



Home > Segurança

Como contribuir com a interoperabilidade da Web de Dados

Por Colaborador Externo | 29 de Junho de 2020 às 14h00



*Por Ana Eliza Duarte e Caroline Burle**

Com o crescimento do uso da Web e das tecnologias de informação e comunicação (TIC), houve também um aumento expressivo no uso e produção de dados que alimentam essas tecnologias. O compartilhamento de dados entre pesquisadores, governos e cidadãos, de forma aberta ou não, requer o fornecimento de metadados. Comunidades diferentes usam distintos padrões de metadados para descrever seus conjuntos de dados.

Nesse sentido, o World Wide Web Consortium (W3C) tem uma área de atuação chamada "Data Activity", a qual se responsabiliza por construir a Web de Dados. O nome da área faz referência ao volume massivo de dados que não para de crescer e, em potencial, a quem souber e puder utilizá-los.

O aumento maciço da quantidade de dados na Web dificulta a recuperação da informação contida nos pacotes de dados, por estes não estarem bem estruturados, descritos e identificados. As pesquisas realizadas na Web para encontrar as informações que buscamos são feitas, majoritariamente, por meio de palavras-chave que se relacionam com a descrição dos pacotes de dados. Se eles não estiverem bem identificados, a busca se torna infrutífera, e isso impede que possamos aumentar o nível de complexidade das nossas pesquisas.

Para que esses dados possam ser utilizados da melhor forma possível, o W3C criou a [Recomendação Boas Práticas de Dados na Web](#) (*Data on the Web Best Practices* – DWBP) que estabelece 35 boas práticas de publicação e consumo de dados na Web, de maneira a se criar um ecossistema autossustentável.

Dentre as práticas sugeridas pelo DWBP, pelo menos sete indicam métodos de descrição e identificação semântica dos dados disponíveis na Web, utilizando ferramentas de tratamento da informação.

Ao seguir essas boas práticas, o Vocabulário do Catálogo de Dados (*Data Catalog Vocabulary* - DCAT), controlado e criado pelo W3C, permite que pacotes de dados da Web sejam descritos e identificados de forma padronizada, de modo que seus metadados possam ser lidos por máquinas. Isso promove interoperabilidade entre os sistemas e evita a criação desnecessária de novos pacotes de dados, pois os já existentes serão reutilizados. O DCAT é projetado, portanto, para facilitar a interoperabilidade entre catálogos de dados publicados na Web.





Além disso, o uso do vocabulário permite maior complexidade nos métodos de busca, pois a relação entre as palavras-chave que os usuários utilizam e os termos do vocabulário será mais estruturada, possibilitando que a recuperação da informação desejada seja otimizada.

A missão de manter e revisar o [DCAT](#) fica sob a responsabilidade de um Grupo de Trabalho do [W3C](#) criado para a padronização da "interoperação de conjuntos de dados", em inglês *Dataset Exchange Working Group (DXWG)*, que possui participação ativa do Centro de Estudos sobre Tecnologias Web (Ceweb.br) do Núcleo de Informação e Coordenação do Ponto BR (NIC.br). O DXWG leva em consideração as solicitações de recursos da comunidade de usuários do DCAT, além de definir e publicar orientações sobre a especificação e o uso de perfis de aplicativos ao solicitar e fornecer dados na Web.

O DCAT, embora bem-sucedido e amplamente utilizado, possui lacunas conhecidas na sua cobertura, como por exemplo, em séries temporais e versões. Essas lacunas devem ser abordadas para que o uso continue a crescer em diferentes comunidades e a variedade de esquemas de metadados seja reduzida.

A maximização da interoperabilidade entre serviços como catálogos de dados, infraestruturas eletrônicas e ambientes de pesquisa virtual, requer não apenas o uso de vocabulários padrão, mas também de perfis de aplicativos. Eles definem como um vocabulário é usado, fornecendo restrições de cardinalidade e/ou listas enumeradas de valores permitidos, para que os dados possam ser validados. Há ainda o desenvolvimento de vários perfis de aplicativos baseados no DCAT, como o [DCAT-AP](#) da Comissão Europeia.

Em vez de limitar o número de padrões de metadados e perfis de aplicativos em uso, os sistemas devem poder expor e receber dados e metadados de acordo com vários padrões, por meio de interfaces transparentes e sustentáveis. Portanto, precisamos de um mecanismo para os servidores indicarem os padrões e perfis de aplicativos disponíveis, e para os clientes escolherem o apropriado. Isso leva ao conceito de negociação de conteúdo por perfil de aplicativo, ortogonal à negociação de conteúdo por formato de dados e linguagem, que já faz parte do HTTP. Um debate sobre negociação de perfis atualmente em desenvolvimento no Internet Engineering Task Force (IETF), com a contribuição do Grupo de Trabalho do Dataset Exchange, baseia-se no [rascunho](#) apresentado no workshop do [SDSVoc](#).

A combinação da definição do DXWG do que se entende por "perfil do aplicativo", em inglês *application profile*, juntamente com a visão do grupo de como clientes e servidores podem interagir de diferentes maneiras com base nesses perfis, fornecerá um meio poderoso para trocar dados em qualquer formato (JSON, RDF, XML etc.) de acordo com as estruturas declaradas contra as quais os dados podem ser validados.





São objetivos do DXWG, portanto, manter a versão 2 do DCAI e estender o padrão para a versão 3 a partir do trabalho realizado até o momento, bem como dar continuidade ao trabalho sobre interoperabilidade de conjuntos de dados que é realizado pelas comunidades. Além de produzir recomendação sobre negociação de conteúdo por perfil, em inglês content negotiation by profile.

O DXWG, assim como qualquer Grupo de Trabalho do W3C, pode receber contribuições externas, que devem ser analisadas pelos editores de cada grupo – dessa forma, qualquer pessoa pode contribuir com o desenvolvimento dos padrões. Convidamos todas as pessoas interessadas a se [envolverem nas discussões!](#)

**Ana Eliza Duarte e Caroline Burle são membros do Ceweb.br/NIC.br.*

Gostou dessa matéria?

Inscreva seu email no Canaltech para receber atualizações diárias com as últimas notícias do mundo da tecnologia.

INSCREVER

COMENTÁRIOS

