

Uma nova abordagem à integração de plataformas colaborativas e pesquisa de informação com recurso à filosofia Linked-Data

Fábio Alves ^{1,2}, Carla Pereira ^{1,2}, Cristóvão Sousa ^{1,2}

1) CIICESI, ESTGF, Instituto Politécnico do Porto, Portugal

80500235@estgf.ipp.pt

2) INESC TEC, Porto, Portugal

csp@estgf.ipp.pt, cds@estgf.ipp.pt

Resumo

O principal objetivo a atingir com o trabalho descrito neste artigo passa pela aplicação da abordagem *Linked-Data* a plataformas colaborativas de gestão de informação estruturada. Assim, como especial atenção para a componente de organização e disponibilização de conteúdos, e considerando que a grande maioria das plataformas colaborativas de apoio à operação de redes apenas permitem a utilização/pesquisa da informação que se encontra internamente armazenada e classificada, pretende-se com esta nova abordagem *Linked-Data* desenvolver os mecanismos necessários para o fácil acesso e utilização de informação disponível em diversas plataformas que contenham tanto informação estruturada como não estruturada. Este artigo descreve a abordagem e arquitetura da solução proposta e apresenta um caso prático de aplicação deste trabalho de investigação numa plataforma colaborativa de dados estruturados.

Palavras chave: *Linked-Data*, Espaços colaborativos, Redes Colaborativas, Gestão de Informação

1. Introdução

Atualmente a falta de informação, não é um problema que se coloque. A problemática passa pelo seu armazenamento e recuperação de forma simples e clara, satisfazendo as necessidades dos utilizadores. No contexto das redes colaborativas, em especial nos espaços colaborativos, todos os anos é realizada uma conferência *SWCS Semantic Web Collaborative Spaces*, onde são apresentadas e discutidas abordagens que melhorem o trabalho colaborativo de redes virtuais ou redes colaborativas. Assim segundo [Champin & Lyon 2012] e [Anadiotis *et al.* 2012], atualmente as principais preocupações passam pela criação de consenso dentro de um espaço colaborativo e também pela criação e manutenção de informação. Assim, numa perspetiva de contribuir para a

investigação nesta área e em particular para o desenvolvimento de novos serviços que poderão ser utilizados, numa primeira fase de teste, pela plataforma H-Know¹, procura-se neste artigo apresentar os primeiros resultados do projeto de criação de uma nova

abordagem e serviço, baseado em mecanismos de *knowledge discovery* (em particular linked- data [Tim Berners-Lee 2006]) para recuperação de informação em contextos de colaboração intra ou inter-organizacionais. A plataforma H-Know, construída no âmbito de um projeto com a mesma designação encontra-se desenvolvida de acordo com o paradigma de espaços colaborativos [Anadiotis *et al.* 2012] que definem a componente colaborativa da mesma, melhorando assim, a performance na criação de relações entre utilizadores (ou criação de redes colaborativas) e a gestão de informação e conhecimento no contexto de atuação da rede colaborativa [Soares & Alves 2012]. Na componente orientada à organização e disponibilização de conteúdos, para a qual este trabalho pretende contribuir, segundo [Carneiro *et al.* 2010] a plataforma é constituída por três ontologias de domínio, duas de domínio social SIOC² e FOAF³ e uma de domínio da construção (neste caso particular no domínio da reabilitação de edifícios). Os mecanismos de pesquisa e recuperação de informação disponíveis na plataforma apenas permitem a utilização/pesquisa de informação armazenada e classificada na própria plataforma. Considerando esta situação uma limitação ao nível de pesquisa de informação relevante para a atuação da rede colaborativa, pretende-se com este trabalho dotar a plataforma de mecanismos que permitam a transferência e fácil acesso e utilização de informação disponível em diversas plataformas que contenham tanto dados estruturados como não estruturados. O recurso a técnicas de *knowledge discovery* (em particular a *Linked-Data*, abordagem que será apresentada em detalhe neste artigo) é aqui considerado uma boa solução para a uniformização da informação, tornando-a mais acessível e organizada. Por exemplo, para o caso concreto da plataforma H-Know, seria interessante que a plataforma que já possui funcionalidades de apoio à operação de redes colaborativas (partilha de informação, criação de ligações entre utilizadores, troca de mensagens e pesquisas por “interesses”) se encontrasse dotada de mecanismos de pesquisa eficientes e eficazes, que permitissem aos seus utilizadores encontrar diversos tipos de informação, tais como decretos de lei, regras ambientais, normas Europeias, entre outros que não se encontram nela armazenados, mas sim disponíveis noutras plataformas e páginas Web. Num cenário de teste da abordagem e serviço proposto neste artigo, a nova abordagem baseada em *Linked-Data* permitirá que a plataforma H-Know comunique com outras do mesmo domínio de atuação.

Considerando os conceitos fundamentais associados a este trabalho de investigação, é apresentada no capítulo 2 uma breve introdução às tecnologias, serviços e standards que suportam este trabalho. Os capítulos 3 e 4 descrevem o trabalho desenvolvido no âmbito deste artigo. Assim, o capítulo 3 dedica-se à apresentação conceptual da proposta de solução do problema, sendo complementado pelo capítulo 4 onde é apresentado um caso prático de implementação da abordagem.

¹ <http://www.h-know.eu/>

² <http://sioc-project.org/>

³ <http://www.foaf-project.org/>

No capítulo 5 são efetuadas as considerações finais, descrevendo as principais conclusões e ainda as expectativas de trabalho futuro.

2. fundamentais

O objetivo deste capítulo é apresentar de forma sucinta os conceitos base deste trabalho. Assim, torna-se relevante a definição dos seguintes conceitos/abordagens/tecnologias: *Linked-Data*, espaços colaborativos, formalismos de representação e serviços.

2.1. Linked-Data

O termo *Linked-Data* é identificado com sendo um conjunto de boas práticas usadas na publicação e conexão de dados na Internet [Tim Berners-Lee 2006]. A utilização de tecnologias e standards como RDF⁴, OWL⁵, é essencial neste tipo de implementações. Segundo Tim Berners-Lee a disponibilização de dados na Internet pode ser identificada e classificada da seguinte forma [Tim Berners-Lee 2006]:

Classificação	Descrição
*	Disponibilização indefinida de dados na Internet
**	Disponibilização de dados na Internet de forma estruturada (a estrutura não é comum nem baseada em standards)
***	Disponibilização de dados na Internet de forma estruturada recorrendo a formatos de dados não proprietários
****	Disponibilização de dados na Internet recorrendo a formatos proprietários ou não, contudo as entidades/itens encontram-se identificadas através de URI's ⁶ .
*****	Disponibilização de dados na Internet com recurso a formatos proprietários ou não, contudo as entidades/itens encontram-se identificadas através de URI's, bem como a ligação a diversas fontes de dados com o mesmo contexto.

Uma vez existindo um conjunto de boas práticas na comunicação e transferência de dados entre plataformas web, torna-se evidente que existe um crescimento exponencial a cada ano que passa de plataformas dotadas de capacidades de comunicação e partilha de informação. Assim, segundo [Binyam Chakilu Tilahun 2013] existem vantagens atualmente identificadas e estudadas, tais como:

- **Processamento/associação:** Existência e/ou criação de datasets públicos que partilham um modelo uniforme de dados baseado em RDF. A utilização destes modelos, permite uma identificação mais clara e expressiva, como por exemplo a definição de predicados (entenda-se por elemento que efetua o relacionamento entre conceitos). Estes modelos tornam a recuperação de dados de uma tarefa mais simples com resultados de melhor qualidade.

- **Coerência:** Um triplo RDF constituído por URI's de diferentes namespaces no *subject* e *object*, dá a entender que estabelece uma ligação entre a entidade que o identifica e o *subject* com a entidade que identifica o *object*.
- **Referenciação:** Utilização de URI's para a referenciação e identificação de dados.
- **Integrabilidade:** Desde que as fontes de dados partilhem modelos de dados RDF a integração semântica e sintática de datasets é facilitada por bibliotecas de integração como por exemplo: RDF2GO⁷. Uma vez identificados os esquemas de dados, existem algoritmos de *matching* de dados entre os diferentes formatos.
- **Oportunidade:** A publicação e atualização em *Linked-Data* é relativamente fácil. Uma vez os dados estando dispersos e identificados por URI's as publicações passam pela disponibilização do URI à plataforma em questão, contudo as atualizações devem ser cuidadas, mas uma vez alterada na fonte, é refletido em todos os locais onde esse URI estiver a ser utilizado.

No entanto em [Champin & Lyon 2012], os autores apontam algumas questões de usabilidade, imprecisão, escalabilidade e dinamismo que consideramos importantes. Na usabilidade os autores referenciam que existe a necessidade de ter conhecimento técnico da linguagem de *queries* estruturada SPARQL⁸. O utilizador para aceder aos dados é obrigado a ter conhecimento da linguagem, bem como da estrutura de dados definida. Mesmo tendo em conta a existência de algumas *frameworks* mais recentes, como RDF2GO que vêm de certa forma ajudar o utilizador nos acessos, porém não deve ser de todo ignorada esta questão. A imprecisão por sua vez, pode ter origem na grande diversidade de domínios que o *Linked-Data* pretende englobar, mas a experiência dos utilizadores, bem como a utilização de *datasets* específicos de domínio ajudam na estruturação da informação.

A escalabilidade e o dinamismo devem ser analisados como um todo, uma vez que, os conteúdos disponíveis na Internet são cada vez mais e diversificados. A constante mudança leva a que “o que hoje é verdade, amanhã já não o seja”, então os domínios começam a crescer e os pontos de ligação são cada vez mais, sendo assim necessário usar técnicas de *reasoning* de ontologias[Nebot & Berlanga 2012].

⁴ <http://www.w3.org/RDF/>

⁵ <http://www.w3.org/2004/OWL/>

⁶ www.w3.org/wiki/URI

⁷ <http://semanticweb.org/wiki/RDF2Go>

⁸ <http://www.w3.org/TR/rdf-sparql-query/>

Segundo[Bechhofer *et al.* 2004] a ideia do *Linked-Data* passa pela aplicação de uma arquitetura geral da *World Wide Web*, de modo a permitir a partilha da estrutura de dados a uma escala global. Com o objetivo de contribuir para este processo, dentro de alguns vocabulários (apresentados na secção seguinte), estão de igual modo a ser implementados *endpoints*, ou seja, serviços que permitem o acesso a dados através da linguagem de *queries* SPARQL. Segundo [Bechhofer *et al.* 2004], existem três tipos de recolha de dados encontrados até ao momento que correspondem à filosofia *Linked-Data*:

Tipo 1: Recorrendo única e simplesmente às capacidades do *SPARQL* e respetivos *endpoints* como capacidade de pesquisa. Este mecanismo é bastante interessante quando é usado um, ou um conjunto bem definido de SPARQL *endpoints* que permitam a catalogação de um conjunto de *queries* bem definidas. Assim, é possível garantir com mais eficiência a qualidade e consistência de dados, contudo a flexibilidade de adaptação a novos serviços requer um esforço superior e integração.

Tipo 2: Esta abordagem, consiste na criação de uma base de dados que efetue uma cópia integral de alguns *datasets* que se encontram definidos manualmente, de acordo com o domínio em questão, como sendo importantes, disponibilizando assim através de um único SPARQL *endpoint* toda a informação referente aos anteriormente selecionados. Com isto é possível reduzir o número de *queries* e ligações a efetuar, não sendo necessário um catálogo de *queries* por SPARQL *endpoint* como definido no tipo anterior. Contudo, o esforço de sincronização das cópias com os originais e a dificuldade na operacionalização são bastante superiores.

Tipo 3: Utilização de um sistema agregador de *queries* que permite a interligação de um diverso conjunto de *datasets*/SPARQL *endpoints* de modo a poder ser efetuada uma *query* ao mediador, sendo o pedido direcionado com total “transparência” para o emissor do pedido. Este sistema permite a existência de um mediador que “faculta” dados íntegros, no entanto o tempo de resposta é superior e existe uma necessidade constante de manter o mediador atualizado e funcional, uma vez que este mecanismo ao contrário do Tipo 2, não possui nenhum mecanismo de *cache* que permita um tempo de resposta mais curto e também uma resposta rápida da *cache* quando o acesso ao original se encontra congestionado.

2.2. Espaços Colaborativos

Os espaços colaborativos são estruturas de suporte às redes colaborativas compostas por um conjunto de funcionalidades que visam o agrupamento de um conjunto de pessoas/organizações que trabalham para alcançarem o mesmo objetivo [Carneiro *et al.* 2010]. Este conceito de espaços colaborativos, engloba um conjunto de conceitos, metodologias e tecnologias diversificadas que ajudam na obtenção dos resultados definidos. De acordo com [Soares & Alves 2012], os espaços colaborativos são criados segundo um ciclo composto por quatro estados (criação, desenvolvimento, finalização e transformação), que ajudam na reutilização de informação. Estes estados são definidos da seguinte forma: 1) O estado “criação”, tal como o próprio nome indica, é quando o espaço é criado e este não é alterado até à definição dos objetivos e formação da equipa de trabalho; 2) Uma vez os objetivos e a equipa inicial definida o estado para passa para “desenvolvimento”, neste estado é expectável que se desenvolvam todos os trabalhos, relatórios e informação necessária ao cumprimento dos objetivos; 3) Obtendo os resultados esperados, o espaço é “finalizado”, com isto as funcionalidades de edição e utilização são bloqueadas, passando a ser possível única e simplesmente a consulta de informação registada neste; 4) e o último estado “transformação”, é aquele que motiva a afirmação apresentada anteriormente - reutilização de informação, isto porque é através deste estado que se torna possível pegar em informação relevante e não confidencial reaproveitando-a para um novo espaço. De acordo com [Soares & Alves 2012], é com base neste estado que os resultados de um espaço colaborativo podem servir para a definição de objetivos de outro.

Nesta fase o leitor pergunta-se: “mas então o espaço colaborativo é construído segundo um ciclo? Mas eu posso usar os espaços colaborativos e não respeitar os ciclos?”, a resposta é “Sim”. É certo que os espaços colaborativos podem ser criados, desenvolvidos e finalizados. Contudo, os autores pretendem neste artigo dar especial ênfase a reutilização de informação que pode essencialmente ser obtida aquando da passagem ao estado “transformação”.

2.3. Formalismos de representação

Nas secções anteriores, foram introduzidos os nomes de alguns formalismos de representação de dados que são essenciais para a compreensão do problema, sendo estes agora descritos brevemente.

RDF

RDF é uma linguagem de representação de dados para a Internet em meta-dados estruturados. É considerada uma linguagem de representação de ontologias leve, que suporta a interoperabilidade dos dados entre aplicações. A terminologia RDF encontra-se associada aos *statements*, (*subject*, *predicate*, *object*), conhecidos como triplos RDF de representação de dados.

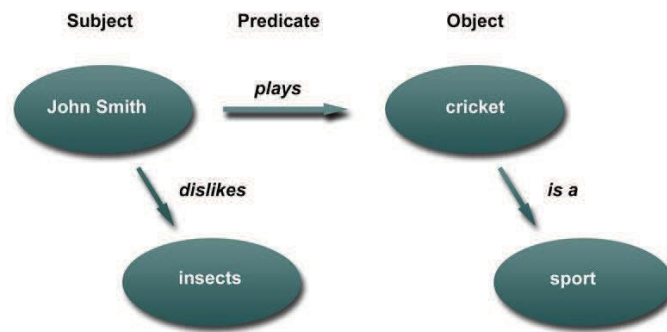


Figura 1: Triplo RDF [Bradley 2011]

Atualmente existem três notações distintas de representação do RDF: RDF Triples, RDF graphs e RDF/XML. No que diz respeito a vocabulários RDF temos [Binyam Chakilu Tilahun 2013]:

FOAF - *The friend of a friend* que é um vocabulário desenvolvido com o objetivo de descrever pessoas e atividades desenvolvidas por estas;

SIOC - *The Semantic Interlinked online Communities* é utilizado para a descrição de organizações e respetivas atividades;

SKOS - *Simple knowledge organization system* que é um vocabulário utilizado na descrição/representação de *thesaurus*, esquemas de classificação e taxonomias;

DCMI - *Metadata Terms (DC Terms)* usado com um propósito geral de relacionamento de meta-dados;

VOID - *Vocabulary of interlinked Datasets* que providencia meta-dados de um grande número de *Datasets*.

OWL

OWL é uma linguagem utilizada no âmbito da Web Semântica que permite a publicação e partilha de ontologias na Internet. Atualmente existem três formalismos distintos de OWL [Eiter *et al.* 2003]:

- OWL Lite: Suporta necessidades de classificação hierárquica e restrições simples. Segundo [Bechhofer *et al.* 2004] o OWL Lite é uma sub-linguagem do OWL DL suportando apenas um subconjunto da linguagem OWL. Neste contexto o subconjunto é definido com sendo um único vocabulário OWL.
- OWL DL (*description logic*): Suporta necessidades de classificação hierárquica e restrições simples (por exemplo, não permite que um recurso seja usado com classe e instância).
- OWL Full: Permite toda a expressividade necessária, contudo o processamento computacional não é garantido, bem como pode não ser possível a obtenção de dados em tempo real. É considerada por combinar as especificidades do OWL DL com a liberdade sintática do RDF.

2.4. Serviços

Na área na qual este trabalho de investigação se insere, a comunidade científica tem desenvolvido um conjunto de serviços que tiram partido das linguagens e vocabulários apresentados na secção anterior, que despertam o interesse não só da comunidade científica, mas também de toda a comunidade empresarial. Ajudando assim a cumprir o objetivo de criação de um esquema global de dados ligados por URI's. Deste modo são apresentados nesta secção algumas *frameworks* que impulsionaram, de certa forma, o trabalho desenvolvido e apresentado neste artigo.

Ldif *Linked data integration Framework*⁹: Esta é uma *framework* desenvolvida com os seguintes objetivos: 1) agregar diversas fontes de dados com diversos vocabulários, podendo estes conter informação proveniente de *datasets* semelhantes e informação replicada, 2) efetuar um mapeamento dos dados recolhidos no esquema de dados local, 3) recorrer a técnicas heurísticas de *matching* na deteção de entidades mapeando-as com *URI's*, 4) os dados são processados de modo a reduzir as ambiguidades, resolução de conflitos de acordo com as políticas definidas no esquema de dados local, e 5) a informação é integrada na base de dados local, ou seja, é integrada na sua totalidade na base de dados de conhecimento local. Neste sentido, a equipa de desenvolvimento apresentou na conferência de *Semantic Technology & Business 2012*, em São Francisco, testes de integração com a DBpedia e FreeBase [Schultz *et al.* 2012].

R2R framework¹⁰: Esta *framework* de certa forma complementa a *framework* apresentada anteriormente, isto porque a *R2R framework* [Bizer & Schultz 2010] implementa *R2R Mapping Language* desenhada para responder aos seguintes requisitos:

- “Perceber” diversos vocabulários: uma vez que as fontes de dados podem ser constituídas por um conjunto diversificado de vocabulários e esta linguagem deverá suportar o mapeamento de termos entre grande parte destes.
- Descoberta e Ligação de dados: Todos os termos são mapeados com um URI para que seja permitido um relacionamento entre esquemas de dados RDF ou OWL. Assim os clientes destes esquemas de dados podem aceder a estes *links*.
- Expressividade: Esta linguagem suporta/providencia flexibilidade estrutural de modo a que os valores obtidos possam sofrer transformações, quer de valor, quer de unidade de medida.
- Suportar diversos níveis de vocabulários *datasets*: Capacidade de distinguir que diferentes fontes de dados podem usar formatos diferentes para representar a mesma propriedade. Por exemplo, um mesmo atributo definido em locais diferentes pode ter significados diferentes como no caso da representação de distâncias em que unidade de medida pode variar caso estejamos a falar dos estados unidos onde o padrão é a milha e na Europa o quilómetro, ou então as ordem dos nomes de um

indivíduo pode ser *FirstName* , *FamilyName* ou *FamilyName, FirstName*.

Apache Solr¹¹ : É um servidor de pesquisas que recorre a REST-API que efetua a indexação da localização de ficheiros em XML, JSON, CSV ou de forma binária através de HTTP. [Smiley & Pugh 2009] enuncia as características deste servidor como sendo capaz de efetuar pesquisas avançadas em todo o texto, otimizar buscas em grandes volumes de dados, possuir interfaces standards de comunicação, autorreplicação e recuperação em caso de falhas, indexação em tempo real, configuração flexível através de XML.

3. Utilização de Linked-Data em plataformas colaborativas de dados estruturados

Sendo o principal objetivo a integração de plataformas colaborativas de dados estruturados, desenhadas para apoiar um domínio de atuação específico de uma rede colaborativa, com recurso ao conceito *Linked-Data*, a abordagem a seguir descrita define as seguintes condições como ponto de partida para aplicação do serviço desenhado:

1. existência de uma ontologia de domínio utilizada na classificação de conteúdos da plataforma;
2. na área social/colaborativa, a representação de atividades e perfis tem obrigatoriamente que ser com recurso aos vocabulários SIOC e FOAF.

A figura 1, representa a arquitetura da solução servindo assim como referencial desta secção.

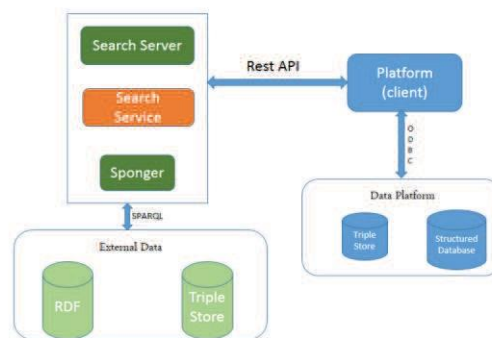


Figura 2:Arquitetura da Solução

⁹ <http://ldif.wbsg.de/>

¹⁰ <http://wifo5-03.informatik.uni-mannheim.de/bizer/r2r/>

¹¹ <http://lucene.apache.org/solr/>

Esta arquitetura encontra-se estruturada em 3 grandes grupos. O grupo 1 corresponde à plataforma de dados estruturados e toda a tecnologia de suporte a ela associada (representada na figura com a cor azul). O grupo 2 encontra-se representado igualmente na figura 1 com as tonalidades de verde. Este grupo representa um conjunto de tecnologias/serviços de suporte à abordagem *Linked-Data*, tais como bases de dados de gestão de conhecimento (*triple stores*), agregadores de conteúdos RDF (SPONGER¹²) e *search server* (apache solr) que efetuam a indexação de conteúdos. O grupo 3 que representa o trabalho desenvolvido no âmbito deste artigo, corresponde à cor laranja da figura e é o *core* deste serviço de pesquisas. Relembrando a secção 2, o serviço SPONGER disponibilizado pela Virtuoso, permite a agregação de diversos tipos de bases de dados, removendo as ambiguidades, tais como URI's diferentes a apontar para o mesmo objeto, tal como o *Search Server* que neste caso será o apache solr¹³ que é o mais popular dentro deste tipo de aplicações, recorre a heurísticas na indexação e recuperação de conteúdos que permitem a devolução da localização da informação em tempo real. Regressando ao foco deste trabalho, o serviço de pesquisas que é o responsável pela resposta a todos os mecanismos de pesquisa implementados na plataforma cliente. Este serviço deverá ser integrado inclusive com a *triple store* local. A comunicação de e para o serviço de pesquisas é efetuada através de uma REST API que disponibiliza à plataforma cliente a possibilidade de efetuar pesquisas locais e remotas em tempo real, recorrendo aos mecanismos de indexação e de pesquisas com métodos heurísticos do apache solr que redireciona os pedidos baseados na sua cache, bem como através do SPONGER que possui N bases de dados a ele associadas. Toda esta lógica de comunicação entre serviços disponibilizada pelos serviços de pesquisas é ainda complementada pelo processamento da informação efetuado pela LDIF *framework* apresentada na secção 2, que remove ambiguidades e duplicados dos dados recolhidos, bem como a interpretação dos diferentes vocabulários utilizados para descrever objetos semelhantes.

¹² <http://virtuoso.openlinksw.com/dataspace/doc/dav/wiki/Main/VirtSponger>

¹³ <http://lucene.apache.org/solr/>

4. Exemplo prático

Para validação do serviço desenvolvido será utilizada a plataforma H-Know. Em termos técnicos a plataforma H-Know encontra-se desenvolvida recorrendo ao CMS Drupal¹⁴ com suporte da base de dados MySql¹⁵. Contudo, foram desenvolvidos serviços de integração com base de dados de conhecimento, nomeadamente com Virtuoso¹⁶ recorrendo a linguagem de *queries* SPARQL. A figura 3 representa a arquitetura da plataforma H-Know.

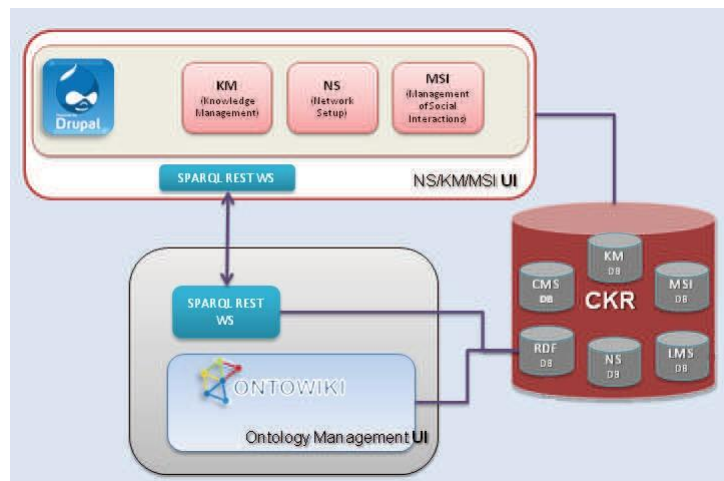


Figura 3: Arquitetura do H-Know

A figura 3 mostra-nos que a plataforma H-Know se encontra estruturada de acordo com três grandes pilares: *Knowledge Management*, *Network Setup* e *Management of Social Interactions*, suportados pela integração dos vocabulários SIOC, FOAF e uma ontologia de domínio da área da Construção civil (neste caso particular na sub-área de reabilitação de edifícios), conforme descrito em [Carneiro *et al.* 2010]. A figura 4 mostra como esta integração será realizada.

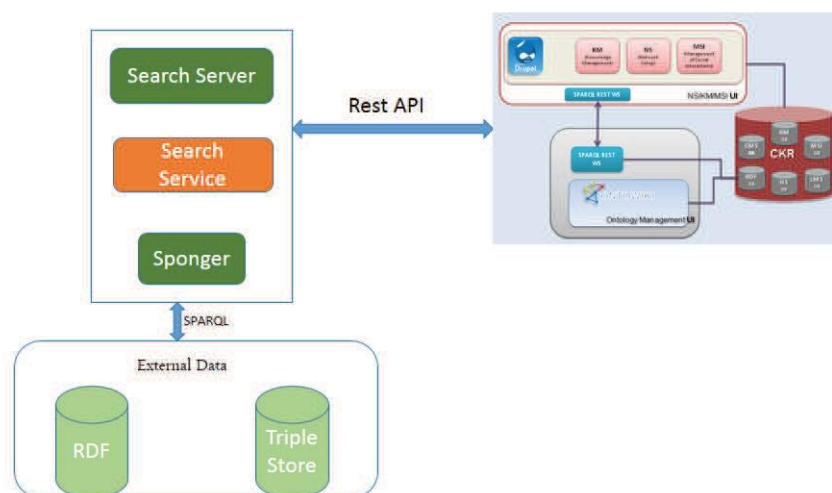


Figura 4: Integração da plataforma H-Know com o serviço desenvolvido

¹⁴ <https://drupal.org/>

¹⁵ <http://www.mysql.com/>

¹⁶ <http://virtuoso.openlinksw.com/>

No H-Know foram desenvolvidos clientes REST, que permitirão a comunicação com o serviço desenvolvido no âmbito deste artigo. Assim, será possível ao utilizador aceder, em tempo real, a informação relevante disponibilizada em plataformas externas. As pesquisas de informação no H-Know são efetuadas essencialmente em dois momentos: 1) quando o utilizador recorre a pesquisa; 2) no mecanismo de sugestões associado ao espaço colaborativo. O mecanismo de sugestões faz uma pesquisa nas fontes de dados locais por conteúdos que se encontrem classificados de forma similar, apresentando-os com conteúdos relacionados no portal. Assim, o espaço colaborativo fica mais enriquecido de conteúdo uma vez que não se restringe apenas a dados armazenados internamente pelos utilizadores.

Cenário Pré-Serviço

Sem considerar os serviços propostos como resultado deste trabalho, as atividades de pesquisa disponíveis na plataforma H-Know são as seguintes:

1. No que diz respeito à pesquisa de conteúdos: é efetuada uma análise à classificação do espaço colaborativo, sendo executado um conjunto de *queries à triple store local* que devolve a localização de conteúdos com classificação semelhante.
2. Relativamente à pesquisa de utilizadores (membros da plataforma), o procedimento é semelhante adicionando o facto de ser possível pesquisar também por perfis de utilizadores. Os pedidos efetuados ao SPARQL *endpoint* da *triple store* local, devolvem a localização dos conteúdos da base de dados local.

Neste cenário os resultados apresentados na tabela 1, correspondem à seguinte situação de pesquisa de conteúdos na plataforma H.Know: um utilizador faz uma pesquisa livre sobre: “eficiência energética” de conteúdos classificados na ontologia do domínio como “processo de construção” e perfis correspondentes a “engenheiro”:

Resultados	
Espaços Colaborativos	Eficiência energética na reabilitação
	Eficiência energética da argamassa X
	Nova Construção (eficiência energética)
Conteúdos	ColocaçãoCapoto.pdf (documento)
	Fachada.jpeg (imagem)
Perfis	João Graça
	António Moura

Tabela 1: Resultados pesquisa H-Know

Cenário Pós-Serviço

Na representação deste cenário com resultados apresentados na tabela 2, partimos da mesma situação de pesquisa, mas o serviço já se encontra ligado através do Sponger a um SPARQL endpoint (<http://semantic.eea.europa.eu/sparql>), disponível pela Agência Europeia para o Ambiente¹⁷.

Resultados	
Espaços	Eficiência energética na reabilitação
Colaborativos	Eficiência energética da argamassa X
	Nova Construção (eficiência energética)
Conteúdos	ColocaçãoCapoto.pdf
	Fachada.jpeg
Perfis	João Graça
	António Moura
Conteúdos (externos)	infrastructure-investments (http://semantic.eea.europa.eu/factsheet.action?uri=http%3A%2F%2Fwww.eea.europa.eu%2Fdata-and-maps%2Findicators%2Finfrastructure-investments)
	renewable-energy-consumption-outlook-from-eea (http://semantic.eea.europa.eu/factsheet.action?uri=http%3A%2F%2Fwww.eea.europa.eu%2Fdata-and-maps%2Findicators%2Frenewable-energy-consumption-outlook-from-eea)

Tabela 2: Resultados pesquisa H-Know mais serviço de pesquisa

¹⁷ <http://semantic.eea.europa.eu>

Os resultados apresentados de conteúdos externos como podemos ver são em inglês, visto que a plataforma H-Know possui a sua ontologia de domínio em 5 línguas diferentes o que permite o mapeamento entre os conceitos da classificação nesse idioma, bem como o mapeamento em diversas línguas da informação que é disponibilizada pela agência europeia para o ambiente.

5. Conclusões e Trabalho Futuro

No decorrer deste trabalho de investigação, na área da partilha de informação em ambientes colaborativos, verificamos que para além da necessidade de comunicação entre plataformas (facto já identificado pela aplicação prática da plataforma H-Know) é também fundamental que a informação partilhada seja útil, relevante e atual para o domínio de atuação da rede colaborativa. Assim, a utilização *de standards* não proprietários, bem como vocabulários standards e tecnologias *OpenSource* permite-nos garantir a interoperabilidade entre as diversas plataformas e também a organização e correta manutenção de conteúdos disponibilizados, procurando que os resultados das pesquisas apresentem qualidade elevada.

Como trabalho futuro pretende-se alargar este serviço para integração com plataformas de dados não estruturados e efetuar testes ao serviço em diferentes plataformas com tecnologias e domínios diferentes do aqui analisado. Estes testes inicialmente poderão começar pela alteração da plataforma H-Know, visto que a base de dados Virtuoso do H-Know é gerida por uma plataforma denominada ontowiki¹⁸ que permite a manutenção/alteração da ontologia do domínio. Assim sendo, a plataforma poderá ser utilizada em domínios distintos.

¹⁸ <http://ontowiki.eu/Welcome>

6. Referências

- Anadiotis, G., Kafentzis, K., Pavlopoulos, I., & Westerski, A. (2012). Building consensus via a semantic web collaborative space. *Proceedings of the 21st International Conference Companion on World Wide Web - WWW '12 Companion*, 1097.
- Bechhofer, S., van Harmelen, F., Hendler, J., Horrocks, I., L. McGuinness, D., F. Patel-Schneider, P., & Andrea Stein, L. (2004). <http://www.w3.org/TR/owl-ref/>.
- Binyam Chakilu Tilahum. (2013). *Linked Data based Health Information Representation , Visualization and Retrieval System on the Semantic Web*.
- Bizer, C., & Schultz, A. (2010). The R2R Framework : Publishing and Discovering Mappings on the Web.
- Bradley, A. (2011). <http://www.seoskeptic.com/seo-semantic-web/>.
- Carneiro, C., Soares, A. L., & Sousa, C. (2010). Integration of domain and social ontologies in a CMS based collaborative platform. *OTM*.
- Champin, P., & Lyon, U. De. (2012). User Assistance for Collaborative Knowledge Construction.
- Eiter, T., Ianni, G., Polleres, A., Schindlauer, R., Tompits, H., Rey, U., & Carlos, J. (2003). Reasoning with Rules and Ontologies.
- Nebot, V., & Berlanga, R. (2012). Building data warehouses with semantic web data. *Decision Support Systems*, 52(4), 853–868.
- Schultz, A., Matteini, A., Isele, R., Mendes, P. N., Bizer, C., & Becker, C. (2012). LDIF - A Framework for Large-Scale Linked Data Integration, 1–3.
- Smiley, D., & Pugh, E. (2009). *Solr 1.4 Enterprise Search Server*.
- Soares, A. L., & Alves, F. (2012). Collaborative Spaces as Mediators for Information Sharing in Collaborative Networks.
- Tim Berners-Lee. (2006). Linked Data- Design issues.