

UPBOX: ARMAZENAMENTO NA NUVEM PARA DADOS DE INVESTIGAÇÃO DA U. PORTO

José Pedro Marques Barbosa

Dissertação realizado sob a orientação da *Prof. Maria Cristina Ribeiro* e do *Prof. João Correia Lopes*
na *Faculdade de Engenharia da Universidade do Porto*

28 de Fevereiro de 2013

1. Motivação

Nas últimas décadas tem-se assistido a um aumento significativo na quantidade e complexidade de dados de investigação, muito devido à evolução de métodos, instrumentos e ferramentas aliados à capacidade de armazenamento digital atual [1].

A preservação e partilha de dados científicos é essencial para reutilização por parte da comunidade. Contudo, ainda existem muitos investigadores que não partilham os seus dados de investigação por vários motivos, entre os quais a complexidade do processo de depósito de dados em repositórios [2, 3].

A Universidade do Porto (UP) desenvolveu um projeto denominado *UPData* cujo objetivo foi o estudo das necessidades de curadoria de dados científicos em múltiplos domínios de investigação. No âmbito deste projeto foi desenvolvido um protótipo de um repositório de dados.

Atualmente, o depósito de dados científicos no repositório experimental da UP é realizado manualmente, através do contacto direto entre o curador e o investigador. Este processo torna-se moroso, pois requer que o investigador prepare o conjunto de dados a submeter e, com ajuda do curador, os anote devidamente, para este proceder ao seu depósito.

2. Objectivos

Este projeto propõe uma nova abordagem à curadoria de dados cujo objetivo é agilizar e automatizar o processo de curadoria e submissão de dados no repositório da UP, aproximando os investigadores do processo de curadoria através de um serviço familiar de gestão e centralização dos seus dados de investigação na nuvem.

Assim sendo, os objetivos deste trabalho são:

- Incluir os investigadores no processo de curadoria;
- Suportar a gestão e anotação dos seus dados de investigação desde o início do projeto;
- Preparar a submissão de dados no repositório da UP.

Esta trabalho visa melhorar o atual fluxo de inserção de dados no repositório da UP, com o obje-

tivo de incentivar investigadores à submissão dos seus dados de investigação. Consequentemente, espera-se uma maior visibilidade da UP no que diz respeito à partilha de dados na *web*.

3. Descrição do Trabalho

A solução proposta, UPBox, é um serviço que suporta a gestão colaborativa de dados de investigação durante todo o processo de investigação, mantendo os investigadores em controlo da organização dos seus conjuntos de dados. Este serviço de armazenamento de dados de investigação na nuvem permite ao investigador anotar os seus dados e, quando pertinente, submete-los para curadoria, com vista a serem disponibilizados no repositório de dados da Universidade do Porto. Esta plataforma simples e familiar aos investigadores servirá, então, como intermediário entre o investigador e o repositório de dados, permitindo agilizar e facilitar todo o processo de depósito.

A UPBox é uma aplicação *web* que permite a um investigador criar áreas de trabalho por projetos, estruturadas em diretórios e partilhá-las com outros investigadores. Está integrada com um sistema de anotação, desenvolvido num projeto paralelo, o *DataNotes*, que permite anotar ficheiros e diretórios. Foi desenvolvido um *web service* que disponibiliza métodos a serem utilizados por futuras aplicações cliente que estendam a UPBox.

A Figura 1 mostra a posição da UPBox no processo de curadoria, bem como a sua integração com outras plataformas. O primeiro contacto do investigador com este processo será com a UPBox com a submissão e gestão dos seus dados de investigação. A qualquer momento o investigador poderá anotar os datasets, sendo para tal direcionado para o *DataNotes*, um sistema de anotação de ficheiros e diretórios. Estas anotações poderão ser criadas e editadas quando oportuno e sempre que se mostre necessário, sendo que no final de cada edição o utilizador é reencaminhado para a UPBox e é recebido um ficheiro com o *backup* da anotação efetuada. Esta anotação servirá para uso futuro, fora do âmbito desta trabalho.

Quando pertinente, o investigador poderá submeter os seus dados e anotações para curadoria, com vista a serem publicados no repositório de dados da UP, através da UPBox. Estes dados serão disponibiliza-

dos ao curador através da ferramenta de curadoria de dados do DSpace para os validar e submeter no repositório. Caso surja alguma dúvida, aquando da curadoria, o investigador será contactado diretamente pelo curador para resolver o problema.

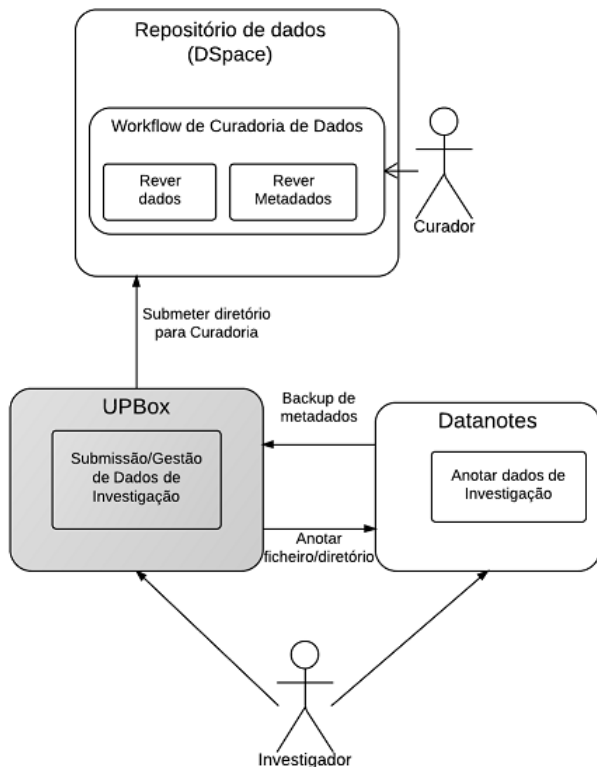


Fig. 1 – Fluxo de trabalho para a curadoria de dados na UP

4. Avaliação

Está prevista a avaliação de todo o processo de depósito de dados no repositório de dados da UP num projeto posterior a este desenvolvimento. Este incluirá testes à gestão de dados na *UPBox*, por parte de investigadores, à anotação de dados no *DataNotes* e respetiva gestão de anotações, ao processo de curadoria de dados, por parte de curadores no repositório de dados da UP e à integração de todas estas plataformas.

Apesar de estarem previstos testes à *UPBox*, realizaram-se testes de usabilidade preliminares ao protótipo desenvolvido. Estes tinham em vista a sua validação, recebendo, assim, *feedback* dos utilizadores e permitindo melhorar alguns aspetos do protótipo para a fase posterior de testes.

O teste foi direcionado a investigadores da UP nas áreas das ciências da informação e de informática, onde lhes foi pedido para realizarem um conjunto de tarefas. Foram utilizadas três métricas para avaliar a usabilidade do protótipo desenvolvido: o tempo de

execução das tarefas, o número de erros encontrado e o número de cliques efetuado, sendo que por vezes poderá ser relevante o percurso para efetuar a tarefa.

Concluiu-se que o protótipo desenvolvido foi bem aceite pelos investigadores devido à sua facilidade de uso e às funcionalidades disponibilizadas. Esperam-se futuros desenvolvimentos que visem a melhoria de algumas funcionalidades, assim como a implementação de outras que melhorem a interação com o utilizador.

5. Conclusões

Neste trabalho desenvolveu-se um protótipo cujo objetivo é suportar a gestão e partilha de dados durante o processo de investigação e permitir a sua anotação. Futuramente espera-se integrar este protótipo com o repositório da UP, de forma a permitir a submissão desses dados e anotações no mesmo de forma simples e transparente.

O protótipo desenvolvido, *UPBox*, consiste numa aplicação *web* que permite o depósito e partilha de dados de investigação na nuvem. A *UPBox* integra com um sistema de anotação, o *DataNotes*, de modo a permitir anotar ficheiros com recurso a vocabulários multidisciplinares. Foi desenvolvido um *web service* que disponibiliza métodos a serem utilizados por futuras aplicações cliente que estendam a *UPBox*.

Estão planeados testes à *UPBox* e a outros sistemas, fora do âmbito deste trabalho, que irão avaliar todo o fluxo de trabalho da inserção de dados no repositório da UP. Mesmo assim, foram realizados testes de usabilidade e questionários para detetar falhas no protótipo desenvolvido e avaliar a sua usabilidade.

Futuramente, com a integração da *UPBox* com o repositório de dados da UP, espera-se que os investigadores se sintam motivados a submeter os seus dados de investigação no repositório da UP com a utilização deste serviço, visto este ter simplificado todo o processo de submissão.

Referências

- [1] Miguel Ferreira. *Introdução à Preservação Digital: Conceitos, estratégias e actuais consensos*. Universidade do Minho. Escola de Engenharia, 2006.
- [2] João Rocha da Silva, Cristina Ribeiro, e João Correia Lopes. *Managing Research Data at the University of Porto: Requirements, Technologies, and Services*. IGI Global.
- [3] João Rocha, Cristina Ribeiro, e João Correia Lopes. *UPData: A Data Curation Experiment at U.Porto using DSpace*. *iPRES 2011: 8th International Conference on Preservation of Digital Objects*, páginas 224–227, 2011.