

UPBox: CLOUD STORAGE FOR RESEARCH DATA OF U. PORTO

José Pedro Marques Barbosa

Dissertation supervised by *Prof. Maria Cristina Ribeiro* and *Prof. João Correia Lopes*
in *Faculty of Engineering of the University of Porto*

February 28th, 2013

1. Motivation

In recent decades there has been a significant increase in the quantity and complexity of research data, much due to the evolution of methods, instruments and tools combined with current digital storage capacity [1].

The preservation and sharing of scientific data is essential for reuse by the community. However, there are still many researchers who do not share their research data for many reasons, including the complexity of the process needed to deposit data in repositories [2, 3].

The University of Porto (UP) developed a project called *UPData* whose objective was to study the needs for curation of scientific data in multiple fields of research. Under this project we developed a prototype of a data repository.

Currently, the deposit of scientific data in the experimental repository of UP is performed manually, through direct contact between the researcher and curator. This process becomes cumbersome, as it requires the researcher to prepare the dataset to be submitted and, with the help of the curator, to properly annotate them, so that he can deposit this data.

2. Goals

This project proposes a new approach to data curation whose goal is to speed up and automate the submission and curating processes in the UP repository, bringing researchers and curation process closer through a familiar service that manages and centralizes their research data in the cloud.

Therefore, the objectives of this work are to:

- Include researchers in the process of curation;
- Support the management and annotation of their research data from the beginning of the project;
- Prepare the submission of data in the repository of UP.

This work aims to improve the current flow of data entry in the repository of UP, in order to encourage researchers to submit their research data. Consequently, we expect a greater visibility of UP with respect to data sharing in the web.

3. Work Description

The proposed solution, UPBox, is a service that supports the collaborative management of research data throughout the research process, keeping researchers in control over the organization of their datasets. This cloud storage service of research data allows the researcher to annotate their data and, where appropriate, submit them for curation in order to be available in the data repository of the University of Porto. This simple and familiar platform will serve as a proxy between the researcher and the data repository, enabling an agile and easy deposit process.

UPBox is a web application that allows researchers to create workspaces for projects, structured directories and share them with other researchers. It is integrated with an annotation system, developed in a parallel project, *DataNotes*, which enables the annotation of files and directories. A web service that provides methods to be used by future client applications that extend the UPBox was developed.

Figure 1 shows the position of UPBox in the process of curation, as well as its integration with other platforms. The researcher's first contact with this process will be with UPBox by submitting and managing their research data. At any time the investigator can annotate the datasets, being redirected to *DataNotes*, a file and directories annotation system. These annotations can be created and edited whenever deemed necessary, and at the end of each edition, the user is redirected to UPBox and a file with the backup of the annotation is received. This annotation will serve for future use, outside the scope of this work.

Where relevant, the researcher may submit his data for curation and annotation, in order to be published in the data repository of UP through UPBox. This data will be available to the curator through the *DSpace's* curating tool, to be validated and inserted in the repository. During the curation process, if anything is unclear, the investigator will be contacted directly by the curator to solve the problem.

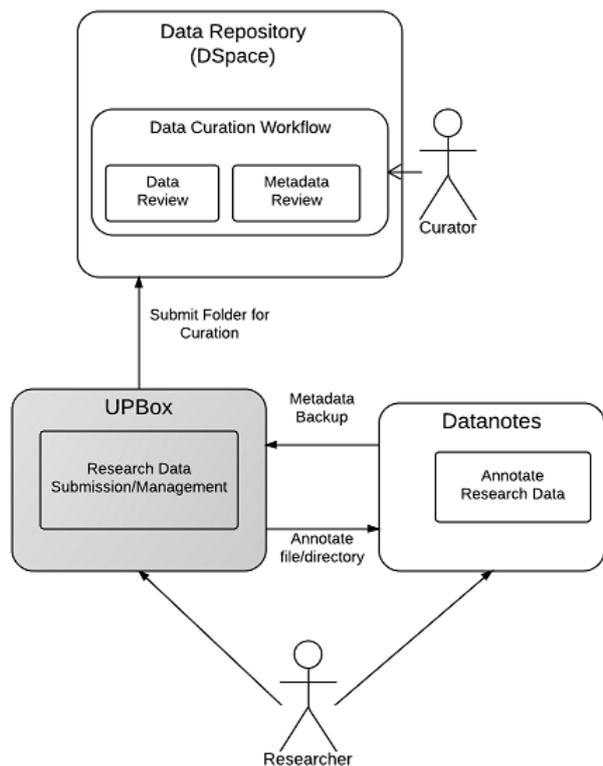


Fig. 1 – Workflow for data curation in UP

4. Evaluation

The evaluation of the whole process of depositing data in the UP data repository is expected to take place in a project following the present development. This evaluation will include tests regarding the management of data in UPBox by the researchers, the annotation of data on DataNotes, the corresponding management of annotations, the process of data curation by the curators, and the integration of all these platforms.

Despite the planned tests, preliminary usability tests were performed. These aimed to validate the developed prototype and to obtain feedback from the target users, to improve some features of the developed prototype.

The test was centered on researchers from UP in the fields of information sciences and information technology, where they were asked to perform a set of tasks. Three metrics to assess the usability of the prototype were used: the execution time of tasks, the number of errors found and the number of clicks made. Additionally, the route taken to perform the task may also

be relevant.

It was concluded that the prototype was well accepted by researchers due to its ease of use and available features. In the future, it is expected that future development will improve the available features, as well as implementation of others.

5. Conclusions

In the presented work, a prototype with the goal of supporting the management and sharing of data during the research process and allowing their annotation was developed. In the future it is expected to integrate this prototype with the repository of UP, to enable the submission of data and annotations in a simple and transparent manner.

The developed prototype, UPBox, is a web application that allows the storage and sharing of research data in the cloud. UPBox integrates with an annotation system, DataNotes, to allow the annotation of files using multidisciplinary vocabularies. A web service that provides methods to be used by future client applications that extend the UPBox was developed.

Tests regarding UPBox and other systems that will evaluate the workflow of inserting data in the UP repository are planned, beside the scope of this study. Still, usability tests and questionnaires were conducted in order to detect failures in the prototype and evaluate its usability.

In the future, with the integration of UPBox in the UP data repository, it is expected that researchers will be motivated to submit their research data in the UP repository using this service, since it has simplified the entire process of submission.

References

- [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, and João Correia Lopes. *Managing Research Data at the University of Porto: Requirements, Technologies, and Services*. IGI Global.
- [3] João Rocha, Cristina Ribeiro, and João Correia Lopes. *UPData: A Data Curation Experiment at U.Porto using DSpace*. *iPRES 2011: 8th International Conference on Preservation of Digital Objects*, pages 224–227, 2011.