
Integração de plataformas institucionais: o caso EUDAT

Ricardo Carvalho Amorim

INESC-TEC/ Faculdade de Engenharia da Universidade do Porto

ricardo.amorim3@gmail.com

Fábio Silva

Faculdade de Engenharia da Universidade do Porto

ffjs1993@gmail.com

Resumo

A qualidade dos dados de investigação é um debate recorrente entre instituições e investigadores. A constante preocupação em garantir que os dados sobrevivem aos seus criadores tem contribuído para o aparecimento de plataformas que, em casos mais avançados, permitem que os investigadores sejam envolvidos diretamente no processo de gestão dos seus dados. A plataforma EUDAT surge como um exemplo de uma uma solução para os inúmeros desafios da investigação, proporcionando ambientes colaborativos, ferramentas de processamento e outras mais valias para instituições que têm dificuldades em adquirir estas ferramentas de forma individual. Contudo, a divulgação de dados de investigação deve ter como princípios não só a sua preservação a longo prazo, mas também a sua potencial reutilização dentro da comunidade científica, pelo que é necessário transpor o conhecimento gerado durante a investigação para os dados associados. O Dendro surge como uma plataforma para apoiar a descrição de dados desde o momento em que são criados, permitindo que os registos produzidos os acompanhem e permitam mais tarde uma interpretação extensiva dos dados. Este caso de estudo apresenta a integração entre o Dendro e o EUDAT com o objetivo de promover um ambiente de gestão de dados abrangente que envolva os diversos responsáveis no processo.

Palavras-chave: Dendro, EUDAT, Gestão de dados de investigação, Descrição de dados, Preservação digital

Integration of institutional platforms: the EUDAT case study

Abstract

Research data quality is a common discussion among both institutions and researchers. To ensure that data outlive their producers, several platforms have recently emerged that, in some scenarios, allow researchers themselves to contribute to their data management. The EUDAT platform is an example of a solution for several challenges in the research environment, providing researchers with collaborative environments, large-scale data processing tools, and other features to promote both their data quality and long-term preservation. Data publication has also to consider aspects beyond long-term preservation, such as ensuring their potential reuse within the research community and thus, it is necessary to create adequate tools to capture knowledge as soon as it is produced. Dendro is a platform for data management from early stages, with a focus on data description, from the moment researchers start their projects. This case study reports on the integration between EUDAT and Dendro, to create an environment for data management involving all the stakeholders in the process.

Keywords: Dendro, EUDAT, Research Data Management, Data description, Data preservation

Introdução

A gestão de dados de investigação apresenta vários desafios para todos os envolvidos no processo. Apesar de grandes instituições já disporem de ferramentas e plataformas para a gestão dos seus dados nas várias fases, o mesmo não acontece com pequenos grupos de trabalho que por vezes não têm suporte para tal e são encarregues de gerir os dados que produzem no decorrer de vários projetos (Comp, 2016).

A preocupação com a preservação e divulgação de dados de investigação de grupos de pequena ou média dimensão tem vindo a promover iniciativas no sentido de criar plataformas para esse efeito, de entre as quais se destaca o EUDAT. Como repositório de dados, o EUDAT procura oferecer um conjunto de serviços para investigadores de vários domínios cujas instituições possam, por exemplo, ter dificuldades em criar e manter infraestruturas para o efeito (Burns, 2013).

Com o objetivo de publicar dados de investigação e, ao mesmo tempo, criar condições para que estes possam ser reutilizados, a descrição de dados apresenta alguns desafios, ao considerar os inúmeros domínios da investigação. A produção de descrições adequadas para os dados de investigação é por vezes uma tarefa que os investigadores já

incluem no seu dia a dia contudo, a falta de ferramentas adequadas faz com que estes recorram frequentemente a suportes que se deterioram rapidamente, e não dão suporte a processos estruturados de descrição. A plataforma Dendro foi desenhada com o princípio de envolver investigadores diretamente na gestão dos seus dados, desde o momento em que começam a produzir dados. A sua arquitetura permite também que equipamentos ou aplicações externas se possam ligar para cobrir ambientes específicos.

Este caso de estudo apresenta a ligação entre o Dendro e o EUDAT, como forma de estabelecer um ambiente abrangente para a gestão de dados. Este tipo de iniciativas permite alargar o leque de possibilidades para instituições que procurem dar resposta às crescentes necessidades de gestão de dados dos seus projetos e de torna-los inteligíveis para a comunidade científica.

Gestão de dados na cauda longa da investigação

A cauda longa da investigação compreende instituições que lidam frequentemente com vários domínios da investigação, em grupos de relativamente pequena dimensão. O suporte tradicionalmente existente para a gestão de dados em grandes instituições, não é aplicável neste cenário onde cada grupo tem requisitos diferentes. Os requisitos existentes para uma adequada gestão de dados podem envolver, entre outros aspetos, a criação de estruturas capazes de lidar vários formatos de representação de informação, com diferenças organizacionais entre domínios, com diferentes metodologias de produção de dados e, ao mesmo tempo, serem capazes de envolver os produtores de dados neste processo (Massimiliano, 2016). Por apresentar inúmeros desafios, a gestão de dados de investigação neste domínio tem vindo a receber bastante investimento com vista a promover a preservação digital e a reutilização de dados que, de outra forma, perderiam o seu valor. Existem, contudo, ainda alguns aspetos que têm vindo a merecer um maior destaque por influenciarem diretamente a forma como estes objetivos são alcançados. A gestão de dados é, por vezes, uma tarefa deixada para as fases finais de cada projeto onde um especialista-frequentemente o responsável pela gestão do projeto-tem a seu cargo a tarefa de organizar, descrever e depositar os recursos gerados numa plataforma destinada para o efeito. Ao concentrar este conjunto de responsabilidades e tarefas num único ponto ou numa única pessoa, o tratamento e depósito dos dados podem não chegar a acontecer, perdendo-se assim o seu valor e mantendo-se apenas a publicação deles resultante (McLure, 2014).

A gestão de publicações é uma área já muito desenvolvida e conta já com inúmeras plataformas que a suportam e garantem a sua continuidade. Para além disso, está já bem documentada, fazendo parte do conhecimento dos investigadores tarefas como a indexação de publicações, a sua divulgação, a sua citação e conseqüente crédito aos autores. O mesmo ponto de vista pode ser aplicado aos dados: um conjunto de dados que possa ser consultado pela comunidade científica pode também ser citado e igualmente dar crédito aos seus autores e instituições envolvidas na sua produção. O desafio é, contudo, distinto uma vez

que a interpretação de dados requer um conhecimento mais profundo sobre a sua origem e contexto, que desta forma não chegam a ser associados ao seu depósito em repositórios. Por outro lado, instituições que se candidatam para financiamentos comunitários têm vindo a deparar-se com recomendações para a criação de um plano para a gestão dos dados resultantes do projeto e, em alguns casos, a sua existência é já um fator de exclusão. Estes planos incluem muitas vezes a especificação de metodologias para o tratamento e descrição de dados, assim como a identificação de um repositório no qual serão depositados estes recursos.

Pela sua natureza, os dados de investigação podem beneficiar de uma descrição com grande pormenor, quando comparados com publicações. Por permitir incluir descrições características do domínio em questão, a descrição de dados de investigação tem sido considerada uma etapa importante no seu processo de gestão. Para um projeto no domínio da biodiversidade, por exemplo, esta descrição pode indicar referências geográficas para os pontos onde foram recolhidas amostras, as espécies identificadas ou mapas com os percursos realizados em cada saída de campo.

Várias plataformas têm surgido para permitir que os depósitos de dados possam incluir igualmente descrições em vários níveis, para facilitar a interpretação por parte da comunidade. Dado o requisito de receber dados de vários domínios, estas plataformas tendem a oferecer funcionalidades que são de interesse para uma grande quantidade de equipas de investigação, delegando funcionalidades específicas para ferramentas individuais que se podem integrar com o repositório em qualquer altura.

A plataforma EUDAT

Criada no âmbito da promoção da gestão de dados de investigação para instituições europeias, o EUDAT apresenta vários módulos que incluem ferramentas colaborativas, gestão de equipas, processamento em larga escala e plataformas de divulgação de dados.

Posicionamento da plataforma

As plataformas para gestão de dados podem ser agrupadas consoante o seu princípio de funcionamento: plataformas que estão sobre o controlo de instituições, sendo instaláveis em servidores institucionais; e plataformas que fornecem um serviço através do qual as instituições podem contratar diferentes tipos de acesso. Ambas apresentam as suas vantagens quer nos custos de implementação, quer no controlo sobre os dados (Amorim, 2015). Por incluir um conjunto de serviços adicionais juntamente com a componente de repositório, a plataforma EUDAT insere-se numa categoria distinta da dos repositórios para dados de investigação. O EUDAT permite que os dados comecem a ser criados diretamente na plataforma, processados e posteriormente divulgados, juntamente com informações que os autores pretendam anexar.

Módulos de gestão e produção de dados

Dentro da plataforma, os diferentes módulos estão interligados e permitem troca de recursos entre si. Esta abordagem surge também como uma solução para acompanhar os processos desde o arranque de cada projeto, permitindo que os investigadores interajam com a plataforma assim que começam a produzir os seus dados.

Para fases iniciais, onde os dados são frequentemente atualizados ou alterados, o módulo B2Drop permite que os investigadores acedam a uma área colaborativa para gerir diariamente os seus resultados. Este módulo mantém ao mesmo tempo versões dos dados, permitindo assim acompanhar as alterações feitas ao longo do tempo.

No final de cada projeto, e caso não existam períodos de embargo a respeitar, os dados envolvidos podem ser divulgados para a comunidade. Nesta fase, o B2Share assume o papel de receber estes recursos—que podem ter origem em outras plataformas para além do B2Drop—e proporcionar meios para os tornar inteligíveis e acessíveis ao exterior. Neste processo, o EUDAT dispõe de interfaces para apoiar o investigador no depósito, assegurando que aspetos como períodos de embargo, identificadores únicos e políticas de acesso são especificados antes de tornar o conjunto de dados público. Nesta fase, o investigador pode também adicionar uma descrição aos seus dados, para facilitar a sua interpretação, contudo, esta descrição tem um carácter facultativo e tende a assumir uma representação não normalizada.

Uma vez que todos os módulos podem comunicar entre si para abranger os vários requisitos envolvidos, o B2Share é também responsável por garantir que os recursos depositados são preservados indefinidamente. Nesta fase o módulo B2Safe encarrega-se de implementar políticas de curadoria e preservação digital incluindo a gestão de identificadores, de cópias de segurança e a verificação frequente da integridade dos dados. Tal como acontece com os outros módulos, o B2Safe é capaz de receber recursos de outras plataformas, reconhecendo automaticamente pacotes de repositórios conhecidos como o Dspace e o FEDORA. Esta abordagem permite responder a plataformas que não tenham capacidade para proporcionar um ambiente de preservação com grandes capacidades de armazenamento.

No lado do investigador externo que procura um determinado conjunto de dados, o B2Find assegura que os conjuntos geridos pela plataforma se encontram devidamente registados e indexados. A sua interface para o exterior foi pensada para permitir pesquisas de dados independentes do seu domínio, origem ou comunidade associada. Este módulo assume particular importância por ser responsável por levar os recursos armazenados e tratados até aos potenciais re-utilizadores, fazendo também a ligação com publicações que possam estar associadas a cada conjunto de dados.

Os módulos propostos na plataforma EUDAT podem assumir um papel importante na gestão de dados em equipas de investigação que não disponham atualmente de meios para o fazer. As fases iniciais de cada projeto estão também abrangidas, permitindo o depósito em áreas privadas e a gestão colaborativa de recursos que, de outra forma, ficariam dependentes dos modelos organizacionais da equipa de investigação. Esta abordagem está em linha com as recentes preocupações no sentido de envolver diretamente os investigadores na gestão dos seus dados até à sua divulgação. Esta é considerada uma prioridade, dado o conhecimento sobre os dados estar naturalmente disponível nesta fase e poder ser capturado sem requerer um grande esforço, existindo ferramentas adequadas para tal.

Metadados uma melhoria na qualidade dos dados

A descrição de dados é uma tarefa que, em alguns casos, já faz parte do quotidiano dos investigadores. O contexto em que os dados são produzidos é tradicionalmente registado em suportes como cadernos de laboratório ou blocos de notas pessoais e inclui geralmente metadados que o investigador considera relevantes para a sua compreensão. Apesar de bastante utilizados, estes suportes podem, a longo prazo, comprometer os registos recolhidos uma vez que tendem a deteriorar-se e a sua preservação está dependente do proprietário. A descrição de dados assume assim um papel importante para a sua compreensão e, como produtores de dados, os investigadores são os melhores candidatos para executar esta tarefa.

A inclusão de metadados no depósito em repositórios pode contribuir para melhorar a visibilidade de um determinado conjunto de dados. Ao permitir que as plataformas existentes indexem o seu conteúdo, por exemplo, as pesquisas podem tornar-se mais eficientes e incluir modelos avançados de pesquisa para melhorar a sua precisão. Um investigador no domínio da Química pode, por exemplo, pesquisar por conjuntos de dados que envolvam um determinado reagente, se este estiver devidamente identificado na folha de metadados. Por outro lado, um outro investigador no domínio da biodiversidade pode restringir os resultados da sua pesquisa a uma determinada cobertura geográfica sem que, para isso seja necessário abrir cada conjunto de dados apresentado. Os esforços no sentido de implementar mecanismos de recolha e produção de metadados pelos investigadores têm feito surgir ferramentas e aplicações desenhadas para serem integradas no seu ambiente de trabalho ou substituir suportes que não permitam a flexibilidade desejada para este efeito (Amorim, 2014). Aplicações neste domínio começam também a valorizar a recolha estruturada de metadados, fazendo com que estes possam ser interpretados, mesmo depois de uma eventual descontinuação da aplicação. Apesar desta vantagem, a descrição estruturada de dados é ainda um tema em desenvolvimento, não fazendo parte de muitos dos repositórios existentes.

Descrição de dados na plataforma Dendro

A plataforma Dendro (Silva, 2014) foi pensada para permitir que a gestão de dados, colaborativa ou não, possa começar com o arranque de cada projeto. Para o investigador, a interface assemelha-se à que já é apresentada em plataformas de armazenamento remoto como Dropbox ou Google Drive. Por outro lado, apresenta também uma interface de preenchimento de campos cujo objetivo é guiar o investigador na descrição dos seus dados, à semelhança do que já é feito em cadernos de laboratório ou blocos de notas. Cada elemento que se adiciona à folha de descrição-descriptor-define um determinado contexto e está associado a um determinado ficheiro ou pasta na plataforma. Apesar de a seleção de descritores depender do investigador, a plataforma é capaz de gerar um perfil de descrição para cada investigador ou equipa de investigação com base em dados recolhidos ao longo das diferentes interações. O resultado consiste na adaptabilidade que a plataforma é capaz de alcançar, independentemente dos domínios que estão envolvidos na sua utilização. O Dendro é capaz de, por exemplo, destacar descritores do domínio da psicologia para grupos de investigação dessa área, sem descurar no entanto, a possibilidade de cada investigador poder individualmente adaptar o perfil às suas necessidades.

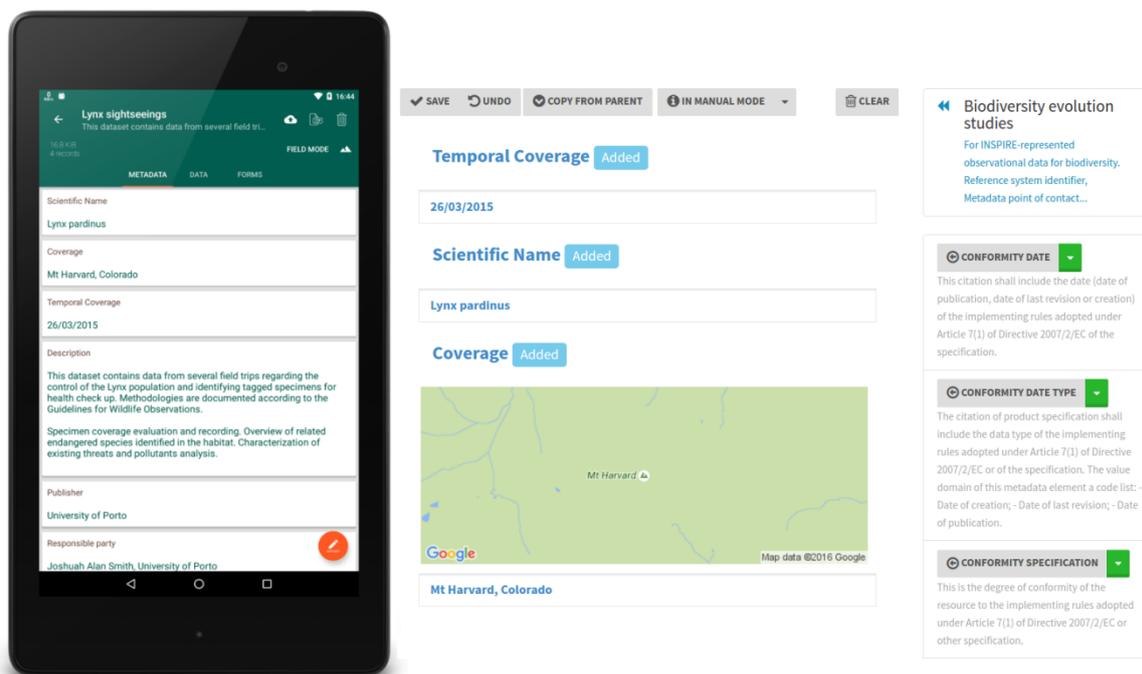


FIGURA 1 VISTA DE DESCRIÇÃO DE UM PROJETO, E INTERFACE EQUIVALENTE NA APLICAÇÃO LABTABLET

Ao construir um perfil adequado para cada domínio, a tarefa de descrição pode tornar-se menos um requisito, e mais uma tarefa guiada, personalizada e por vezes automatizada. Dentro de cada pasta, os investigadores podem copiar a folha de descrição completa e replicá-la para outro local do projeto, alterando apenas as descrições que sejam necessárias. A gestão colaborativa permite também que, se um investigador determinar que para os seus projetos determinados descritores devem ser destacados, estes possam ser

sugeridos aos seus colaboradores. As descrições feitas seguem também um determinado modelo de metadados, promovendo assim a sua interpretação por outros sistemas. Especialistas de informação contribuem aqui para a seleção de modelos de metadados pertencentes a normas existentes no domínio. No domínio da biologia marinha, por exemplo, o perfil da norma INSPIRE já especifica detalhadamente vários níveis de descrição para ajudar os investigadores a descreverem os seus dados. Por outro lado, no mesmo domínio, mas com maior abrangência, o perfil Darwin Core inclui alguns descritores que cobrem conceitos como georeferenciação, identificação de espécies e outros elementos relacionados. Na sua utilização, os investigadores podem selecionar livremente descritores de vários perfis, se considerarem que a sua inclusão é relevante para apoiar a interpretação dos seus dados.

O Dendro pertence a uma categoria de repositórios com um maior foco em capturar dados e os respetivos metadados assim que estes são produzidos. Consiste assim numa plataforma intermédia, que se integra no ambiente de trabalho dos investigadores, podendo ser complementado por ferramentas que produzam dados para estes serem automaticamente depositados. Nesta categoria podem ser considerados, por exemplo, cadernos de laboratório eletrónicos como o LabTablet e outros equipamentos laboratoriais. Dada a sua natureza, o Dendro pode em qualquer altura ser ligado a outras plataformas para exportar os recursos no final de cada projeto. A Figura 1 apresenta o mesmo projeto em duas interfaces distintas: na plataforma Dendro (lado direito) e na aplicação LabTablet (lado esquerdo). Ambas as plataformas têm como objetivo auxiliar na descrição de dados, independentemente do domínio em questão e neste caso concreto o seu funcionamento em conjunto pretende dar resposta a ambientes que exijam, por exemplo, uma maior mobilidade. Os investigadores podem tirar partido das funcionalidades que um dispositivo móvel tem, em particular dos sensores que caracterizam estes dispositivos – georeferenciação, temperatura, entre outros –, e sincronizar os conteúdos com o Dendro em qualquer altura. Depois de encerrarem cada projeto, os investigadores podem passar para a fase em que depositam os seus dados numa plataforma escolhida para o efeito. Idealmente, nesta fase, o investigador ou a sua equipa devem apenas escolher o repositório destino para os seus dados e proceder com a exportação.

Interligação Dendro–EUDAT

A plataforma Dendro foi desenhada com base no princípio de ser uma plataforma para gestão intermédia de dados de investigação. A partir deste ponto, o objetivo é fazer com que dados e metadados possam ser exportados em conjunto para um repositório que tenha em consideração os aspetos relacionados com a preservação digital e a disseminação de dados. Para este efeito, e dadas as capacidades já implementadas da plataforma EUDAT, foi estabelecida a ligação das duas plataformas para que, no final de cada projeto, seja

possível exportar os pacotes de dados e metadados produzidos pelos projetos de investigação e assegurar a sua disseminação e preservação.

The screenshot shows a web interface for a data repository. On the left, there's a 'Folder' section with icons for upload, download, refresh, and share. Below it is a 'Selection' section with 'COPY FROM PARENT' and 'IN MANUAL MODE' buttons. A list of files is shown: 'Up to carrental', 'definitions.pdf', and 'instances.xlsx'. The main content area displays a record for 'Vehicle reservation assignment in car rental' by Sebastião Oliveira from the University of Porto. It includes an abstract, a list of collections (Generic), and a table of files:

Name	Date	Size	Download
files.zip	10 Jul 2016	2.1 MB	Download
metadata.rdf	10 Jul 2016	1.9 kB	Download
metadata.txt	10 Jul 2016	1.7 kB	Download

Below the table, there are 'Export' options (BibTeX, MARC, MARCXML, DC, EndNote, NLM, BibWorks) and 'Metadata' information including a PID, Publication (University of Porto), and Licence (Attribution-ShareAlike 4.0 International).

FIGURA 2 EXEMPLO DE EXPORTAÇÃO DE UM CONJUNTO DE DADOS PARA O REPOSITÓRIO EUDAT

A implementação recorreu à arquitetura com base em extensões, adicionando a opção para o repositório EUDAT no formulário de exportação de datasets. Neste ponto, os investigadores podem escolher vários repositórios—Dspace, CKAN e Figshare são alternativas também disponíveis—sendo reencaminhados para uma interface específica de cada um para acompanhar a submissão. Nesta fase são pedidas credenciais de acesso para depósito, sendo que estas podem ser automaticamente configuradas se os depósitos forem feitos pela instituição e não por investigadores de forma individual. A exportação é feita através do módulo B2Share, que mais tarde se encarrega da sua disseminação para os restantes módulos. Após o depósito nesta plataforma, é enviada uma notificação com o identificador resultante, e o endereço onde se pode aceder ao depósito feito. A partir deste momento, os dados podem ser consultados pela comunidade, e citados de igual forma. A Figura 2 apresenta o resultado de um depósito na plataforma. Para a comunidade científica passa a ser possível aceder aos dados depositados, assim como à sua descrição produzida pelos próprios investigadores do seu domínio. À semelhança do que é feito com publicações, estes conjuntos de dados também podem ser citados, dispondo de um identificador único criado para o efeito.

Divulgação antecipada de dados de investigação

A publicação de dados é ainda um tema em debate quer para instituições, quer para investigadores. Em alguns casos, a divulgação de dados não pode ser feita sem que terminem restrições de acesso como períodos de embargo. Para estes casos, algumas plataformas já permitem manter os dados privados até que esse período termine, respeitando assim quaisquer imposições legais que possam existir e, ao mesmo tempo, garantindo que os dados já se encontram depositados e prontos a serem divulgados. Por outro lado, é também frequente encontrar projetos em que a divulgação dos dados pode

ocorrer desde as fases iniciais para promover aspetos como a visibilidade do projeto ou da instituição para o exterior. A integração entre o Dendro e o EUDAT teve em consideração estas restrições, e cobre assim três dos possíveis cenários quando se aborda a publicação de dados de investigação:

- **Publicação no final do projeto:** o investigador responsável ou a própria instituição encarregam-se de proceder à exportação dos dados do Dendro para o EUDAT, assim que o projeto termina;
- **Publicação imediata:** permite criar projetos na plataforma Dendro e assegurar que estes estão acessíveis desde o início. Investigadores externos que queiram consultar os dados podem fazê-lo diretamente na plataforma;
- **Publicação de metadados:** os metadados sobre o projeto são disseminados para a comunidade científica. Investigadores interessados em aceder aos dados podem solicitar autorização para o fazer ao responsável pelo projeto após se registarem na plataforma.

Os três cenários visam responder a necessidades particulares que podem existir em vários domínios, assegurando que os investigadores e instituições envolvidas têm o controlo sobre os seus dados e recebem crédito pela sua disponibilização.

Para os casos em que se pretende arrancar com o processo de divulgação desde as fases iniciais (quer para os dados, quer para os metadados), a plataforma Dendro tira partido de mecanismos de disseminação já existentes. A disseminação é feita através do protocolo OAI-PMH, que estipula um conjunto de regras para a ligação entre repositórios e indexadores, facilitando a troca de informações entre os mesmos. A divulgação de dados no lado do EUDAT é feita pelo módulo B2Find. À semelhança dos restantes módulos, este é capaz de se ligar a diferentes componentes do EUDAT para indexar conteúdos que estejam disponíveis. Por outro lado, o B2Find também é capaz de recolher informação de repositórios que sejam compatíveis com este protocolo, redireccionando o utilizador que pesquisa nesta interface para o destino correspondente ao conjunto de dados que pretende consultar.

Simulação de transportes: um caso de teste

A ligação das duas plataformas permite que vários cenários para a divulgação de dados sejam abrangidos. Como uma iniciativa de teste e divulgação destas ferramentas, foram contactados vários grupos de trabalho para identificar conjuntos de dados que estivessem prontos a serem depositados e divulgados. O procedimento de avaliação consistia em duas fases: uma entrevista inicial para avaliar o domínio do investigador e os seus hábitos de gestão de dados; e uma atividade de descrição destes dados na plataforma Dendro e consequente exportação para a plataforma EUDAT como passo final.

Como caso de teste foi selecionado um investigador no domínio da simulação de transportes que dispunha de dados prontos a depositar e que tinha servido de base para publicações na sua área. Durante a entrevista inicial, o investigador revelou que não dispunha de quaisquer indicações sobre como gerir os seus dados, ficando encarregue de o fazer, se pretendesse que estes fossem preservados. Relativamente à sua publicação, estes dados poderiam ser acedidos por outros investigadores, dado que serviam como base para testes de algoritmos desenvolvidos neste domínio, como por exemplo de alocação ou reserva de recursos. Quando questionado sobre como é partilhada a informação neste domínio, o investigador revelou desconhecer plataformas para o efeito, o que implica que para todos os projetos, seria necessário obter parcerias com empresas que pudessem ceder dados devidamente anonimizados para testes neste domínio. Na entrevista seguinte, dados de uma experiência feita durante o projeto mais recente foram depositados e descritos pelo investigador, adicionando descrições que este considerava importantes para facilitar a interpretação pela comunidade científica. No final deste processo de descrição, os dados foram exportados para a plataforma EUDAT, estando agora disponíveis para a comunidade e podendo igualmente serem citados (Oliveira, 2016). No seguimento deste contacto, outros investigadores procederam também ao depósito de conjuntos de dados disponíveis, mantendo agora um registo do seu depósito, tal como acontece com as suas publicações (Camacho, 2016) (Vieira, 2016).

Conclusões

Dependendo da arquitetura de cada plataforma, a integração de plataformas para a gestão de dados pode trazer benefícios para a sua preservação e divulgação. Fruto do crescente interesse em propor ferramentas que se integrem no dia a dia dos investigadores, plataformas como o EUDAT têm vindo a ser apresentadas e melhoradas. Apesar de todas as funcionalidades que oferecem, estas plataformas ainda revelam alguma dificuldade em lidar com a qualidade dos dados, mais concretamente com a sua descrição. A existência de plataformas como o Dendro, com uma preocupação em capturar e descrever os dados assim que estes são produzidos podem contribuir para facilitar a sua interpretação e reutilização pela comunidade científica. A ligação entre as duas plataformas revelou ser uma solução para investigadores em vários estágios dos seus projetos e, em particular, foi capaz de dar resposta a cenários em que os projetos já terminaram, mas os dados ainda se mantêm sob responsabilidade dos investigadores.

Referências bibliográficas

(Amorim, 2015) AMORIM, Ricardo Carvalho *et al.* – A Comparative Study of Platforms for Research Data Management: Interoperability, Metadata Capabilities and Integration Potential.

New Contributions in Information Systems and Technologies Advances in Intelligent Systems and Computing. 2015) 101-111. doi:10.1007/978-3-319-16486-1_10.

(Amorim, 2014) AMORIM, Ricardo Carvalho *et al.* – LabTablet: Semantic Metadata Collection on a Multi-domain Laboratory Notebook. Communications in Computer and Information Science Metadata and Semantics Research. 2014 193-205. doi:10.1007/978-3-319-13674-5_19.

(Burns, 2013) BURNS, C. Sean, Amy Lana, John Budd – Institutional repositories: exploration of costs and value. D-Lib Magazine 2013 19.1.

(Camacho, 2016) CAMACHO, Rui – Comparative Study of Classification Algorithms Using Molecular Descriptors in Toxicological Data Bases. University of Porto 2016 <http://hdl.handle.net/11304/fc173a9d-18c7-4949-ba72-15f8f5e5806d>

(Comp, 2016) AMORIM, Ricardo Carvalho *et al.* – A comparison of research data management platforms: architecture, flexible metadata and interoperability. Universal Access in the Information Society. 2016 1-12. doi:10.1007/s10209-016-0475-y

(Massimiliano, 2016) ASSANTE, Massimiliano, Leonardo Candela, Donatella Castelli, Alice Tani. – Are Scientific Data Repositories Coping with Research Data Publishing? Data Science Journal 15 2016 doi: <http://doi.org/10.5334/dsj-2016-006>.

(McLure, 2014) MCLURE Merinda, Level Allison V., Cranston Catherine L, Oehlerts Beth, Culbertson M. – Data curation: A study of researcher practices and needs. portal: Libraries and the Academy. 2014 14(2) 139-64. doi:10.1353/pla.2014.0009

(Oliveira, 2016) OLIVEIRA, Beatriz – Vehicle reservation assignment in car rental. University of Porto 2016, <http://hdl.handle.net/11304/12dce69e-6355-4353-b021-8887fed6528c>

(Silva, 2014) João Rocha, Cristina Ribeiro, João Correia Lopes – Ontology-based multi-domain metadata for research data management using triple stores. Proceedings of the 18th International Database Engineering & Applications Symposium. 2014 105-114. doi:10.1145/2628194.2628234

(Vieira, 2016) VIEIRA, Bruno – TweepProfiles Twitter data. University of Porto 2016 <http://hdl.handle.net/11304/93e8b144-5ada-46bb-974f-786b21b1ab06>