

Capítulo

3

Dados abertos conectados para a Educação

Judson Bandeira¹, Thiago Ávila¹, Williams Alcantara¹, Armando Barbosa¹,
Ig Bittencourt¹, Seiji Isotani²

¹Instituto de Computação – Universidade Federal de Alagoas (UFAL)
57.072-900 – Maceió – AL – Brasil

²Instituto de Ciências Matemáticas e de Computação – Universidade de São Paulo(USP)
13.566-590 – São Carlos – SP– Brasil

{jmb,thiago.avila,wla,abs,ig.ibert}@ic.ufal.br, sisotani@icmc.usp.br

Abstract

Nowadays, large amounts of educational data are published on the web in several formats. Most of these data is published in unstructured formats that are understood only by humans. The Linked Data add semantic to a knowledge base, allowing linking different databases by establishing relationships between data published in different sources, also representing them in machine comprehensible ways. The publication of Linked Open Data enables the automatic processing of large amounts of data, the running of inferences, and the reuse of these data, enabling the construction of knowledge bases that may support the improvement of management and the educational development.

Resumo

Atualmente, grandes quantidades de dados educacionais são publicados na web em diversos formatos. Grande parte desses dados é publicada em formatos não estruturados, que são compreendidos apenas por humanos. Os Dados Conectados agregam semântica a uma base de conhecimento, permitindo conectar bases de dados diferentes mediante o estabelecimento de relações entre dados publicados em diferentes fontes e, ainda, representá-los de forma que sejam compreensíveis por máquina. A publicação de dados abertos conectados possibilita o processamento automático de grande quantidade de dados, a execução de inferências e o reuso desses dados, viabilizando a construção de bases de conhecimento que apoiarão a melhoria da gestão e o desenvolvimento educacional.

1. Introdução

Com o crescimento da web e o uso massivo de tecnologias da informação e comunicação, a quantidade de dados gerados e disponibilizados tem crescido nos últimos anos. Esses dados, quando abertos, contribuem para a implantação da transparência, utilizada como uma ferramenta importante no combate à corrupção, e a ampliação da participação da sociedade no processo de desenvolvimento de novas soluções em diversas áreas do conhecimento. No contexto educacional, os dados abertos são importantes em diversas atividades, como por exemplo, podem ser utilizados no desenvolvimento de soluções tecnológicas que auxiliem na tomada de decisão de professores e gestores escolares, bem como na ampliação da oferta e produção de novos conhecimentos que sirvam de base para o desenvolvimento e aprimoramento de recursos educacionais.

Entretanto, observa-se que os dados educacionais são publicados, predominantemente, em formato não estruturado — o que limita a sua descrição e reutilização. Assim, o processo de consumo tem ocorrido mediante muito esforço e custo devido à baixa qualidade dos dados disponibilizados.

Para a melhoria na educação é importante acompanhar a efetividade de investimentos feitos nessa área, desde investimentos na obtenção de recursos físicos para a organização estrutural de um ambiente escolar até investimentos em qualificação de professores e gestores. Uma das soluções possíveis para analisar a efetividade é o acompanhamento do desempenho de indicadores escolares, como o Índice de Desenvolvimento da Educação Básica (IDEB), que avalia requisitos de qualidade da escola e permite mensurar fatores que impactam na qualidade do ensino. Ademais, o uso de dados é um insumo para diversas outras atividades de cunho educacional, como a produção de materiais didáticos, livros, publicações, dentre outros objetos de apoio ao aprendizado. Com o avanço do ensino a distância e ferramentas de educação digital, a produção e o consumo de dados educacionais crescem significativamente.

Entretanto, o desenvolvimento de soluções tecnológicas torna-se muito caro devido aos formatos em que os dados são disponibilizados, juntamente com a qualidade insatisfatória dos dados educacionais publicados. Diversos conceitos foram desenvolvidos nos últimos anos como uma alternativa para a melhoria da qualidade de dados e, recentemente, foram estabelecidos os Dados Abertos Conectados, facilitando o compartilhamento de conhecimento na Web por meio da redução da barreira para a publicação e acesso a dados como parte de um espaço de informação global [Bizer e Berners-Lee 2009].

A utilização de Dados Abertos Conectados abre uma gama de possibilidades para o processamento automático de dados, contribuindo para a descoberta de novos conhecimentos que podem ser úteis para resolver problemas relacionados à qualidade e a confiabilidade dos dados educacionais, à integração e integridade de bases de dados, à dificuldade de recuperação e descoberta de dados disponíveis em diversas bases, entre outros.

2. Dados Educacionais

No contexto educacional, dados são muito utilizados durante o desenvolvimento de estudos e pesquisas para prover o melhoramento e a inovação em metodologias, processos e ferramentas utilizadas pela sociedade. Em países como o Brasil, importantes bases de dados educacionais, como o IDEB, Censo Escolar e Censo do Ensino Superior,

estão dispostas na web em formatos como *pdf*, *xls*, *csv*, *doc*, etc. Essas bases de dados fornecem informações como o número de matrículas, o total de alunos, o movimento escolar, características básicas da instituição, equipamentos e edificações existentes, além de dados sobre o pessoal técnico e administrativo e as características dos docentes [Rigotti & Cerqueira 2004]. Esses dados educacionais são importantes, pois refletem o estado atual da educação, e o seu acesso e processamento podem auxiliar na tomada de decisão dos gestores escolares.

São várias as aplicações dos dados educacionais, podendo ser utilizados: i) por gestores educacionais para o planejamento de metas e objetivos a serem alcançados pela educação; ii) para avaliar a efetividade de medidas adotadas no contexto educacional; iii) durante o desenvolvimento de pesquisas; e iv) por soluções da indústria que atuam no campo educacional, pois o uso de dados educacionais abertos agrega valor e proporcionam inovação. Com base nisso, os dados educacionais são importantes para avaliar e propor melhorias na educação.

Neste contexto, é importante considerar a crescente utilização da tecnologia como apoio à educação. Dessa forma, surgem várias abordagens para melhoria do aprendizado do aluno, como Jogos sérios, Gamificação, Ambientes Virtuais de Aprendizagem e Sistemas Tutores. Contudo, a ampliação do uso de tecnologias impulsiona a produção de dados em diversos sistemas e bancos de dados armazenados em estruturas descentralizadas e pouco integradas, dificultando o aproveitamento desses dados para a produção de novos conhecimentos [Alcantara et al. 2015].

Neste contexto, surge a necessidade de abertura e aprimoramento na oferta de dados visando o seu reuso e a produção de melhores conteúdos educacionais.

3. Dados Abertos

Dados abertos são dados que estão sempre disponíveis e acessíveis (em formato conveniente); que podem ser reusados e redistribuídos (garantindo a legibilidade por máquina); e com participação universal (qualquer grupo ou pessoa). [Handbook 2012]

Os dados abertos permitem que pessoas e organizações utilizem informações públicas livremente para apoiar a tomada de decisão, subsidiar a produção de novos conhecimentos, gerar aplicativos, fazer análises e outros produtos [Bandeira et al. 2014]. Para que um conjunto de dados seja considerado aberto, ele precisa permitir que o cidadão acesse com facilidade, e o utilize ou redistribua sem restrições. Ademais, os dados precisam ser facilmente encontrados em um lugar indexado, sem impedimento de leitura por máquinas ou restrições legais [Neves 2013].

Com dados abertos disponíveis, abrem-se possibilidades para a sociedade que vão desde a análise mais profunda das informações públicas por meio da correlação de diferentes bases de dados, até o desenvolvimento de soluções tecnológicas que fazem uma leitura frequente de bases de dados públicas para gerar oportunidades de negócio e outros benefícios à sociedade [Neves, 2013 apud Eaves, 2009].

Entretanto, apesar do crescimento da disponibilidade de dados abertos, a atual oferta de dados na web ainda tem ocorrido em formatos que impõem limitações quanto a sua reutilização, pois em sua maioria são consumidos apenas por humanos, não permitindo que sejam reutilizados de forma automatizada por agentes de software [Wood et al. 2013].

4. Dados Conectados

Como resposta aos desafios apresentados na seção anterior, o conceito de dados conectados emerge visando orientar as organizações a disponibilizarem os dados em formatos legíveis por máquina. Dados Conectados referem-se a um conjunto de boas práticas para publicação e conexão de dados estruturados na web utilizando padrões internacionais do *World Wide Web Consortium - W3C*, permitindo o estabelecimento de uma rede de dados que se conectam e auto enriquecem [Heath, 2011; Wood et al. 2013].

Para se adequar aos requisitos de Dados Conectados, os dados precisam obedecer a quatro princípios: (1) Devem ser usados Identificadores Universais de Recursos (*Universal Resource Identifier - URIs*) como nomes para as coisas a serem publicadas; (2) Devem ser usadas URIs HTTP para que os usuários possam localizar estes nomes; (3) Quando a URI for encontrada, ela deve prover informação útil, usando padrões como o RDF (*Resource Description Framework*) ou o SPARQL (*SPARQL Protocol and RDF Query Language*); (4) As URIs devem incluir hiperlinks para outras URIs, para que os usuários possam descobrir novos recursos que se relacionem à URI que estejam buscando [Berners-Lee 2006].

4.1. Boas Práticas para Publicação de Dados Conectados

No âmbito do grupo de trabalho para o estudo de Dados Conectados Governamentais², o W3C estabeleceu uma série de boas práticas para facilitar o desenvolvimento e fornecimento de dados governamentais abertos como Dados Abertos Conectados. Apesar de estas práticas terem sido concebidas inicialmente para o ambiente governamental, elas podem ser aplicadas a outros segmentos de negócio, como a educação.

Anterior à existência de normas e padrões internacionais para intercâmbio de dados na Web, o esforço para construir aplicações usando técnicas tradicionais de gerenciamento de dados era muito maior. Considerando o aumento da oferta de dados na Web, as melhores práticas estão evoluindo e a adoção dos Dados Abertos Conectados vem ao encontro dessas tendências. As boas práticas para publicação de Dados Conectados estabelecidas pelo W3C são descritas resumidamente abaixo [W3C 2014]:

- Preparar as partes interessadas (*stakeholders*): Etapa que visa preparar as partes interessadas, em especial os decisores, sobre o processo de publicação e manutenção de dados abertos conectados;
- Selecionar conjuntos de dados: Etapa que visa à seleção dos dados que serão publicados, priorizando os que possuem maior potencial para reuso, podendo ser combinado com outros. Dessa forma os usuários poderão utilizá-los e reusá-los para diversas finalidades;
- Modelar os dados: Essa etapa visa estabelecer uma melhor representação dos objetos de dados e como eles serão utilizados por aplicações de forma independente a sua origem;
- Especificar uma licença apropriada: Visa o estabelecimento das condições de uso (e não uso) dos dados que serão publicados. O reuso dos dados ocorre com maior

²Government Linked Data (GLD) Working Group - http://www.w3.org/2011/gld/wiki/Main_Page

frequência quando existe clareza sobre a sua origem, propriedade (autoria) e demais condições relacionadas ao uso destes dados;

- Estabelecer bons identificadores universais (URIs) para dados conectados: O núcleo dos dados conectados se baseia em um planejamento bem feito de identificação e referenciamento dos dados na web, baseados em URIs HTTP. Devem ser estabelecidos os requisitos para objetos de dados, suporte a diversos idiomas, alteração de dados ao longo do tempo e estratégia de persistência;
- Utilizar vocabulário padrão: Descrever objetos com vocabulários previamente definidos, sempre que possível. Quando necessário, ampliar tais vocabulários de acordo com a necessidade. Podem ser criados novos vocabulários (somente quando necessário) seguindo as melhores práticas para este fim;
- Converter Dados para Dados Conectados: Esta etapa visa estabelecer as condições para a conversão de conjuntos de dados (*dataset*) para dados conectados. Pode ser entendida como uma etapa de enriquecimento de dados. É uma atividade comumente realizada com o auxílio de ferramentas, scripts ou processos automatizados;
- Prover acesso automatizado aos dados: Devem ser providos diversos meios e formatos para que ferramentas de busca e outros recursos de processamento e consumo automatizado possam utilizar os dados mediante recursos padrões da Web;
- Anunciar para o público: Os novos dados publicados devem ser divulgados para que a comunidade possa fazer o devido uso. É importante destacar que ao se publicar um dado conectado é automaticamente estabelecido um contrato social implícito com seus usuários;
- Estabelecer um contrato social para os dados publicados: Esta etapa refere-se à necessidade de se reconhecer a responsabilidade em garantir a disponibilidade, manutenção e atualização dos dados publicados. Por serem conectados, tais dados precisam ter a garantia que vão ficar disponíveis onde a organização publicadora estabelece e que a disponibilidade será mantida ao longo do tempo.

Para apoiar a adoção destas práticas de publicação, serão apresentadas na próxima seção algumas ferramentas voltadas à publicação e também ao consumo de Dados Conectados.

4.2. Ferramentas para Publicação e Consumo de Dados Conectados

A produção de dados conectados requer um conjunto de requisitos de qualidade, sendo alguns deles obtidos a partir do cumprimento de boas práticas para a sua publicação. Para o estabelecimento de vocabulários e ontologias, são necessárias ferramentas de engenharia de ontologias ou ferramentas de organização de palavras-chave. Para o enriquecimento de dados, são utilizados geradores, sincronizadores, compactadores, armazenadores e publicadores de triplas RDF, além de conversores de dados e ambientes de programação. Para o provimento de acesso aos dados por máquina, são utilizados servidores de triplas que oferecem *endpoints SPARQL*, sendo estes últimos ambientes de

execução de consultas e consumo aos dados conectados. A Figura 1 apresenta uma visão geral das principais ferramentas para publicação e consumo de dados conectados.

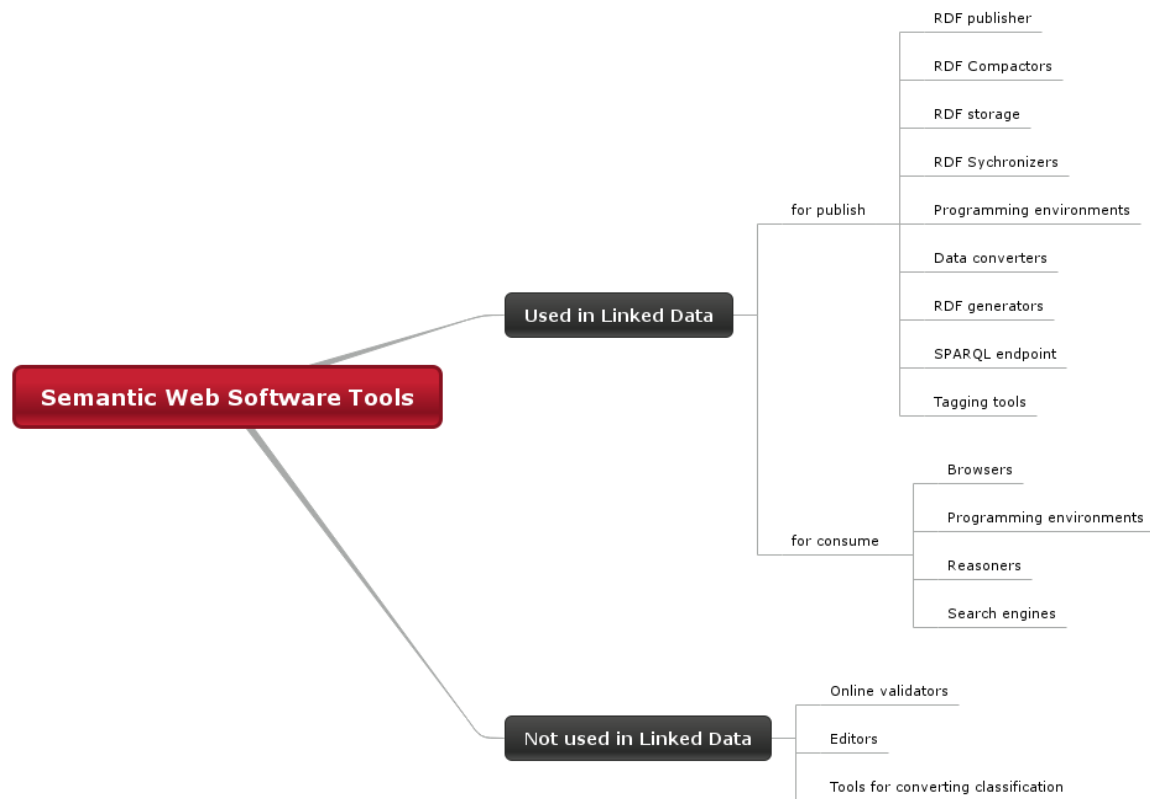


Figura 1. Mapa mental sobre ferramentas e os processos de publicação e consumo de dados conectados.

Para fins de consumo de dados conectados, a necessidade de uso de várias ferramentas é menor. Normalmente são utilizados ambientes de desenvolvimento de software, ferramentas de buscas em dados conectados e navegadores web.

A Figura 1 mostra a relação entre os tipos de ferramentas e como estas se relacionam na publicação e consumo de dados conectados. Uma lista contendo a descrição, bem como o link para a fonte dessas ferramentas pode ser encontrada do site do W3C³⁴. É importante salientar que dados conectados podem ser enriquecidos semanticamente com a utilização de ontologias. Para isso, é necessário a utilização de editores de ontologias, que não estão relacionadas explicitamente com ambos os processos, além dos editores outras ferramentas que não estão ligadas ao consumo ou publicação de dados conectados podem ser utilizadas para aprimorar o uso de ontologias.

5. Exemplos de Uso de Dados Conectados

Esta seção abordará sobre os casos reais em que são usados os dados conectados.

³ <http://www.w3.org/wiki/SemanticWebTools>

⁴ <http://www.w3.org/2001/sw/wiki/Tools>

5.1. DBPedia

A Wikipédia, existente desde 2001, armazena mais de 14 milhões de artigos em centenas de línguas e dialetos (com quase 900 mil artigos em português) e é considerada a maior enciclopédia digital do mundo, sendo editada e ampliada por centenas de colaboradores de todo o mundo. Devido ao seu tamanho e abrangência, ela já é considerada como um dos principais repositórios de conhecimento da humanidade, cujo conteúdo é reutilizado sistematicamente.

Assim como diversas outras grandes bases de dados na web, a Wikipédia passou a enfrentar grandes desafios no que tange a recuperação e reuso de seus conteúdos. Nesta direção, considerando a evolução dos conceitos e técnicas de ofertas de dados na web, como os Dados Conectados, surgiu o projeto DBPedia. A DBPedia é um projeto colaborativo para extração de dados da Wikipédia tornando-os disponíveis e recuperáveis na Web, permitindo a realização de consultas sofisticadas conectando diferentes conjuntos de dados existentes na Wikipédia. Dentre outros benefícios, a DBPedia permite novos mecanismos de navegação, conexão de dados e aprimoramento da Wikipédia.

A base de conhecimento da DBPedia, versão em inglês, descreve mais de 4 milhões de entidades classificadas em uma ontologia. Essa base de conhecimento contempla mais de 1,4 milhões de pessoas, cerca de 735 mil lugares, descrição de mais de 400 mil conteúdos multimídia (álbuns musicais, filmes, etc.), 241 mil organizações (contemplando 58 mil empresas e 41 mil instituições de ensino), dentre outros conteúdos relevantes. A *DBPedia* completa, em todos os idiomas atualmente contemplados, já contém 38,3 milhões de entidades armazenadas, contendo 25,2 milhões de links para imagens, 29,8 milhões de links para páginas externas e 80,9 milhões de links para categorias da Wikipédia. Além disso, está conectada com cerca de 50 milhões de outros conjuntos de dados conectados. Em 2014, a DBPedia alcançou um volume de três bilhões de informações estruturadas em triplas RDF [DBPedia 2015].

A produção de dados conectados da *DBpedia* é desenvolvida mediante diversas etapas, que são a leitura, interpretação, extração de dados a partir da Wikipédia, enriquecimento dos dados com o apoio de ontologias e geração de triplas RDF, que são armazenadas em um servidor de triplas. A partir desse servidor é disponibilizado um endpoint SPARQL⁵. Os dados conectados são consumidos e visualizados por páginas HTML (*HyperText Markup Language*) geradas a partir das triplas que proporcionam uma nova e rica experiência de navegação em torno dos dados da Wikipédia. A Figura 2 apresenta uma visão geral do framework de geração de dados conectados da *DBPedia*.

⁵<http://live.dbpedia.org/sparql>

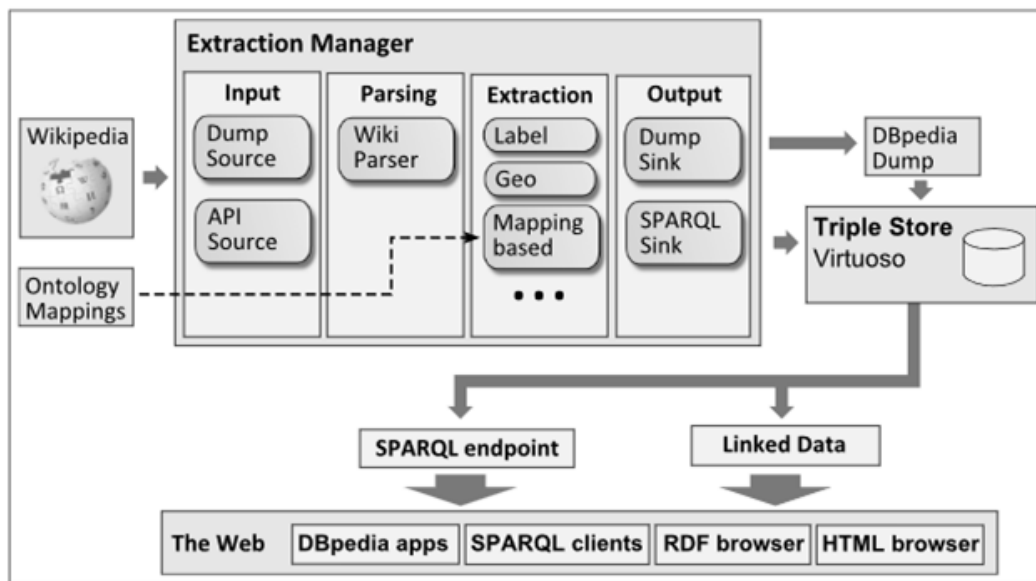


Figura 2. Visão geral do framework de geração e disponibilização de dados conectados da DBpedia [Lehmann et al. 2015]

Através desse framework, a base da *DBpedia* possibilita a execução de consultas relevantes e curiosas, como: (a) “Quais as universidades localizadas nos Estados Unidos da América?” ou (b) “Qual o total de estudantes da Universidade de Massachusetts?”. As Figuras 3 e 4 apresentam um exemplo de consulta SPARQL para executar a consulta (a) e uma página HTML gerada a partir de triplas RDF sobre a Universidade de Massachusetts que respondem a pergunta (b).

```

PREFIX geo: <http://www.w3.org/2003/01/geo/wgs84_pos#>
PREFIX d: <http://dbpedia.org/ontology/>

SELECT ?uni ?link ?lat ?long ?endowment
WHERE {
  ?s foaf:homepage ?link;
    rdf:type <http://schema.org/CollegeOrUniversity>;
    rdfs:label ?uni.
  OPTIONAL {?s geo:lat ?lat;
               geo:long ?long.
             ?s d:endowment ?endowment.}
  FILTER (LANGMATCHES(LANG(?uni), 'en'))

  {?s dbpedia2:country "U.S."@en .}
  UNION
  {?s dbpedia2:country "U.S." .}
  UNION
  {?s d:country :United_States .}
}
ORDER BY ?s
  
```

Figura 3: Consulta SPARQL na DBpedia para consultar as Universidades localizadas nos Estados Unidos da América [Taylor 2013]

About: University of Massachusetts Boston

An Entity of Type : Public university, from Named Graph : <http://dbpedia.org>, within Data Space : www.dbpedia.org



The University of Massachusetts Boston, also known as UMass Boston, is an urban public research university and the third-largest campus in the five-campus University of Massachusetts system. The university is located on 177 acres (0.72 km²) on what used to be known as the Columbia Point peninsula in the City of Boston, Massachusetts, United States, but became known as Harbor Point in the 1980s after development. UMass Boston is the only public university in Boston. Students are primarily from Massachusetts but also from other parts of the United States and from foreign countries.

| Property | Value |
|---|---|
| dbpedia-owl:abstract | <ul style="list-style-type: none"> The University of Massachusetts Boston, also known as UMass Boston, is an urban public research university and the third-largest campus in the five-campus University of Massachusetts system. The university is located on 177 acres (0.72 km²) on what used to be known as the Columbia Point peninsula in the City of Boston, Massachusetts, United States, but became known as Harbor Point in the 1980s after development. UMass Boston is the only public university in Boston. Students are primarily from Massachusetts but also from other parts of the United States and from foreign countries. |
| dbpedia-owl:city | <ul style="list-style-type: none"> dbpedia:Boston |
| dbpedia-owl:country | <ul style="list-style-type: none"> dbpedia:United_States |
| dbpedia-owl:facultySize | <ul style="list-style-type: none"> 800 (xsd:integer) |
| dbpedia-owl:mascot | <ul style="list-style-type: none"> The Beacon |
| dbpedia-owl:numberOfStudents | <ul style="list-style-type: none"> 15741 (xsd:integer) |
| dbpedia-owl:numberOfUndergraduateStudents | <ul style="list-style-type: none"> 11866 (xsd:integer) |
| dbpedia-owl:officialSchoolColour | <ul style="list-style-type: none"> Pantone 7462 , and White |
| dbpedia-owl:state | <ul style="list-style-type: none"> dbpedia:Massachusetts |
| dbpedia-owl:thumbnail | <ul style="list-style-type: none"> http://en.wikipedia.org/wiki/Special:FilePath/UMASSBOSTON_ID_blue.v2.png?width=300 |
| dbpedia-owl:type | <ul style="list-style-type: none"> dbpedia:Public_university |

Figura 4. Página sobre a Universidade de Massachusetts gerada a partir dos dados conectados da DBpedia [DBpedia 2015]⁶

A *DBpedia*, devido a sua relevância quanto à oferta de Dados Conectados no mundo já é fonte de conhecimento para cerca de 40 projetos e diversos casos reais de uso, sendo utilizada por grandes empresas ao redor do mundo, como a BBC [Kobilarov et al. 2009].

5.2. LODSpringer

A *Springer* é uma editora com presença global (aproximadamente em 25 países do mundo) especializada em publicações científicas. Atualmente, é responsável por mais de 2.900 periódicos (*journals*) e mais de 200.000 livros, em diversas áreas do conhecimento englobando temática nas ciências exatas, agrárias, da saúde e humanas.

Decorrente desse expressivo volume de informações gerenciadas, a Springer deparou-se com necessidades de melhoria da sua gestão interna e de fortalecimento da área de marketing e vendas, bem como, de aprimorar a gestão do seu relacionamento com seus clientes. A partir dessas necessidades, a *Springer* passou a analisar o seu principal ativo, que é a sua imensa base de informações e conhecimento científico e, a partir desse ativo, visualizou diversas oportunidades para: (i) Identificar e melhorar o desenvolvimento de produtos; (ii) Identificar clientes potenciais por meio da análise de estatísticas de usuários, e ainda, (iii) Oferecer novos produtos e deixar artigos e livros científicos cada vez mais “descobriáveis”, ampliando a sua disseminação e uso pela comunidade científica internacional. Foi estabelecida a seguinte questão-chave para nortear a busca por soluções para essas necessidades e motivações: “Como podemos deixar nossos dados mais enriquecidos para melhorar a usabilidade e o reuso de conteúdos?” [Birukou 2014].

Como resposta a essa questão-chave, a Springer estabeleceu o projeto *LODSpringer* que visa estabelecer uma base de dados conectados a partir do acervo de conteúdos existente na editora. Como projeto piloto, visa, dentre outras atividades, adicionar aos

⁶ http://www.dbpedia.org/page/University_of_Massachusetts_Boston

metadados dos conteúdos diversas informações das conferências em que os artigos são publicados de forma estruturada e seguindo os princípios dos dados conectados. A Figura 5 apresenta um exemplo de enriquecimento dos dados dos anais de uma conferência internacional sobre Web Semântica.



Figura 5: Enriquecimento de dados dos anais da conferência The Semantic Web ISWC 2013 [Birukou 2014]

Até o mês de maio/2015, a *LODSpringer*⁷ já estava composta de cerca de 8000 anais de conferências científicas de diversas áreas da Computação, publicadas em livros como: *LNCS* (*Lecture Notes in Computer Science*), *LNBIP* (*Lecture Notes in Business Information Processing*), *CCIS* (*Communications in Computer and Information Science*), *IFIP-AICT* (*Advances in Information and Communication Technology*) e *LNICST* (*Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*). Até o final de 2015 estão previstas a incorporação de mais de 2000 anais de congressos a *LODSpringer*. Todos esses dados conectados estão acessíveis atualmente através de um *endpoint SPARQL* (<http://lod.springer.com/sparql-form/index.html>), onde diversas consultas podem ser realizadas sobre as conferências editoradas pela Springer, conforme o exemplo contido no Quadro 1.

```
PREFIX spr-p: <http://lod.springer.com/data/ontology/property/>
PREFIX spr-c: <http://lod.springer.com/data/ontology/class/>
SELECT ?conference
WHERE {graph ?g {
  ?conference a spr-c:Conference .
  ?conference spr-p:confCity "Paris"@en
}}
```

Quadro 1. Consulta que exhibe todas as conferências realizadas na cidade de Paris na *LODSpringer*

Além disso, foi disponibilizado a ferramenta *LODLive*⁸ que permite a qualquer usuário fazer consultas interativas visualizando as conexões entre o conteúdo da *LODSpringer*. A Figura 6 apresenta um exemplo de conexão entre as conferências, respectivas edições e anais representadas graficamente na *LODSpringer*.

⁷ <http://lod.springer.com>

⁸ <http://lod.springer.com/live/>

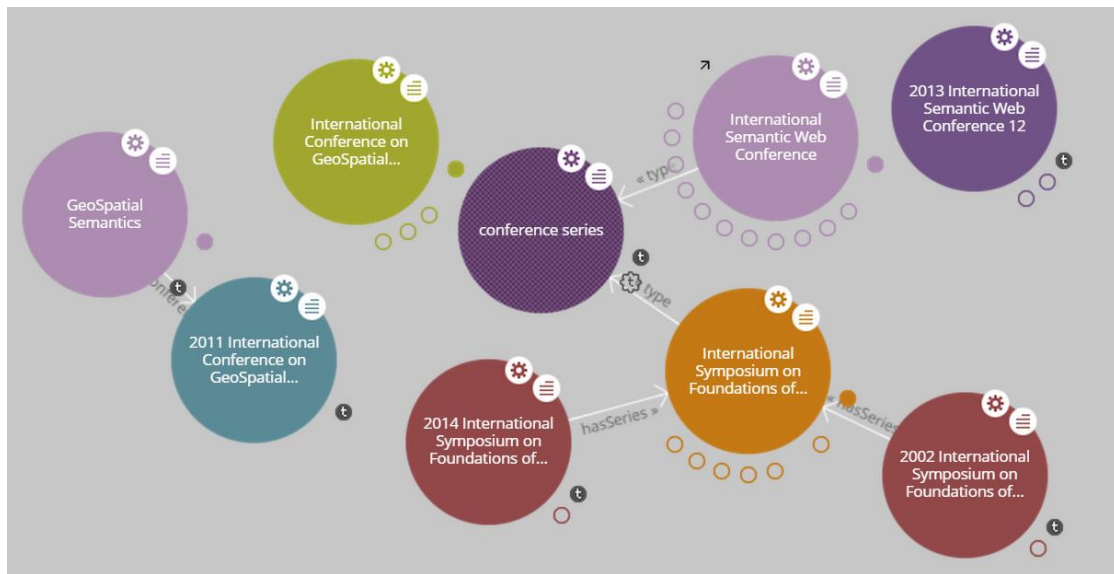


Figura 6: Representação de conferências sobre Web Semântica, Dados Geoespaciais, Inteligência Artificial e Gestão do Conhecimento na LODSpringer.

Com essa iniciativa inovadora baseada em Dados Conectados, a Springer busca atingir diversos resultados relevantes como: (i) Aumentar o fluxo de uso dos documentos, submissões, o atendimento aos conferencistas; (ii) Prover colaboração com a comunidade de pesquisa internacional; (iii) Prover colaboração com a comunidade de bibliotecas (iniciativas de bibliotecas conectadas) oferecendo às bibliotecas melhores serviços de dados; (iv) Melhorar o planejamento interno da editora, com a identificação de novas conferências relevantes; (v) Produzir e disponibilizar informações enriquecidas sobre um artigo ou conferência; (vi) Gerar valor adicional para autores e organizadores de conferências; e (vii) ampliar o seu quantitativo de leitores, citações e downloads [Burikou 2014]. A Figura 7 mostra um exemplo de captura de dados de proveniência de uma determinada conferência armazenada na *LODSpringer*, permitindo ao usuário obter diversas informações sobre a sua procedência e relevância.

| Provenance information for RDF data: rdf/conference/semweb2014 | |
|---|---|
| URI: | http://lod.springer.com/data/rdf/conference/semweb2014 |
| Property | Value |
| dc:creator | ■ Springer (en) |
| dc:date | ■ 2015-04-01 (xsd:date) |
| dc:description | ■ Meta-data about conference proceedings published by Springer. Includes the following series: LNCS (Lecture Notes in Computer Science), LNBIP (Lecture Notes ... »more» (en) |
| void:inDataset | ■ < http://lod.springer.com/data/dataset/springer/conferences > |
| rdfs:label | ■ Provenance information for RDF data: rdf/conference/semweb2014 (en) |
| dc:publisher | ■ Springer (en) |
| dc:rights | ■ CC0 (en) |
| rdf:type | ■ void:Dataset |

Figura 7: Informações de proveniência da International Semantic Web Conference na LODSpringer.

Como benefícios futuros do uso dos dados conectados, o projeto espera contribuir para a transparência na ciência (*Open Science*), contribuir para a investigação e avaliação de certificação de entidades relacionadas com a pesquisa científica (Conferências, Universidades e Grupos de pesquisa) visando um cenário de “Ciência conectada”, onde serão adicionadas novas conexões a uma grande nuvem de conhecimento científico conectado na decorrência da produção de novos dados conectados [Burikou 2014].

5.3. Semântica.Globo.com

O Globo.com é um portal e provedor de Internet pertencente ao grupo de mídia brasileiro Globo. Possui aproximadamente 500 mil assinantes, além de hospedar quase 700 sites próprios e filiados. O Globo.com é o 6º website mais acessado do Brasil, e o 112º no mundo, segundo dados da *Alexa*, e é estruturado em cinco pilares de conteúdo: notícias, esportes, entretenimento, tecnologia e vídeos. Para cada área de conteúdo, existe um portal específico vinculado ao Globo.com [Wikipédia 2015].

Como um grande portal de mídia, o Globo.com e seus portais vinculados acumulam um volume expressivo de conteúdo multimídia, composto por textos, imagens e vídeos. Com a missão de organizar e distribuir as informações das Organizações Globo, foi instituído o projeto Semântica.Globo.com, que está estruturando a base multimídia da organização utilizando conceitos de Web Semântica, Ontologias e Dados Conectados.

Dentre os desafios existentes com o uso desses conceitos estão: a geração automatizada de conteúdo relacionado, a identificação e representação de entidades únicas que são objetos de produção de conteúdos pelos portais do Globo.com, a

estruturação da base de conteúdos, agilidade na recuperação de dados, aprimoramento da qualidade dos dados (eliminação de conteúdos duplicados, validação de conteúdos, integração com bases externas), simplicidade na autorização e autenticação de acesso aos conteúdos, melhoria do desempenho dos portais, dentre outros [Medeiros 2013 (a)]. A Figura 8 e a Figura 9 representam o modelo de integração entre os dados e como os dados são representados, respectivamente, na infraestrutura da Globo.com.

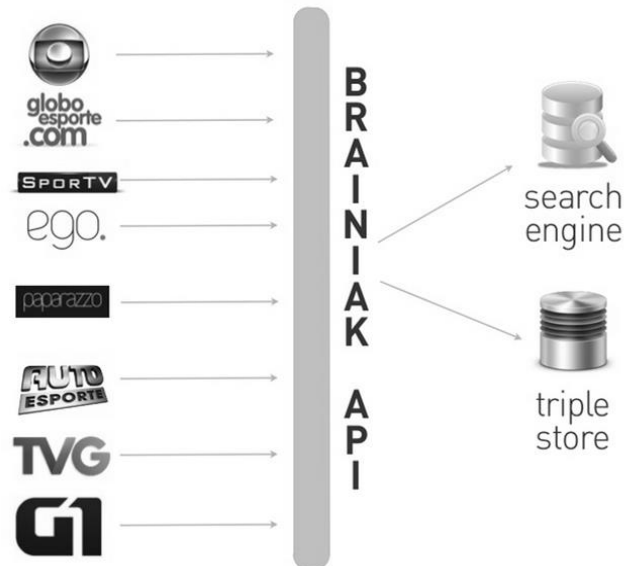


Figura 8: Modelo de integração dos dados do Globo.com [Medeiros 2013 (b)]

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.
@prefix foaf: <http://xmlns.com/foaf/0.1/>.
@prefix dbpedia: <http://dbpedia.org/ontology>

<http://semantica.globo.com/esportes/Jogador/Neymar>
  rdf:type          dbpedia:SoccerPlayer ;
  rdfs:label        "Neymar" ;
  foaf:name          "Neymar da Silva Santos Júnior" ;
  dbpedia:birthDate "05/02/1992" .
```

Figura 9: Exemplo de representação RDF, em formato *Turtle*, do Globo.com [Medeiros 2013 (b)]

O projeto Semântica.Globo.com estima, num futuro breve, suportar a navegação baseada em inferências nos portais Globo, interoperabilidade com grandes bases externas como a *DBPedia* e Catálogos de Dados Governamentais, enriquecimento de conteúdos permitindo a geração de conteúdos relacionados de forma automatizada, dentre outros benefícios [Medeiros 2013 (a)].

5.4. BBC

A BBC é a maior empresa de transmissão de conteúdo de mídia no mundo. Sua missão é levar informação, educação e diversão à vida das pessoas.

Portais de transmissão de conteúdo de mídia publicam uma grande quantidade de dados, como texto, áudio e vídeo. Historicamente, tendem a criar subdivisões em seus portais de acordo com cada contexto, gerando sites de domínios específicos para cada

um. Dessa forma, esses sites não são integrados entre si, dificultando a navegação e a busca por um determinado conteúdo. Por exemplo, era possível obter informação sobre um programa de TV a partir de uma página web, mas não era viável obter também mais informações sobre os artistas que participam do programa, a partir de outras páginas web. Sendo praticamente impossível obter todas as informações publicadas sobre um determinado artista.

Além disso, a BBC possuía uma grande diversidade de programas, cerca de 1500, e não conseguia manter informações atualizadas na web sobre todos. Criar sites específicos para cada um tornava difícil a manutenção e, muitas vezes, esses websites acabavam sendo deixados sem manutenção ou até mesmo removidos.

Em 2007, a BBC lançou a plataforma *BBC Programmes*⁹, cujo objetivo era definir um identificador único na web para cada programa, com uma associação entre as páginas HTML e *feeds* em formato legível por máquina (RDF/XML, JSON e XML), permitindo que outros grupos de trabalho dentro da própria BBC pudessem conectar essas páginas a novos sites, ou sites já existentes de programas. [Raimond et al. 2010]

A BBC também tem aplicado técnicas de dados conectados para enriquecer a experiência de busca do usuário, utilizando o *BBC Controlled Vocabulary* (CV) para classificar os termos de busca e fornecer os melhores resultados de conteúdo para um determinado assunto. Cada termo do BBC CV é conectado a um recurso associado ao *DBpedia*, permitindo assim, que a BBC aproveite os metadados disponibilizados pelo *DBpedia* [Kobilarov et al. 2009; Raimond et al. 2010], conforme ilustrado na Figura 10.

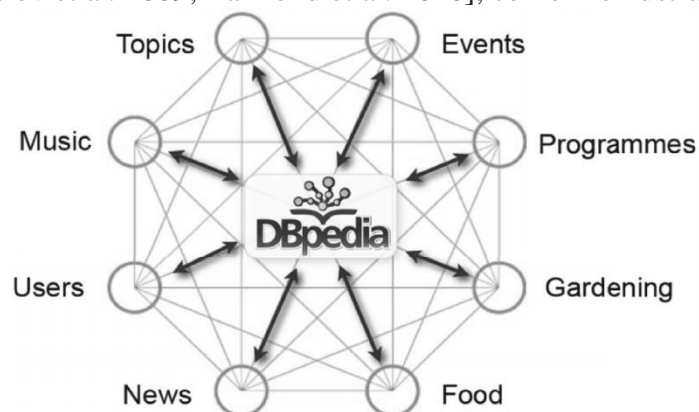


Figura 10 - Conectando termos entre BBC e DBpedia [Kobilarov et al. 2009]

5.5. Boa Moradia

O Boa Moradia é um software destinado a pessoas que objetivam mudar de cidade e estão à procura de um imóvel para alugar ou comprar, utilizando dados abertos conectados para recomendar imóveis em um bom lugar para viver.

Para realizar tal propósito, a aplicação utiliza informações como a cidade de destino e preferências do usuário. As preferências são informações fornecidas pelo usuário e descrevem o que o usuário espera que exista no novo local para morar. Por exemplo, “desejo morar próximo de um hospital, escola e padaria”. Quanto ao seu funcionamento, na página inicial do sistema, o usuário informa a cidade e em seguida os locais que estejam próximos ao seu novo imóvel (por exemplo: hospital, escola e padaria). Para recomendar as moradias, o software inicialmente estabelece um ranking de qualidade de vida dos bairros existentes na cidade a partir da utilização de dados governamentais sob

⁹ <http://www.bbc.co.uk/programmes>

perspectivas de educação, saúde e segurança. Para isso, os dados são enriquecidos, conectados e incorporados à base de conhecimento do software. Além deste ranking de qualidade de vida, a aplicação considera as preferências informadas pelo usuário, e apresenta resultados semânticos e personalizados.

Os dados utilizados para esse fim são os dados abertos governamentais, como os da prova Brasil e ENEM, que são fornecidos pelo INEP. Esses dados representam: taxa de rendimento das escolas e Unidades Básicas de Saúde (UBS) fornecidos pelo Portal Brasileiro de Dados Abertos¹⁰, malhas censitárias e cadastro nacional de endereços fornecidos pelo Instituto Brasileiro de Geografia e Estatística - IBGE, dentre outros dados de Governos locais, como o Alagoas em Dados e Informações, da Secretaria de Planejamento de Alagoas¹¹.

Os dados providos em diversos formatos por esses portais são transformados em RDF. Os códigos utilizados para identificar estados, cidades, bairros, escolas e UBS, são utilizados para formar as URIs utilizadas pelo sistema, conectando e fazendo equivalência entre dados de diferentes fontes. Abaixo, o Quadro 2 apresenta uma consulta SPARQL que será executada na base de conhecimento do Boa Moradia. A consulta busca por sete bairros da cidade representada pela URI “http://nees.com.br/linkn/onto/locality/city/2704302”.

```
PREFIX loc: < http://nees.com.br/linkn/onto/locality/>
SELECT *
FROM <http://nees.com.br/linkn/data/br/al/address/>
WHERE {
  <http://nees.com.br/linkn/onto/locality/city/2704302> loc:hasLocation ?neighborhood.
  ?neighborhood a loc:Neighborhood ;
                loc:haslocalityName ?name ; loc:hasLocalityCode ?code .
}
ORDER BY ?name
LIMIT 7
```

Quadro 2. Consulta SPARQL que exibe sete bairros do município de Maceió no BoaMoradia.

A Tabela 1 apresenta os resultados obtidos a partir da execução da consulta SPARQL do Quadro 2.

Tabela 1. Resultado da consulta SPARQL que exibe sete bairros do município de Maceió no BoaMoradia.

| Neighborhood | Name | Code |
|---|-----------------|--------------|
| http://nees.com.br/linkn/onto/locality/neighborhood/270430205001 | Antares | 270430205001 |
| http://nees.com.br/linkn/onto/locality/neighborhood/270430205002 | Barro Duro | 270430205002 |
| http://nees.com.br/linkn/onto/locality/neighborhood/270430205003 | Bebedouro | 270430205003 |
| http://nees.com.br/linkn/onto/locality/neighborhood/270430205004 | Benedito Bentes | 270430205004 |
| http://nees.com.br/linkn/onto/locality/neighborhood/270430205005 | Bom Parto | 270430205005 |
| http://nees.com.br/linkn/onto/locality/neighborhood/270430205006 | Canaã | 270430205006 |

¹⁰ <http://dados.gov.br/>

¹¹ <http://www.dados.al.gov.br>

| | | |
|---|--------|--------------|
| http://nees.com.br/linkn/onto/locality/neighborhood/270430205007 | Centro | 270430205007 |
|---|--------|--------------|

Após estabelecer os conceitos sobre dados educacionais, dados abertos, dados conectados e apresentação de diversos casos reais de uso de dados conectados em empresas e projetos de conteúdo em larga escala, a próxima seção contextualizará o uso de dados abertos conectados no contexto educacional.

6. Dados Abertos Conectados aplicados à Educação

Esta seção apresenta benefícios, vantagens e limitações que surgem na aplicação de dados abertos conectados no contexto da educação. Além disso, serão apresentadas iniciativas de aplicação de dados abertos conectados para a educação.

6.1 Benefícios

Apesar de existirem desafios relacionados à utilização de dados abertos conectados na educação, tais como a necessidade de maior estímulo à publicação de dados educacionais em formato aberto, diversificar a oferta de dados em diferentes formatos, aplicar dados conectados a repositórios educacionais abertos, dentre outros [Alcantara et al. 2015], a utilização de dados conectados em educação pode trazer benefícios, tais como [Dietze et al. 2013]:

- **Integração de serviços educacionais:** Por meio da utilização de descrições RDF para descrever os principais elementos de serviços e interfaces de programação (APIs) é possível descobrir serviços e executá-los de forma integrada;
- **Integração de dados educacionais:** Por meio do enriquecimento dos dados com vocabulários RDF e adicionando links para recursos externos.

6.2 Vantagens

Além dos benefícios citados, a utilização de dados conectados em educação gera vantagens como [Vega-Gorgojo et al. 2015]:

- **Reuso de dados:** Existência de grande quantidade de dados conectados para diversas necessidades de aplicação;
- **Enriquecimento:** Enriquecimento de dados educacionais por meio da utilização de vocabulários e de conexões com outros dados.
- **Disponibilidade:** Prover a interoperabilidade de sistemas através de uma interface uniforme para acessar os dados educacionais conectados.
- **Integração:** A união de dados educacionais é facilitada pelo uso da Web Semântica.
- **Descoberta de dados:** A exploração das conexões torna possível encontrar novos dados.
- **Internacionalização:** A oferta de conjuntos de dados em múltiplos idiomas oferece meios para a internacionalização de aplicações, bem como dos dados.
- **Precisão:** As anotações presentes em dados conectados são mais precisas que anotações presentes em documentos baseados em texto.

6.3 Limitações

Foram identificadas algumas limitações na aplicação dos dados abertos conectados no contexto da educação, que são os que seguem [Vega-Gorgojo et al. 2015]:

- Qualidade dos dados: Lidar com informações incorretas, incompletas ou inesperadas a partir da Web dos dados;
- Controle dos dados: Falta de controle de fontes de dados, sendo possível haver indisponibilidade de algumas fontes;
- Privacidade dos dados: Preocupações com exposição de informações, as quais possam levar a descoberta de informações sigilosas através da conexão de dados;
- Proveniência dos dados: Necessidade de conhecimentos para saber como diferentes fontes contribuem para a procedência de um determinado conjunto de dados;
- Licenciamento dos dados: Conhecimento sobre questões legais para reuso de dados conectados;
- Aceitação de vocabulários: Necessidade de entendimento dos vocabulários disponíveis para o domínio da educação;
- Perda de informação: Possibilidade de perda de informações decorrentes de mapeamentos imprecisos de dados;
- Disponibilidade de vocabulário: Indisponibilidade de vocabulários necessários a subdomínios específicos no contexto educacional;
- Custo de enriquecimento: Custo inicial para enriquecimento e estabelecimento de conexões entre os dados educacionais;
- Custo de publicação: Custo inicial para estruturar o processo e as ferramentas de publicação de dados conectados;
- Custo computacional: Dependendo do volume da base de conhecimento, o potencial de interatividade sobre dados conectados pode demandar um maior custo computacional para processamento e disponibilidade dos dados;
- Interação: Prover novos recursos de interação e visualização de dados conectados.

6.4 Iniciativas

Apesar de ser uma área recente, já existem algumas iniciativas que visam estimular a produção e uso de dados abertos conectados na educação. Por exemplo, na União Europeia, o projeto *Linked Universities*¹² consiste de uma aliança de universidades europeias engajadas em disponibilizar seus dados públicos como dados conectados. Neste contexto, o projeto identifica que ainda existem poucas universidades disponibilizando seus dados de forma conectada, utilizando tecnologias como RDF e SPARQL para permitir acesso a recursos educacionais como publicações, cursos, materiais didáticos, etc., e que essas iniciativas ainda encontram-se desconectadas uma das outras.

O projeto *Linked Universities* tem como objetivo geral criar um ambiente de compartilhamento de dados conectados na web, oriundos de diferentes instituições,

¹² <http://linkeduniversities.org>

contribuindo para a ampliação da Web dos Dados. Especificamente o projeto visa alcançar os seguintes resultados:

- Identificar, apoiar e desenvolver uma base comum de vocabulários para dados conectados que seja utilizável pelas universidades, compartilhando conceitos comuns como cursos, qualificações, material didático, etc;
- Descrever metodologias e ferramentas de reuso e compartilhamento para apresentar os dados conectados em universidades;
- Apoiar iniciativas voltadas à apresentação de dados conectados nas universidades através de experiências de compartilhamento e reuso.

Atualmente, o projeto conta com 11 Universidades integrantes de diversos países europeus, como o Reino Unido, Alemanha, Grécia, Finlândia, Turquia, Suécia, República Tcheca e Espanha.

Ademais, no site do projeto são apresentados os repositórios de dados conectados, acessíveis através de catálogos de dados e *endpoints SPARQL*, conforme ilustram as Figuras 11 e 12.

The screenshot shows the 'Open Data Service' page of the University of Southampton. The page has a dark sidebar on the left with a search bar and a list of navigation links: Home, 5★ Data, Frequently Asked Questions, Apps & Tools (with sub-links: Open Data Map, What Week Is It?, iSolutions Workstations, Room Finder), Data Catalogue, Events Calendar, Catering, Bus Information, University Map, Places, Phonebook, and Academic Programmes. The main content area is titled 'Open Data Service' and 'Data Catalogue'. It includes a paragraph explaining that the page serves as a list of all datasets, available in Turtle, RDF/XML, HTML, and SPARQL formats. Below this, there are three sections: 'Academic Courses' with the URI <http://id.southampton.ac.uk/dataset/courses> and a description 'Academic courses and programs for 2014 onwards'; 'Academic Sessions' with the URI <http://id.southampton.ac.uk/dataset/university-timetable> and a description 'Key dates within the University of Southampton's timetable, specifically semester start and end dates.'; and 'Apps using our data' with the URI <http://id.southampton.ac.uk/dataset/apps> and a description 'A catalog of known websites, web widgets, phone apps and other tools using our datasets.'

Figura 11. Catálogo de dados da Universidade de Southampton, Reino Unido

UNIVERSITY OF
Southampton

SPARQL Endpoint

+ Show Help and Examples

```
1 PREFIX soton: <http://id.southampton.ac.uk/ns/>
2 PREFIX foaf: <http://xmlns.com/foaf/0.1/>
3 PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
4 PREFIX geo: <http://www.w3.org/2003/01/geo/wgs84_pos#>
5 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
6 PREFIX org: <http://www.w3.org/ns/org#>
7 PREFIX spacerel: <http://data.ordnancesurvey.co.uk/ontology/spatialrelations/>
8 PREFIX ep: <http://eprints.org/ontology/>
9 PREFIX dct: <http://purl.org/dc/terms/>
10 PREFIX bibo: <http://purl.org/ontology/bibo/>
11 PREFIX owl: <http://www.w3.org/2002/07/owl#>
12 PREFIX oo: <http://purl.org/openorg/>
13
14 SELECT * WHERE {
15   ?s ?p ?o .
16 } LIMIT 10
```

Send Query Reset

Options

Output format (if supported by query type):
HTML Table ▼

jsonp/callback (for JSON results)
[]

Show results inline:
☒

Change HTTP method: GET POST

Figura 12. Endpoint SPARQL da Universidade de Southampton, Reino Unido

Outra iniciativa é a *LinkedEducation.org*¹³: é uma plataforma aberta que visa promover o uso de dados conectados educacionais mediante o incentivo ao compartilhamento de recursos e informações relacionadas à educação. O projeto provê um fórum para que os pesquisadores e profissionais que atuam no âmbito da Web e tecnologias semânticas possam compartilhar e discutir a respeito de conjuntos de dados, esquemas, aplicações e eventos para identificar melhores práticas. Desta maneira, considerando que a interconexão de dados é um dos princípios fundamentais dos dados para a promoção da interoperabilidade, *LinkedEducation.org* visa identificar as melhores práticas e potenciais relações entre recursos de dados a fim de contribuir para uma Web educacional bem interligada de dados.

O *website* da iniciativa permite aos interessados terem acesso a diversos projetos, ferramentas, esquemas de dados, vocabulários, ontologias, bem como eventos e atividades relacionadas à publicação e uso de dados conectados na educação. Através da ferramenta *LinkedUp Data Repository*¹⁴, vinculada à iniciativa, é possível acessar e interagir com diversos catálogos de dados e *endpoints SPARQL* de projetos que ofertam dados conectados para a educação, como o Education.data.gov.uk (Governo do Reino Unido), Instituto de Estatísticas da UNESCO, Banco Mundial, Universidade de Oxford, Conselho Nacional de Ciência e Tecnologia da Itália, dentre outras, conforme ilustra a Figura 13.

¹³<http://www.linkededucation.org>

¹⁴<http://data.linkededucation.org/linkedup/catalog/browse/>

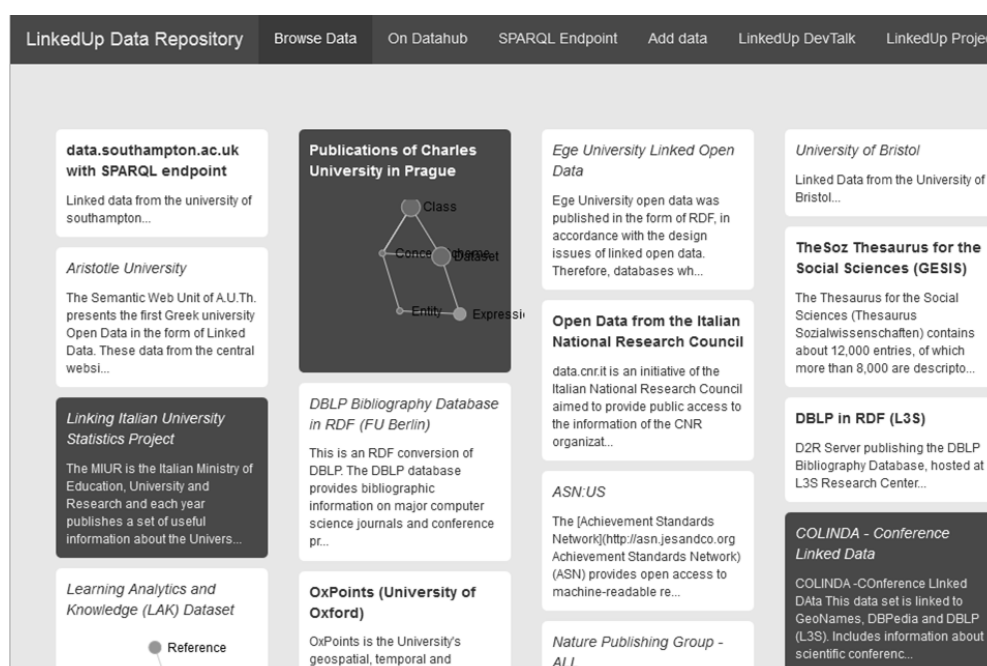


Figura 13. Ferramenta LinkedUp Data Repository, que visualiza fontes de dados conectados para a educação mapeados por LinkedEducation.org

7. Conclusão

O presente artigo discutiu sobre o potencial dos dados abertos conectados para promover o aprimoramento da educação com ênfase na produção, compartilhamento e reuso de dados e informações. Mesmo ainda sendo um conceito de vanguarda e tendo algumas limitações e desafios, a existência de iniciativas concretas em empresas e, principalmente, em diversas instituições de caráter educacional apresentam os dados abertos conectados como um poderoso recurso para o desenvolvimento do conhecimento global e, conseqüentemente, da educação.

Considerando o referencial teórico, as iniciativas analisadas e a vivência dos autores, podemos sugerir que os dados abertos conectados podem impulsionar atividades de ensino, pesquisa, extensão e a gestão educacional, conforme tópicos abaixo:

ENSINO

- Contribuir com o aprimoramento do material didático das disciplinas, pois os professores poderão reutilizar os conteúdos dos anos anteriores e atualizá-los para o ano vigente. Além disso, enriquecer o conteúdo apontando para outros recursos, como vídeos, áudios, etc;
- Contribuir com o aprendizado dos alunos, que terão acesso ao material didático, avaliações e registros de atividades das disciplinas em anos anteriores;
- Contribuir com a interdisciplinaridade, provida através das conexões para conteúdos relacionados no material didático do estudante, utilizando recursos educacionais como vídeos, áudios, mapas e etc.

PESQUISA

- Contribuir com a continuidade e o aprimoramento de pesquisas, que poderão reutilizar experiências e resultados de pesquisas relacionadas que ficarão disponíveis em repositórios de dados abertos conectados;
- Facilitar a integração de pesquisas na mesma instituição e entre várias instituições de ensino, decorrente do fácil acesso e descoberta de dados e resultados das mesmas;
- Facilitar a validação e procedência de referências bibliográficas, permitindo a navegação interativa entre as pesquisas.

EXTENSÃO

- Incentivar novos projetos de extensão como aprimoramento aos projetos de pesquisa existentes e disponíveis nos repositórios de dados conectados;
- Fortalecer núcleos de inovação tecnológica e incubadoras que terá uma rica oferta de conhecimento decorrente das pesquisas que subsidiará o reuso desse conhecimento para o desenvolvimento de novos produtos e serviços a serem realizados por empreendedores em parceria com os pesquisadores responsáveis pelas pesquisas;
- Incentivar a inovação, permitindo o desenvolvimento de produtos e produção de patentes a partir de pesquisas existentes e disponíveis nos repositórios de dados conectados.

GESTÃO

- Facilitar o acompanhamento, por gestores, do andamento das pesquisas e projetos desenvolvidos pelos grupos de pesquisa e unidades acadêmicas;
- Permitir que o cidadão/sociedade tenha maior envolvimento com os projetos, pesquisas e conhecimento desenvolvido pela Universidade, decorrente da possibilidade de acesso aos dados e informações;
- Permitir que novas aplicação sejam geradas para melhorar a gestão escolar em todos os níveis;
- Gerir o desenvolvimento dos estudantes de forma independente, bem como analisar o desenvolvimento de instituições de ensino.

Para que a aplicação de dados conectados possa se tornar uma realidade mais aplicável ao contexto educacional, em maior escala, recomenda-se o desenvolvimento de pesquisas e projetos que incorporem esses conceitos em soluções educacionais voltadas a educação básica e superior. Ainda, podem ser desenvolvidas investigações sobre o uso de dados conectados para o ensino, pesquisa, extensão e gestão conforme as possibilidades descritas nesse artigo.

Referências

Alcantara, W., Bandeira, J., Barbosa, A., Lima, A., Ávila, T., Bittencourt, I. & Isotani, S. (2015). Desafios no uso de Dados Abertos Conectados na Educação Brasileira. Anais do DesafiE - 4º Workshop de Desafios da Computação Aplicada à Educação. CSBC 2015. Recife: Sociedade Brasileira de Computação.

- Bandeira, J., Alcantara, W., Barbosa, A., Ávila, T., Bittencourt, I. & Isotani, S. (2014). Dados Abertos Conectados na Educação Brasileira. Jornada de Atualização em Tecnologia da Informação. Anais do III Simpósio Brasileiro de Tecnologia da Informação - SBTI 2014.
- Berners-Lee, T. (2006). Design issues: Linked data. <http://www.w3.org/DesignIssues/LinkedData.html>, Julho de 2015.
- Birukou, A. (2014). Linked Data Initiatives at Springer Verlag, <http://pt.slideshare.net/birukou/20140310-springerlod-stpetersburg>, Julho de 2015.
- DBpedia. (2015). About DBpedia, <http://wiki.dbpedia.org/about>, Julho de 2015.
- Dietze, Stefan, et al. "Interlinking educational resources and the web of data: A survey of challenges and approaches." *Program* 47.1 (2013): 60-91.
- Eaves, D. (2009). The Three Laws of Open Government Data, <http://eaves.ca/2009/09/30/three-law-of-open-government-data/>, Outubro de 2014.
- Heath, T. (2011). Linked Data — Welcome to the Data Network. IEEE Internet Computing archive. Volume 15 Issue 6. Pages 70-73
- Kobilarov, G., Scott, T., Raimond, Y., Oliver, S., Sizemore, C., Smethurst, M. & Lee, R. (2009). Media meets semantic web—how the BBC uses DBpedia and linked data to make connections. In *The semantic web: research and applications* (pp. 723-737). Springer Berlin Heidelberg.
- Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P., Hellmann S., Morsey, M., Van Kleef, P., Auer, S. & Bizer, C. (2015). DBpedia – A Large-scale, Multilingual Knowledge Base Extracted from Wikipédia. *Semantic Web Journal*, Vol. 6 No. 2, pp 167–195.
- Medeiros, I. (2013-a). Linked Data at Globo.com. Web of Linked Entities - WWW 2013, <http://pt.slideshare.net/icaromedeiros/linked-data-at-globocom>, Julho de 2015.
- Medeiros, I. (2013-b). Web Semântica. Curso Novas Mídias, UFRJ, <http://pt.slideshare.net/icaromedeiros/apresentacao-ufrj-icaro2013>, Julho de 2015.
- Neves, O. M. C. (2013). Evolução Das Políticas De Governo Aberto No Brasil. Anais do VI Congresso Brasileiro de Gestão Pública – CONSAD. Brasília, Brasil, <http://consadnacional.org.br/wp-content/uploads/2013/05/092-EVOLU%C3%87%C3%83O-DAS-POL%C3%8DTICAS-DE-GOVERNO-ABERTO-NO-BRASIL.pdf>, Outubro de 2014.
- Raimond, Y., Scott, T., Sinclair, P., Miller, L., Betts, S. & Mcnamara, F. (2010). Case study: use of semantic web technologies on the BBC web sites. W3C Semantic Web Use Cases and Case Studies.

- Rigotti, J. I. R. & Cerqueira, C. A. (2004). As bases de dados do INEP e os indicadores educacionais: conceitos e aplicações. RIOS-NETO, ELG e RIANI, J. de LR Introdução à Demografia da educação. Campinas: ABEP.
- Taylor, J. (2015). DBpedia SPARQL Query US Universities. Stack Overflow, <http://stackoverflow.com/questions/17056476/dbpedia-sparql-query-us-universities>, Julho de 2015.
- Vega-Gorgojo, G., Asensio-Pérez, J. I., Gómez-Sánchez, E., Bote-Lorenzo, M. L., Munoz-Cristobal, J. A. & Ruiz-Calleja, A. (2015). A Review of Linked Data Proposals in the Learning Domain. Journal of Universal Computer Science, 21(2), 326-364.
- Wikipédia. (2015). Verbete sobre o portal Globo.com, <https://pt.wikipedia.org/wiki/Globo.com>, Julho de 2015.
- Wood, D., Zaidman, M., Ruth, L. & Hausenblas, M. (2014). Linked Data: Structured data on the Web. New York: Manning Publications Co.
- W3C. World Wide Web Consortium. (2014). Best Practices for Publishing Linked Data, <http://www.w3.org/TR/ld-bp/>, Julho de 2015.