



Universidade do Porto

FEUP Faculdade de
Engenharia

UPBox

Armazenamento na Nuvem para Dados de
Investigação da U. Porto

Orientando: José Barbosa

Orientadora: Prof. Cristina Ribeiro

Co-Orientador: Prof. João Correia Lopes

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]

Repositórios de dados

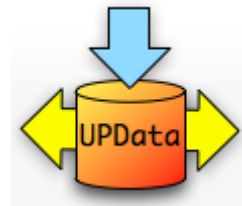
Instituições disponibilizam infraestruturas para manter os seus dados científicos:

- Dados volumosos e de domínios específicos
 - National Center for Biotechnology Information (NCBI)
- Dados pequenos de múltiplos domínios
 - DSpace e EPrints

Contexto

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

- UPData



Problemas

Gestão de dados pouco integrada no processo de investigação



- Depósito requer contacto entre o curador e o investigador;
- Curador externo ao grupo de investigação;
- Os dados não chegam ao repositório.

Objetivo

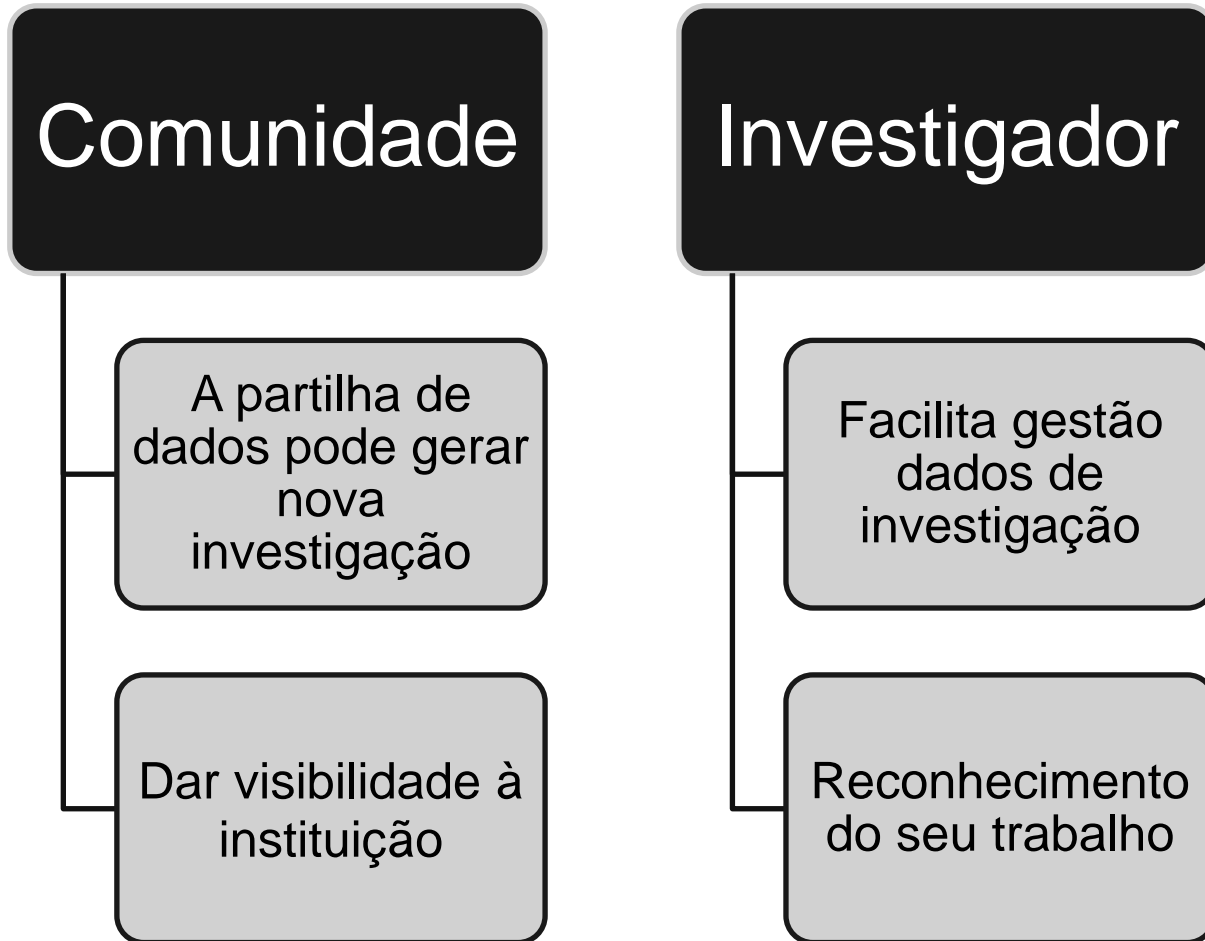
Incluir investigadores no processo de curadoria

Como?

Através de um sistema colaborativo baseado em serviços na nuvem:

- Suportar gestão de dados desde o início do projeto;
- Integrar com sistema de anotação;
- Preparar submissão de dados nos repositórios;

Motivação



Solução

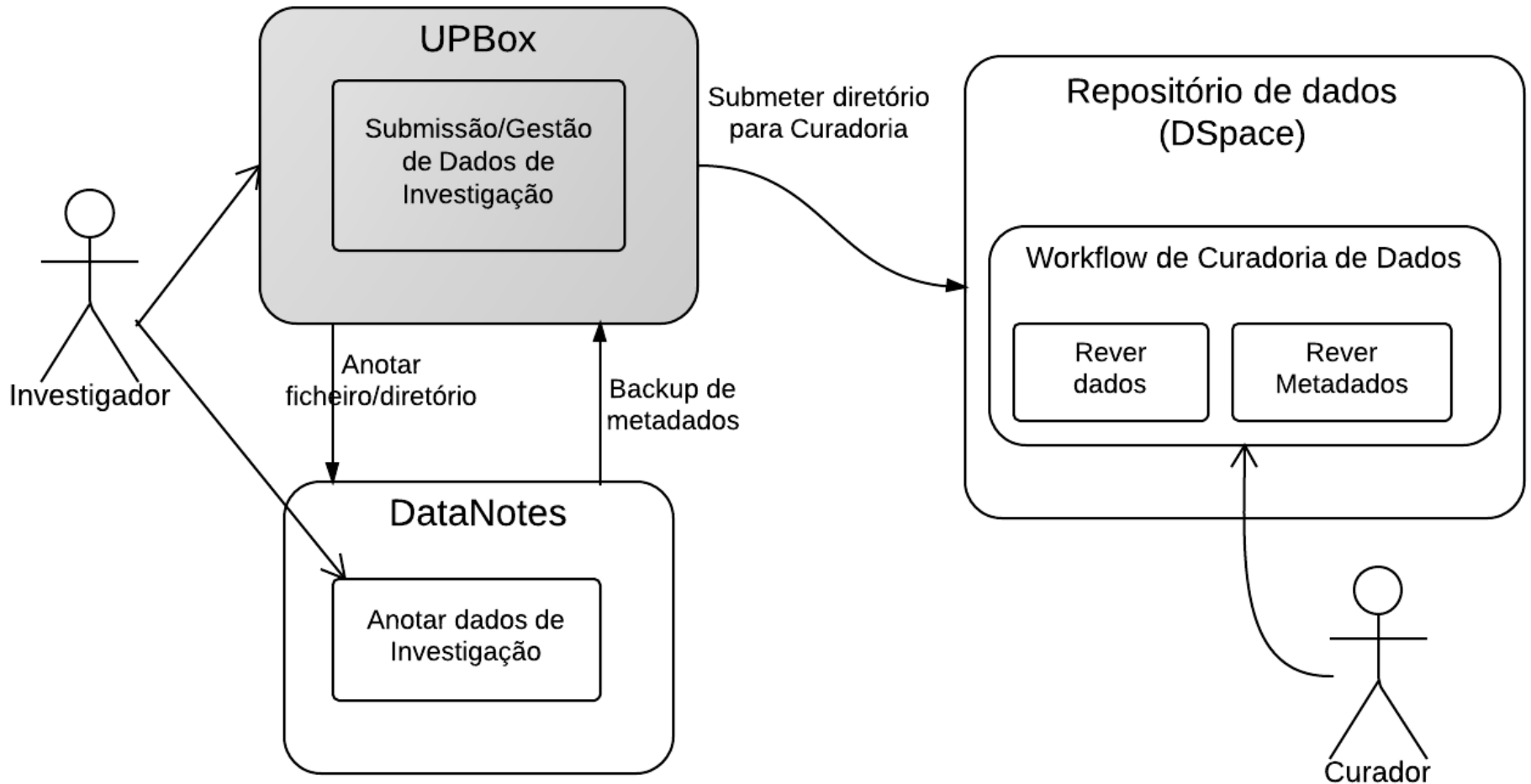
Investigadores utilizam email e DropBox para gerir e partilhar dados de investigação



UPBox:

- Fácil de utilizar;
- Ambiente controlado pela instituição;
- Suporta o grupo de investigação na gestão de dados do dia-a-dia.

Especificação



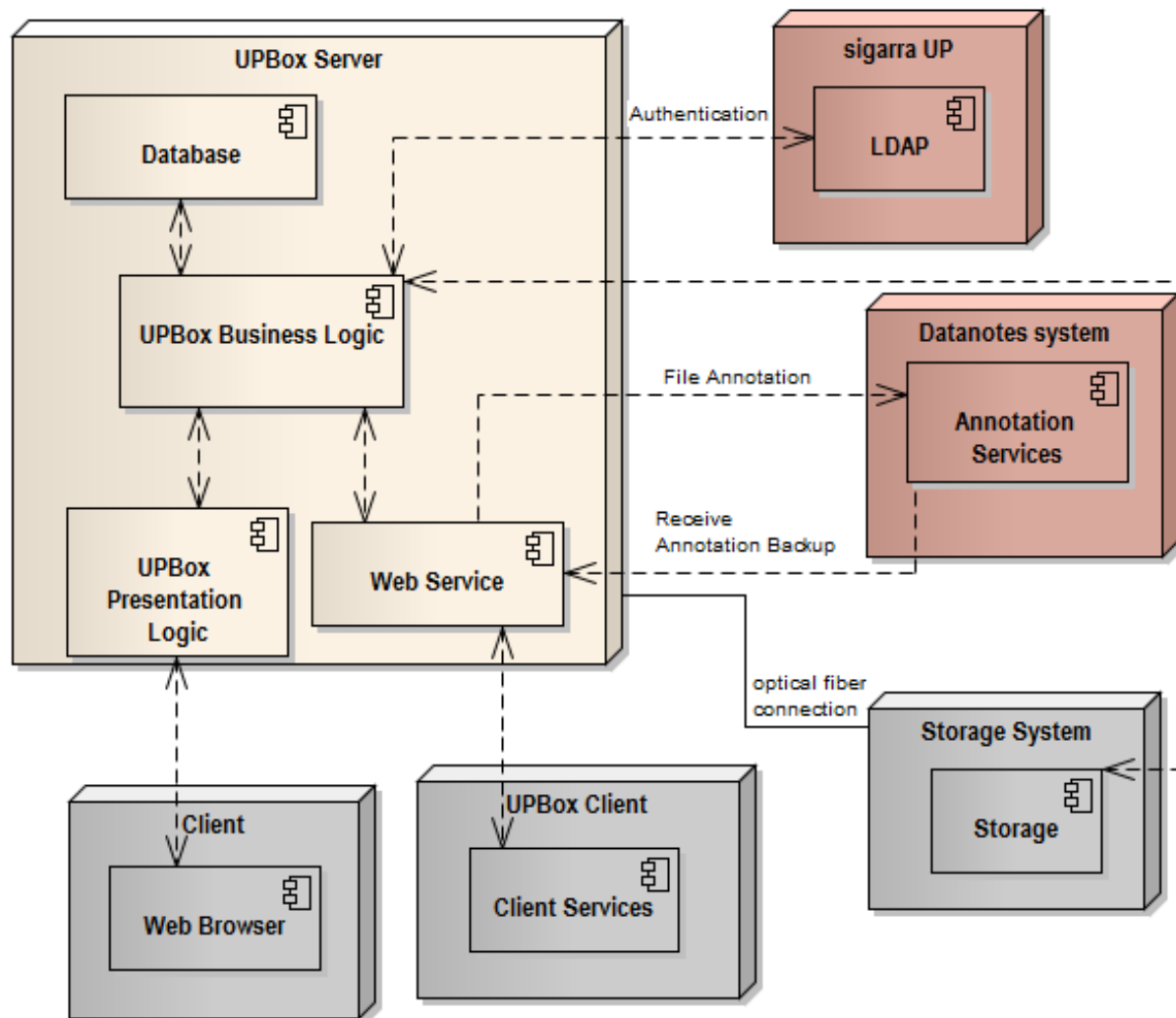
Requisitos funcionais

Identificador	Descrição
RF01	Autenticação com credenciais da UP.
RF02	Gestão de projetos e de acesso aos mesmos.
RF03	Descarregar ficheiros armazenados.
RF04	Gestão de ficheiros armazenados.
RF05	Carregamentos de múltiplos de ficheiros.
RF06	Compressão de ficheiros.
RF07	Descompressão de ficheiros.
RF08	Navegação em diretórios idêntica a sistemas de ficheiros.
RF09	Deteção de erros em criação de diretórios e upload de ficheiros.
RF10	Obtenção de dados através de API.
RF11	Comunicação com o <i>Datanotes</i> para anotação de ficheiros.
RF12	Receção de backups de anotações do <i>Datanotes</i> .
RF13	Marcação de diretórios para curadoria.

Requisitos não funcionais

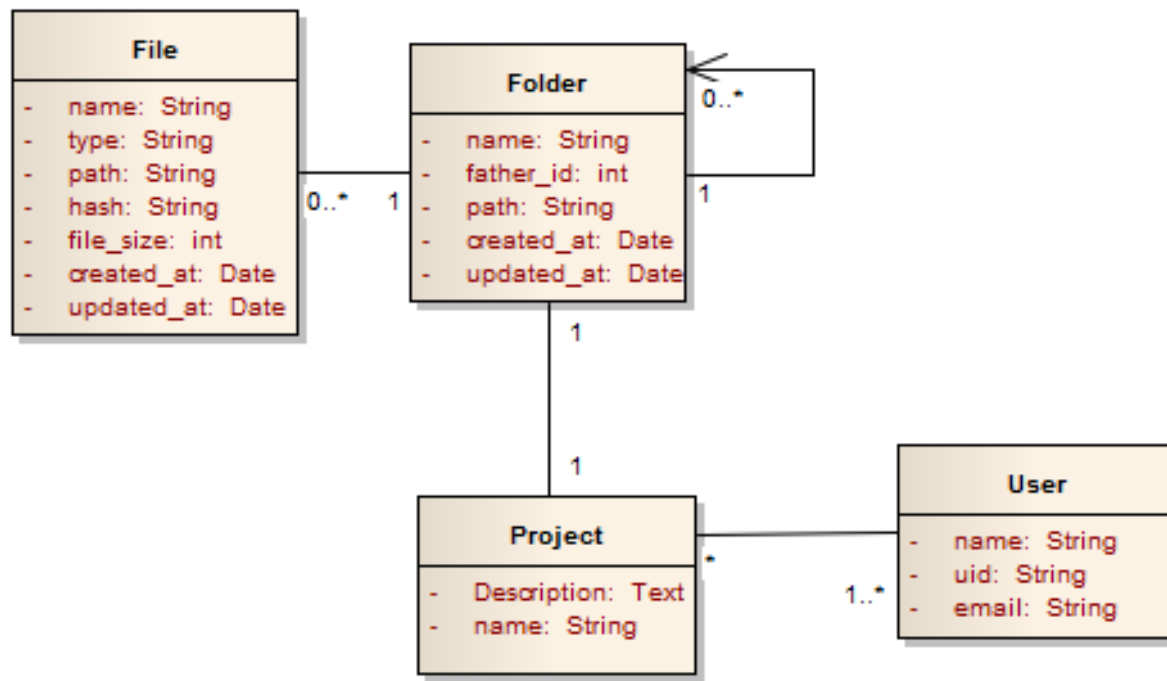
- Usabilidade e simplicidade;
- Interoperabilidade;
- Segurança e privacidade;
- Dados mantidos em servidores da UP.

Arquitetura



Arquitetura

Modelo conceptual do domínio



Tecnologias

nginx server

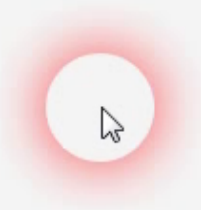
UPBox

PostgreSQL

Debian

Ruby on Rails
HTML5 + CSS3
JavaScript

Demonstração



Please sign in

Sign in

Avaliação

- Teste de usabilidade:
 - Utilidade;
 - Facilidade de utilização.
- Público alvo:
 - Investigadores de eng. informática;
 - Investigadores de ciências de informação.

Avaliação

Tarefas

1. Autenticação;
2. Criar um projeto;
3. Partilhar projeto;
4. Adicionar diretórios/ficheiros ao projeto;
5. Efetuar um download;
6. Remover diretório/ficheiro.

Métricas:

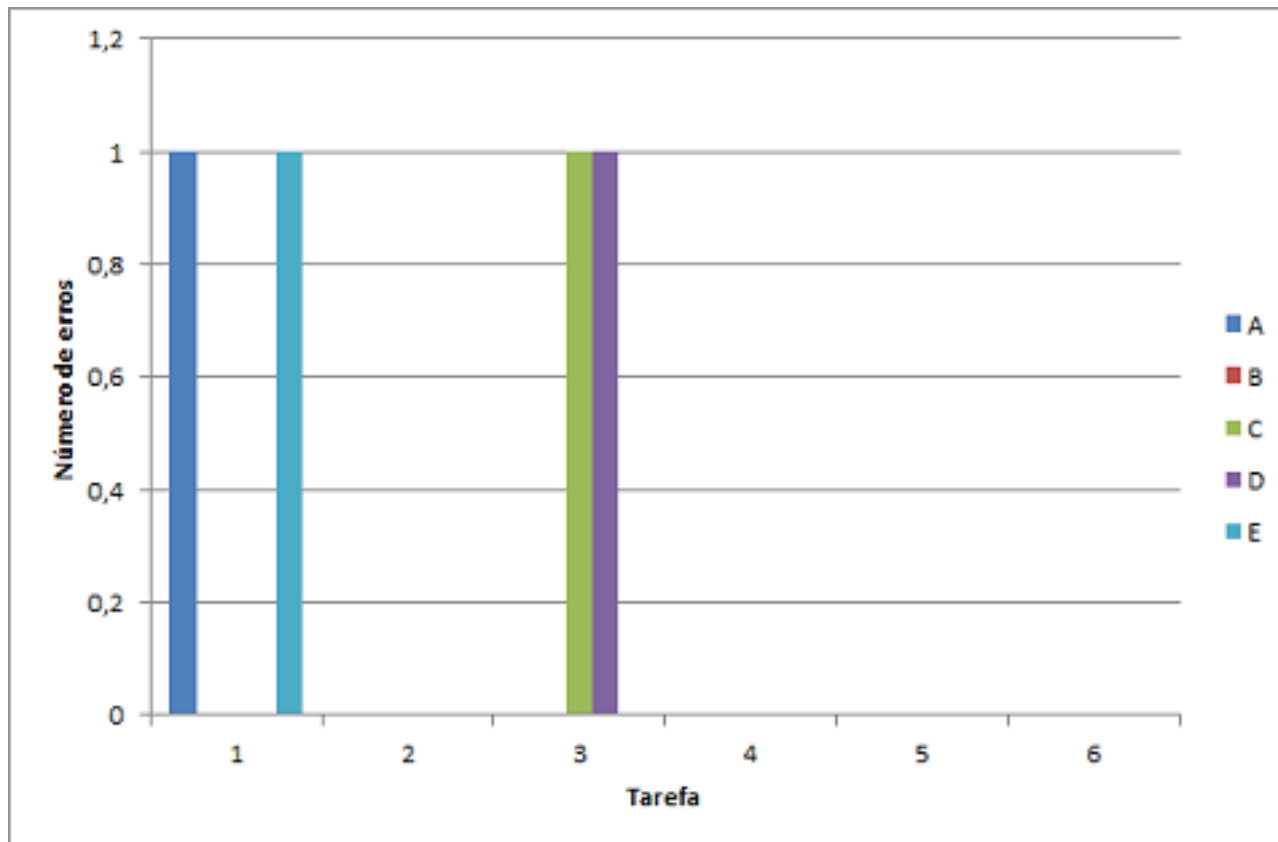
- Tempo;
- Número de erros;
- Número de cliques.

Avaliação

Questionário informal

1. Detetar problemas de usabilidade;
2. Recolher informações sobre facilidade de uso;
3. Avaliar expectativas, sugestão de novas funcionalidades.

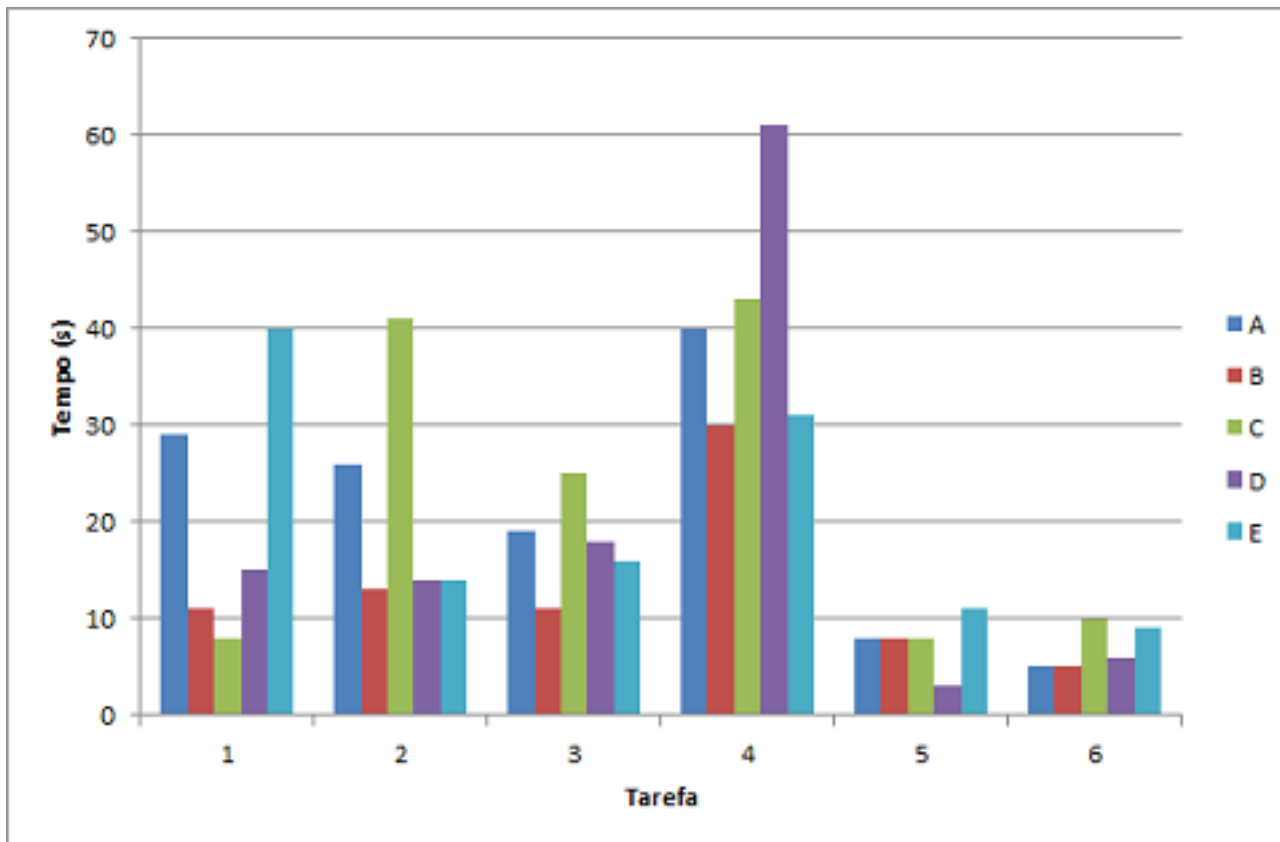
Resultados



Tarefas

1. Autenticação;
2. Criar projeto;
3. Partilhar projeto;
4. Adicionar diretórios/ficheiros ao projeto;
5. Download;
6. Remover diretório/ficheiro.

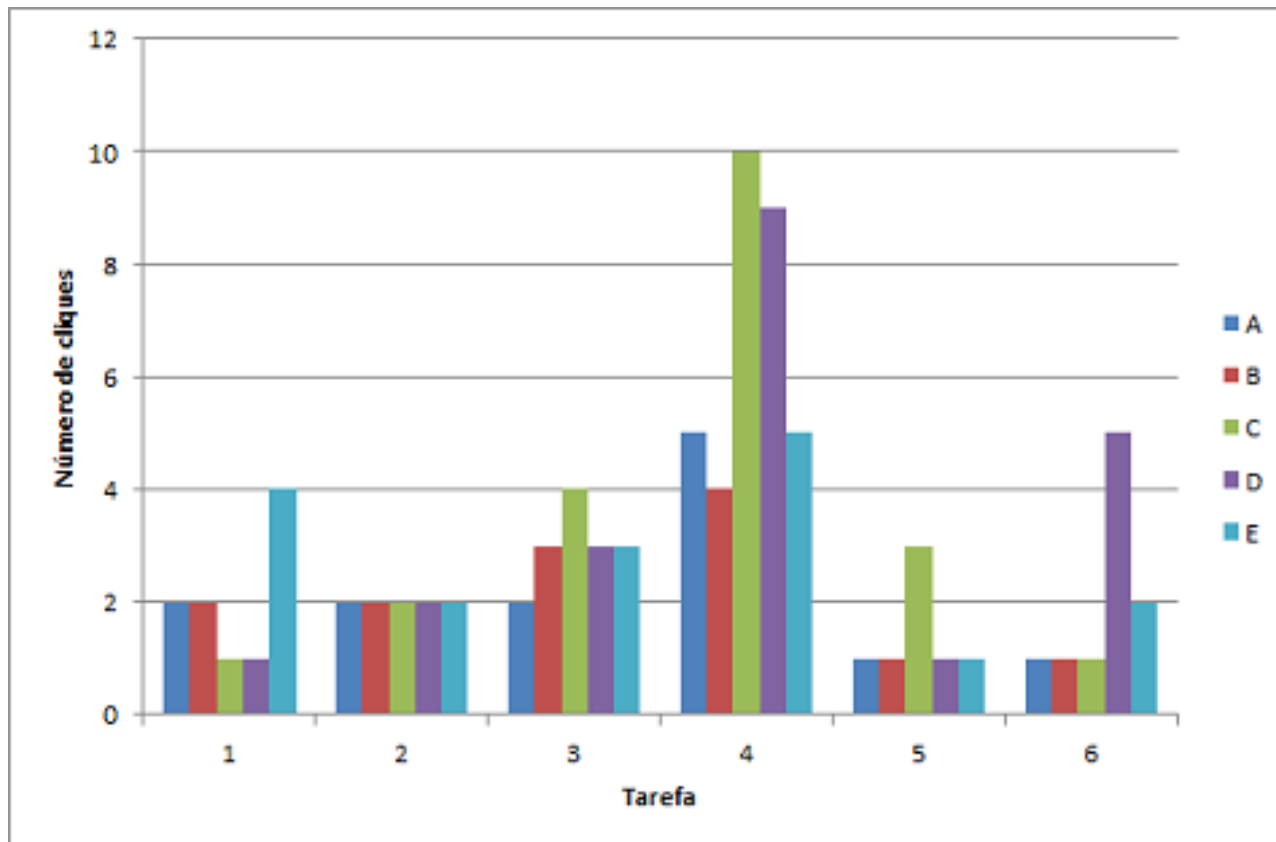
Resultados



Tarefas

1. Autenticação;
2. Criar projeto;
3. Partilhar projeto;
4. Adicionar diretórios/ficheiros ao projeto;
5. Download;
6. Remover diretório/ficheiro.

Resultados



Tarefas

1. Autenticação;
2. Criar projeto;
3. Partilhar projeto;
4. Adicionar diretórios/ficheiros ao projeto;
5. Download;
6. Remover diretório/ficheiro.

Resultados

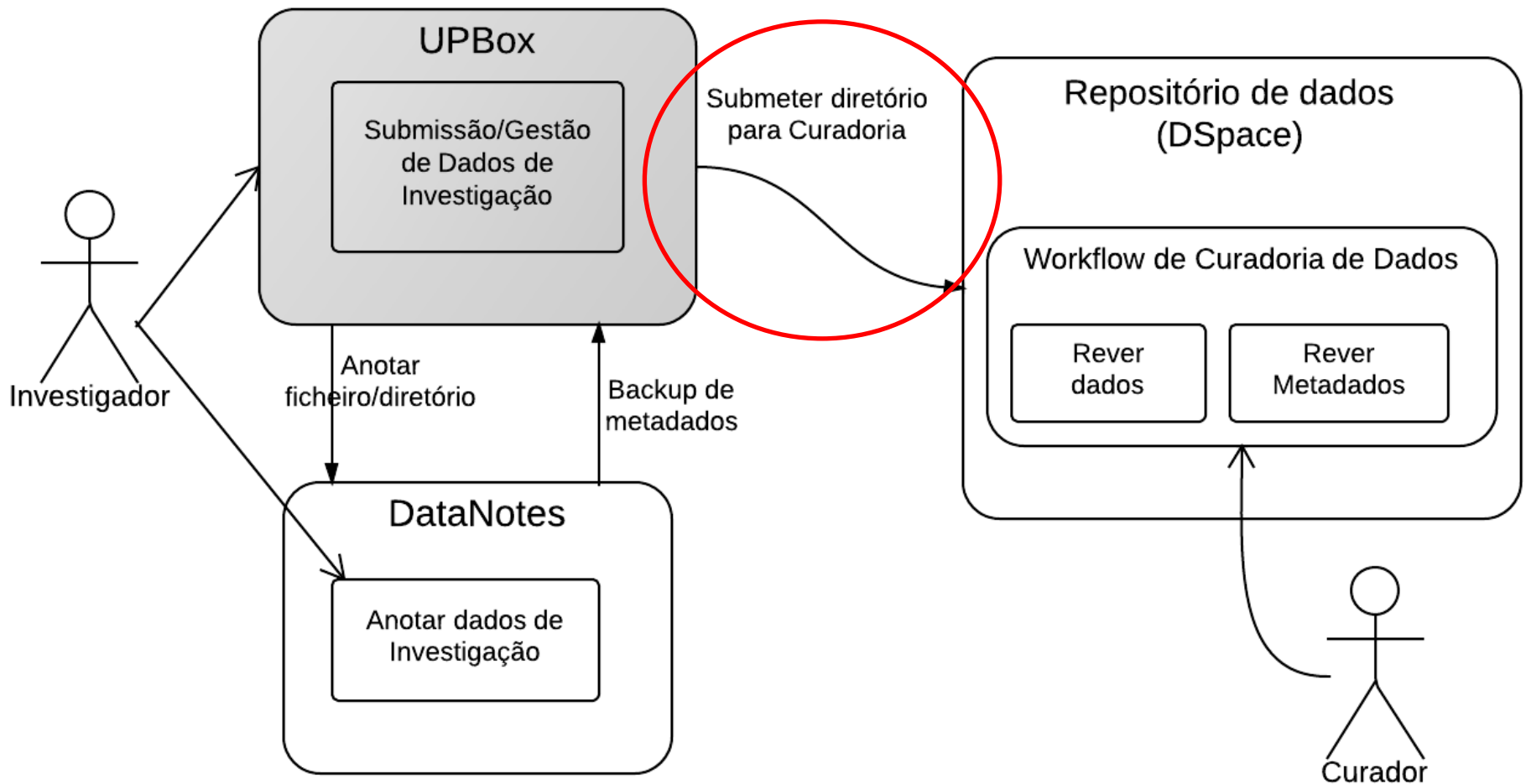
Considerações dos investigadores recolhidas no questionário informal:

- Protótipo usável que satisfaz as necessidades de gestão de dados;
- Rever gestão de colaboradores;
- Permitir upload/download de diretórios;
- Expandir UPBox a cliente desktop;

Conclusão

- Sistema de armazenamento que suporta a gestão de dados de investigação desde o início do projeto;
- A plataforma satisfaz as necessidades de gestão de dados e integra sistema de anotação;
- Espera-se agilizar a submissão de dados no repositório e aumentar o número de submissões.

Trabalho Futuro



Trabalho Futuro

- Desenvolvimento de clientes Desktop que complementem a UPBox;
- Rever gestor de colaboradores;
- Adicionar funcionalidades:
 - Upload de diretórios;
 - Download de diretórios;

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.