



Curso de formação:

**Publicação de  
dados de  
biodiversidade  
através do GBIF**

# OpenRefine

Rui Figueira

[ruifigueira@isa.ulisboa.pt](mailto:ruifigueira@isa.ulisboa.pt)

Nó Português do GBIF,  
Instituto Superior de Agronomia,  
Universidade de Lisboa



## OpenRefine



<http://openrefine.org/>

Programa open source para a limpeza e transformação de dados;

Inicialmente chamada Freebase Gridworks, e depois Google Refine, actualmente é desenvolvida e mantida pela comunidade open source;

Apesar de mostrar os dados em tabela, é diferente nos objectivos e funcionalidades em comparação com uma folha de cálculo.

## Principais características



<http://openrefine.org/>

- Permite a aplicação de filtros simples e compostos;
- as células não contém fórmulas, apenas valores;
- as fórmulas são usadas para aplicar transformações aos dados;
- as transformações aplicam-se a todas as linhas seleccionadas, de uma só vez;
- possui algoritmos para agrupar valores próximos, com escrita ou fonética semelhante  
Ex: Figueira, Figueria
- Possibilita “undos” ilimitados;
- permite o uso de APIs e obter dados online;
- todas as acções são traduzidas para um script;
- permite a implementação de scripts em GREL;
- importa e exporta vários formatos;
- fácil arquivo e transferência de projectos.

## Demonstração



<http://openrefine.org/>

- Abrir OpenRefine e importar dados;

- Possibilita “undos” ilimitados;

- permite o uso de APIs e obter dados online;

- todas as acções são traduzidas para um script;

- permite a implementação de scripts em GREL;

- importa e exporta vários formatos;

- fácil arquivo e transferência de projectos.

# Publicação de dados de biodiversidade através do GBIF

## Obrigado pela atenção

Nó Português do GBIF  
Instituto Superior de Agronomia  
Herbário  
Tapada da Ajuda  
1349-017 Lisboa, Portugal

Tel: (+351) 213653165  
email: [node@gbif.pt](mailto:node@gbif.pt)  
<http://www.gbif.pt>

O Nó Português é acolhido no ISA com o apoio da FCT.



Esta apresentação é publicada segundo a licença CC-BY-SA

