



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/>

## OpenRefine

*“Uma ferramenta gratuita, em código aberto e poderosa para trabalhar com dados desordenados”*

O que é:

- Programa open source para a limpeza e transformação **massiva** 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.

# OpenRefine



<http://openrefine.org/>

## OpenRefine

*“Uma ferramenta gratuita, em código aberto e poderosa para trabalhar com dados desordenados”*

### O que não é:

- não é uma base de dados, não serve para inserir ou gerir informação;
- não é uma folha de cálculo, não serve para produzir análises;

# OpenRefine



<http://openrefine.org/>

**Faceting**

**Clustering**

**Reconciling**

## OpenRefine

*“Uma ferramenta gratuita, em código aberto e poderosa para trabalhar com dados desordenados”*

### Principais características

- 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

# OpenRefine



<http://openrefine.org/>

**Faceting**

**Clustering**

**Reconciling**

## OpenRefine

*“Uma ferramenta gratuita, em código aberto e poderosa para trabalhar com dados desordenados”*

### Principais características

- Possibilita “undos” ilimitados;
- permite o uso de APIs e obter dados online;
- todas as acções são registadas para um script;
- permite a implementação de scripts em GREL e suporta regular expressions;
- importa e exporta vários formatos;
- fácil arquivo e transferência de projectos.

# OpenRefine



<http://openrefine.org/>

**Faceting**

**Clustering**

**Reconciling**

## OpenRefine

*“Uma ferramenta gratuita, em código aberto e poderosa para trabalhar com dados desordenados”*

### Recursos

Download e tutoriais em vídeo

<http://openrefine.org/>

Documentação

<https://github.com/OpenRefine/OpenRefine/wiki>

GREL

<https://github.com/OpenRefine/OpenRefine/wiki/General-Refine-Expression-Language>

# OpenRefine



<http://openrefine.org/>

**Faceting**

**Clustering**

**Reconciling**

## OpenRefine

*“Uma ferramenta gratuita, em código aberto e poderosa para trabalhar com dados desordenados”*

### Demonstração

- importar um ficheiro de dados
- aplicar um filtro de texto
- aplicar filtros numéricos
- fazer correções
- identificar valores duplicados
- desfazer acções (“undos”)

# 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.



INSTITUTO  
SUPERIOR D  
AGRONOMIA  
*Universidade de Lisboa*

**FCT** Fundação  
para a Ciência  
e a Tecnologia

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

