

Análise de Repositórios de REAs em Relação ao Uso dos Padrões de Linked Open Data¹

Luciano Bernardes de Paula², William Simão de Deus³, Ellen Francine
Barbosa⁴

Resumo

Recursos Educacionais Abertos (REAs) são uma parte muito importante para a Educação Aberta. Esses recursos são compartilhados de forma aberta, possibilitando seu uso, reuso e adaptação livres. Um ponto importante para tornar eficiente a busca por REAs em repositórios pela Web e como o metadado deve expor dados sobre o recurso. Uma iniciativa global para definir padrões para propiciar acesso a dados abertos na Web e o Linked Open Data. Este trabalho apresenta uma análise, em duas etapas, de 16 repositórios de REAs a respeito do uso dos padrões de LOD em seus metadados. Ao final foi possível constatar que a grande maioria dos repositórios analisados, com poucas exceções, não utilizam os padrões definidos pelo LOD e, quando utilizam, o fazem de forma superficial, sem explorar seu potencial.

Abstract

Open Educational Resources (OERs) are a very important part of Open Education. These resources are openly shared, allowing their use, reuse, and adaptation. An important aspect to make the search for OERs in repositories on the Web efficient is how the metadata of these resources should expose data about them. A global initiative to define standards to allow access to open data on the Web is the Linked Open Data (LOD). In this paper it is presented an analysis, in two stages, of 16 repositories of OERs about the use of LOD standards in their metadata. At the end it was possible to verify that the majority of the analysed repositories, with few exceptions, do not use standards defined by LOD and, when they do, they do it superficially without exploring its potential.

¹ Trabalho publicado e apresentado no Simpósio Brasileiro de Informática na Educação (SBIE 2020).

² Pós-Graduando em Computação Aplicada à Educação, USP, lucianobpaula@usp.br.

³ Doutorando, Instituto de Ciências Matemáticas e de Computação - Universidade de São Paulo (ICMC-USP), williamsimao@usp.br.

⁴ Professora Doutora, Instituto de Ciências Matemáticas e de Computação - Universidade de São Paulo (ICMC-USP), francine@icmc.usp.br.

Cite as: Paula, L. & Deus, W. & Barbosa, E. (2020). Análise de Repositórios de REAs em Relação ao Uso dos Padrões de Linked Open Data. Anais dos Trabalhos de Conclusão de Curso. Pós-Graduação em Computação Aplicada à Educação Instituto de Ciências Matemáticas e de Computação. Universidade de São Paulo.

1. Introdução

Os Recursos Educacionais Abertos (REAs) são um importante parte do conceito de Educação Aberta. Segundo [Atkins et al. 2007], REAs são recursos disponibilizados em qualquer mídia, que podem ser utilizados para ensinar, aprender e pesquisar. Esses recursos são publicados sob licença aberta que permite seu uso gratuito, possibilitando que estes sejam readequados para qualquer propósito. Isso inclui cursos completos, materiais de cursos, textos, vídeos, *software* e quaisquer outras mídias.

Atualmente, muitos desses recursos são digitais e, com a popularização da Web, são armazenados e disponibilizados em repositórios digitais de REAs, nos quais é possível buscar por recursos sobre um determinado assunto [Navarrete and Luján-Mora 2015].

Para tornar eficiente essa busca via Web, é muito importante o uso de metadados que sejam completos em informações.

Considerando isso, na Web atual, há uma iniciativa global para definir padrões para compartilhamento de dados abertos, o Linked Open Data (LOD). Por meio dos princípios e padrões do LOD, é possível atingir níveis relevantes de compartilhamento de qualquer dado que seja aberto, como é o caso dos REAs, o que propicia experiências ricas para aqueles que queiram aprender e pesquisar sobre diversos assuntos⁵.

Existem diversas vantagens de se utilizar padrões do LOD em repositórios de REAs. Uma vantagem que pode ser citada é que a utilização desses padrões enriquece a forma como o metadado dos REAs podem ser interpretados, dando significado as relações que o recurso possa ter com outros. Outra vantagem que pode ser considerada é o fato que o uso desses padrões possibilita também a interligação de recursos a outras bases que, inclusive, podem não ser de REAs. Esses aspectos são apresentados em [Navarrete and Luján-Mora 2015].

Para ilustrar, é possível citar um caso de uso hipotético, baseado nos conceitos definidos por Tim Berners-Lee em [Berners-Lee 2006]: um professor, ao buscar por REAs sobre uma região do Brasil pode, por meio das ligações no metadado dos recursos encontrados, chegar até os *sites* oficiais das prefeituras das capitais dos estados da região pesquisada, ou até mesmo em uma base de dados geográficos, na qual poderá obter informações sobre o relevo e clima daquela região. Caso todos esses recursos estejam interligados via metadado e utilizando os padrões LOD, todo processo se torna transparente para o usuário, sendo executado pelo próprio sistema de busca, uma vez que os dados são bem estruturados em formatos legíveis por máquinas.

No entanto, o potencial do uso dos padrões de LOD em repositório de REAs ainda não é explorado, como visto em [Nahhas et al. 2018]. Os autores declararam que as abordagens existentes se encontram de modo disperso, sem objetivos claros que possam ajudar a entender o impacto de conceitos do Linked Data, que considera não só os dados abertos, mas qualquer tipo de dado, para os REAs. Ou seja, os benefícios dessa abordagem ainda não são explorados.

Como dito em [Navarrete and Luján-Mora 2015], outra vantagem é que os padrões de LOD podem trazer uma maior homogeneidade em como os metadados dos repositórios devem ser e quais informações devem conter. Essa padronização, derivada da homo-

⁵ Benefícios do LOD - <https://www.w3.org/2005/Incubator/lld/wiki/Benefits> - acessado em 28/09/2020.

geneidade, propicia uma maior interoperabilidade entre os repositórios e seus recursos, possibilitando sua utilização automatizada por uma ampla gama de aplicações.

Dessa forma, este trabalho tem como objetivo apresentar uma análise sobre o atual nível de uso dos padrões de LOD em repositórios de REAs e apresentar as vantagens que a utilização desses padrões propicia nesse contexto. Para isso foi feita, em uma primeira etapa, uma análise de repositórios de REAs conhecidos para verificar se esses utilizam padrões de LOD e, em caso positivo, como os utilizam. Em uma segunda etapa, foram feitas buscas por repositórios de REAs no *site* Linked Open Data Cloud⁶, que possui um catálogo de bases que estão interligadas seguindo os padrões do LOD, com o objetivo de identificar aqueles que seguem totalmente esses padrões. Todo o procedimento realizado em ambas etapas e descrito em seções mais a` frente no texto.

Este trabalho encontra-se estruturado da seguinte forma: na Seção 2 e apresentada uma síntese dos trabalhos encontrados na literatura que são relacionados com este; na Seção 3 e apresentado o que são repositórios de REAs e sua relação com metadados que descrevem os recursos; na Seção 4 e apresentado o que é LOD, seus padrões e a

classificação de cinco estrelas para bases de dados; na Seção 5 e apresentada a metodologia utilizada na análise dos repositórios das duas etapas; na Seção 6 são apresentados os resultados encontrados a partir da análise feita; na Seção 7 e apresentada uma discussão sobre os resultados, impactos e limitações do trabalho e a Seção 8 concentra as considerações finais do estudo.

É importante citar que este trabalho foi publicado no Simpósio Brasileiro de Informática na Educação (SBIE 2020), na Trilha 1: “Abordagens Metodológicas e Tecnológicas para Educação Básica e Superior”, evento que ocorreu de 24 a 28 de novembro de 2020, de forma *online*. A referência do trabalho publicado e indicada em [Paula et al. 2020]. Uma vez que o citado simpósio e considerado relevante para a comunidade científica brasileira, a publicação deste trabalho neste simpósio atesta a sua qualidade e a importância do assunto para os interessados em REAs e seus repositórios.

2. Trabalhos relacionados

Ao pesquisar a literatura, por meio de conhecidos repositórios de artigos científicos, tais como IEEEExplore, ACM Digital Library e Science Direct, foram encontrados trabalhos que investigam tópicos discutidos neste estudo. Entretanto, uma característica comumente encontrada nesses trabalhos foi a ausência de uma análise específica sobre o uso de LOD em repositórios de REAs atuais. Os trabalhos que mais se relacionam com este visam o uso dos padrões de LOD para melhorar a interoperabilidade de REAs em repositórios, porém não analisa os existentes.

Por exemplo, um dos problemas relatados na literatura a respeito da interoperabilidade entre repositórios de REAs e a falta de padronização dos metadados. Em [Koutsomitropoulos and Solomou 2018] foi apresentada a possibilidade de se utilizar ontologias, que são utilizadas em LOD, para facilitar a interligação de diferentes bases de dados de REAs.

Em [Pereira et al. 2018] e apresentada a possibilidade de se utilizar dados de redes sociais para recomendar objetos educacionais para usuários, o que demonstra o potencial

⁶LOD Cloud - <https://lod-cloud.net/> - acessado em 28/09/2020.

do LOD nos metadados de repositórios de REAs. Como limitação, usuários devem ter contas ativas em redes sociais para o uso da abordagem e nem analisam o uso dos mesmos em repositórios existentes.

[Rajabi et al. 2014] apresentaram como expor metadados seguindo o padrão IEEE LOM [IEEE 2002]. Para isso, foi feita uma relação entre esse metadado e os padrões do LOD. Nesse sentido, nota-se que as generalizações dos resultados foram restritas somente aos repositórios que adotam o padrão IEEE LOM.

Em [Sicilia et al. 2010] foi apresentado um exemplo prático de um portal de repositórios de REAs interligado utilizando LOD, apresentando as vantagens que essa ação propicia. Em [Kawase et al. 2013] foi abordada a questão de outra forma, apresentando como foi construída uma plataforma que coleta dados sobre objetos de aprendizagem, tratando a heterogeneidade com o uso de LOD. Em ambos os trabalhos, no entanto, notou-se a limitação dos acervos e o total de áreas analisadas. Complementarmente, [Navarrete and Luján-Mora 2015] realizou uma contribuição teórica sobre as vantagens de se usar LOD em repositórios de REA.

[Vega-gorgojo et al. 2015] conduziu uma revisão sistema 'tica da literatura a respeito de trabalhos que unem o conceito de LOD com aprendizagem e objetos educacionais. O trabalho também apresentou recomendações a respeito do uso dos padrões de LOD em repositórios. Dentre os estudos citados, não há nenhum que faz uma análise dos repositórios de REAs em relação ao uso dos padrões de LOD.

Por fim, [Lama et al. 2011] apresentou um sistema automatizado para interligar um repositório de REAs com a DBpedia, um portal que possui os metadados, seguindo os padrões do LOD, dos artigos da Wikipedia. Este trabalho não considerou a interligação do repositório com outros além do da DBpedia.

O trabalho apresentado aqui diferencia-se dos demais, principalmente, pelo fato de apresentar uma análise de repositórios conhecidos em relação ao uso dos padrões de LOD. Considerando a amostra de artigos obtida, nenhum outro trabalho encontrado apresentou tal análise prática.

3. Repositórios de REAs e metadados

Atualmente existem diversos repositórios de REAs disponíveis na Web. Esses repositórios possibilitam que os REAs sejam acessados via Web, pesquisados por assunto e palavras-chave, independente se o recurso é um texto, um vídeo, uma imagem, etc.

Uma parte importante a ser considerada para os repositórios de REAs são os metadados relacionados com os recursos armazenados. Um metadado pode ser definido como um dado sobre outro dado [Valiente et al. 2015]. Por exemplo, um arquivo de vídeo digital pode possuir um arquivo metadado que explica o seu conteúdo, por meio de informações como título, autor, assunto, tipo de mídia, etc.

A função do metadado em um REA é de grande importância, pois possibilita a obtenção de informações sobre o recurso sem que seja necessário visualizá-lo por completo. Como no exemplo anterior, um REA em forma de vídeo, sem metadados adequados, faz com que haja a necessidade de o usuário assisti-lo para entender quais assuntos são abordados no recurso, o que não aconteceria caso o metadado fosse completo em

informações.

Tudo isso e ainda mais importante ao se considerar a busca por REAs de forma automatizada, executada por máquinas. Uma simples busca por palavras-chave em um repositório de REA e um exemplo de busca automatizada e essa possuirá resultados melhores caso o metadado dos recursos sejam relevantes e definam características do recurso de forma abrangente.

4. Linked Open Data - LOD

Tim Berners-Lee, considerado o criador da Web, definiu o conceito de dados interligados (*linked data*) [Berners-Lee 2006]. Esse conceito é baseado em quatro princípios:

i) uso de URIs (*Uniform Resource Identifier*) para dar nomes a recursos publicados na Web;

ii) uso de URIs via HTTP (*Hypertext Transfer Protocol*) para que esses nomes possam ser consultados na Web;

iii) ao acessar um URI, deve ser provido informações úteis, utilizando RDF (*Resource Description Framework*) e SPARQL (uma linguagem de busca em bases de dados de arquivos RDF);

iv) e preciso incluir *links* para outros URIs, dessa forma é possível descobrir novos recursos.

Considerando dados abertos, ou seja, aqueles que podem ser acessados, consumidos e reutilizados de forma livre, Tim Berners-Lee também definiu o conceito de dados abertos interligados, ou Linked Open Data (LOD), que tem como objetivo interligar dados abertos, formando uma rede de metadados que conecta conceitos e recursos. Atualmente, no *site* Linked Open Data Cloud é possível verificar a quantidade de bases de dados abertos interligadas dessa forma.

Os padrões de LOD possibilitam a interligação de recursos que, sem esses padrões, se encontrariam isolados uns dos outros. Por meio da interligação dos metadados dos recursos, buscas por um item pode levar ao descobrimento de outros relacionados ao primeiro, tornando a experiência do usuário na obtenção desses dados muito mais rica. Além disso, o intuito do uso desses padrões é permitir que essas buscas sejam automatizadas, sendo executadas de forma transparente para os usuários.

Entretanto, há desafios para a adoção completa de todos os padrões do LOD, como visto em [Fayyaz et al. 2018]. Alguns dos desafios para a adoção do LOD em larga escala passam por questões técnicas, por exemplo, quando há grandes bases de dados que devem ser adequadas aos padrões. As ferramentas disponíveis para adequar bases de forma automática, além de exigir conhecimento especializado, nem sempre suportam bases heterogêneas, o que é uma realidade em repositórios de REAs.

Na próxima subseção são apresentados brevemente os padrões do LOD.

4.1. Padrões de LOD

O LOD indica como padrão para metadados o *framework* RDF. Um arquivo do tipo RDF é composto por diversas triplas a respeito do recurso que se queira descrever. As triplas seguem o padrão como mostrado a seguir.

< su jeito > < predicado > < ob jeto >

Por exemplo, uma página em uma URL U_x pode ser a respeito de um assunto A_y , portanto, uma tripla genérica poderia ser como a apresentada a seguir, seguindo a ordem *< su jeito > < predicado > < ob jeto >*.

< U_x > < sobre > < A_y >

Como exemplo, um REA sobre a Região Sudeste do Brasil poderia possuir a seguinte tripla.

< REA > < sobre > < Sudeste >

Sendo que as três partes da tripla são *links* para outros recursos. A parte do sujeito (*< REA >*) seria um *link* para o próprio REA que esta sendo descrito, a parte do predicado (*< sobre >*) seria um *link* para a definição na ontologia que descreve esse tipo de relação entre um REA e outro recurso e a parte do objeto (*< Sudeste >*) pode indicar um *link* para um terceiro recurso (seja REA ou não) sobre a Região Sudeste.

E por meio das triplas que o RDF permite a interligação de metadados que estejam armazenados em diferentes repositórios. Seguindo o exemplo acima, a relação *< sobre >* pode possuir sinônimos definidos em outras ontologias, utilizadas por outras bases de dados para definir os seus metadados. O mesmo pode ser dito sobre o objeto *< Sudeste >* que pode existir também em outro repositório. Dessa forma, uma busca na base inicial pode ser expandida também para uma consulta em outras bases que possuam esse outro tipo de relação e objeto. O mesmo pode ser dito em relação a um sujeito, caso o mesmo REA esteja em diferentes repositórios.

Ha diversos formatos compatíveis com o RDF como, por exemplo, Turtle, N-triples, N-quads, RDF/XML, entre outros, que podem ser usados em metadados de forma a cumprir os requisitos do LOD.

O metadado em RDF de um recurso pode ser disponibilizado de diversas formas: em um arquivo RDF separado; em anotações embarcadas no código HTML da página, seguindo o padrão RDFa; exposto em um *endpoint* SPARQL, ou ate mesmo na forma de *dataset* com os metadados para *download*. há também vocabulários que podem ser expressados utilizando RDF/XML ou equivalente, como e o caso do Dublin Core⁷.

Os padrões do LOD também indicam o SPARQL⁸ como a linguagem de consulta para bases de dados de arquivos RDF. Essa linguagem permite buscas sobre as triplas que compõem os arquivos RDF. Seguindo o exemplo acima, uma busca por recursos sobre a Região Sudeste, poderia ser feita como mostrado a seguir, utilizando SPARQL.

```
SELECT ?sujeito
WHERE
{ ?sujeito < sobre > < Sudeste > }
```

No exemplo acima são retornados todos os sujeitos (?sujeito) ou, considerando repositórios de REAs, todas as URIs que indicam REAs, e que tenham em seus metadados

⁷Dublin Core - <https://www.dublincore.org/> - acessado em 28/09/2020.

⁸SPARQL - <https://www.w3.org/TR/rdf-sparql-query/> - acessado em 09/10/2020.

a relação < *sobre* > com o objeto < *Sudeste* >. Esse exemplo é somente ilustrativo, não apresentando todos os detalhes como a definição dos *namespaces* utilizados para as consultas, que indicam onde as *tags* foram definidas, entre outros.

Utilizando SPARQL, é possível realizar buscas combinando quaisquer um dos três elementos das triplas contidas em arquivos RDF. Por exemplo, é possível realizar uma busca que retorne todos os sujeitos que tenham qualquer tipo de relação (parte do predicado) com o objeto < *Sudeste* >, ou retornar todos os tipos de relação que existam com esse mesmo objeto, e assim por diante.

4.2. Cinco estrelas do LOD

Tim Berners-Lee elaborou uma classificação de cinco estrelas, acumulativas, para aferir quanto um dado está disponibilizado seguindo os padrões de LOD [Berners-Lee 2006]. Essa classificação, baseada nos princípios do conceito de dados interligados, inclui o uso de licenças públicas e considera que os dados são publicados na Web. Ao se utilizar os padrões e publicar os dados sob uma licença aberta, pode-se considerar que o padrão de LOD está sendo usado. É possível que organizações que não possuam dados abertos também utilizem essa classificação, entretanto, é dito que seguem os padrões de Linked Data (LD). Uma vez que este trabalho tem como foco o uso desses padrões para REAs, que são essencialmente abertos, esta pesquisa considerou o uso dos padrões de LOD.

O significado da classificação de 1 a 5 estrelas é apresentado na Tabela 4.1.

Estrelas	Descrição
*	Dados publicados na Web em qualquer formato sob uma licença aberta.
**	Dados publicados de forma estruturada e em um formato que seja legível por máquinas.
***	Dados publicados de forma estruturada na Web e utilizando um formato não proprietário.
****	Dados publicados na Web no formato RDF.
*****	O metadado no formato RDF indica URIs de outras bases de dados relacionadas.

Tabela 4.1. Descrição da classificação de 5 estrelas do LOD.

Seguindo essa classificação, definida em [Berners-Lee 2006], um possível cenário de exemplo pode ser descrito como a seguir:

- Um *site* que apresente uma imagem de uma tabela com valores receberia 1 estrela, pois cumpre somente este quesito.
- Caso essa mesma tabela seja publicada em algum formato legível por máquinas, por exemplo uma planilha ou outro arquivo similar, receberia mais uma estrela, totalizando 2 estrelas.
- Caso o formato utilizado não seja proprietário, ou seja, não esteja sob algum tipo de licença fechada, o *site* receberia a terceira estrela.
- Caso a tabela seja publicada com metadado, que a descreve, utilizando o padrão RDF, seja em um arquivo separado ou mesmo embutido no código da página que contém a tabela, o *site* receberia a quarta estrela.
- E, caso o metadado, utilizando RDF, indique outros recursos relacionados com os dados apresentados na tabela e esses estejam em outras bases de dados, a quinta estrela é obtida.

Ou seja, a partir da descrição da classificação de cinco estrelas e possível fazer um paralelo a respeito de como deve ser um repositório de REA para que esse obtenha a nota máxima seguindo essa escala, como definida em [Berners-Lee 2006]. Essa descrição é feita a seguir.

- Um repositório que publique REAs na Web, cumpre com o critério da primeira estrela.
- Caso o REA publicado nesse repositório esteja em um formato legível por máquinas, permitindo que, por meio de consultas via Web, esse recurso seja encontrado, o critério da segunda estrela é obtido.
- Caso o formato utilizado no REA não seja proprietário, a terceira estrela é recebida.
- Caso o repositório disponibilize também metadado sobre o REA e esse esteja definido seguindo o padrão RDF, o repositório seria classificado com quatro estrelas.
- Para a obtenção da quinta estrela, o metadado do REA, utilizando RDF, deve ligá-lo a outros recursos, que podem ser REAs ou não, e esses estejam em outros repositórios.

5. Metodologia

Para atingir o objetivo proposto neste trabalho foram desenvolvidas duas etapas distintas de pesquisa. Em ambas etapas, a metodologia utilizada foi a pesquisa exploratória que, segundo [Gil 2002], tem como objetivo proporcionar uma maior familiaridade com o problema em questão, o que possibilita a construção de hipóteses. As etapas são descritas a seguir.

- Etapa 1: análise de repositórios de REAs com o intuito de verificar se esses utilizam os padrões do LOD e como é feito esse uso.
- Etapa 2: análise dos repositórios de REAs encontrados por meio de buscas no *site* Linked Open Data Cloud. Esta etapa teve como intuito verificar se há repositórios de REAs listados na nuvem do LOD, mesmo que não sejam conhecidos.

5.1. Análise dos repositórios

Na Etapa 1, os repositórios foram acessados e tiveram suas coleções e estruturas analisadas. Os repositórios considerados nesta etapa foram: MIT OpenCourseWare, OER Commons, EduCapes, Repositório Aberto, MetaFinder, SkillsCommon, Merlot, Xpert, iOER e OAsis.

Durante essa etapa, as seguintes informações foram observadas para cada repositório:

- *Como são compartilhados os metadados dos REAs disponíveis no repositório?*
- *Como os repositórios estão utilizando padrões recomendados pelo LOD?*

Para extrair essas informações, foi realizado o seguinte procedimento:

- i) Busca manual pelo metadado do REA encontrado no repositório, caso este metadado seja disponibilizado em um arquivo separado;
- ii) Utilização de dois extratores automáticos *online* de metadado descritos usando RDF/RDFa e outros padrões relacionados. Os extratores utilizados foram: RDFa 1.1

Distiller and Parser⁹ e Graphite Browser¹⁰. Esses dois extratores funcionam de forma equivalente: por meio de sua página principal, é possível inserir um URL e os esses apresentam as triplas encontradas embarcadas na página encontrada no endereço ou mesmo em arquivo separado, residente no mesmo endereço da página;

iii) Busca manual no código HTML da página de apresentação do REA por termos chave que indiquem metadado em RDF/RDFa e outros padrões relacionados. Neste caso, foi analisado o código fonte da página, buscando termos como “rdf”, entre outros, na tentativa de se visualizar se há anotações relacionadas com metadado.

É importante citar que, em relação aos metadados reconhecidos pelos extratores automáticos, foi feita a conferência dos resultados obtidos, pois em alguns casos esses não eram relacionadas às características específicas do REA em si, mas de informações genéricas sobre o recurso (por exemplo: somente título e tipo de mídia) e/ou sobre a página do repositório. Dessa forma, o resultado dos extratores foi validado para certificação de que o metadado obtido era ou não sobre características relevantes do REA.

A partir das informações obtidas após os passos do procedimento terem sido finalizados, cada repositório foi classificado de acordo com as classificações apresentadas na Seção 4.2, ou seja, a classificação de cinco estrelas do LOD.

Na Etapa 2 foram feitas buscas no *site* Linked Open Data Cloud por dados sobre REAs. Para isso, foram testados diferentes termos de busca, como mostrado a seguir:

i) Inicialmente, foi feita uma busca pelos termos *educational resources*, o que resultou em 4 bases.

ii) Após a primeira busca, foi feita uma segunda utilizando o termo *educational*, com o intuito de se obter resultados mais amplos. Nessa busca foram obtidos 17 resultados, sendo que os 4 resultados da consulta anterior estavam contidos nesses.

iii) Tornando a busca ainda mais abrangente, foi feita uma consulta utilizando o termo *education*. Dessa vez, foram obtidos 28 resultados, sendo que os 17 resultados da pesquisa anterior estavam contidos dentre esses. Foi estabelecido, portanto, que esse seria o resultado investigado.

Ao fim, foram analisadas as 28 bases resultantes, na tentativa de se entender quais delas possuem, de fato, relação com REAs. Essa análise resultou em 6 bases que eram, de fato, relacionadas a REAs. É importante citar que uma das bases resultantes nesta última consulta não foi obtida em nenhuma das duas consultas anteriores, o que valida a estratégia de se obter resultados mais amplos.

A lista com os repositórios analisados nas Etapas 1 e 2 e apresentada na Tabela 5.2¹¹, na qual são apresentados o nome do repositório, seu *link* de acesso e em qual das duas etapas este foi analisado.

A seguir serão apresentados e discutidos os resultados das análises dos repositórios, tanto para a Etapa 1 quanto para a Etapa 2.

⁹RDFa 1.1 Distiller - <https://www.w3.org/2012/pyRdfa/> - acessado em 28/09/2020.

¹⁰Graphite Browser - <http://graphite.ecs.soton.ac.uk/browser/> - acessado em 28/09/2020.

¹¹*Links* acessados em 01/08/2020.

Repositório	Links	Etapas
MIT OpenCourseWare	https://ocw.mit.edu/	1
OER Commons	http://oercommons.org/	1
EduCapes	https://educapes.capes.gov.br/	1
Repositório Aberto	https://repositorioaberto.uab.pt/	1
Mason OER MetaFinder	https://oer.deepwebaccess.com/oer/desktop/en/search.html	1
SkillsCommons	https://skillscommons.org	1
Merlot	https://www.merlot.org/merlot/	1
Xpert	https://www.nottingham.ac.uk/xpert/	1
Oasis	http://oasis.col.org/	1
iOER	http://ioer.ilsharedlearning.org/	1
Dados abertos da Open University	https://data.open.ac.uk	2
Didactalia	http://didactalia.net/	2
IneveryCREA	http://ineverycrea.com/	2
IneveryCREA Argentina	http://ineverycrea.com.ar	2
mEducator	http://meducator.open.ac.uk/	2
Organic Edunet Linked Open Data	http://data.organic-edunet.eu/	2

Tabela 5.2. Repositórios analisados nas Etapas 1 e 2.

6. Resultados

Nesta seção os resultados obtidos nas duas etapas são apresentados e discutidos. E

feita

uma análise de cada repositório consultado e sua classificação.

6.1. Resultado da análise dos principais repositórios (Etapa 1)

A seguir, é apresentada uma descrição das notas definidas para cada tipo de repositório, bem como uma justificativa para essa classificação. É importante citar que o formato PDF (*Portable Document Format*) foi considerado, neste trabalho, como legível por máquinas (*machine-readable*). Embora exista essa questão na literatura, atualmente há estudos que já consideram o uso deste formato para esse fim, dada sua grande adoção, como pode ser visto em [Adobe 2017]. Na Tabela 6.3 é apresentado um sumário dos pontos analisados em cada repositório na Etapa 1.

Repositório	Metadado específico	Dataset ou endpoint SPARQL	Machine-readable
MIT OpenCourseWare	Naõ	Naõ	Sim
OER Commons	Sim	Naõ	Sim
EduCapes	Sim	Naõ	Sim
Repositório Aberto	Sim	Naõ	Sim
Oasis	Sim	Naõ	Naõ
SkillsCommons	Naõ	Naõ	Naõ
Mason OER MetaFinder	Naõ	Naõ	Sim
Merlot	Naõ	Naõ	Naõ
Xpert	Naõ	Naõ	Naõ
iOER	Naõ	Naõ	Naõ

Tabela 6.3. Análise dos repositórios da Etapa 1.

A seguir, a análise de cada repositório da Etapa 1 é detalhada, explicando todos os aspectos que foram considerados em relação à classificação dos repositórios segundo

a escala de estrelas do LOD.

MIT OpenCourseWare: * * * Este repositório é focado em cursos, de diversas áreas, e os recursos derivados desses cursos são disponibilizados como REAs. Não foi encontrado nenhum tipo de metadado específico em relação aos recursos disponibilizados, sendo que as características dos REAs disponibilizados, como título, autor, etc., somente são apresentados em suas respectivas páginas, dentro do repositório. Não foi possível encontrar *dataset* ou *endpoint* SPARQL para consulta. Dessa forma, esse repositório possui 3 estrelas na classificação do LOD, uma vez que possui recursos publicados sob licença aberta, em formato legível por máquinas e não proprietário (PDF, por exemplo), porém, não utiliza os padrões do LOD e não interliga seus metadados com outras bases.

OER Commons: * * * Neste repositório foi possível encontrar uma tabela no formato “.xlsx” que descreve os campos utilizados no metadado de seus REAs, assim como o vocabulário utilizado, porém, não são utilizados padrões de LOD. Não foi possível encontrar nenhum *dataset* dos metadados disponível para *download*. Este repositório possui um *endpoint* para acesso a uma base de metadados, porém para acessar o *endpoint*, é necessário o pagamento de licença para obter um *token* de autenticação¹². Dessa forma, é possível considerar que esse repositório possui 3 estrelas na classificação do LOD, uma vez que possui recursos publicados sob licença aberta, em formato legível por máquinas e não proprietário. Porém, não utiliza os padrões do LOD e não há interligação dos metadados com outras bases, uma vez que usa formato e vocabulário próprios.

EduCapes: * * * Neste repositório foi possível identificar que, para cada REA apresentado, há metadado no padrão Dublin Core disponível, de forma embarcada na página que apresenta o recurso. Porém, é interessante citar que, manualmente, no HTML da página do REA, há mais informações a respeito do recurso, em forma de metadado, do que as reconhecidas pelos extratores automáticos. É provável que isso se deve a algum problema de formatação das páginas em relação às anotações, o que impossibilita o reconhecimento por ferramentas automáticas. Não foram encontradas formas de se consultar a base de metadados, seja para *download*, seja como *endpoint* SPARQL para consulta. No repositório, há REAs disponibilizados em formato legível por máquinas (PDF), utilizando licenças abertas. Dessa forma, esse repositório é classificado com 3 estrelas.

Repositório Aberto: * * * O metadado dos REAs são disponibilizados seguindo o padrão Dublin Core, porém, ocorreu a mesma situação que o repositório EduCapes: manualmente, no código HTML da página do REA, há mais informações em forma de metadado, do que as reconhecidas pelos extratores automáticos. É provável que os dois repositórios utilizem a mesma plataforma de *software* para publicação dos REAs, por isso existe a mesma inconsistência. Não foram encontradas formas de se consultar a base de metadados, seja para *download*, seja como *endpoint* para consulta. No repositório, há REAs disponibilizados em formato legível por máquinas (PDF), utilizando licenças abertas. Dessa forma, este repositório também pode ser classificado com 3 estrelas.

Oasis: * * * Os REAs encontrados neste repositório possuem metadado seguindo o padrão Dublin Core embarcado na página HTML. Não foi encontrado *endpoint* SPARQL ou formas de se realizar buscas pelo metadado. Assim como o EduCapes e o Repositório Aberto, há mais metadado encontrado manualmente do que o reconhecido pelos extra-

¹²*Endpoint* do OERCommons - <https://www.oercommons.org/api/search> - e requisitado um *token* de autenticação para continuar - acessado em 28/09/2020.

tores. É possível que haja a mesma situação em relação ao formato das páginas que os outros repositórios citados. Os REAs são disponibilizados em PDF e, dessa forma, esse repositório recebe 3 estrelas.

SkillsCommons: * Neste repositório, não foram encontrados metadados que fossem específicos para os REAs manualmente ou com o auxílio dos extratores. O formato dos REAs disponibilizados eram, em sua maioria em formato proprietário (por exemplo “.doc”) ou comprimido (por exemplo “.zip”). Não foi possível encontrar metadado específico sobre o REA tanto manualmente quanto utilizando os extratores. Dessa forma, esse repositório recebe apenas 1 estrela.

MetaFinder, Merlot, Xpert e iOER: não se aplica Tratam-se de portais que agregam informações sobre outros repositórios de REAs. Dessa forma, esses portais não possuem nenhum tipo de metadado específico para os REAs encontrados, pois somente redirecionam o usuário para o repositório original. O uso ou não dos padrões de LOD dependem do repositório original do recurso.

6.2. Resultado dos repositórios encontrados no LOD Cloud (Etapa 2)

Nesta etapa foram feitas buscas no *site* Linked Open Data Cloud em busca de bases relacionadas com REAs. É importante citar que foram consideradas todas as bases relacionadas com REAs e seus metadados, não necessariamente aquelas que são repositórios dos recursos, de fato. A síntese dos resultados é apresentada a seguir:

Dados abertos da Open University: * * * * * Neste repositório são compartilhados os metadados dos REAs disponibilizados pela Open University. São dados de publicações, cursos e materiais multimídia produzidos pela universidade. Todos os recursos são disponibilizados seguindo os padrões de LOD e são, na grande maioria, compartilhados sob licença aberta. Os metadados são disponibilizados nos formatos CSV, JSON, XML e há um *endpoint* SPARQL para consulta.

Didactalia: * * * * * O Didactalia é uma comunidade global e um repositório que proporciona a professores e estudantes a oportunidade de criar, compartilhar e encontrar REAs. É citado que o uso dos padrões de Linked Open Data e a criação de um ecossistema de comunidades de aprendizagem no contexto de aprendizado social são o conceito central do Didactalia. Cada recurso possui um arquivo RDF específico, que pode ser obtido na sua própria página, rico em triplas com várias informações sobre o próprio recurso.

IneveryCREA e IneveryCREA Argentina: * * * * * IneveryCREA é uma comunidade de professores e profissionais da educação que tem como objetivo a criação, o desenvolvimento e o compartilhamento de conteúdo e REAs na Espanha e na Argentina. Cada página HTML que disponibiliza um conteúdo possui uma versão com metadados embarcados, relacionados ao recurso apresentado.

mEducator: * * * * * O mEducator publica dados sobre repositórios outros repositórios de REAs. Usa um RDF *schema* para interligar recursos educacionais com *datasets* da nuvem do LOD. Portanto não se trata de um repositório, mas sim um local para encontrar *links* para *datasets* de metadados de repositórios de REAs. Durante a concepção deste trabalho, o *endpoint* para consulta se encontrava *offline* e a última atualização em

sua seção de notícias foi feita em 2013.

Organic Edunet Linked Open Data: * * * * * Esse repositório contém metadados, que utilizam os padrões de LOD, de uma coleção de REAs disponibilizados pelo portal Organic.Edunet. Os tópicos incluem agricultura orgânica e agroecologia.

7. Discussão

Como apresentado, praticamente os repositores rios analisados na Etapa 1 receberam 3 estrelas na classificação de LOD. Isso se deve ao fato de nenhum deles apresentarem os metadados dos REAs disponíveis utilizando padrões do LOD, e somente disponibilizando os recursos de forma isolada. Em uma avaliação ainda mais rígida, alguns deles, inclusive, receberiam uma estrela a menos, uma vez que o formato utilizado para disponibilização dos recursos educacionais e o formato PDF, até recentemente não indicado para compartilhamento de dados abertos.

Exceto os repositórios encontrados por meio do *site* LOD Cloud, todos os outros analisados que utilizavam de alguma forma os padrões do LOD faziam de forma superficial, ou seja, com metadados com poucas informações ou que não apresentavam características importantes dos REAs, por exemplo, o seu assunto.

Outro ponto negativo que pode ser destacado é a falta de *endpoints* para consulta SPARQL nesses repositórios que possuem metadados usando padrões relacionados com o LOD, o que impossibilita buscas automatizadas sejam feitas em suas bases.

Um aspecto importante a ser explorado é que muitos REAs podem ser arquivos de vídeo e/ou áudio, que naturalmente não são padrões estruturados nem legível por máquinas, do ponto de vista do LOD. Nesta situação, o papel do metadado é ainda mais relevante. Como pode ser observado, a maioria dos repositórios analisados possuem pouca informação disponível via metadados.

É relevante apontar que, tanto uma maior riqueza em detalhes importantes do REA em seu metadado, quanto a exposição de *endpoints* SPARQL para consulta aos repositórios possibilitam a busca automática, com qualidade, dos recursos. Esse tipo de busca resulta em uma experiência completa para o usuário que consulte esses repositórios em busca de algum recurso. Por exemplo, citando o caso de uso apresentado na Seção 1, o resultado obtido pelo usuário poderia ter sido gerado a partir de uma única busca no repositório inicial, uma vez que a busca poderia ser expandida para outras bases, seja de REAs ou não.

Nota-se, então, que os repositórios de REAs possuem um amplo caminho a percorrer para serem considerados “abertos” sob o ponto de vista de LOD. Como apresentado, muitos repositórios não possuem mecanismos que facilitam o compartilhamento e acesso aos metadados de seu acervo, como os *endpoints* para consulta aos metadados. Indiretamente, isso faz com que cada repositório tenha uma coleção restrita e fechada em si. Logo, agregadores ou buscadores descentralizados de REAs encontram dificuldades para acessar este acervo e disponibilizá-lo para os usuários.

O uso dos padrões de LOD e as interligações entre repositório e outras bases seriam muito vantajoso, pois auxiliaria na disseminação do uso dos REAs, aumentando sua

abrangência e potencial de (re)utilização para os usuários. Essas interligações naturalmente criariam *subclouds*¹³ dentro da nuvem do LOD, aumentando a interoperabilidade e visibilidade dos repositórios e recursos, e permitindo também consultas com escopo bem definido, por exemplo, que retornem somente outros REAs, ou somente recursos educacionais, etc.

Um aspecto importante a ser considerado é que, uma vez que a definição de REA é ampla, o mesmo ocorre com a definição de repositório de REA. Dessa forma, há repositórios que armazenam e compartilham REAs e seus metadados, há aqueles que são portais para REAs disponibilizados em outros repositórios, e há também casos híbridos. Esse fato faz com que, apesar de existir exemplares de cada um desses tipos nos repositórios analisados neste trabalho, esses representam apenas uma fração dos repositórios atuais e, dessa forma, os resultados não podem ser generalizados para todos os repositórios existentes, sendo necessária a análise individual de cada um.

8. Conclusões

Neste trabalho foi apresentada uma análise do uso de padrões de LOD em repositórios de REAs. O uso desses padrões abre inúmeras possibilidades de se obter informações ricas sobre o assunto apresentado no REA, propiciando uma busca completa sobre o recurso, inclusive entre bases de dados distintas. Nos resultados, foi possível verificar que a maioria dos repositórios mais conhecidos não utilizam esses padrões ou, quando utilizam, os utilizam de forma superficial, fornecendo somente algumas anotações semânticas gerais sobre o recurso compartilhado, em vez de disponibilizar informações mais detalhadas via *endpoint* SPARQL, por exemplo. Por outro lado, os repositórios encontrados via LOD Cloud fazem um bom trabalho a esse respeito, porém ainda são em número reduzido.

A partir deste trabalho de conclusão de curso foi gerado o artigo [Paula et al. 2020] que foi publicado no SBIE 2020, importante evento sobre informática e educação, e este foi aceito para publicação. A partir deste resultado é possível concluir que o tema tratado aqui é relevante, atestando sua importante contribuição para a área. Em relação ao trabalho publicado no evento e este, o aqui apresentado possui maiores detalhes nas Seções 1 e 4, uma maior discussão na Seção 7 e mais trabalhos referenciados.

Como trabalhos futuros pretende-se prosseguir na exploração dos resultados apresentados neste estudo, assim como todo o potencial que o LOD propicia para repositórios de REAs e para a disseminação desses recursos dentro da nuvem do LOD.

¹³LOD Subclouds - <https://lod-cloud.net/#subclouds> - acessado em 09/10/2020.

Referencias

- Adobe (2017). Open data institute study shows role for pdf with open data. *The Blog Adobe*. <https://theblog.adobe.com/open-data-institute-study-shows-role-for-pdf-with-open-data/> - acessado em 28/09/2020.
- Atkins, D. E., Brown, J. S., and Hammong, A. L. (2007). A review of the open educational resources (OER) movement: Achievements, challenges, and new opportunities. *Creative common, online*.
- Berners-Lee, T. (2006). Linked data. *W3C Standards*. <https://www.w3.org/DesignIssues/LinkedData.html> - acessado em 12/12/2020.
- Fayyaz, N., Ullah, I., and Khusro, S. (2018). On the current state of linked open data: Issues, challenges and future directions. *International Journal on Semantic Web and Information Systems*, 14:110–128.
- Gil, A. C. (2002). *Como elaborar projetos de pesquisa*. Editora Atlas, 4th edition.
- IEEE (2002). IEEE Standard for Learning Object Metadata. *IEEE Std 1484.12.1-2002*, pages 1–40.
- Kawase, R., Fisichella, M., Niemann, K., Pitsilis, V., Vidalis, A., Holtkamp, P., and Pereira Nunes, B. (2013). Openscout: Harvesting business and management learning objects from the web of data. pages 445–450.
- Koutsomitropoulos, D. A. and Solomou, G. D. (2018). A learning object ontology repository to support annotation and discovery of educational resources using semantic thesauri. *IFLA Journal*, 44(1):4–22.
- Lama, M., Vidal, J. C., Garcia, E. O., Bugarin, A., and Barro, S. (2011). Semantic linking of a learning object repository to dbpedia. In *2011 IEEE 11th International Conference on Advanced Learning Technologies*, pages 460–464.
- Nahhas, S., Bamasag, O., Khemakhem, M., and Bajnaid, N. (2018). Added values of linked data in education: A survey and roadmap. *Computers*, 7:45.
- Navarrete, R. and Lujan-Mora, S. (2015). Use of linked data to enhance open educational resources. In *2015 International Conference on Information Technology Based Higher Education and Training (ITHET)*, pages 1–6.
- Paula, L. B., Deus, W. S., and Barbosa, E. F. (2020). Análise de repositórios de REAs em relação ao uso de padrões de Linked Open Data. *Simpósio Brasileiro de Informática na Educação (SBIE 2020)*.
- Pereira, C., Campos, F., Stroële, V., David, J. M., and Braga, R. (2018). Broad-rsi – educational recommender system using social networks interactions and linked data. *Journal of Internet Services and Applications*, 9.
- Rajabi, E., Miguel-Angel, S., and Salvador, S.-A. (2014). Interlinking educational resources to web of data through iee lom. *Computer Science and Information Systems*.
- Sicilia, M., Ebner, H., Sanchez-Alonso, S., Alvarez, F., Abian, A., and Barriocanal, E. (2010). Navigating learning resources through linked data: a preliminary report on the re-design of organic.edunet. volume 717.

-
- Valiente, M.-C., Sicilia, M.-A., Garcia-Barriocanal, E., and Rajabi, E. (2015). Adopting the metadata approach to improve the search and analysis of educational resources for online learning. *Computers in Human Behavior*, 51:1134 – 1141. Computing for Human Learning, Behaviour and Collaboration in the Social and Mobile Network Era.
- Vega-gorgojo, G., Asensio-Perez, J., Gomez-Sanchez, E., Bote-Lorenzo, M., Muñoz-Cristobal, J., and Ruiz-Calleja, A. (2015). A review of linked data proposals in the learning domain. *Journal of Universal Computer Science*, 21:326–364.