

---

# LINKED OPEN DATA NO CONTEXTO ACADÊMICO: IDENTIFICAÇÃO E ANÁLISE DE VOCABULÁRIOS UTILIZADOS NA ACADEMIA E NA PESQUISA CIENTÍFICA

*Linked Open Data for Academic Contexts: identification and analysis of vocabularies used by academic institutions and scientific researches*

---

Vitor Rozsa (1) Moisés Lima Dutra (2), Januário Albino Nhacuongue (3)

Universidade Federal de Santa Catarina (UFSC), vitor.rozsa@hotmail.com (1) moises.dutra@ufsc.br(2)

januanita@hotmail.com (3)

## Resumo

O presente trabalho tem como objetivo identificar quais são os vocabulários Linked Data utilizados na descrição de recursos no contexto acadêmico. Esta é uma pesquisa descritiva, pois caracteriza os vocabulários identificados; utiliza uma abordagem quantitativa para definir indicadores de popularidade; e realiza a coleta de dados por meio de uma abordagem documental. Os vocabulários são identificados por meio de *Web Sites* e conjuntos de dados (*datasets*) no contexto acadêmico. Indicadores de popularidade são criados baseados na presença do vocabulário no Linked Open Vocabularies (LOV) e da quantidade de ocorrência dos vocabulários em *datasets* tanto dentro quanto fora do contexto acadêmico. Os *datasets* utilizados para a identificação de vocabulários foram encontrados em iniciativas Linked Data para o contexto acadêmico. Ao todo, foram identificados 102 vocabulários utilizados na descrição de dados e metadados para o contexto acadêmico, relacionados a variados tópicos: Pessoas e Organizações, Cursos e Formação, Recursos Acadêmicos e Institucionais, Projetos, Pesquisas, Coleções, Recursos de Multimídia, Páginas Web e Redes Sociais, Localização, Tempo e Eventos, Dados Estatísticos, Metadados e Outros. Assim, percebemos que a CI pode contribuir em diferentes vertentes relacionadas ao Linked Data, tais como na criação e disseminação de diretrizes sobre o uso de metadados, conciliadas a partir de práticas e estudos de usuário.

**Palavras-chave:** Linked Education; Linked Science; Linked Data; Vocabulários abertos; Web de Dados; Ontologia

## 1 Introdução

A ciência e a educação superior representam atividades fundamentais para o processo de transformação e evolução da sociedade, uma vez que a pesquisa científica produz novos conhecimentos e a academia dissemina conhecimentos e instrui os cientistas e os futuros profissionais. Estas atividades compõem o contexto acadêmico e situam-se próximas umas das outras, considerando que a pesquisa científica na maior parte das ve-

## Abstract

This work aims to identify the Linked Data vocabularies used to describe resources in academic contexts. It is a descriptive research, based on a quantitative approach to define indicators of popularity, as well as a documental approach for data gathering. We have identified vocabularies by means of academic-context-based Web Sites and datasets. We have proposed popularity indicators based on the presence of vocabularies in the LOV dataset (Linked Open Vocabularies). It has also been taken into account the amount of occurrence of vocabularies in datasets both inside and outside the academic contexts. We have identified the main datasets used to identify vocabularies in Linked Data initiatives for academic contexts. We have found 102 vocabularies used to describe data and metadata for academic contexts, related to various topics such as People and Organizations, Courses and Training, Academic and Institutional Resources, Projects, Research, Collections, Multimedia Resources, Web Pages and Social Networks, Location, Time and Events, Statistical Data, Metadata, and Others. In the end, we can conclude that Information Science can contribute in different aspects for the Linked Data approach, such as in the creation and dissemination of guidelines for metadata use, according to users' studies and practices.

**Keywords:** Linked Education; Linked Science; Linked Data; Open Vocabularies; Web of Data; Ontology

zes é realizada dentro das próprias universidades, que ambas utilizam a informação como insumo fundamental e que o efeito de suas atividades ocorre através da produção e disseminação do conhecimento. Nesse contexto, torna-se primordial a divulgação de informações tanto relacionadas à produção científica (ex.: resultados de pesquisas, experimentos e pesquisadores envolvidos) quanto às instituições de ensino superior (ex.: cursos, professores e materiais educacionais).

Tendo em vista seu emprego na geração de conhecimento, a informação pode ser entendida como um re-

curso. Segundo Silva e Ribeiro (2011), contrariamente ao paradigma custodial, patrimonialista, historicista e tecnicista que sobrevalorizava a guarda e memória do acervo ou patrimônio institucional, o paradigma pós-custodial, informacional e científico trata a informação como recurso. Daí a necessidade da sua valorização como fenômeno humano e social e, conseqüentemente, possuidor de dinamicidade no acesso e uso. Na Ciência da Informação (CI), esta mudança paradigmática suscita novas indagações e exercícios hermenêuticos (Silva, 2010). Nesta mudança, o profissional da informação possui um papel preponderante nas tarefas de reuni-la, organizá-la e disseminá-la, com o intuito de possibilitar a geração de conhecimento. Silva (2010) classifica a mediação pós-custodial inerente a estas tarefas em institucional, distribuída e/ou partilhada, e cumulativa.

A mediação institucional é realizada por bibliotecários, arquivistas e demais profissionais da informação nas tradicionais instituições culturais e é partilhada por profissionais empenhados na disponibilização de acervos. A mediação distribuída ocorre em certos tipos de serviços e mídias digitais. Enquanto o mediador localiza, digitaliza, seleciona e disponibiliza o conteúdo, outros profissionais fornecem aplicações para que os usuários possam atuar em conteúdos e *links* adicionais. Por último, a mediação cumulativa exige uma inovação para desenvolver iniciativas tecnológicas visando tornar os recursos acessíveis (Silva, 2010). No contexto das mediações apontadas por Silva (2010), as atividades dos profissionais da CI podem incidir sobre políticas para o fomento de vocabulários abertos no contexto acadêmico. Nos últimos anos, uma das principais ferramentas modernas utilizadas para atender as necessidades de organizar, armazenar, disseminar e acessar a informação no contexto acadêmico é a Web de Dados, por intermédio da adoção dos princípios do Linked Data (Berners-Lee, 2006) e suas tecnologias relacionadas. A Web de Dados é a noção na qual os dados estão disponíveis em um meio comum, estão fortemente relacionados e são publicados de acordo com determinados padrões, facilitando sua identificação e uso. Os dados publicados na Web de Dados encontram-se disponíveis em um “espaço global”, o que permite sua descoberta e utilização pelas pessoas e aplicações diversas (Bizer, Heath e Berners-Lee, 2009).

Para que conjuntos de dados (do inglês, *datasets*) tornem-se parte da Web de Dados, é necessária a conformidade com os princípios do Linked Data, que são de uma série de práticas e diretrizes para publicar e estruturar dados na Web (Bizer; Heath; Berners-Lee, 2009). A interligação entre os dados de diferentes fontes estabelece a conexão em um único espaço global de dados, enquanto o uso dos padrões propostos e de um modelo de dados comum permite a criação de aplicações genéricas que são capazes de operar sobre todo este espaço de dados, o que resume a essência do Linked Data (Heath e Bizer, 2011).

Embora possamos identificar vantagens para a operacionalização da informação na adoção dos princípios do Linked Data que possuem potencial para transformar a ciência e a educação superior, estes princípios não são amplamente empregados no contexto acadêmico (D'Aquin et al., 2013a; Linked Education, 2017). A colaboração científica busca há muito tempo prover maior compartilhamento e reuso de conhecimento e dados (Keßler, D'Aquin e Dietze, 2013), porém a maioria dos cientistas ainda vê a falta de acesso aos dados gerados por outros pesquisadores ou instituições como um grande impedimento ao progresso da ciência (Purohit et al., 2016). Ainda assim, é possível observar que o uso de recursos científicos e acadêmicos disponíveis por meio do Linked Data tem aumentado (D'Aquin et al., 2013a; Daga et al., 2016), justificando o interesse na continuação de iniciativas e demandando trabalhos que visem a publicação destes recursos na Web de Dados.

Alguns dos fatores que tornam localizáveis os dados disponíveis na Web de Dados é o reuso de vocabulários reconhecidos e sua interligação com dados em outros *datasets*. De acordo com D'Aquin et al. (2013a), os principais mecanismos que conectam diferentes *datasets* são a criação de ligações explícitas entre os recursos de *datasets* e o reuso de vocabulários. Os vocabulários permitem a definição de conceitos, propriedades e relações entre diferentes conceitos e são utilizados na estruturação e contextualização dos dados na Web de Dados para que possam ser compreendidos e processados automaticamente. Em vista disso, existem iniciativas que centralizam vocabulários ou facilitam sua identificação como, por exemplo, o Linked Universities, Linked Education, OLE, Projeto LinkedUp, Open University (descritos em detalhes na seção seguinte) Linked Open Vocabularies (LOV) e LODStats (descritos em detalhes na seção 3).

Diante da relevância do papel dos vocabulários Linked Data no contexto da Web de Dados, objetiva-se com este trabalho identificar quais são os vocabulários utilizados na descrição de recursos no contexto acadêmico. Pautada neste objetivo, esta é uma pesquisa descritiva, pois busca identificar e caracterizar os vocabulários Linked Data empregados no contexto acadêmico. Utiliza uma abordagem quantitativa, pois analisa a quantidade de ocorrências (usos) dos vocabulários nos diferentes *datasets*. Em relação aos procedimentos de coleta de dados, os vocabulários estudados neste trabalho são identificados por meio da investigação de Websites e *datasets* relacionados ao Linked Data e voltados para o contexto acadêmico, sendo assim uma pesquisa documental. Posto que as atividades de pesquisa científica estão amplamente ligadas às universidades, implicando na ocorrência da construção de *datasets* científicos pelas próprias instituições acadêmicas, tratamos as atividades científicas e acadêmicas relacionadas ao

Linked Data dentro de um mesmo contexto, o acadêmico.

A pesquisa realizada neste trabalho está dividida em duas macro-etapas: (i) identificação e análise dos vocabulários sugeridos pelas iniciativas; e (ii) identificação e análise dos vocabulários para o contexto acadêmico empregados em *datasets*. A primeira macro-etapa envolve: (a) a busca de vocabulários explicitamente sugeridos nos portais das iniciativas; e (b) a verificação da presença destes vocabulários na ferramenta centralizadora de vocabulários LOV. A segunda macro-etapa envolve: (a) a identificação de *datasets* no contexto acadêmico a partir das iniciativas; (b) a identificação dos vocabulários empregados na descrição destes *datasets*; (c) a coleta de informações referentes à autoria e reuso do vocabulário; e (d) o agrupamento destes vocabulários em categorias temáticas de modo a facilitar sua visualização e busca.

Na seção 2 são descritas as principais iniciativas na publicação de Linked Data no contexto acadêmico. Na seção 3, descreve-se o papel dos vocabulários na Web de Dados e ferramentas relevantes para a obtenção de informações sobre os vocabulários. Na seção 4 são apresentados os trabalhos relacionados. Na seção 5 são apresentados os vocabulários identificados por intermédio das iniciativas descritas na seção 2 e é realizada a discussão sobre os resultados obtidos. E, finalmente, na seção 6 são realizadas as considerações finais deste trabalho.

## 2 Linked Open Data no Contexto Acadêmico

O cenário Linked Data ligado ao contexto acadêmico evolui a partir da disponibilização de novos *datasets* relevantes para este contexto. A disponibilização, ou seja, a criação e publicação de *datasets* é provocada pela necessidade de tornar acessíveis estes dados para consumo das pessoas. Este processo é favorecido principalmente a partir da criação de iniciativas e projetos que tenham por objetivo reunir esforços e facilitar a troca de experiências, disseminação de conhecimentos, ferramentas e vocabulários voltados para o Linked Data.

No contexto acadêmico existem iniciativas que seguem e disseminam os princípios do Linked Data. Estas iniciativas incentivam a produção de *datasets* relacionados a instituições e materiais educacionais. Entre estes, estão informações de repositórios educacionais abertos e dados produzidos pelas universidades (Dietze et al., 2013), além de *datasets* relacionados a informações científicas sobre, por exemplo, pesquisadores, publicações e dados de pesquisa (Taibi et al., 2016). Pode-se visualizar exemplos destas iniciativas na forma de ações independentes e de alianças, como o Open University, a Linked Universities e o Linked Education, de grupos para o desenvolvimento de padrões, como o Open Linked Education Data Community Group, e de

projetos de incentivo ao uso de Linked Data, como o projeto LinkedUp.

Nesta seção são descritas algumas das iniciativas mais expressivas nos últimos anos voltadas para a evolução do Linked Data no contexto acadêmico. A tabela 1 sintetiza as iniciativas Linked Data no contexto acadêmico que são exploradas nesta pesquisa.

<i>Iniciativa</i>	<i>Descrição</i>
Linked Universities	Aliança de iniciativas (especialmente estabelecidas em universidades) que buscam disponibilizar publicamente seus dados por meio do Linked Data.
Linked Education	Plataforma <i>online</i> voltada para a promoção do Linked Data para propósitos acadêmicos.
Open Linked Education Data Community Group	Grupo que visa o desenvolvimento de padrões para o Linked Data no contexto acadêmico.
Projeto LinkedUp	Projeto que visou a exploração e uso de dados disponíveis na Web com propósitos acadêmicos.
Open University	Universidade de ensino à distância baseada no Reino Unido que busca tornar seus dados públicos mais acessíveis, reutilizáveis e exploráveis.

Tabela 1. *Iniciativas Linked Data no Contexto Acadêmico.*

### 2.1 Linked Universities

O Linked Universities é uma aliança de iniciativas que buscam tornar seus dados públicos, por meio do Linked Data. Esta aliança visa a junção dos esforços de diversas frentes, especialmente estabelecidas em universidades, de modo a disseminar conhecimentos obtidos individualmente. Deste modo, o Linked Universities fomenta o compartilhamento de práticas e a colaboração no desenvolvimento de plataformas Linked Data sob uma ótica compartilhada, a criação de uma Web de dados de Universidades.

A partir desta visão, o Linked Universities traça seus objetivos (Linked Universities, 2014): (i) identificar, manter e desenvolver vocabulários Linked Data comuns, utilizáveis pelas universidades para conceitos comuns; (ii) descrever os procedimentos e compartilhar ferramentas com o objetivo de disponibilizar Linked Data nas universidades; e (iii) apoiar, por meio do compartilhamento e reuso de experiências, iniciativas para a disponibilização de dados nas universidades.

O Linked Universities mantém um portal Web como espaço colaborativo para disponibilização de recursos para o desenvolvimento de Linked Data nas universidades. Neste portal, é possível encontrar informações sobre os membros da comunidade do Linked Universities, vocabulários para anotação de dados, ferramentas,

publicações, *datasets* e SPARQL EndPoints (interfaces de consulta manual ou automática dos *datasets*) relacionados com os esforços da aliança.

## 2.2 Linked Education

O Linked Education se trata de uma plataforma *online* voltada para a promoção do Linked Data para propósitos acadêmicos. Os principais esforços desta plataforma estão direcionados para o compartilhamento de recursos e informações relacionadas ao Linked Data para o contexto acadêmico, tendo em vista facilitar e promover o compartilhamento de dados e recursos acadêmicos em escala Web.

O Linked Education disponibiliza um portal para pesquisadores e profissionais do Linked Data discutirem e trocarem informações sobre *datasets*, vocabulários, aplicações e eventos e para identificarem melhores práticas relacionadas com o Linked Data no contexto acadêmico. A partir disso, o Linked Education descreve o seu principal objetivo como sendo (Linked Education, 2016, tradução livre): “identificar melhores práticas bem como potenciais ligações entre recursos individuais para contribuir para uma Web de Dados educacionais bem interconectada”. Em seu portal Web, é ainda possível encontrar informações sobre *datasets* e vocabulários para a educação, ferramentas e aplicações, notícias e eventos relacionados ao Linked Data educacional.

## 2.3 Open Linked Education Data Community Group

O Open Linked Education Data Community Group (1) (OLED) é um grupo mantido pelo W3C, criado como ponto central de trabalho tendo em vista o desenvolvimento de padrões para o Linked Data no contexto acadêmico. O grupo é utilizado para a realização de discussões, desenvolvimento de testes e para permitir a conexão entre a comunidade internacional de especialistas na Web. O OLED fornece acesso a recursos a partir de sua Wiki. Entre estes recursos estão vocabulários para a representação de recursos acadêmicos, exemplos de modelagem de aspectos particulares na educação, padrões de design na modelagem de *datasets* para a educação e *datasets* que empreguem os padrões sugeridos pela Wiki.

Os esforços do OLED estão direcionados para (Open Linked Education Data Community Group, 2015): (i) identificar, a partir de iniciativas Linked Data, as práticas utilizadas para compartilhar recursos relacionados à educação na Web; (ii) identificar as melhores práticas na criação de *datasets* para o contexto acadêmico e documentá-las; e (iii) facilitar a adoção destas práticas por meio da interação com a comunidade e da evidência dos benefícios do reuso destas práticas na modelagem de dados no desenvolvimento de aplicações.

Por fim, é de se notar que, de acordo com a ideia de que existem diversos aspectos dentro de um contexto

que não podem ser atendidos satisfatoriamente por apenas um único vocabulário, o OLED salienta que não é seu objetivo criar uma ontologia para a educação, uma vez que este é um extenso contexto. Mas é objetivo do OLED fornecer padrões comuns que utilizam vocabulários existentes na representação de dados relacionados ao contexto acadêmico.

## 2.4 Projeto LinkedUp

O LinkedUp é um projeto que visa a exploração e uso de dados disponíveis na Web com propósitos acadêmicos. O projeto foi mantido pelo *7th Framework Programme for Research and Technological Development* (FP7), tendo iniciado em novembro de 2012 e durado 2 anos. Entre suas principais atividades em seu período ativo estavam a identificação e promoção de casos de sucesso e inovadores que exploram a Web de Dados em cenários acadêmicos como parte de aplicações e ferramentas; e a curadoria de *datasets* no contexto acadêmico.

Entre os principais objetivos do projeto, tendo sempre em vista o caráter aberto dos dados, estavam (LinkedUp, 2017): (i) identificar cenários inovadores e robustos de ferramentas que integram e analisam dados no setor educacional; (ii) coletar, anotar e descrever dados na Web de Dados com relevância para a educação para viabilizar seu uso por terceiros; (iii) fornecer uma *framework* completo para a avaliação de aplicações em larga escala na Web de Dados, levando em conta aspectos educacionais; e (iv) demonstrar e promover os benefícios das tecnologias da Web de Dados na educação e fornecer uma plataforma de testes para aplicações que utilizam Linked Data reutilizável neste domínio.

Como principais resultados destes objetivos, esperava-se (LinkedUp, 2017): (i) fornecer *datasets* de qualidade garantida para a educação; (ii) um *framework* de avaliação de propósito genérico para aplicações baseadas na Web de Dados; (iii) a criação de aplicações inovadoras de gestão da informação na Web; (iv) o cruzamento entre os setores públicos e privados; e (v) a transferência de tecnologias inovadoras para a gestão da informação na Web.

Uma das principais contribuições do LinkedUp está na criação do Linked Education Catalog, um catálogo de *datasets* reunidos tendo em vista a construção de aplicações no contexto acadêmico. Este catálogo está disponível no serviço de hospedagem de *datasets* Datahub.io, no qual os *datasets* relacionados estão reunidos em um grupo denominado “Linked Education Cloud”. Ao todo, há um total de 36 *datasets* reunidos no grupo Linked Education Cloud. O propósito inicial do catálogo fora fornecer *datasets* para atividades relativas ao LinkedUp, mas está disponível para acesso e uso por meio do Datahub.io.

## 2.5 Open University

A Open University (OU) é uma universidade pública de ensino à distância e pesquisa, baseada no Reino Unido. A partir do final de 2010, como resultado do projeto LUCERO, foi criado o ponto central de Linked Data da OU, o “data.open.ac.uk”, aqui referido como portal Linked Data da OU.

No portal Linked Data da OU são fornecidas informações sobre os *datasets* gerados pela OU, informações sobre a construção dos *datasets* (como termos e propriedades e os vocabulários utilizados), endereço dos SPARQL endpoints e exemplos de aplicações baseadas nos *datasets* disponíveis. O objetivo em expor dados no portal Linked Data da OU é tornar os dados públicos disponíveis mais acessíveis, reutilizáveis e exploráveis (Daga et al., 2016). Assim, a OU disponibiliza diferentes *datasets* em 6 tipos de categorias (The Open University, 2017): recursos educacionais abertos, produção científica, mídias sociais, dados organizacionais, dados de projetos de pesquisa e metadados sobre seus *datasets*.

Por fim, é interessante notar que a partir de uma visão mais ampla estas iniciativas estão interligadas, abrindo espaço para contribuições mútuas. Essa interligação ocorre pela participação dos mesmos membros em diferentes projetos, que aproveitam suas experiências prévias para contribuir em novas situações, pela indicação dos mesmos vocabulários e *datasets* acadêmicos por diferentes portais, reforçando a importância daquele vocabulário ou *dataset* no contexto acadêmico, e pelas referências (e.g. por meio de notícias e *posts* em fóruns) que uma iniciativa pode fazer a outra, revelando novas fontes de informação aos seus usuários. Percebe-se assim que esta colaboração, situada num nível mais amplo entre as iniciativas analisadas, tem apenas a agregar ao desenvolvimento do Linked Data para o contexto acadêmico.

## 3 Vocabulários Linked Data

Para que a integração de recursos na Web por meio do Linked Data seja possível, é necessário tanto prover um modelo de dados comum quanto permitir a descrição de conceitos e a forma como estes estão relacionados. Na Web de Dados, enquanto o Resource Description Framework (RDF) fornece um modelo de dados genérico e abstrato para a descrição de recursos por meio de triplas contendo *sujeito*, *predicado* e *objeto*, os vocabulários são utilizados para descrever termos e conceitos específicos no domínio e seus diferentes tipos de relacionamento (Heath e Bizer, 2011). Assim, um modelo de dados genérico viabiliza a interoperabilidade sintática, permitindo a integração de dados de fontes diversas, enquanto o compartilhamento de vocabulários viabiliza a interoperabilidade semântica, permitindo a conceitualização formal dos dados e sua interpretação pelas aplicações.

Dentro de um mesmo contexto é possível encontrar situações diversas das quais os dados de um *dataset* são derivados. Por exemplo, uma instituição de ensino pode prover informações sobre pessoas (e.g. alunos, professores e servidores), recursos educacionais (e.g. textos, áudio e vídeos), projetos de pesquisa (e.g. temas de pesquisa, pesquisadores e publicações) e recursos físicos (e.g. móveis, salas e laboratórios). Cada situação pode apresentar uma grande quantidade de termos específicos, exigindo o emprego de diferentes vocabulários na descrição destes termos.

Uma vez que os vocabulários possibilitam a significação dos dados em nível semântico, eles representam uma peça essencial na construção da Web de Dados. A relevância de seu papel, somado a diversidade de contextos e situações que exigem representações específicas, tem por consequência a proliferação dos vocabulários disponíveis na Web (D'Aquin e Noy, 2012). Em geral, a grande quantidade de diferentes conceitos presentes em um *dataset* exige a combinação de diferentes vocabulários para a descrição destes dados. Esta diversidade motiva a busca por vocabulários que podem não ser encontrados (ou não existem), gerando a demanda da construção de novos vocabulários. Quando estes novos vocabulários não são disponibilizados em locais “visíveis” e acessíveis na Web, como portais específicos ou bibliotecas de vocabulários, eles não são reutilizados e, conseqüentemente, podem ter seus termos duplicados em um novo vocabulário.

Sugere-se enfaticamente o reuso ao invés da criação de novos vocabulários que dupliquem os termos de vocabulários já existentes. O reuso aumenta a probabilidade de os dados serem consumidos por aplicações que estão baseadas em vocabulários mais conhecidos, de forma que estas não necessitam de etapas extras de processamento dos dados ou de adaptações na aplicação para reconhecer os diferentes termos (Heath e Bizer, 2011). Quando diferentes fornecedores utilizam o mesmo vocabulário para descrever seus dados, a integração destes dados é amplamente facilitada (D'Aquin e Noy, 2012). Em vista disso, a busca por vocabulários para a descrição de conceitos e relações é uma importante etapa durante a construção de um *dataset*.

De acordo com Heath e Bizer (2011), alguns dos critérios a serem considerados na seleção de um vocabulário para reuso estão, de maneira geral, relacionados com: (i) sua popularidade; (ii) um processo claro e ativo de governança; (iii) sua cobertura de termos no *dataset* em questão; e (iv) seu grau de expressividade de modo que atenda às necessidades do *dataset* em questão.

Ainda, caso não existam vocabulários que atendam às necessidades do *dataset* a ser descrito, os autores supracitados descrevem algumas das melhores práticas na criação de vocabulários dedicados. Entre elas estão a busca pela suplementação de vocabulários já existen-



Core, do que vocabulários de domínios específicos do contexto acadêmico. Como este baixo reuso de vocabulários é indesejável, pois dificulta a utilização de dados de diferentes *datasets*, os autores realizaram um teste no qual resumiram os diferentes termos encontrados em um conjunto reduzido de vocabulários, tendo como resultado uma maior coesão da rede de dados educacionais no contexto trabalhado - nos *datasets* da pesquisa - e maior facilidade em utilizar e recuperar informações de forma conjunta nestes *datasets*.

O trabalho de d' Aquin et al. (2013) revela uma preocupação relacionada com a finalidade da publicação de *datasets* na Web de Dados, que é o uso dos dados por aplicações diversas. A questão da finalidade – o propósito do *dataset* - deve ser considerada durante a construção dos *datasets*, na escolha dos vocabulários utilizados para a estruturação e descrição dos dados que serão disponibilizados, de forma a facilitar sua descoberta e uso pelas aplicações.

No trabalho de Taibi et al. (2016), os autores buscaram avaliar o escopo e cobertura do Linked Data em relação ao contexto acadêmico. Para isso, identificaram os tipos de recursos e tópicos acadêmicos cobertos pelo Linked Data, os principais tópicos cobertos e os tópicos sub-representados no contexto acadêmico. Seus principais resultados estão relacionados a descoberta dos aspectos mais adequados para caracterizar o conteúdo dos *datasets*. Um destes aspectos é considerar o tipo de recurso descrito além da categoria atribuída a ele. Além disso, os autores ainda identificaram uma tendência à descrição de disciplinas voltadas para a Ciência da Computação e Ciências da Vida, sendo as Ciências Sociais Aplicadas uma disciplina sub-representada no contexto acadêmico.

Neste trabalho (Taibi et al., 2016) ainda são citadas as contribuições do LinkedUp e comenta-se sobre o LinkedUp Catalog, composto tanto por *datasets* criados para a academia, quanto *datasets* criados em outros cenários, mas considerados exploráveis no contexto acadêmico. A partir da ideia de exploração de *datasets* tanto dentro quanto fora do contexto acadêmico, os autores levantam duas questões: quais *datasets* e recursos podem ser empregados no contexto acadêmico; e como os *datasets* e recursos devem ser descritos para facilitar sua busca e exploração. Esta última questão é relevante para o presente trabalho uma vez que, segundo Taibi et al. (2016), *datasets* mal descritos apresentam alta dificuldade de compreensão para o usuário e podem ser facilmente ignorados. Desta forma, ressaltam a importância na escolha dos termos e propriedades, ou seja, dos vocabulários utilizados na descrição dos dados.

Por fim, em pesquisa realizada por Schaible et al. (2014), os autores solicitaram aos publicadores e consumidores de Linked Data que respondessem um questionário com o objetivo de identificar quais são as es-

tratégias preferidas de reuso de vocabulários. Os autores argumentam que, além da recomendação geral de utilizar vocabulários consolidados e reconhecidos, não há recomendações estabelecidas sobre como selecionar os vocabulários a serem reutilizados. Isto posto, entender-se-ia que os projetistas de *datasets* contam mais com sua experiência e intuição do que com qualquer outro recurso. Como resultados de sua pesquisa, os autores identificaram que é considerada uma melhor abordagem utilizar termos de vocabulários reconhecidos e de domínios específicos do que definir novos termos proprietários e interligá-los com classes e propriedades externas. Além disso, pôde ser identificado que é preferível, pelos fornecedores de dados, o uso de termos de vocabulários mais frequentemente utilizados a termos de vocabulários menos utilizados. Quanto a estes aspectos, foi possível perceber que informações sobre o domínio de aplicação do vocabulário e sobre a quantidade de *datasets* utilizando o vocabulário são consideradas relevantes na decisão de qual vocabulário reutilizar. Também foi possível perceber que a popularidade do vocabulário como um todo é mais importante do que a frequência com que são utilizados seus termos e propriedades. E, por fim, identificou-se a preferência de se manter uma mistura equilibrada entre vocabulários populares e específicos de domínio, de modo a fornecer uma estrutura clara dos dados e facilitar seu consumo.

## 5 Vocabulários para o Contexto Acadêmico

Nesta seção são descritas as ações empregadas e resultados obtidos na identificação de vocabulários voltados para o contexto acadêmico. Entre estas ações estão principalmente a busca por listas explícitas de vocabulários em portais na Web e a análise de *datasets*. As principais fontes analisadas foram as iniciativas descritas na seção 2 e na seção 3, contando com: Linked Universities, Linked Education, OLED, Projeto LinkedUp e Open University.

Inicialmente, são listados os vocabulários explicitamente sugeridos pelas iniciativas analisadas para uso no contexto acadêmico. Em seguida, os vocabulários identificados no passo anterior são buscados dentro dos vocabulários disponibilizados no projeto LOV, de forma a fornecer um primeiro subsídio para a compreensão da contribuição desta ferramenta em relação à centralização de vocabulários para o contexto acadêmico (seção 5.1). Depois, são listados os vocabulários identificados a partir dos *datasets* analisados (seção 5.2) e, finalmente, é realizada a discussão sobre os resultados obtidos a partir destas ações (seção 5.3).

### 5.1 Vocabulários Identificados em Portais

Entre as iniciativas estudadas, foi possível identificar sugestões de vocabulários para o contexto acadêmico nos portais do Linked Universities, Linked Education e na *wiki* do OLED. Os vocabulários identificados estão

listados no Quadro 1 (apêndice), no qual são apresentados um total de 28 vocabulários que estão divididos em 10 categorias, definidas a partir de informações encontradas nos próprios portais nos quais os vocabulários foram identificados. Esta primeira amostra revela a pluralidade de áreas específicas que demandam vocabulários para a descrição de conceitos e suas relações. Os nomes dos vocabulários e as categorias nas quais estão inseridos fornecem a informação necessária para identificar sua fonte original. Realizaram-se buscas na ferramenta Google para cada vocabulário mencionado no Quadro 1. O termo de busca utilizado foi composto pelo nome do vocabulário mais a palavra “vocabulary”. Por meio deste procedimento (realizado em 23/02/2017) todos os vocabulários foram identificados na primeira página de resultados da ferramenta de busca.

Como complemento desta primeira ação, também buscou-se cada vocabulário presente no Quadro 1 dentro do projeto LOV. No portal Web do projeto LOV há uma ferramenta que permite a busca de vocabulários, que foi utilizada nesta tarefa (Figura 2). No total, foram identificados 20 vocabulários presentes no Quadro 1 por meio da ferramenta de busca de vocabulários. Vale ainda mencionar que dentro desta mesma ferramenta é possível navegar pelos vocabulários por meio de marcadores que separam as categorias de vocabulários. Atualmente, para o marcador “Academia” (no original, *Academy*) são recuperados 10 vocabulários. Mas também há marcadores para domínios gerais ou específicos que podem ser relevantes para a descrição de *datasets* acadêmicos como, por exemplo, Geografia, Biologia, Multimídia e Pessoas.

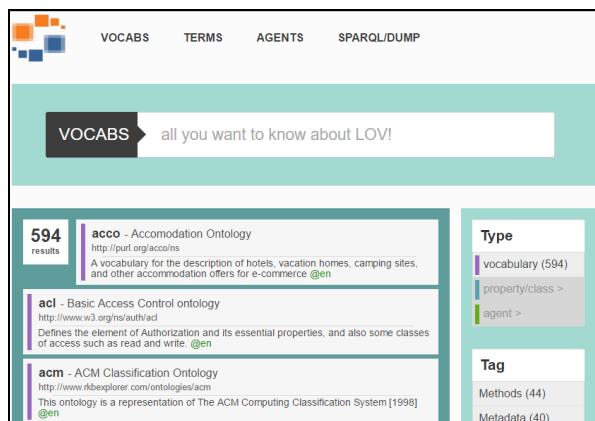


Figura 2. Ferramenta de Busca do Projeto LOV. (Linked Open Vocabularies, 2017b).

A partir do Quadro 1 e dos resultados obtidos na ferramenta de busca, é possível observar que de um total de 28 (100%) vocabulários pesquisados, 20 (71,4%) foram encontrados no portal LOV. Ou seja, aproximadamente três em cada quatro vocabulários sugeridos nas listas específicas estavam disponíveis neste recurso centralizador. É interessante notar que apenas quatro dos 28 vocabulários pesquisados estavam com o mar-

cador “Academia”. Este baixo número de vocabulários marcados reforça a importância de referências específicas que indiquem quais são os vocabulários passíveis de serem utilizados em determinados contextos.

## 5.2 Vocabulários Identificados em *Datasets*

Para todas as iniciativas pesquisadas foi possível identificar *datasets* exploráveis no contexto acadêmico. Assim, inicialmente identificou-se *datasets* válidos para serem utilizados na análise. Em seguida, identificou-se os vocabulários empregados na descrição destes *datasets*. Então, buscou-se coletar informações específicas de cada vocabulário e, por fim, agrupou-se os vocabulários em categorias.

Os *datasets* foram identificados a partir das iniciativas Linked Data no contexto acadêmico, obtendo-se um total de 36 *datasets*. Evitou-se duplicar informações quando se identificou o mesmo *dataset* listado por diferentes iniciativas. Diante disso, vale mencionar que o projeto LinkedUp e o portal de dados abertos da OU forneceram a maior parte de *datasets* únicos - não repetidos em outras fontes - para esta análise.

O procedimento de identificação dos vocabulários contou com a descoberta e análise de recursos relacionados aos *datasets* que fazem referência aos vocabulários empregados na descrição dos dados. Para cada *dataset* selecionado, procurou-se identificar recursos como o *dump* das triplas do *dataset*, SPARQL EndPoint, arquivos de metadados e documentações *online*. Em alguns casos, o *dump* de um *dataset* contém os vocabulários utilizados em sua descrição. O mesmo pode ocorrer caso exista disponível um SPARQL EndPoint para o *dataset*, que lista os vocabulários específicos para acessar as triplas daquele *dataset*. Arquivos de metadados de um *dataset* também podem fornecer informações sobre os recursos utilizados na descrição do *dataset*. A Figura 3 (em apêndice) contém exemplos do *dump* de um *dataset* (Figura 3a), de um SPARQL EndPoint (Figura 3b) e de um arquivo de metadados (Figura 3c). Por fim, a documentação *online*, em alguns casos, pode listar os vocabulários empregados em sua descrição. Nesse contexto, também foram suprimidos *datasets* indisponíveis e que não forneciam pelo menos uma das formas mencionadas de identificar os vocabulários empregados em sua descrição.

Desta forma, foram identificados 102 vocabulários utilizados na descrição de dados e metadados para o contexto acadêmico. Alguns vocabulários não apresentaram recursos adequados ao entendimento de seu propósito. Entre as deficiências estão: vocabulário inacessível, documentação insuficiente ou inexistente e vocabulário em idioma diferente do português, inglês e espanhol. Estes vocabulários não foram incluídos na análise. Além disso, os vocabulários RDF, RDFS, OWL, Semantic Web Rule Language (SWRL) e Semantic Web Rule Language Built-Ins (SWRLB) foram igno-



rados por se tratarem de recursos básicos na descrição de *datasets*, sendo que os três primeiros puderam ser observados na grande maioria dos *datasets* analisados. Para cada um dos vocabulários incluídos na análise foram identificadas determinadas características com diferentes propósitos.

Na tabela 2 são descritas as características recuperadas dos vocabulários analisados. As características descritas estão divididas entre os grupos “Características Gerais” e “Características de Popularidade”. A escolha das características observadas foi influenciada pelos resultados da pesquisa de Schaible, Gottron e Scherp (2014), descrito na seção de trabalhos relacionados. Com o primeiro grupo, “Características Gerais”, buscou-se fornecer informações que provêm uma visão geral do vocabulário, uma vez que a autoridade que publica e/ou mantém o vocabulário pode indicar qualidade e o tipo de documentação do vocabulário, e a disponibilidade do vocabulário para consulta pode representar facilidade de compreensão e uso do vocabulário.

Com o segundo grupo, “Características de Popularidade”, espera-se fornecer informações acerca da popularidade do vocabulário. Alguns vocabulários são criados para atender a determinados *datasets* e por isso são mais específicos e menos reutilizáveis. De forma inversa, vocabulários menos específicos possuem maiores condições de serem reutilizados e por isto são mais populares e podem ser observados em diferentes *datasets*. Assim sendo, espera-se auxiliar na identificação do direcionamento do vocabulário por meio de informações como sua presença ou não no LOV, pela quantidade de *datasets* acadêmicos (aqueles analisados neste trabalho) e *datasets* gerais (encontrados no LODStats) que o reutilizam.

<i>Característica</i>	<i>Descrição</i>
<i>Características Gerais</i>	
Nome	Identificação do vocabulário.
Autoridade	Identificação da autoridade (organização ou autores) que criou e/ou mantém o vocabulário. No caso de múltiplos autores, apenas um dos autores foi citado por motivos de simplificação
Documentação (Doc)	Tipo de documentação fornecida para o vocabulário. Foram identificados 4 tipos de documentação: <i>online</i> (página Web), documento PDF para <i>download</i> , artigo científico e <i>intra</i> vocabulário.
Disponibilidade (Dis)	Vocabulário disponível para consulta <i>online</i> ou <i>download</i> em algum formato (RDF/XML, N3, Turtle, etc)
<i>Características de Popularidade</i>	
Presença no LOV (Lov)	Vocabulário listado no LOV.
<i>Datasets</i> Acadêmicos (Aca)	Número de <i>datasets</i> no contexto acadêmico nos quais o vocabulário foi identificado.
<i>Datasets</i> em geral (Ger)	Número de <i>datasets</i> presentes no LODStats nos quais o vocabulário foi identificado.

Tabela 2. *Características recuperadas dos vocabulários.*

Finalmente, identificou-se o propósito e atribuiu-se uma categoria a cada vocabulário, de modo a fornecer informações sobre o contexto no qual o vocabulário pode ser empregado. Foram criadas 13 categorias para dividir os vocabulários por contexto de aplicação. Preferiu-se criar novas categorias e mais gerais do que aquelas identificadas nos portais, de modo a possibilitar o agrupamento de uma quantidade maior de vocabulários com propósitos similares. As categorias criadas foram: Pessoas e Organizações, Cursos e Formação, Recursos Acadêmicos e Institucionais, Projetos, Pesquisas, Coleções, Recursos de Multimídia, Páginas Web e Redes Sociais, Localização, Tempo e Eventos, Dados Estatísticos, Metadados e Outros, e são descritas em mais detalhes a seguir. Além disso, as categorias foram reunidas em grupos distribuídos em cinco tabelas para facilitar a apresentação e discussão dos resultados.

Na Tabela 3(em apêndice) está representado o primeiro grupo, contendo as categorias de Pessoas e Organizações, Cursos e Formação e Recursos Acadêmicos e Institucionais. A categoria de Pessoas e Organizações apresenta vocabulários utilizados, por exemplo, na descrição de informações gerais sobre discentes, docentes, servidores, gestores, instituições acadêmicas e organizações em geral, e também na descrição do relacionamento entre estas entidades. A categoria Cursos e Formação apresenta vocabulários utilizados na descrição

de cursos, suas características e relações com outros recursos acadêmicos. E a categoria Recursos Acadêmicos e Institucionais apresenta vocabulários utilizados na descrição de recursos gerais e de aprendizagem no contexto acadêmico.

Já na Tabela 4 (em apêndice) são apresentados os vocabulários relacionados com as categorias Projetos, Pesquisas e Coleções. Os vocabulários na categoria Projetos são utilizados na descrição das propriedades de projetos de *software* e projetos de pesquisa científica, de grupos de pesquisa, editais para projetos e seus resultados, e também na descrição da estrutura de financiamento de projetos de pesquisa. A categoria Pesquisas engloba vocabulários voltados para a descrição de informações relacionadas à pesquisa científica como, por exemplo, pesquisadores, comunidades de pesquisa, artigos e citações. E a categoria Coleções contém vocabulários utilizados na descrição de recursos de coleções de livrarias, arquivos e museus. Vale notar que a maior parte dos vocabulários nesta última categoria estão voltados mais para a descrição de registros bibliográficos.

Na Tabela 5 (em apêndice) são apresentados os vocabulários agrupados nas categorias Recursos de Multimídia e Páginas Web e Redes Sociais. Uma vez que os suportes de informação são diversos, os *datasets* no contexto acadêmico utilizam variados vocabulários para representar recursos de multimídia. A categoria Recursos de Multimídia engloba vocabulários mais gerais para descrição de mídias como áudio e vídeo, e vocabulários mais específicos para a descrição de, por exemplo, vídeos no Youtube, músicas e *podcasts*. Já a categoria Páginas Web e Redes sociais engloba vocabulários utilizados na descrição de recursos acadêmicos na forma de páginas Web e para a descrição de comunidades *online*.

Na Tabela 6 (em apêndice) estão representadas as categorias Localização, Tempo e Eventos e Dados Estatísticos. A primeira categoria, Localização, envolve vocabulários para a descrição de coordenadas espaciais, informações geoespaciais e relações espaciais. A segunda categoria, Tempo e Eventos, apresenta vocabulários que permitem a descrição de datas, horários, intervalos de tempos, eventos e situações. E a terceira categoria, Dados Estatísticos, apresenta vocabulários utilizados na descrição de conceitos de dados estatísticos relacionados à academia.

Por fim, na Tabela 7 (em apêndice) são apresentadas as categorias Metadados e Outros. A categoria Metadados engloba vocabulários utilizados na descrição de informações que auxiliam na descoberta de outras informações mais específicas no *dataset*. Nesta categoria há vocabulários para adição de, por exemplo, informações para catalogação e arquivamento de *datasets*, licença de direitos autorais, exemplos e notas de uso, metadados sobre pesquisas e questionários. A categoria Outros

foi criada para agrupar os vocabulários que não foram encaixados nas demais categorias. Nessa categoria há vocabulários, por exemplo, para a descrição de produtos e negócios, para a classificação de áreas acadêmicas e para a modelagem de medidas e suas leituras.

### 5.3 Discussão

Um dos principais meios utilizados pelos usuários para a compreensão e seleção dos vocabulários é sua documentação. A documentação de um vocabulário pode ser externa, quando criado um documento específico com informações sobre o vocabulário, ou interna, quando estas informações são inseridas diretamente no código do vocabulário. Quanto a este quesito, foram considerados como documentados os vocabulários para os quais foi possível identificar informações relacionadas como, por exemplo, a origem, propósito, uso e exemplos.

Foram identificadas documentações na forma de páginas Web, documentos para *download*, artigos científicos e dentro do próprio vocabulário. Para determinados vocabulários foi encontrada pouca ou nenhuma informação. Vale comentar que, em alguns casos, os vocabulários são traduções em RDF de padrões em outros formatos, permitindo assim que o projetista obtenha maiores esclarecimentos sobre o propósito e uso do vocabulário uma vez consultada a documentação do padrão no formato original. Além disso, a possibilidade de consultar o vocabulário exportado para algum formato reconhecido (RDF, Turtle, N3, etc.) pode ser relevante para alguns projetistas no entendimento deste vocabulário, especialmente quando não há documentações externas disponíveis. Neste quesito, foi possível acessar a exportação de aproximadamente 80% dos vocabulários.

Durante a análise, foram identificados vocabulários específicos criados por diferentes organizações para descrever seus *datasets*. Porém, é difícil afirmar, sem um exame mais preciso de seus recursos, quando estes vocabulários estão duplicando termos e propriedades que poderiam ser aproveitados a partir de vocabulários estabelecidos. Por meio das colunas que indicam a presença do vocabulário no LOV e a quantidade de *datasets* acadêmicos e gerais que o utilizam pode-se obter uma indicação da reusabilidade do vocabulário. Ou seja, quanto maior o número de ocorrências, maior será a chance de ser reutilizado em outros *datasets* diferentes daquele para o qual foi criado especificamente.

Corroborando com os resultados de D'Aquin et al. (2013b), é possível observar por meio das tabelas que os vocabulários mais utilizados não são específicos para o contexto acadêmico, mas sim para descrever conceitos gerais e relacionamentos. Os vocabulários mais utilizados nos *datasets* analisados fornecem termos e propriedades para a descrição de recursos acadêmicos e institucionais (DCTERMS e DCE), pessoas

(FOAF), tópicos e conceitos (SKOS) e metadados (Vocabulary of Interlinked Datasets). Ainda de acordo com os resultados do trabalho dos autores mencionados anteriormente, também foram identificados vocabulários específicos para o contexto acadêmico como, por exemplo, para pesquisas (CiTO), pessoas e organizações (VIVO), cursos e formação (MLO, LSC e LOM) e registros bibliográficos (FaBio e BiBo). Vale mencionar que parte dos *datasets* analisados pertencem ao catálogo LinkeUp, que se trata da mesma fonte utilizada na pesquisa dos autores citados, o que pode ter contribuído para a similaridade destas observações mencionadas.

Em relação à quantidade de vocabulários utilizados por *dataset* no contexto acadêmico, é possível observar a partir das tabelas que apenas uma pequena parte dos vocabulários é altamente reutilizada, por volta de 7% (reuso maior ou igual a 10) do total. Por volta de 32% (reuso maior ou igual a 2 e menor que 10) dos vocabulários são moderadamente reutilizados, ou seja, um terço. E a maior parte, por volta de 61% dos vocabulários, é utilizada em apenas um dos *datasets*.

Como citado anteriormente, pode ser precipitado afirmar, sem um exame acurado, que vocabulários com poucas recorrências - como os 61% dos vocabulários utilizados em *datasets* no contexto acadêmico - foram criados para propósitos específicos e possuem baixa possibilidade de reuso. De forma a dispor de mais informações, a análise anterior foi repetida identificando-se a taxa de reuso dos vocabulários para os *datasets* presentes no LODStats.

Assim, a análise sobre o reuso de vocabulários utilizados nos *datasets* em geral (encontrados no LODStats) permite observar resultados similares ao reuso de vocabulários nos *datasets* de cunho acadêmico. De modo a proporcionar uma análise mais acurada, identificou-se a diferença na quantidade dos vocabulários mais reutilizados nos *datasets* no contexto acadêmico (DCT, com 30 ocorrências) e no LODStats (WGS84 Geo Positioning, com 411), chegando na proporção de 13,7 para 1. Este valor foi utilizado para definir os limites na identificação dos vocabulários altamente, moderadamente e pouco reutilizados. Assim, para os *datasets* no LODStats, observou-se que apenas uma pequena quantidade de vocabulários, aproximadamente 4% (reuso maior ou igual a 137), é altamente reutilizada. A quantidade de vocabulários moderadamente reutilizados também é baixa, por volta de 7% (reuso maior ou igual a 27 e menor que 137). E a maioria dos vocabulários são pouco reutilizados, por volta de 89% (reuso menor que 27).

Convém salientar que dos vocabulários identificados nesta pesquisa e buscados no LODStats aproximadamente 59% (60 vocabulários) não retornaram ocorrências, ou seja, não foram identificados casos de uso fora do contexto acadêmico. Este resultado pode ser inter-

pretado de diferentes maneiras. Por um lado, estes vocabulários estariam fortemente relacionados ao contexto acadêmico e há poucos *datasets* que descrevem dados deste tipo presentes no LODStats. E por outro lado, como parecem mostrar os dados anteriores sobre o baixo reuso de vocabulários, estes vocabulários não são realmente utilizados por motivos diversos, como desconhecimento, falta de informações sobre o vocabulário, alta especificidade dos termos e propriedades, entre outros.

Quanto às estratégias de reuso de vocabulários, foi possível identificar que há uma média de 8 vocabulários por *dataset*. Este valor parece estar de acordo com os resultados da avaliação citada anteriormente, realizada por Schaible, Gottron e Scherp (2014), que identificou que os projetistas de *datasets* têm preferência por manter uma mistura equilibrada entre *datasets* populares e específicos de domínio. Além do mais, estes autores comentam sobre uma das possíveis estratégias de reuso ser considerar a autoridade que publicou o vocabulário. Uma análise das autoridades relacionadas aos vocabulários permite identificar um total de 56 autores/publicadores (dois vocabulários foram desconsiderados neste valor por sua origem não ter sido identificada). Além disso, os vocabulários mais reutilizados estão relacionados com autoridades amplamente reconhecidas como o W3C (ex.: GEO, SKOS e vCard), DCMI (ex.: DC, DCT e DCMI-TYPE), Foaf Project (ex.: FOAF) e DERI (ex.: SIOC e VoID).

Em relação aos resultados encontrados, de um total de 102 (100%) vocabulários pesquisados, 58 (56,8%) foram encontrados no portal LOV. Ou seja, aproximadamente 1 em 2 vocabulários empregados nos *datasets* estavam disponíveis neste recurso centralizador, o que evidencia a abrangência deste portal e reforça a sua importância para a comunidade do Linked Data.

Por fim, outra análise importante a respeito do LOV é relacionar o grau de reuso dos vocabulários no contexto acadêmico a sua presença no LOV. Desta forma, foi possível identificar que todos os vocabulários (100%) altamente reutilizados podem ser encontrados no portal LOV. Para os vocabulários moderadamente reutilizados, a maioria deles podem ser encontrados no LOV (aproximadamente 82%). E para os vocabulários pouco reutilizados, uma parte significativa (aproximadamente 39%), pode ser encontrada no LOV. Estes resultados reforçam a ideia de que a popularidade dos vocabulários pode ser passível de medição. Neste trabalho, as medidas de popularidade utilizadas estão relacionadas à presença ou ausência dos vocabulários no portal LOV e a quantidade de *datasets* que os utilizam. Neste sentido, quanto mais popular o vocabulário, maior será sua presença em listas e portais e maior será seu uso nos *datasets*.

## 6 Considerações Finais

O direcionamento deste trabalho esteve em estimular o reuso de vocabulários na construção de *datasets* para a academia e ciência por meio da identificação de iniciativas e vocabulários empregados nestes contextos. Desta forma, investigou-se cinco iniciativas das quais foi possível identificar informações e recursos para o Linked Data no contexto acadêmico, bem como perceber a existência de relações implícitas entre essas iniciativas, seja pela participação dos mesmos profissionais em diferentes iniciativas, seja por fornecerem direcionamentos concordantes na forma de ferramentas e informações para os interessados no Linked Data. A partir dos portais Web das iniciativas, foram identificados um total de 28 vocabulários sugeridos a serem utilizados na descrição de dados acadêmicos e científicos. Além disso, a partir da análise de um total de 36 *datasets* para o contexto acadêmico, identificou-se 102 vocabulários empregados na descrição destes *datasets*, que foram adicionalmente explorados tendo em vista disponibilizar mais informações para sua compreensão e reuso.

Vale mencionar que uma parte significativa dos *datasets* identificados foi suprimida da análise devido à indisponibilidade de recursos que auxiliassem na identificação de seus vocabulários. Em relação a isto, a identificação dos vocabulários na parte restante dos *datasets* representou um desafio, uma vez que os recursos disponíveis para estes eram variados, entre arquivo de *dump*, arquivo de *metadados*, documentação Web e, em alguns casos, SPARQL endpoint. Este aspecto inviabilizou (devido ao custo e complexidade) a automação da descoberta destes vocabulários e exigiu trabalho manual dos pesquisadores para atender a cada caso específico. Esta heterogeneidade e indisponibilidade na forma de acessar os *datasets* pode ainda ser vista como um obstáculo na criação e manutenção de aplicações Linked Data no contexto acadêmico.

A construção de vocabulários abertos, devidamente estruturados e documentados, configura-se importante estratégia para fundamentar a compreensão e o reuso de dados e informações do tipo Linked Data. Conforme se procurou mostrar ao longo deste trabalho, alguns vocabulários analisados carecem de mais especificações. Sendo a Ciência da Informação um campo interdisciplinar, acreditamos que a sua contribuição se torna necessária em diversas vertentes. Pode-se exemplificar citando-se a experiência dos profissionais da informação na criação e disseminação de diretrizes sobre o uso de metadados, conciliadas a partir de práticas e estudos de usuário, o que pode vir a ser importante para a estruturação dos vocabulários. Deste modo, poder-se-ia eliminar barreiras como a falta de informações sobre o vocabulário, garantindo o reuso e, conseqüentemente, a sua popularidade. Essa popularidade poderia ser associada a linguagens que permitissem uma maior familia-

ridade dos usuários com os vocabulários, obedecendo à terminologia de domínios específicos de conhecimento.

No âmbito dos *datasets*, os vocabulários são utilizados para descrever termos e conceitos específicos em determinado domínio e seus diferentes tipos de relacionamento, preservando a semântica. As propriedades do padrão de metadados Dublin Core fornecem uma estrutura de descrição de conjuntos de dados no vocabulário VoID. Neste ponto, o olhar da CI e as lições aproveitadas das discussões da comunidade Dublin Core Metadata Initiative (DCMI) podem enriquecer a aplicação dos mecanismos descritivos nos vocabulários.

Se por um lado, o Linked Data suscita a contribuição dos profissionais da CI em torno dos seus princípios, por outro lado, impõe novos desafios, como por exemplo, sobre o desenvolvimento de modelos de conjuntos de dados bibliográficos. A ausência de um modelo universal para representar metadados bibliográficos com o Linked Data dificulta o reuso destes recursos em repositórios que agreguem conteúdo a partir de várias fontes distintas. Desta forma, o desafio recai sobre a construção de novos vocabulários ou no reuso dos termos de vocabulários de outros domínios, para se representar metadados bibliográficos. Uma adequada representação dos metadados, de acordo com o Linked Data, poderia ser a garantia de interoperabilidade entre várias plataformas.

Muitas questões são ressaltadas quando se trata da relação entre o Linked Data e a CI. Estas questões reforçam a importância do nosso trabalho e enaltecem a necessidade do comprometimento da área em prol das tecnologias emergentes para a descoberta e o reuso da informação.

Como trabalhos futuros, pretende-se aprofundar a análise dos vocabulários, buscando identificar e compreender seus termos e propriedades, podendo revelar duplicações existentes e eleger os vocabulários mais abrangentes a serem reutilizados para a descrição de recursos acadêmicos e científicos. Igualmente, pretende-se analisar vocabulários cujos termos possam ser reutilizados na descrição de recursos em Unidades de Informação e de aplicações da área CI, de maneira geral.

## Notes

(1) O acrônimo “OLED” não é oficial, mas foi utilizado por motivos de simplificação.

## Referências

Auer, S. et al. (2012). LODStats - An Extensible Framework for High-Performance Dataset Analytics. // Proceedings of the 18th international conference on Knowledge Engineering and Knowledge Management. Galway, Irlanda, Outubro 08-12, 2012. 353-362

- Berners-Lee, T. (2006). Linked Data - Design Issues, 2006. <https://www.w3.org/DesignIssues/LinkedData.html>. (2017-03-16).
- Bizer, C.; Heath, T.; Berners-Lee, T. (2009). Linked Data - The Story So Far // *International Journal On Semantic Web And Information Systems* 5:3 (2009) 1-22.
- Daga, E. et al. (2016). The Open University Linked Data – [data.open.ac.uk](http://data.open.ac.uk) // *Semantic Web* 7:2 (fev 2016) 183-191.
- Dietze, S. et al. (2013). Interlinking educational resources and the web of data: A survey of challenges and approaches // *Program: electronic library and information systems* 47:1 (2013) 60-91.
- D'Aquin, M. et al. (2013a). Assessing the educational linked data landscape // *Proceedings Of The 5th Annual Acm Web Science Conference On - Websci '13, Paris, França - Maio 02 - 04, 2013*. 43-46
- D'Aquin, M. et al. (2013b). LinkedUp veni competition: Linked and Open Data for Education // *Ceur Workshop Proceedings, Geneva, set. 2013*.
- D'Aquin, M.; Noy, N. F. (2012). Where to publish and find ontologies? A survey of ontology libraries // *Web Semantics: Science, Services and Agents on the World Wide Web* 11 (mar. 2012) 96-111.
- Heath, T.; Bizer C. (2011). *Linked Data: Evolving the Web into a Global Data Space* (1st edition). Morgan & Claypool Publishers, 2011. 1-136
- Keßler, C.; d'Aquin, M.; Dietze, S. (2013). Linked Data for science and education // *Semantic Web* 4:1 (2013) 1-2.
- LinkedUp (2017). About LinkedUp, 2017. <https://linkedup-project.eu/about/>. (2017-02-28).
- Linked Education (2017). Home Page: Welcome to Linked Education, 2017. <http://linkededucation.org/>. (2017-02-28).
- Linked Open Vocabularies (2017a). Linked Open Vocabularies (LOV): Vocabularies in LOV, 2017a. <http://lov.okfn.org/dataset/lov/>. (2017-02-28).
- Linked Open Vocabularies (2017b). Linked Open Vocabularies (LOV): VOCABS, 2017b. <http://lov.okfn.org/dataset/lov/vocabs/>. (2017-02-28).
- Linked Universities (2014). Home Page: Welcome to LinkedUniversities.org, 2014. <http://linkeduniversities.org/>. (2017-02-28).
- Open Linked Education Data Community Group (2015). Main Page: Introduction to the Open Linked Education Data Community Group, 2015. <https://www.w3.org/community/opened/wiki/>. (2017-02-28).
- Purohit, S. et al. (2016). Effective Tooling for Linked Data Publishing in Scientific Research // *IEEE Tenth International Conference On Semantic Computing (ICSC)* (2016) 24-31.
- Schaible, J.; Gottron, T.; Scherp, A. (2014). Survey on Common Strategies of Vocabulary Reuse in Linked Open Data Modeling // *European Semantic Web Conference (ESWC)* (2014) 457-472.
- RKB Explorer (2017). SPARQL Query Interface, 2017. <http://seek.rkbexplorer.com/sparql/>. (2017-02-28).
- Silva, A. M. (2010). Mediação e mediadores em Ciência da Informação // *Revista Prisma*. 9 (2010).
- Silva, A. M.; Ribeiro, F. (2011). *Paradigmas, serviços e mediações em Ciência da Informação*. Recife: Nectar, 2011.
- Taibi, D. et al. (2016). Educational Linked Data on the Web - Exploring and Analysing the Scope and Coverage // *Open Data For Education*. (2016) 16-37.

The Open University (2017). Home Page: Linked Data from The Open University, 2017. <http://data.open.ac.uk/>. (28-02-2017).

W3C (2016). W3C DATA ACTIVITY: Building the Web of Data, 2016. <https://www.w3.org/2013/data/>. (28-02-2017).

---

Copyright: © 2017. Rozsa (et al.) This is an open-access article distributed under the terms of the Creative Commons CC Attribution-ShareAlike (CC BY-SA), which permits use, distribution, and reproduction in any medium, under the identical terms, and provided the original author and source are credited.

---

Received: 2017-03-16. Accepted: 2017-05-25

## Apêndices

Quadro 1

<i>Vocabulários</i>	<i>Linked Universities</i>	<i>Linked Education</i>	<i>OLED</i>
<i>Citações</i>			
Citation Counting And Context Characterization Ontology	-	X	-
Citation Typing Ontology	-	X	-
<i>Comunidades e Publicações Acadêmicas</i>			
Publishing Roles Ontology	-	X	-
Publishing Status Ontology	-	X	-
Publishing Workflow Ontology	-	X	-
The Bibliographic Ontology	X	X	X
VIVO Ontology	X	-	X
<i>Cursos e Formação</i>			
Metadata for Learning Opportunities	X	-	X
ReSIST Courseware Ontology	-	X	X
RDF Binding IEEE LOM	-	X	X
Teaching Core Vocabulary	X	-	X
XCRI Course Advertising Profile	X	-	-
<i>Instituições</i>			
Academic Institution Internal Structure Ontology	X	X	-
Bowlogna Ontology	X	X	X
Core organization ontology	X	-	-
<i>Pesquisas</i>			
CERIF Ontology	X	-	-
Linked Science Core Vocabulary	X	-	-
<i>Recursos Acadêmicos e Institucionais</i>			
Asset Description Metadata Schema	-	X	-
Discourse Relationships Vocabulary	-	X	-
ISO Metadata for Learning Resources	-	-	X
Learning Resource Metadata Initiative	-	-	X
mEducator Educational Resource Schema	-	X	X
<i>Recursos de Multimídia</i>			
W3C Ontology for Media Resources	X	-	-
<i>Redes Sociais</i>			
Atom Activity Streams RDF Mapping	-	X	-
<i>Registros Bibliográficos</i>			
Bibliographic Reference Ontology	-	X	-
Document Components Ontology	-	X	-
FRBR-Aligned Bibliographic Ontology	-	X	-
<i>Tópicos Acadêmicos</i>			
Mathematics Subject Classification	X	-	-

Quadro 1. *Vocabulários sugeridos pelas iniciativas.*

Figura 3

```

<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF
  xmlns:adhoc="http://vocab.ox.ac.uk/ad-hoc-data-ox/"
  xmlns:aiiso="http://purl.org/vocab/aiiso/schema#"
  xmlns:cerif="http://spi-fm.uva.es/neologism/cerif#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:dcat="http://www.w3.org/ns/dcat#"
  xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:exif="http://www.w3.org/2003/12/exif/ns#"
  xmlns:fn="http://www.w3.org/2005/xpath-functions#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:geo="http://www.w3.org/2003/01/geo/wgs84_pos#"
  xmlns:gr="http://purl.org/goodrelations/v1#"
  xmlns:lyou="http://purl.org/linkingyou/"
  xmlns:mlo="http://purl.org/net/mlo/"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:oxcap="http://purl.ox.ac.uk/oxcap/ns/"
  xmlns:oxnotation="https://data.ox.ac.uk/id/notation/"

  <rdf:Description rdf:about="http://oxpoints.oucs.ox.ac.uk/id/23232325">
    <dc:title rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
      Bevington Road
    </dc:title>
    <rdf:type rdf:resource=
      "http://ns.ox.ac.uk/namespace/oxpoints/2009/02/owl#Carpark"/>
    <skos:prefLabel rdf:datatype=
      "http://www.w3.org/2001/XMLSchema#string">
      Bevington Road
    </skos:prefLabel>
  </rdf:Description>

```

(a) Exemplo de *dump* de um *dataset*.

**SPARQL Query Interface**

This interface permits queries to be made over the information held within the repository, using the **SPARQL Query Language**.

Result format:

Query:

```

PREFIX id: <http://seek.rkbexplorer.com/id/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>

PREFIX property: <http://seek.rkbexplorer.com/id/property/>
PREFIX toolType: <http://seek.rkbexplorer.com/id/tooltype/>
PREFIX tool: <http://seek.rkbexplorer.com/id/tool/>
PREFIX taskType: <http://seek.rkbexplorer.com/id/taskType/>
PREFIX task: <http://seek.rkbexplorer.com/id/task/>
PREFIX artifact: <http://seek.rkbexplorer.com/id/artifact/>
PREFIX artifactType: <http://seek.rkbexplorer.com/id/artifactType/>
PREFIX oc: <http://www.gsic.uva.es/ontologies/ontoolcoleModel.owl#>
PREFIX dbpedia: <http://dbpedia.org/ontology/>

SELECT * WHERE { ?s rdfs:label ?o } LIMIT 10

```

(b) Exemplo de um SPARQL EndPoint.

```

<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xml:base="http://www.purl.org/meducator/ns/"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:sioc="http://rdfs.org/sioc/ns#"
  xmlns:cc="http://creativecommons.org/ns#"
  xmlns:status="http://www.w3.org/2003/06/sw-vocab-status/ns#"
  xmlns:skos="http://www.w3.org/2004/02/skos/core#"
  xmlns:xs="http://www.w3.org/2001/XMLSchema"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <rdf:Description rdf:about="http://purl.org/meducator/ns/">
    <rdfs:label>mEducator Metadata Description
  Scheme</rdfs:label>
    <dcterms:title>Namespace for mEducator's Metadata
  Description Scheme, Version 1.1 </dcterms:title>
    <dc:contributor>Participants in mEducator's TRG Working
  Group </dc:contributor>
    <dc:description>This schema is used to described medical
  educational resources that will be shared across
  academic institutions in Europe</dc:description>

```

(c) Exemplo de um arquivo de metadados.

Figura 3. Exemplos de recursos relacionados aos datasets e que apresentam vocabulários. (Figura (b) captura de tela elaborada a partir de RKB Explorer (2017)).

Tabela 3

				Características Gerais			Características de Popularidade		
Nome		Autoridade		Doc	Dis	Lov	Aca	Ger	
Pessoas e Organizações									
Friend of a Friend		FOAF Project		Web	Sim	Sim	27	248	
Academic Institution	Internal Structure	Talis Information Ltd.		Web	Sim	Sim	10	1	
Core Organization Ontology		W3C		Web	Sim	Sim	6	30	
An Ontology for vCards		W3C		Web	Sim	Sim	5	3	
VIVO Core Ontology		VIVO Project		Web	Sim	Sim	3	1	
BIO: A vocabulary for biographical information		Ian Davis		Web	Sim	Sim	1	17	
SOTON Ontology		University of Southampton		Web	Sim	Não	1	15	
School-ont		education.data.gov.uk		Não	Não	Não	1	1	
NEPOMUK Contact Ontology		OSCAF Ontologies		Web	Sim	Sim	1	0	
ResumeRDF		rdfs.org		Web	Sim	Não	1	0	
Time indexed person role		Aldo Gangemi		Web	Sim	Não	1	0	
OxPoints ontology		University of Oxford Vocabs.		Web	Não	Não	1	0	
People Ontology		Open University		Não	Não	Não	1	0	
UNIVCAT Vocabulary		Open Data / Pompeu Fabra University		Não	Não	Não	1	0	
Cursos e Formação									
Metadata for Learning Opportunities		Scott Wilson		PDF	Não	Não	4	0	
ReSIST Courseware Ontology		Bene Rodriguez		Voc.	Sim	Sim	3	2	
Teaching Core Vocabulary Specification		Linked Science		Web	Sim	Sim	3	0	
XCRI Course Advertising Profile		Scott Wilson		Não	Não	Não	3	0	
Open Learn Ontology		Open University		Não	Não	Não	2	0	
South Africa Open University Ontology		Open University		Não	Não	Não	2	0	
Learning Object Metadata Ontology		Open Discovery Space project		Voc.	Sim	Sim	1	0	
Key Information Set Ontology		Open University		Não	Sim	Não	1	0	
Recursos Acadêmicos e Institucionais									
DCMI Metadata Terms		DCMI		Web	Sim	Sim	30	217	
Dublin Core Metadata Element Set		DCMI		Web	Sim	Sim	16	298	
Simple Knowledge Organization System		W3C		Web	Sim	Sim	11	112	
DBpedia Ontology		Dbpedia		Web	Sim	Sim	7	5	
DCMI Type Vocabulary		DCMI		Web	Sim	Sim	1	0	
GEM Element Set Qualifier Vocabulary		GEM Consortium		Voc.	Sim	Não	1	0	
Achievement Properties	Standards Network	Core University of Washington		Art.	Sim	Não	1	0	
mEducator Metadata Description Scheme		mEducator's TRG Working Group		Não	Sim	Não	1	0	
Ontoolcole		Guillermo Vega-gorgojo		Art.	Não	Não	1	0	

Tabela 3. Vocabulários para descrição de pessoas, organizações, cursos, formação e recursos acadêmicos e institucionais.



Tabela 4

<i>Características Gerais</i>			<i>Características de Popularidade</i>				
<i>Nome</i>	<i>Autoridade</i>	<i>Doc</i>	<i>Dis</i>	<i>Lov</i>	<i>Aca</i>	<i>Ger</i>	
<b>Projetos</b>							
Description of a Project vocabulary	Edd Dumbill	Web	Sim	Sim	2	8	
Academic Research Project Funding Ontology	University of Oxford Vocabs.	Web	Sim	Não	1	0	
<b>Pesquisas</b>							
AKT Reference Ontology	AKT Project	Não	Sim	Sim	2	27	
AKT Support Ontology	AKT Project	Não	Sim	Sim	2	23	
Semantic Web for Research Communities	York Sure	Art.	Sim	Sim	2	5	
AKT Extension Ontology	-	Não	Não	Não	1	3	
CERIF Ontology 1.3	EuroCRIS	Web	Sim	Sim	1	0	
Citation Typing Ontology	SPAR Ontologies	Web	Sim	Sim	1	1	
ArticleTypes Ontology	Macmillan Publishers Ltd.	Web	Sim	Não	1	0	
A vocabulary for describing academic researchers	University of Oxford Vocabs.	Voc.	Sim	Não	1	0	
<b>Coleções</b>							
The Bibliographic Ontology	Bruce D'Arcus	Web	Sim	Sim	9	10	
Expression of Core FRBR Concepts in RDF	Ian Davis	Web	Sim	Sim	1	3	
ISBD elements	Metadata Management Associates	Web	Sim	Sim	1	2	
Eprints Ontology	Eprints	Web	Sim	Sim	1	0	
Nature.com Core Ontology	Macmillan Publishers Ltd.	Web	Sim	Sim	1	0	
RDA Group 2 Elements (obsoleto)	DCMI/RDA Task Group	Web	Sim	Sim	1	0	
British Library Terms RDF schema	British Library, Metadata Services	Voc.	Sim	Sim	1	0	
Blogs Ontology (blogs)	Macmillan Publishers Ltd.	Web	Sim	Não	1	0	
Journals Ontology	Macmillan Publishers Ltd.	Web	Sim	Não	1	0	
PublishStates Ontology	Macmillan Publishers Ltd.	Web	Sim	Não	1	0	
Relations Ontology	Macmillan Publishers Ltd.	Web	Sim	Não	1	0	
ReviewStates Ontology	Macmillan Publishers Ltd.	Web	Sim	Não	1	0	
SeverityLevels Ontology	Macmillan Publishers Ltd.	Web	Sim	Não	1	0	
SummaryTypes Ontology	Macmillan Publishers Ltd.	Web	Sim	Não	1	0	
Eprints Terms	UKOLN, Jisc	Web	Não	Não	1	0	

Tabela 4. *Vocabulários descrição de para projetos, pesquisas e coleções.*

Tabela 5

<i>Características Gerais</i>		<i>Características de Popularidade</i>				
<i>Nome</i>	<i>Autoridade</i>	<i>Doc</i>	<i>Dis</i>	<i>Lov</i>	<i>Aca</i>	<i>Ger</i>
<b>Recursos de Multimídia</b>						
Ontology for Media Resources	W3C	Web	Sim	Sim	5	0
Media RDF Vocabulary	Digital Bazaar, Inc.	Web	Não	Sim	3	0
The Timeline Ontology	Yves Raimond	Web	Sim	Sim	1	4
Exif data description vocabulary	W3C	Web	Sim	Sim	1	1
Music Ontology	Yves Raimond	Web	Sim	Sim	1	0
Yovisto Ontology	Yovisto	Web	Sim	Não	1	0
MPEG-7 Ontology	Rhizomik	Não	Sim	Não	1	0
Audio RDF Vocabulary	Digital Bazaar, Inc.	Web	Não	Não	1	0
Video RDF Vocabulary	Digital Bazaar, Inc.	Web	Não	Não	1	0
Podcast Ontology	Open University	Não	Não	Não	1	0
<b>Páginas Web e Redes Sociais</b>						
Semantically-Interlinked Online Communities	DERI	Web	Sim	Sim	5	62
RDF Site Summary	RSS-DEV Working Group	Web	Sim	Sim	1	5
Linking-You Vocabulary	University of Southampton	Web	Sim	Sim	1	1
RSS 2.0 Enclosures	-	Não	Sim	Não	1	0
Audioboo Ontology	Open University	Não	Não	Não	1	0
Youtube Ontology	Open University	Não	Não	Não	1	0

Tabela 5. *Vocabulários para descrição de recursos de multimídia, páginas Web e Redes Sociais online.*

Tabela 6

<i>Características Gerais</i>		<i>Características de Popularidade</i>				
<i>Nome</i>	<i>Autoridade</i>	<i>Doc</i>	<i>Dis</i>	<i>Lov</i>	<i>Aca</i>	<i>Ger</i>
<b>Localização</b>						
WGS84 Geo Positioning	W3C	Web	Sim	Sim	10	411
GeoNames ontology	GeoNames	Web	Sim	Sim	4	6
Spatial Relations Ontology	Ordnance Survey	Web	Sim	Sim	1	11
<b>Tempo e Eventos</b>						
The Event Ontology	Yves Raimond	Web	Sim	Sim	4	2
Time Ontology	W3C	Web	Sim	Sim	2	18
Linking Open Descriptions of Events	Ryan Shaw	Web	Sim	Sim	2	0
RDF Calendar	W3C	Web	Sim	Sim	1	1
The Time Interval Pattern	Valentina Presutti	Voc.	Sim	Sim	1	0
Time Indexed Situation	Aldo Gangemi	Não	Sim	Sim	1	0
Extended Date Time Format Ontology	Open University	Não	Não	Não	1	0
<b>Dados Estatísticos</b>						
Data Cube	W3C	Web	Sim	Sim	2	77
The Statistical Core Vocabulary (obsoleto)	LiDRC	Web	Sim	Sim	2	1
SDMX-RDF vocabulary	SDMX	Web	Sim	Sim	1	68
LOIUS Ontology (baseado em voc. obsoleto)	Giovanni Pirrotta	Art.	Não	Não	1	0

Tabela 6. *Vocabulários para descrição de localização, tempo, eventos e dados estatísticos.*

Tabela 7

<i>Características Gerais</i>		<i>Características de Popularidade</i>				
<i>Nome</i>	<i>Autoridade</i>	<i>Doc</i>	<i>Dis</i>	<i>Lov</i>	<i>Aca</i>	<i>Ger</i>
<b>Metadados</b>						
Vocabulary of Interlinked Datasets	DERI	Web	Sim	Sim	16	56
Creative Commons Rights Expression Language	Creative Commons	Web	Sim	Sim	7	3
OU Ontology	Open University	Não	Não	Não	5	0
Nice Tag Ontology	Alexandre Monnin	Web	Sim	Sim	3	0
SemWeb Vocab Status ontology	W3C	Web	Sim	Sim	2	6
VANN: A vocabulary for annotating vocabulary descriptions	Ian Davis	Web	Sim	Sim	2	6
Data Catalog Vocabulary	W3C	Web	Sim	Sim	2	3
Tag ontology	Richard Newman	Voc.	Sim	Sim	2	0
DDI-RDF Discovery Vocabulary	DDI Alliance	Web	Sim	Sim	1	0
<b>Outros</b>						
GoodRelations	Martin Hepp	Web	Sim	Sim	2	3
XQuery, XPath, and XSLT Functions and Operators	W3C	Sim	Não	Não	2	0
BBC Ontology	BBC	Web	Sim	Sim	1	0
ISO 639 RDF Representation	Scott Martin	Web	Sim	Não	1	0
Subjects Ontology	Macmillan Publishers Ltd.	Web	Sim	Não	1	0
Ontology for modelling metering	University of Oxford Vocabs.	Voc.	Sim	Não	1	0
Univ-Bench	SWAT / Lehigh University	Não	Sim	Não	1	0

Tabela 7. *Vocabulários para descrição de metadados e temas gerais.*