

# TALLER GBIF.ES CALIDAD EN BASES DE DATOS DE BIODIVERSIDAD

Madrid, 2-4 Diciembre 2015



Pablo Muñoz Rodríguez  
Unidad de Coordinación – GBIF.ES  
Nodo Nacional de Información en  
Biodiversidad en España

# OpenRefine



*“A power tool for working with messy data”*

Follow us on: [Github](#) [Twitter](#)

Google™ Custom Search   x

# OPEN Refine



*A free, open source, powerful tool for working with messy data*

- [Home](#)
- [Download](#)**
- [Documentation](#)
- [Community](#)
- [Post archive](#)

OpenRefine News: [October 2015](#)

OpenRefine News: [September 2015](#)

OpenRefine News: [August 2015](#)

OpenRefine News: [July 2015](#)

OpenRefine News: [June 2015](#)

[Mapping OpenRefine Ecosystem](#)

[2014 survey results](#)

[A Governance Model for OpenRefine](#)

[Using OpenRefine: a manual](#)

[OpenRefine History](#)

## Welcome!

OpenRefine (formerly Google Refine) is a powerful tool for working with messy data: cleaning it; transforming it from one format into another; and extending it with web services and external data.

Please note that since October 2nd, 2012, Google is not actively supporting this project, which has now been rebranded to OpenRefine. Project development, documentation and promotion is now fully supported by volunteers. Find out more about the [history of OpenRefine](#) and how you can [help the community](#).

## Using OpenRefine - The Book



**Using OpenRefine**, by Ruben Verborgh and Max De Wilde, offers a great introduction to OpenRefine. Organized by recipes with hands on examples, the book covers the following topics:

1. Import data in various formats
2. Explore datasets in a matter of seconds
3. Apply basic and advanced cell transformations
4. Deal with cells that contain multiple values
5. Create instantaneous links between datasets
6. Filter and partition your data easily with regular expressions
7. Use named-entity extraction on full-text fields to automatically identify topics
8. Perform advanced data operations with the General Refine Expression Language

## Introduction to OpenRefine

### 1. Explore Data

OpenRefine can help you explore large data sets with ease. You can find out more about this functionality by watching the video below and going through these articles



Follow us on:

Github

Twitter

Google™ Custom Search

Search ×

# Download OpenRefine

## [Home](#)

## [Download](#)

## [Documentation](#)

## [Community](#)

## [Post archive](#)

[OpenRefine News:](#)

[October 2015](#)

[OpenRefine News:](#)

[September 2015](#)

[OpenRefine News:](#)

[August 2015](#)

[OpenRefine News: July](#)

[2015](#)

[OpenRefine News: June](#)

[2015](#)

[Mapping OpenRefine](#)

[Ecosystem](#)

[2014 survey results](#)

[A Governance Model for](#)

[OpenRefine](#)

[Using OpenRefine: a](#)

[manual](#)

[OpenRefine History](#)

## OpenRefine Core

### Google Refine 2.5 - Stable version

- **Windows kit**, Download, unzip, and double-click on *google-refine.exe*. If you're having issues with the above, try double-clicking on *refine.bat* instead.
- **Mac kit**, Download, open, drag icon into the Applications folder and double click on it. **NOTE:** If you have issues installing Refine on Mac, please refer to [issue 590](#) - OpenRefine 2.5 for mac support java 6 and 7 only
- **Linux kit**, Download, extract, then type `./refine` to start. **NOTE:** OpenRefine 2.5 for linux support java 6 and 7 only

See also the [installation instruction](#)

### OpenRefine 2.6 - Development version

Download OpenRefine [2.6-beta1](#) (link for all releases).

### All releases

All previous releases are available [here](#).

## Extensions

The following extension are available for OpenRefine:

- [Crowdsourcing extension](#) - by Zemanta
- [DBpedia extension](#) - by Zemanta
- [History tools, pivot tool and scatterplot tool using D3](#) - by VIB-BITS
- [Diff plugin](#) - by VIB-BITS
- [LMF Extension](#) (to be migrated to Apache Marmotta after summer) - by Salzburg Research
- [Named-Entity Recognition](#) - by Ruben Verborgh ([Free Your Metadata](#))
- [RDF extension](#) - by DERI
- [Stats extension](#):
  - [Stats extension](#) - by the Chicago Tribune for Refine 2.1

# OpenRefine Documentation

## [Home](#)

## [Download](#)

## [Documentation](#)

## [Community](#)

## [Post archive](#)

[OpenRefine News:](#)

[October 2015](#)

[OpenRefine News:](#)

[September 2015](#)

[OpenRefine News:](#)

[August 2015](#)

[OpenRefine News: July](#)

[2015](#)

[OpenRefine News: June](#)

[2015](#)

[Mapping OpenRefine](#)

[Ecosystem](#)

[2014 survey results](#)

[A Governance Model for](#)

[OpenRefine](#)

[Using OpenRefine: a](#)

[manual](#)

[OpenRefine History](#)

## Documentation for User

### Official Documentation

- [FAQ](#)
- [Official Documentation Wiki](#)
- [OpenRefine Mailing List](#)
- [File a Bug or Feature Request](#)

### Tutorials

- [Curated Tutorial List](#)

### Using OpenRefine - The Book



**Using OpenRefine** by Ruben Verborgh, Max De Wilde offers anyone a great introduction to OpenRefine. Organized by recipes with hands on examples, the book covers the following topics:

1. Import data in various formats
2. Explore datasets in a matter of seconds
3. Apply basic and advanced cell transformations
4. Deal with cells that contain multiple values
5. Create instantaneous links between datasets
6. Filter and partition your data easily with regular expressions
7. Use named-entity extraction on full-text fields to automatically identify topics
8. Perform advanced data operations with the General Refine Expression Language

## Documentation for Developers

- [Wiki for Developers](#)
- [Developer Discussion List](#)

Follow us on:

[Github](#)

[Twitter](#)

Google™ Custom Search

x

## OpenRefine Docs

[Home](#)

[Download](#)

[Documentation](#)

[Community](#)

[Post archive](#)

[OpenRefine News:  
October 2015](#)

[OpenRefine News:  
September 2015](#)

[OpenRefine News:  
August 2015](#)

[OpenRefine News: July  
2015](#)

[OpenRefine News: June  
2015](#)

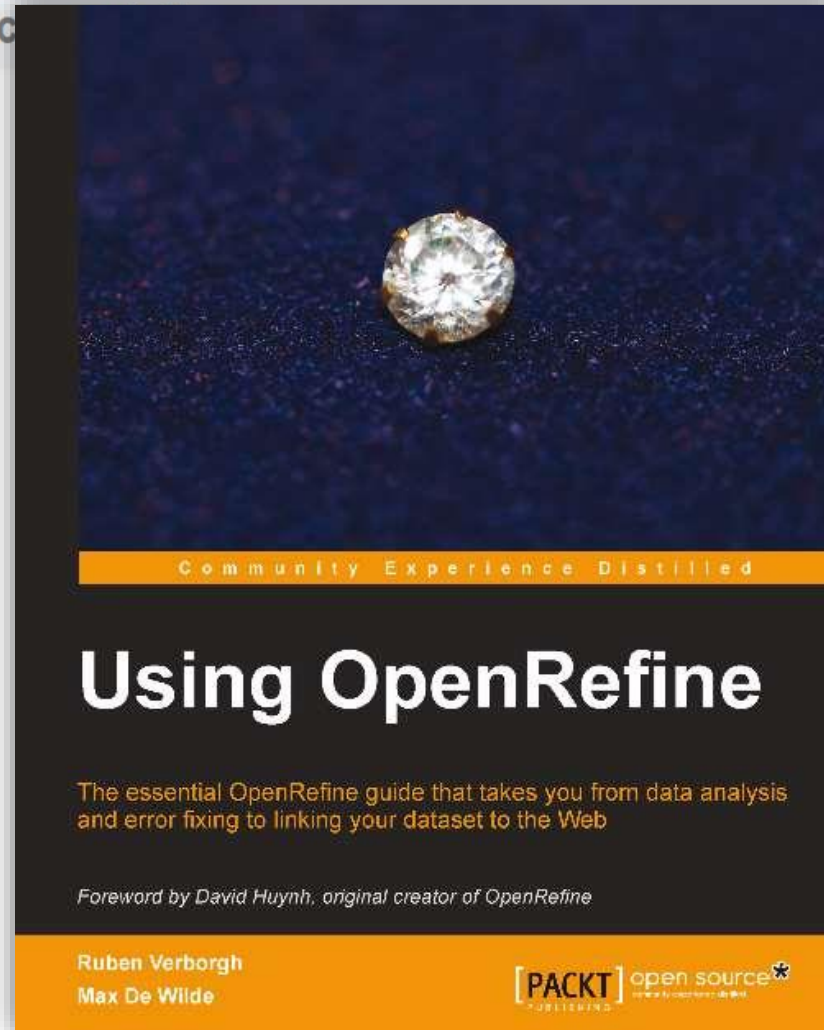
[Mapping OpenRefine  
Ecosystem](#)

[2014 survey results](#)

[A Governance Model for  
OpenRefine](#)

[Using OpenRefine: a  
manual](#)

[OpenRefine History](#)



## Documentation for Developers

- [Wiki for Developers](#)
- [Developer Discussion List](#)

anyone a great introduction to  
the book covers the following

# GREL: Enlaces de interés

GitHub OpenRefine

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

General Refine Expression Language

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

Extensiones

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

# GREL: Algunas funciones básicas

Textos entre “ ” o /~/

Add column based on column ScientificNameAuthor

New column name

On error  set to blank  store error  copy value from original column

Expression  Language  No syntax error.

Preview History Starred Help

row	value
	cells[\"Genus\"].value+\" \" +cells[\"Species\"].value+\" \" +cells[\"ScientificNameAuthor\"].value
1.	L. Thymus pulegioides L.
2.	L. Thymus pulegioides L.
3.	L. Thymus serpyllum L.
4.	L. Thymus pulegioides L.
5.	L. Thymus pulegioides L.
6.	L. Thymus pulegioides L.

OK Cancel

- Reemplazar valores en una columna

**value.replace("sentencia\_original","sentencia\_nueva")**

- Concatenar textos de columnas diferentes:

**cell["columnaX"].value + cell["columnaY"].value + cell["columnaZ"].value**

- Cambiar unidades, p.e. metros a kilómetros

**value/factor de conversión**