



UPBox

Solução de armazenamento na nuvem para dados científicos na U.Porto

Orientando: José Barbosa

Orientadora: Prof. Cristina Ribeiro

Co-Orientador: Prof. João Correia Lopes

Mestrado Integrado em Eng.
Informática e Computação

Apresentação Final (PDIS)

10-07-2012

Contexto

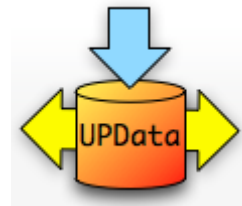
Aumento da geração de dados de investigação

“A partilha de dados e de recursos serão a chave para a resolução dos novos problemas da ciência e da engenharia” [1]

Contexto

Projeto de curadoria de dados na Universidade do Porto [2]:

- UPData



Problemas

Os dados de investigação são armazenados por contacto direto entre o curador e o investigador:

- Processo ineficaz;
- Desmotivação de investigadores;
- Necessidade de mais recursos para curadoria.

Objetivo

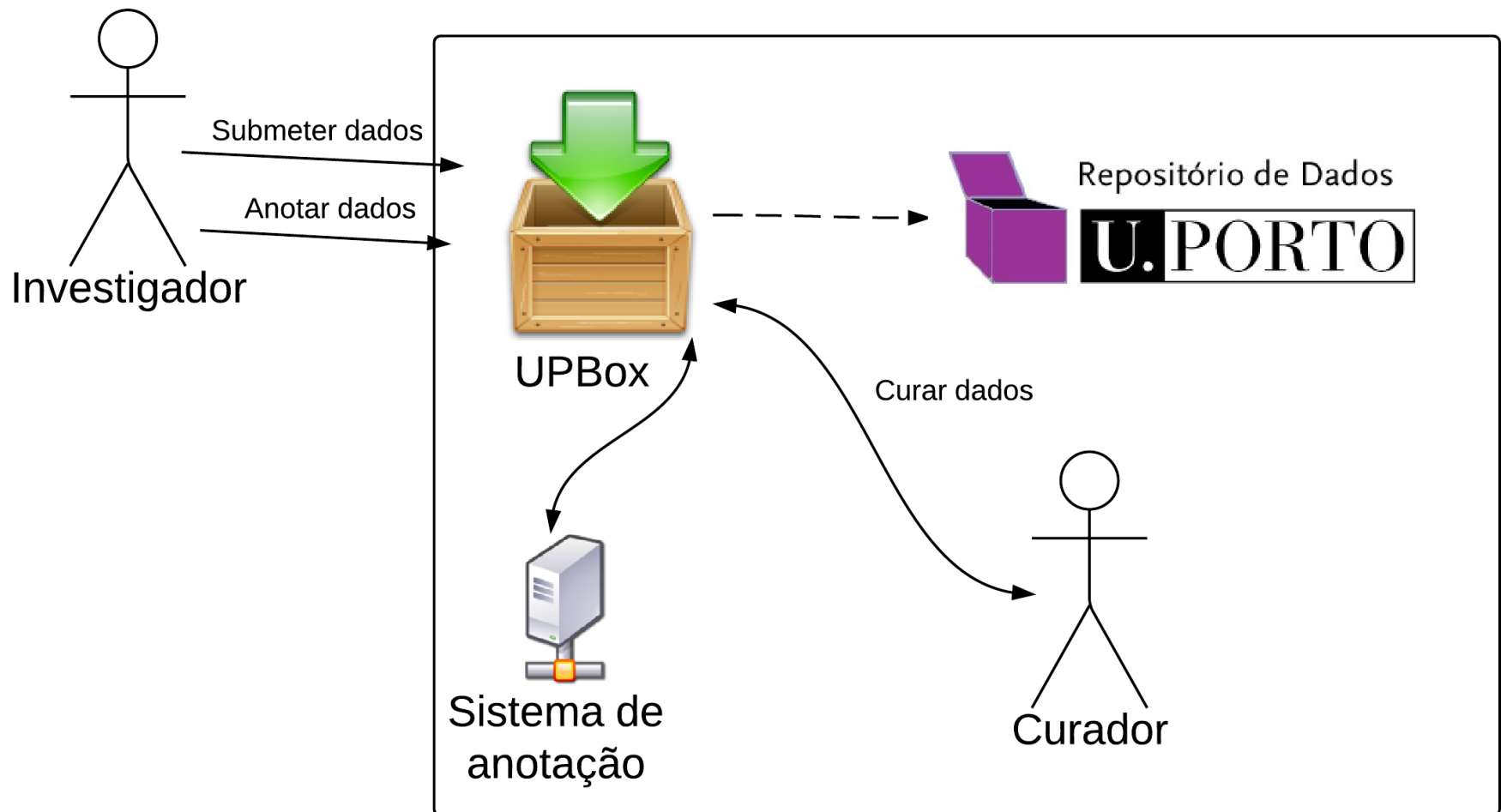
Incluir investigadores no processo de curadoria

Como?

Através de um sistema colaborativo baseado em serviços na nuvem:

- Agilizar o processo de submissão de dados nos repositórios;
- Estimular participação de investigadores no processo de curadoria;
- Auxiliar investigadores na gestão e centralização de dados de investigação;
- Integração com sistema de anotação.

Objetivo



Motivação



Estado da Arte

Repositórios de dados de investigação:

1. Curadoria por organizações científicas:

- NCBI (National Center for Biotechnology Information).
- ICPSR (Inter University Consortium for Political and Social Research);
- DANS (Data Archiving and Networked Services);
- UK Data Archive (Ciências Sociais e Humanidades).

Estado da Arte

Repositórios de dados de investigação:

2. Curadoria por Universidades:

- Datashare (Universidade de Edimburgo).

3. Curadoria por organismos oficiais:

- ANDS (Australian National Data Service).

Estado da Arte

Serviços de armazenamento na nuvem

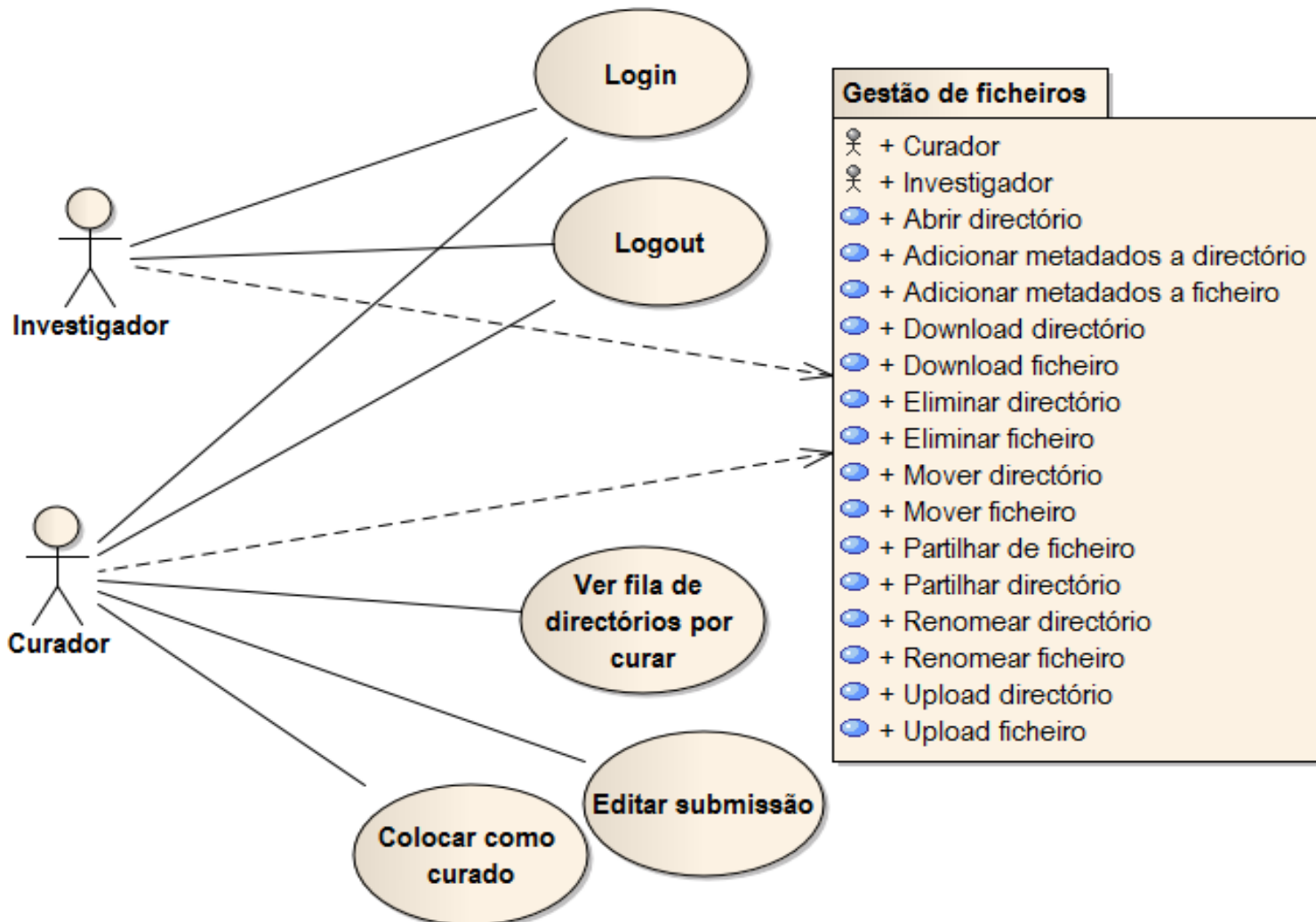
- ownCloud, Dropbox, Ubuntu One
 - Centralização de ficheiros;
 - Gestão de ficheiros remota.
- FTPbox
 - Centralização de ficheiros com diretório local.

Ideia

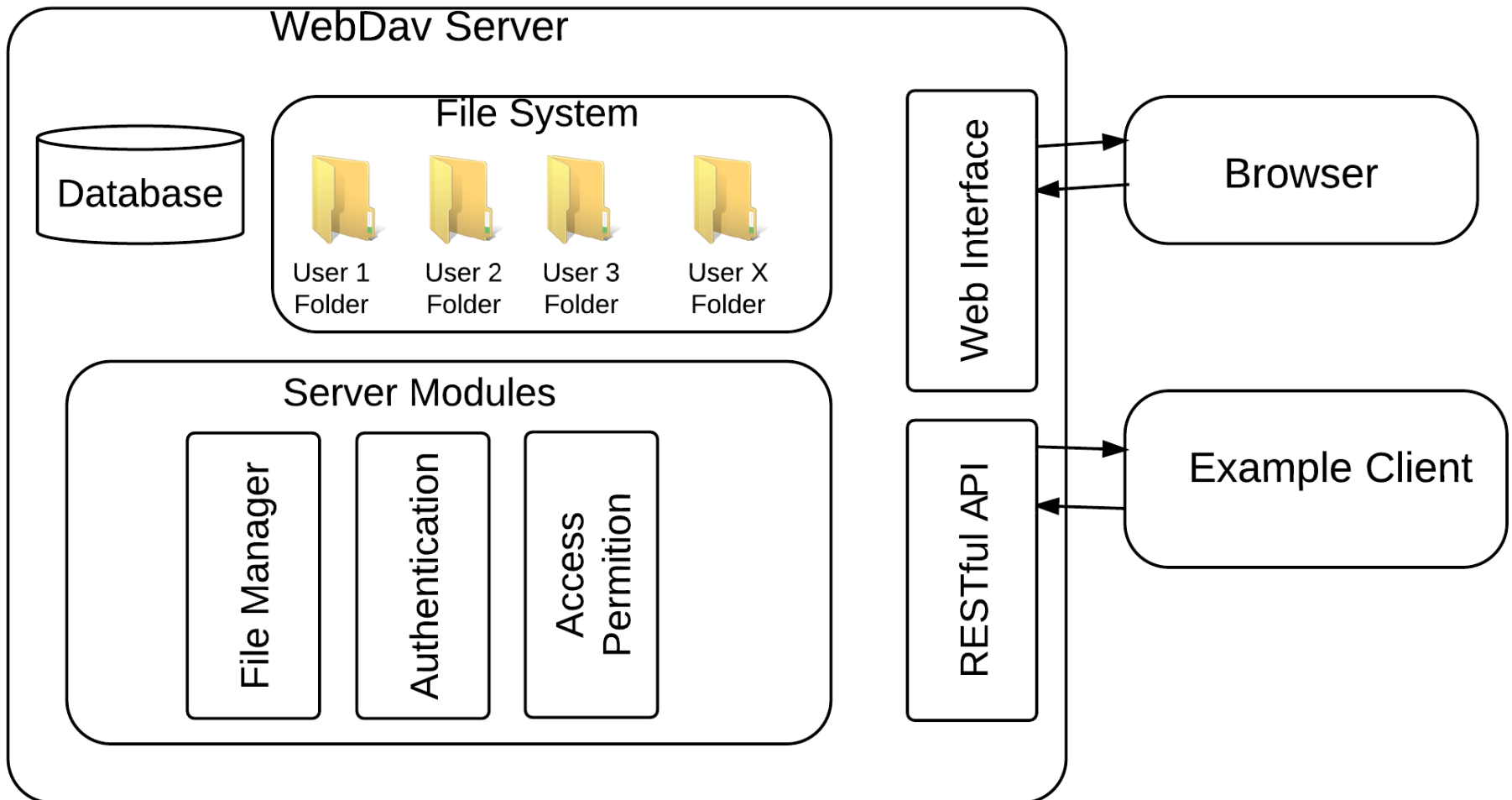
UPBox!

- ... sistema simples e familiar
- ... centralizar dados de investigação
- ... anotar dados de investigação
- ... permitir a participação do investigador no processo de curadoria [2]
- ... agilizar a preservação nos repositórios de dados

Requisitos da solução



Proposta de solução



Porquê WebDav?

Ao contrário de FTP oferece [3]:

- Gestão de ficheiros remota;
- Atribuir metadados a documentos;
- Gerir permissões de acesso;
- Permite transferência de ficheiros segura.

Metodologia

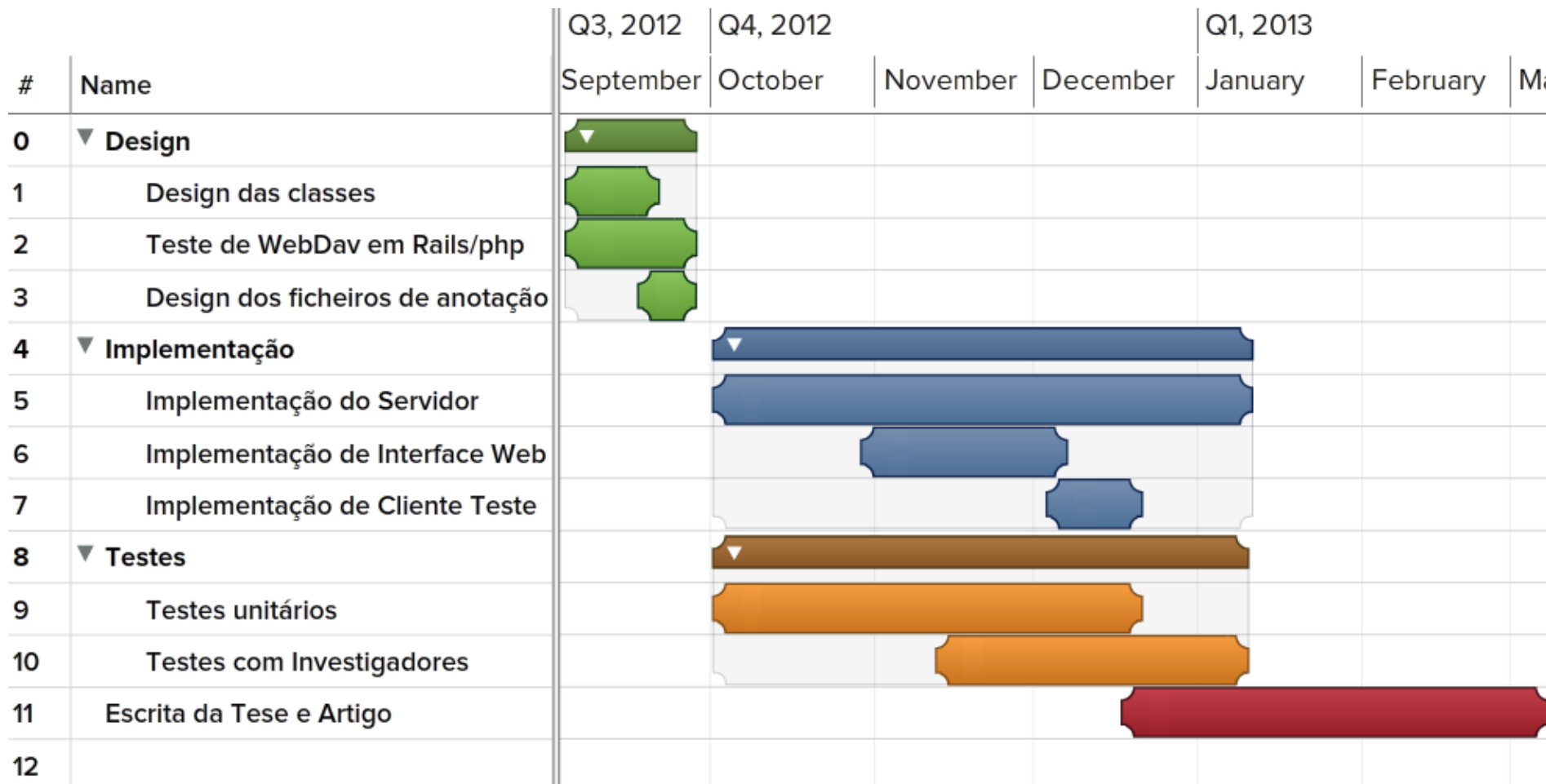
Desenvolvimento iterativo:

- Método ágil;
- Testes contínuos;
- Feedback de investigadores;
- Design simples.

Prova do conceito:

- Implementação do servidor com interface web;
- Implementação de aplicação simples para teste da API;
- Testes com investigadores.

Planeamento



Referências

- [1] Hey, T. (2003). The Data Deluge: An e-Science Perspective. *Grid computing*, (January 2003), 1-17. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1002/cbdv.200490137/abstract>
- [2] Ribeiro, C., Eugénia, M., Fernandes, M., & Porto, U. (2011). Data Curation at U . Porto : Identifying current practices across disciplinary domains by, 14-17.
- [3] Dridi, F., & Neumann, G. (n.d.). How to implement Web-based groupware systems based on WebDAV. *Proceedings. IEEE 8th International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises (WET ICE'99)*, 114-119. IEEE Comput. Soc. doi:10.1109/ENABL.1999.805185
- [4] Ribeiro, C., Saraiva, R., Rodrigues, E., Fernandes, M., Marques Gomes, C., & Carvalho, J. (2010). Os Repositórios de Dados Científicos: Estado da Arte.