Previdência Social – BEPS, que possui periodicidade mensal, contém dados referentes ao sistema previdenciário, os quais foram tratados, organizados e disponibilizados em séries agregadas antes de compor os dois instrumentos. Ambos, conforme citado anteriormente, estão disponíveis pela internet através do sítio da Previdência Social (<http://www.previdencia.gov.br/>), além dos exemplares publicados em outras mídias, como por exemplo o uso de cadernos impressos.

Como forma de representar os grandes grupos de dados analisados, durante as reuniões de planejamento, foi apresentado o seguinte modelo de classes que elucida o conjunto de dados a serem publicados:

---

<sup>11</sup> Disponível em <http://www3.dataprev.gov.br/infologo/>

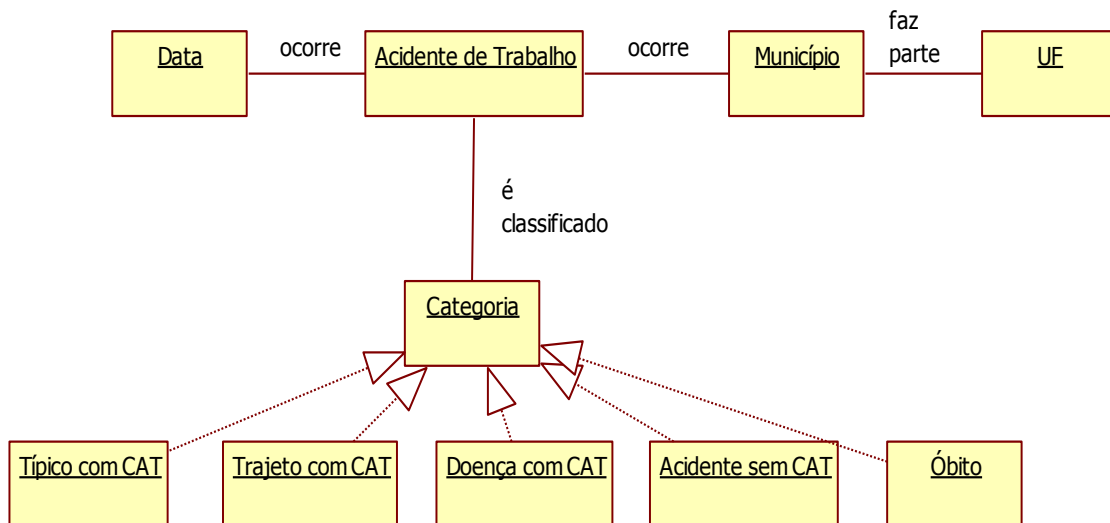


Figura 1: Modelo de Classes<sup>12</sup> sobre evento de acidente de trabalho

Com o escopo para publicação definido, o debate sobre o desenvolvimento do trabalho foi na direção de examinar as leis e princípios para a publicação de dados abertos. Nesta direção, foi necessário identificar um recorte de dados mais apropriado ao processo de publicação, pois o esforço para estruturação dos dados e de sua semântica precisava ser adequado aos prazos estabelecidos para a conclusão do projeto. Assim, foi possível chegar à representação a seguir como o modelo final que será alvo de representação para publicação:

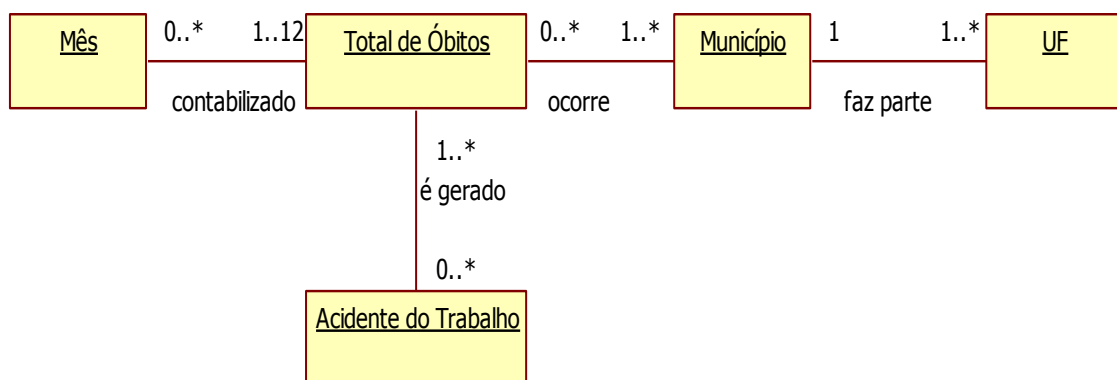


Figura 2: Modelo de Classes<sup>13</sup> adotado com respectivas cardinalidades (ocorrências).

<sup>12</sup> Para facilitar a compreensão do modelo, a leitura pode ser traduzida por sentenças: Acidente de Trabalho ocorre em data; Acidente de Trabalho ocorre em Município; Acidente de Trabalho é classificado em Categoria; São tipos de Categoria: Acidente Típico com CAT (Comunicação de Acidente de Trabalho), Acidente de Trajeto com CAT, Acidente que causa Doença com CAT, Acidente que causa Doença sem CAT e Acidente que causa Óbito. Neste modelo preliminar as cardinalidades (ocorrências) não foram identificadas.

<sup>13</sup> De forma análoga à Figura 1 a leitura deste modelo também pode ser traduzida em sentenças: Total de Óbitos é contabilizado de um até 12 Meses e um Mês pode ter zero ou vários Óbitos; Total de Óbitos ocorre em nenhum Município ou em vários Municípios e um Município pode ter um ou vários Totais de Óbitos; Um



Com o entendimento do conjunto de dados, partiu-se para o processo de extração<sup>14</sup> destes dados das bases operacionais do Anuário Estatístico, para em seguida dar início ao processo de transformação e carga para publicação.

Apesar da possibilidade de uso de ferramentas automatizadas para apoiar estas ações (tais como D2R, Triplify e Virtuoso RDF), estas etapas estão em curso e demandam um grande esforço, uma vez que é a partir deste ponto que serão geradas as representações em RDF, formato definido para ser usado pelos dados a serem disponibilizados, incluindo a representação semântica destes dados. Estas atividades são conhecidas como geração de triplas<sup>15</sup> RDF ou ainda “triplificação”<sup>16</sup> dos dados. Estas atividades estão previstas para serem concretizadas até o final de 2011.

Outra atividade de fundamental importância e que complementa este projeto, é a organização, revisão e disponibilização de todo o vocabulário relacionado ao conteúdo a ser publicado, sendo o mesmo associado a uma estrutura de metadados, a qual terá como componentes, além das definições, as informações sobre a composição dos dados (em caso de agregações) e suas respectivas origens, por exemplo.

#### 4 CONCLUSÕES

A Web alterou radicalmente a forma de compartilhar conhecimentos, reduzindo as barreiras para a publicação e para o acesso aos documentos. Um dos maiores fatores de sucesso do ambiente Web é a liberdade que seus usuários têm em criar e disseminar informação, logo, o grande desafio é como trabalhar com este ambiente descentralizado e quase que anárquico, buscando um modelo de organização de informações.

Ligações entre documentos através de estruturas de hipertexto permitem aos usuários percorrer o espaço informacional onde os documentos estão publicados, por meio dos assim chamados navegadores Web, enquanto motores de busca associados com analisadores de ligações de documentos fazem com que os usuários localizem os documentos requeridos

---

Município faz parte de uma UF e uma UF pode ter um ou vários Municípios; Total de Óbitos é gerado por um Acidente de Trabalho ou por vários Acidentes e um Acidente pode ter nenhum Óbito ou então vários Óbitos totalizados.

<sup>14</sup> Este processo faz uso de dinâmica para seleção de registros para publicação segundo uma janela temporal pré-determinada, com o uso de linguagem SQL (Linguagem de consultas para Bancos de Dados Relacionais - Structured Query Language).

<sup>15</sup> Uma tripla RDF é composta por: [Sujeito ou Recurso, Predicados ou Propriedades, Valores (o conteúdo do objeto)] e para geração da parte semântica destas triplas é essencial o uso de ontologias, incluindo ontologias de fundamentação de forma a facilitar o intercâmbio de dados (GUIZZARD, 2005).

<sup>16</sup> O fluxo para triplificação consiste em etapas para extração, alinhamento, diferenciação e complementação para carga (BIZER et. al., 2009). Outra referência para o assunto é o tutorial para triplificação, disponível em: <http://www.inf.puc-rio.br/~psalas/tutorial.html>



(JACOBS & WALSH, 2004 *apud* BERNES-LEE, 2009). Esta combinação foi a principal responsável pelo sucesso da Internet e pelo seu pleno constante incremento (BERNES-LEE, 2009).

Porém, com a adoção de novas aplicações, como as redes sociais, associadas a plataformas tecnológicas mais avançadas, como estruturas semânticas para intercâmbio de dados entre ambientes automatizados, novas culturas vão sendo estabelecidas sobre a rede mundial de computadores, como o compartilhamento de informações e dados e a construção de ações colaborativas (AUER et. al., 2007). Dotar a Web de facilidades para permitir o processamento automatizado das informações e conteúdo é um dos caminhos naturais para trabalharmos com a miríade de dados que estão presentes atualmente na Internet.

Neste cenário é que são estabelecidas as bases para o conceito de Dados Abertos. A concepção para estruturar Governo Eletrônico passa a incorporar como algo necessário para a sua plena implementação, os aspectos referentes a Dados Abertos Governamentais.

Assim, em 2009, Barack Obama (OBAMA, 2009) redefiniu o conceito de Governo Eletrônico, ao definir que o mesmo, bem como toda a estrutura do novo governo norte-americano, deveria atender fundamentalmente a três princípios: transparência, participação e colaboração. Deste modo, o Governo Eletrônico norte-americano foi obrigado a se reelaborar, incorporando em definitivo o conceito de Dados Abertos, e mais além, o de Dados Abertos Ligados.

Alinhado a este momento, é que a Dataprev, através da iniciativa de implementar um projeto baseado no conceito de Dados Abertos, insere o Governo Brasileiro no contexto mundial de iniciativas de igual natureza e inicia um processo arrojado de desenvolvimento de competências tecnológicas de natureza estratégicas, além de prestar um enorme serviço ao cidadão comum, ao disponibilizar dados previdenciários num formato acessível e intercambiável.

Os próximos passos a serem trilhados no projeto aqui apresentado, serão a publicação em ambiente Web e a divulgação para a comunidade e sociedade em geral do Anuário Estatístico com Dados de Acidentes de Trabalho, juntamente com todo o vocabulário associado ao mesmo. A publicação do mesmo terá sua divulgação, além das estratégias de veiculação tradicionais, associada ao registro em mecanismos de buscas convencionais da Web e estendida a outros mecanismos, especialmente orientados ao ambiente de dados abertos, tal como, o *Comprehensive Knowledge Archive Network* (CKAN).



O processo de aferição dos resultados alcançados passará pelo uso de redes sociais<sup>17</sup>, blogs<sup>18</sup> e outros canais de comunicação, além de avaliações com membros da sociedade civil organizada (ONG's, comunidades de software e representantes de organizações internacionais, tal como o escritório Brasil do consórcio W3C). Este é um dos primeiros esforços conduzidos pela Dataprev para incremento da participação da sociedade no desenvolvimento do nosso país. Outras ações precisarão ser desenvolvidas para tornar mais eficaz a publicação de dados abertos de governo.

Tanto o uso de ontologias de fundamentação quanto o incremento da adoção de padrões de interoperabilidade no processo de desenvolvimento de novos sistemas, transformará os sistemas computadorizados em verdadeiras fontes de informação para a sociedade, fazendo com que surja aquele que o pesquisador norte-americano Kingsley Idehen chamou de *citizen analyst* ou analista cidadão.

## REFERÊNCIAS

1. AUER, S. et. al. **DBpedia: A Nucleus for a Web of Open Data**. 2007. The Semantic Web, Lecture Notes in Computer Science, 2007, Volume 4825/2007, 722-735. SpringerLink.com.
2. BERNES-LEE, T. **Putting Government Data online**. 2009. Disponível em <http://www.w3.org/DesignIssues/GovData.html> com download em 01 de julho de 2011.
3. BIZER, C, HEATH, T., BERNES-LEE, T. **Linked Data. The story so far**. Disponível em <http://eprints.ecs.soton.ac.uk/21285/1/bizer-heath-berners-lee-ijswis-linked-data.pdf> com acesso em junho de 2011.
4. BIZER, C. et. al. DBpedia - A crystallization point for theWeb of Data. **Web Semantics: Science, Services and Agents on theWorldWideWeb** 7 (2009) 154–165
5. CHAVES, M. B. F. Apresentação sobre ações de Governo para Dados Abertos. Disponível em [http://www.w3c.br/conferenciaegov/06\\_COI-PR\\_Mirian.pdf](http://www.w3c.br/conferenciaegov/06_COI-PR_Mirian.pdf) com acesso em julho de 2011.
6. DADOS Abertos para a Democracia na Era Digital. Brasília: Fundação Alexandre Gusmão, 2011. 84p. Disponível em [http://www.consegi.gov.br/sobre\\_consegi/livro-consegi-dados-abertos-para-a-democracia-na-era-digital?set\\_language=pt-br&cl=pt-br](http://www.consegi.gov.br/sobre_consegi/livro-consegi-dados-abertos-para-a-democracia-na-era-digital?set_language=pt-br&cl=pt-br)
7. DAVIES, T. Open data, democracy and public sector reform. **A look at open government data use from data.gov.uk**. 2010. Disponível em

---

<sup>17</sup>Disponível em <http://twitter.com/#!/dataprev>

<sup>18</sup> Disponível em <http://portal.dataprev.gov.br/category/blog/>



- <http://practicalparticipation.co.uk/odi/report/wp-content/uploads/2010/08/How-is-open-government-data-being-used-in-practice.pdf> com download em 20 de maio de 2011.
8. GUIZZARDI, G. **Ontological Foundations for Structural Conceptual Models**. Enschede: Netherlands. Telematica Instituut Fundamental Research Series, No. 015 (TI/FRS/015). 2005. Disponível em <http://www.inf.ufes.br/~gguizzardi/OFSCM.pdf> com acesso em junho de 2011.
  9. GRAY, J. et. al. **Unlocking the Potential of Aid Information**. 2009. Disponível em <http://www.unlockingaid.info/wp-content/uploads/2010/02/UnlockingAidInformation.pdf> com download em 13 de junho de 2011.
  10. MANUAL dos Dados Abertos: Governo. Disponível em [http://www.w3c.br/pub/Materiais/PublicacoesW3C/Manual\\_Dados\\_Abertos\\_WEB.pdf](http://www.w3c.br/pub/Materiais/PublicacoesW3C/Manual_Dados_Abertos_WEB.pdf) com download em 01 de julho de 2011.
  11. OBAMA, B. **Memorandum on Transparency and Open Government**. 2009. Disponível em [\(www.whitehouse.gov/the\\_press\\_office/Transparency\\_and\\_Open\\_Government/\)](http://www.whitehouse.gov/the_press_office/Transparency_and_Open_Government/) com download em 23 de março de 2011.
  12. VASSILAKIS, C., LEPOURAS, G. **An ontology for e-government public services**. 2006. Disponível em <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.76.2496&rep=rep1&type=pdf> com download em 13 de junho de 2011.