

INTEROPERABILIDADE EM DADOS ABERTOS LEGISLATIVOS

Fabricio Rocha de Sousa*

Palavras-chaves: Dados Abertos. Integração. Interoperabilidade. Automação.

RESUMO

Depois da popularização dos *websites* na internet desde a década de 1990, os primeiros anos do século XXI trouxeram um significativo aumento da distribuição de dados puros, voltados não à leitura humana, mas ao processamento por máquina (SANTOS, 2018). Uma das diversas aplicações que esta técnica possibilita é o cruzamento de informações de naturezas e fontes diversas, o que ficou popularmente conhecido como *mashup* (ZANG; ROSSON; NASSER, 2008).

Paradoxalmente, entretanto, a utilização desses dados modelados para máquinas permanece sendo um processo pouco automatizado, quase artesanal e extremamente dependente de intervenções humanas, o que a inviabiliza para usuários sem domínio técnico e a torna difícil mesmo para usuários com formação e/ou experiência nas tecnologias utilizadas (ZANG; ROSSON; NASSER, 2008) – problemas que se multiplicam ao se utilizar diferentes fontes de dados, como se faz em um *mashup*.

Tal processo é tipicamente composto por diferentes etapas – como localização, obtenção (parametrizada ou não), desserialização, interpretação e processamento – que decorrem em necessidades e dificuldades inerentes, frequentemente apontadas por usuários e autores:

- Para se obter um conjunto de dados, é preciso descobrir ou conhecer previamente onde eles se encontram e como obtê-los. Isso quase sempre exige um conhecimento não-técnico sobre qual instituição é a fonte responsável por tais informações (ALONSO, 2008) e os endereços na internet de seus portais, arquivos e/ou APIs (*Application Programming Interface*, ou Interface de Programação de Aplicações).
- Como não há uniformidade ou padronização, formal ou apenas já consolidada, em relação ao fornecimento de dados para processamento por máquina (DIFRANZO, 2011), cabe ao usuário descobrir os métodos de acesso disponíveis, os parâmetros de seleção que eventualmente possam ser utilizados e os valores cabíveis a cada um, e o(s) formato(s) de serialização dos dados fornecidos.
- As estruturas de dados são predefinidas pela instituição que as fornece e também não são necessariamente padronizadas (DIFRANZO, 2011), e é preciso que os usuários as interpretem sintática e semanticamente para, então, implementar em forma de código para computador a lógica necessária e específica para o processamento de cada estrutura, o que diminui a possibilidade de reaproveitamento desta lógica, mesmo com outros dados semelhantes.
- Há uma forte dependência de documentação – isto é, informação para leitura humana – para o acesso e o uso dos dados, mas tal documentação tende a ser, quando existente, defasada em relação ao desenvolvimento e à expansão dos conjuntos de dados existentes (ZANG; ROSSON; NASSER, 2008).

* Câmara dos Deputados. E-mail: fabricio.rocha@camara.leg.br



Desde a década de 1990, diversas propostas tecnológicas foram apresentadas para enfrentar separadamente pelo menos alguns dos problemas apontados. Algumas delas foram efetivamente implementadas e oficialmente transformadas em padrões por entidades como *World Wide Web Consortium* (W3C) e *Internet Engineering Task Force* (IETF).

Um estudo preliminar indica que uma combinação simples e consistente de algumas dessas tecnologias, complementadas por poucos desenvolvimentos e padronizações ainda inexistentes ou desconhecidas, poderia hipoteticamente tornar muito mais automatizada, fácil e ampla a utilização e a integração de informações de diferentes serviços de dados abertos – sem que eles precisem ser alterados.

A mais simples medida remonta ao sistema de hipertexto ENQUIRE, antecessor da *web* (BERNERS-LEE, 1980). Os descritores de relação são termos predefinidos e padronizados para uso junto a *links*, que informam, de maneira processável por máquina, o que um objeto apontado por uma URL significa em relação ao documento ou à rede em que o *link* se encontra: parte, conjunto, agente, produto, versão ou mesmo uma representação alternativa do mesmo recurso – o que ajudaria na tarefa de localização de serviços de dados abertos.

Além de um "endereço" na internet, o acesso a um recurso também pode depender de um método específico, de informação sobre os formatos de serialização disponíveis e de parâmetros e seus possíveis valores para seleção, filtragem e adequação dos dados. Tecnologias criadas para descrição de APIs em portais, como OpenAPI (Swagger) e RAML, podem fornecer esses dados de maneira processável por computador. Tais informações de acesso a um recurso poderiam ser fornecidas se a URL do recurso fosse acessada com o método *HTTP OPTIONS*, ainda hoje apenas parcialmente especificado.

A interpretação das estruturas de dados fornecidas também pode ser deixada a cargo de *softwares* se, juntamente com os dados propriamente ditos, forem fornecidos documentos de descrição sintática e semântica para processamento automatizado. Formatos de descrição sintático-estrutural abundam para formatos XML, mas ainda não são oficialmente definidos padrões equivalentes para os populares formatos JSON e CSV.

A descrição semântica das estruturas de dados difere muito da descrição dos próprios dados e é mais simples. Este destaque é importante porque o modelo conceitual RDF (*Resource Description Framework*), o mais popular e adequado para esta tarefa, é quase sempre exemplificado como forma de descrição dos dados em si. O processamento semântico permitiria superar diferenças de línguas, de estruturas e de formatos de serialização. Também parecem faltar, contudo, padrões para descrição semântica, nos mesmos formatos usados nos fornecimentos dos dados, para dados em XML e CSV.

Para descrições semânticas, é preciso um conjunto de termos que explique o que há em cada campo de dados e as formas com que tais dados se relacionam – isto é, uma ontologia. Já existem diversas ontologias criadas para uso com o modelo RDF, e nesta arquitetura elas não se excluem – pelo contrário, se complementam e se interligam. Para dados abertos legislativos, uma nova ontologia poderia aproveitar e aprofundar outras já criadas para este domínio, como a alemã OParl e a OCD italiana.

Com tais medidas, a capacidade intelectual humana deixaria de ser tão necessária nas fases de obtenção e tratamento dos dados, e poderia se concentrar no reconhecimento e na análise das

informações obtidas, sob a ótica de negócio. Definições de uma ontologia podem incluir informações para esta etapa. Uma ontologia legislativa poderia ser facilmente transformada em um portal multilíngue com explicações sobre agentes e processos legislativos de diferentes casas parlamentares e países, graças a uma tecnologia já de uso comum chamada *content negotiation*: quando acessados por uma aplicação, os endereços da ontologia forneceriam dados puros; quando acessados por um navegador web, forneceriam páginas HTML.

O aprofundamento dos estudos e sugestões aqui apresentadas se faz especialmente oportuno no momento em que a Câmara dos Deputados é designada sede do Centro de Dados Abertos da União Interparlamentar (IPU), que por ser uma entidade política internacional pode viabilizar junto a organizações como o W3C as definições técnicas ainda pendentes, e propor aos parlamentos dos países membros uma adoção consistente e uniforme das mesmas recomendações tecnológicas.

REFERÊNCIAS

ALONSO, José Manuel. Improving access to Government through better use of the Web. **W3C Blog**. 2011. Disponível em: https://www.w3.org/blog/2008/05/better-web-access-to-government/. Acesso em: 17 mai. 2019.

BERNERS-LEE, Timothy John. **Enquire Manual v.1.1**. Geneva: CERN, 1980. Disponível em: http://infomesh.net/2001/enquire/manual. Acesso em: 20 mai. 2019.

DIFRANZO, Dominic, *et al.* The Web Is My Back-end: Creating Mashups with Linked Open Government Data**.** *In:* WOOD, David. **Linking Government Data**. New York: Springer, 2011. Disponível em: https://link.springer.com/chapter/10.1007/978-1-4614-1767-5_10. Acesso em: 17 mai. 2019.

SANTOS, Wendell. **Research Shows Interest in Providing APIs Still High.** ProgrammableWeb. 2018. Disponível em: https://www.programmableweb.com/news/research-shows-interest-providing-apis-still-high/research/2018/02/23. Acesso em: 21 mai. 2019.

ZANG, Nan; ROSSON, Mary Beth; NASSER, Vincent. Mashups: Who? What? Why? *In:* CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS. 2008, Florence. **Resumos** [...]. New York: ACM, 2008. p. 3171-3176. Disponível em: http://mashup.pubs.dbs.uni-leipzig.de/files/p3171-zang.pdf. Acesso em: 17 mai. 2019.