DATA ON
THE WEB

Capacitación

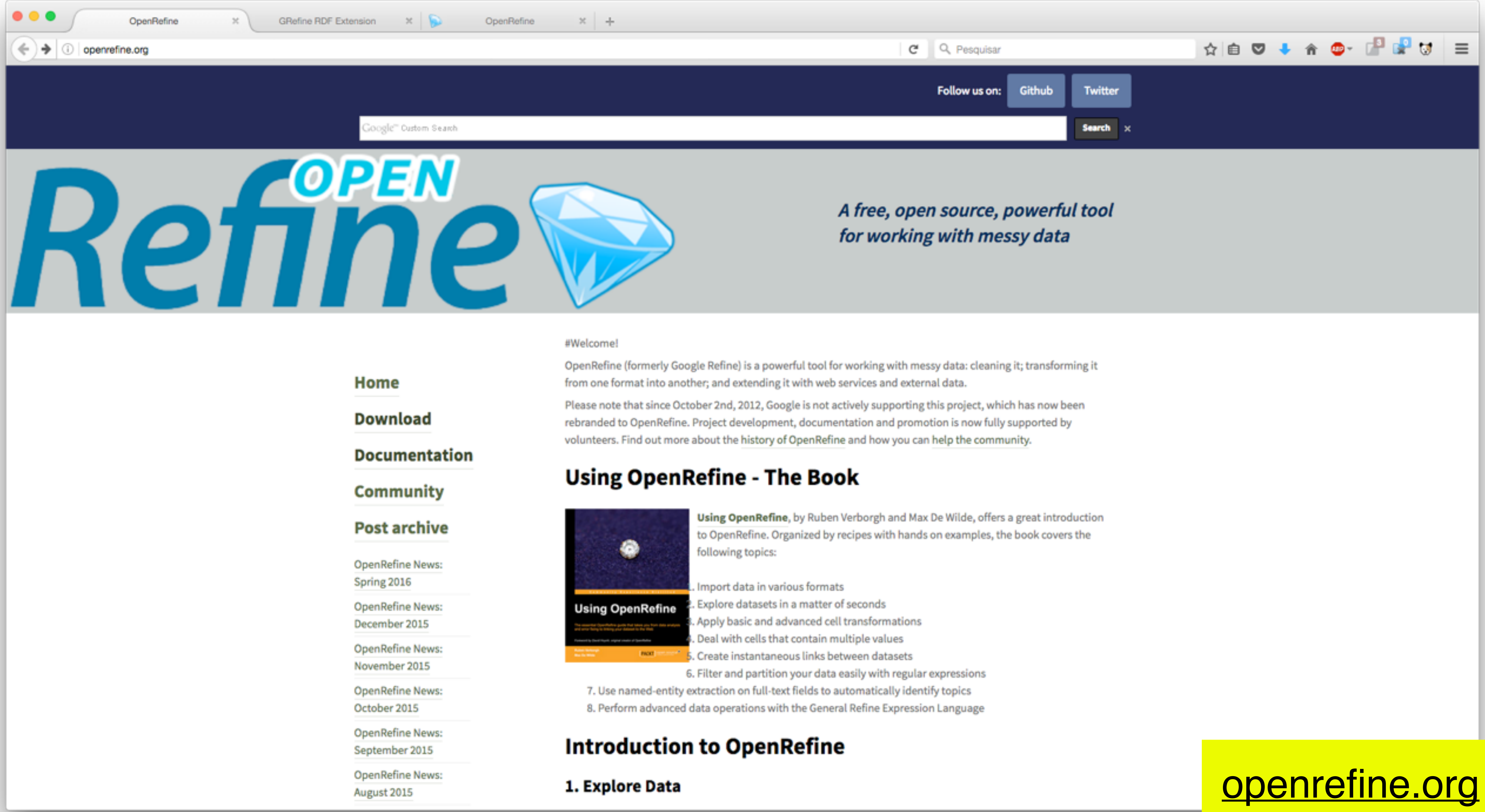
Mejores Prácticas de Datos en la Web

Mini curso: Open Refine

Newton Calegari
@newtoncalegari

ceweb.br

OPEN
Refine



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

OpenRefine News: Spring 2016

OpenRefine News: December 2015

OpenRefine News: November 2015

OpenRefine News: October 2015

OpenRefine News: September 2015

OpenRefine News: August 2015

#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

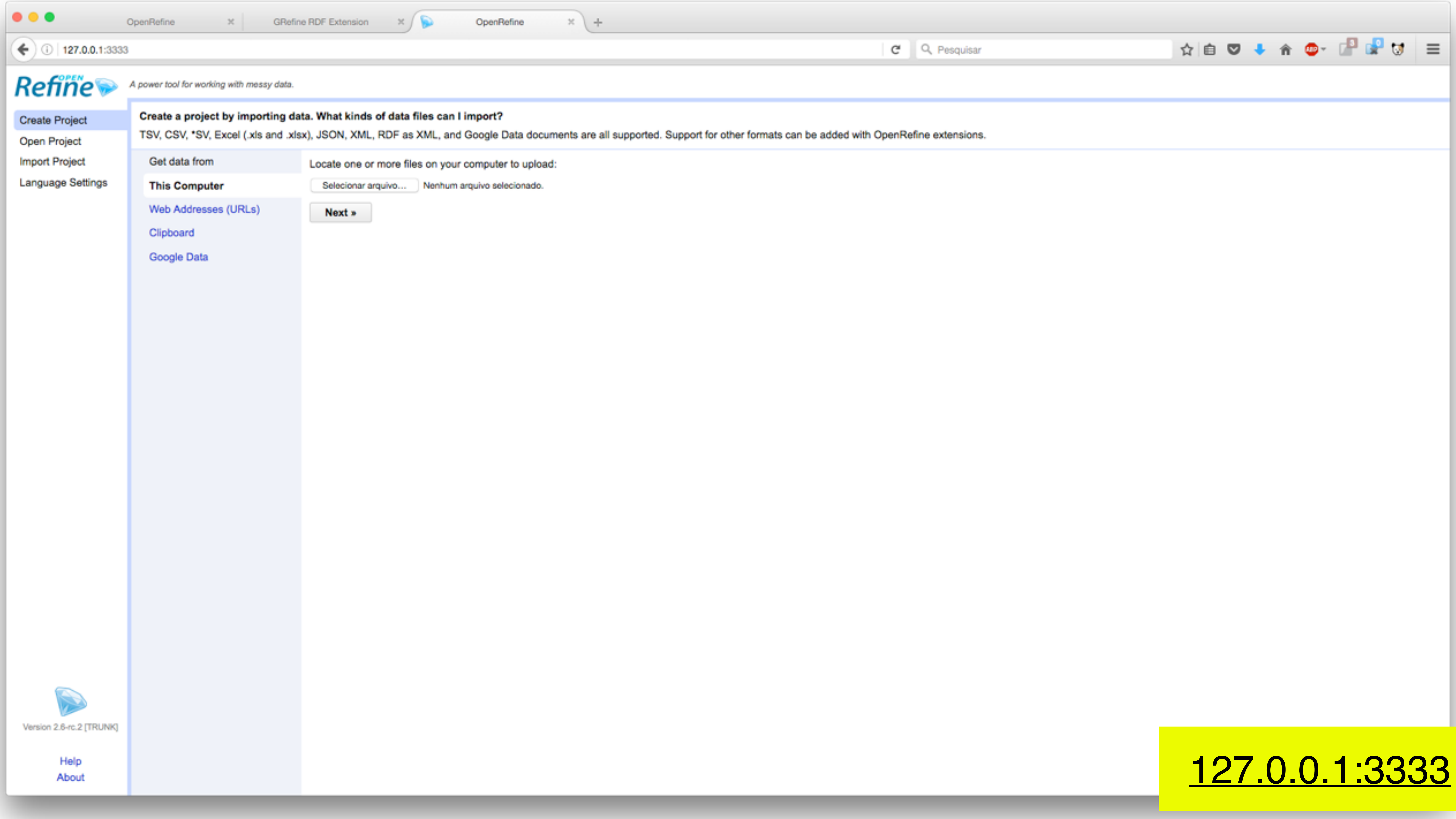
Using OpenRefine, by Ruben Verborgh and Max De Wilde, offers a great introduction to OpenRefine. Organized by recipes with hands on examples, the book covers the following topics:

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.org



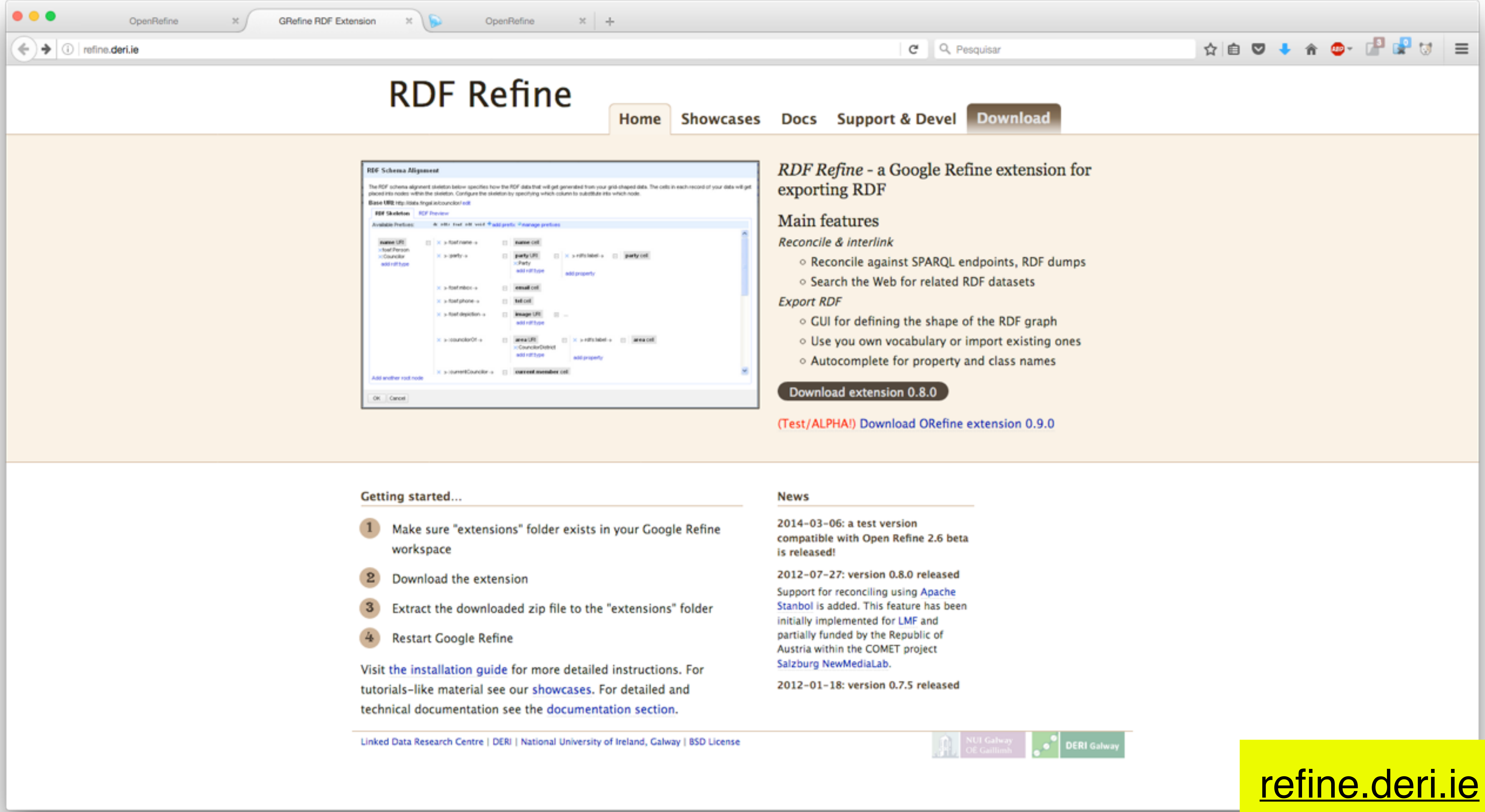
A power tool for working with messy data.

Create a project by importing data. What kinds of data files can I import?

TSV, CSV, *SV, Excel (.xls and .xlsx), JSON, XML, RDF as XML, and Google Data documents are all supported. Support for other formats can be added with OpenRefine extensions.

Locate one or more files on your computer to upload:

127.0.0.1:3333



RDF Refine

Home | Showcases | Docs | Support & Devel | Download

RDF Refine - a Google Refine extension for exporting RDF

Main features

Reconcile & interlink

- Reconcile against SPARQL endpoints, RDF dumps
- Search the Web for related RDF datasets

Export RDF

- GUI for defining the shape of the RDF graph
- Use you own vocabulary or import existing ones
- Autocomplete for property and class names

Download extension 0.8.0

(Test/ALPHA!) Download ORefine extension 0.9.0

Getting started...

1. Make sure "extensions" folder exists in your Google Refine workspace
2. Download the extension
3. Extract the downloaded zip file to the "extensions" folder
4. Restart Google Refine

Visit the installation guide for more detailed instructions. For tutorials-like material see our showcases. For detailed and technical documentation see the documentation section.

News

2014-03-06: a test version compatible with Open Refine 2.6 beta is released!

2012-07-27: version 0.8.0 released

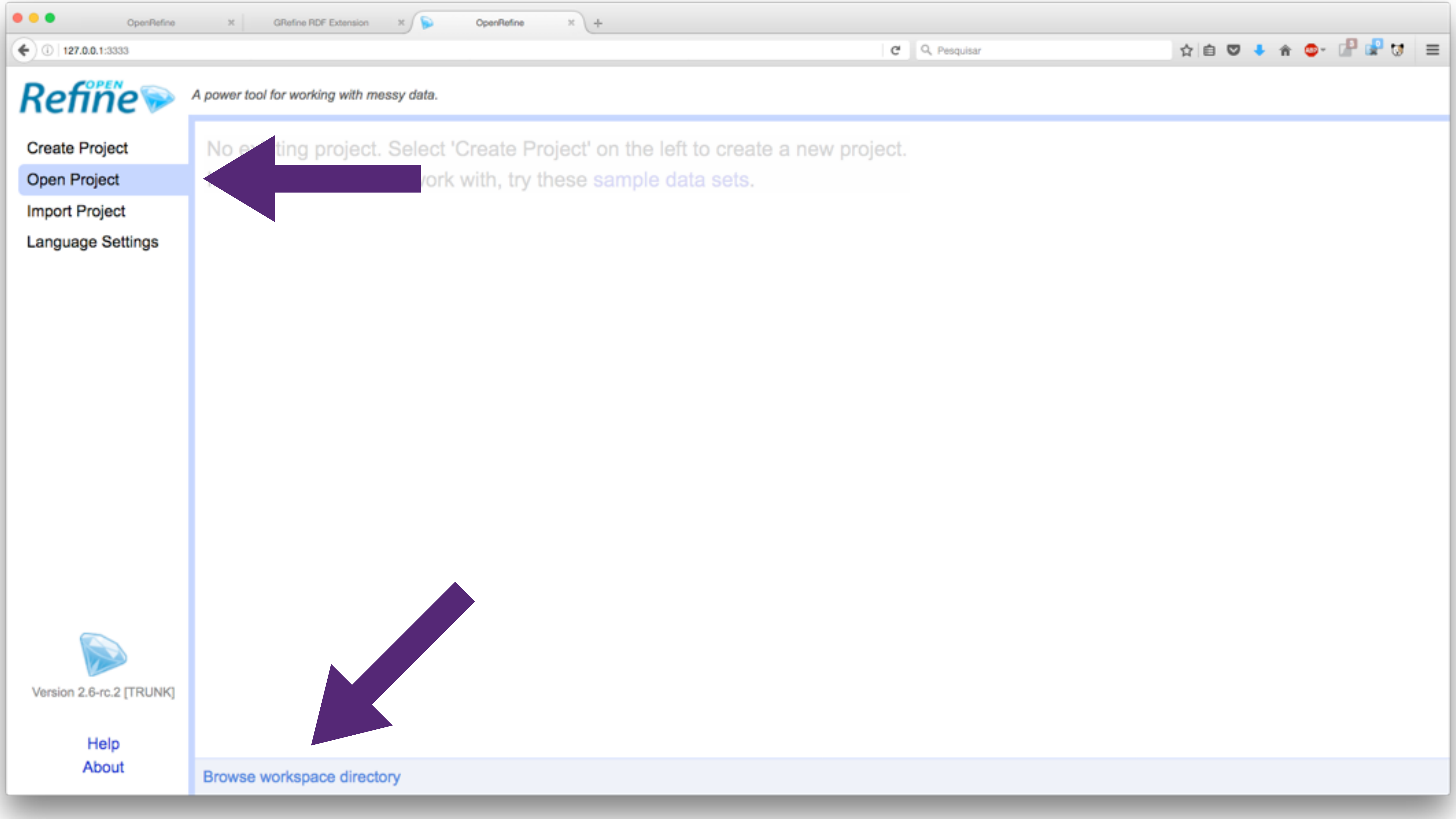
Support for reconciling using Apache Stanbol is added. This feature has been initially implemented for LMF and partially funded by the Republic of Austria within the COMET project Salzburg NewMediaLab.

2012-01-18: version 0.7.5 released

Linked Data Research Centre | DERI | National University of Ireland, Galway | BSD License

refine.deri.ie

Installing the RDF Extension



A power tool for working with messy data.

No e ting project. Select 'Create Project' on the left to create a new project.

ork with, try these sample data sets.

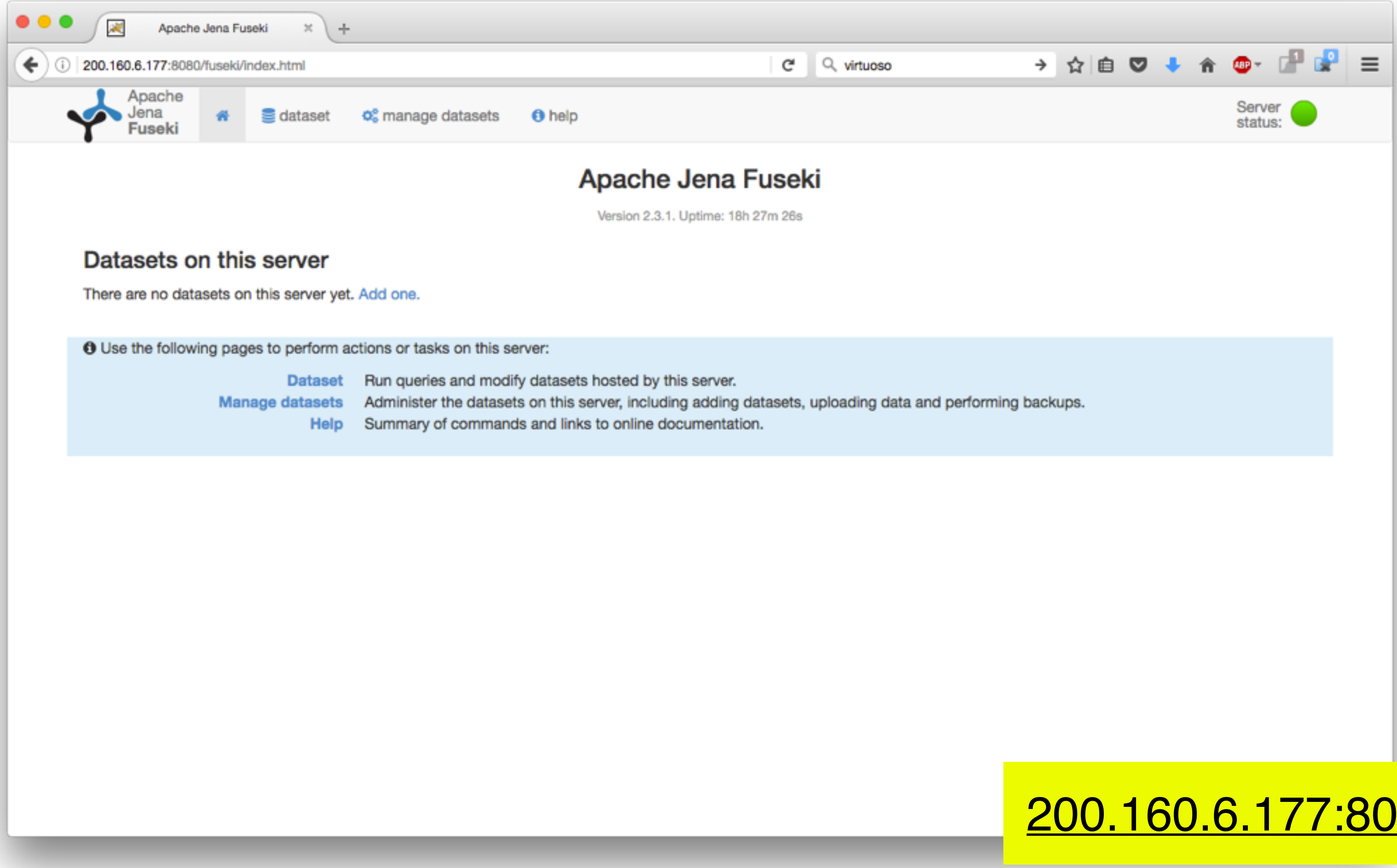
Top 100 books - CSV

newtoncalegari.com.br/dwbp-costa-rica/

Mapping vocabulary to data

title	schema:name
author	schema:author
year	schema:datePublished
country	schema:locationCreated
language	schema:inLanguage
pages	schema:numberOfPages

Uploading to RDF Store



Apache Jena Fuseki

Version 2.3.1. Uptime: 18h 27m 26s

Datasets on this server

There are no datasets on this server yet. Add one.

Use the following pages to perform actions or tasks on this server:

Dataset	Run queries and modify datasets hosted by this server.
Manage datasets	Administer the datasets on this server, including adding datasets, uploading data and performing backups.
Help	Summary of commands and links to online documentation.

200.160.6.177:8080/fuseki/

SPARQL - basics

```
SELECT ?s  ?p  ?o
WHERE {
  ?s  ?p  ?o
}
```

SPARQL - basics

```
PREFIX schema: <http://schema.org/>
SELECT ?title ?author ?country
WHERE {
  ?book schema:author      ?author .
  ?book schema:name        ?title  .
  ?book schema:locationCreated ?country .
  FILTER(?country = 'Brazil')
}
```