

Working with Data using OpenRefine

Owen Stephens

Schedule

- Introduction to OpenRefine (slides)
- Basic OpenRefine functions I (demonstrations and follow-along)
- Basic OpenRefine functions II (demonstrations and follow-along)
- Advanced OpenRefine functions (demonstrations and follow-along)

“a tool for working with
messy data”

<http://openrefine.org>

OpenRefine can help when...

- you have data in a simple tabular format
- there are inconsistencies in how the data is formatted
- there are inconsistencies in where data appears
- there are inconsistencies in terminology used in the data

OpenRefine can help you...

- Get an overview of a data set
- Resolve inconsistencies in a data set
- Help you split data up into more granular parts
- Match local data up to other data sets
- Enhance a data set with data from other sources

For example...

Data you have	Desired data
1st January 2014	2014-01-01
01/01/2014	2014-01-01
2014-01-01	2014-01-01
Jan 1 2014	2014-01-01

For example...

Data you have	Desired data
London	London
London]	London
London,]	London
london	London

For example...

Data you have	Desired data							
	Institution	Library name	Address 1	Address 2	Town/City	Region	Country	Postcode
University of Wales, Llyfrgell Thomas Parry Library, Llanbadarn Fawr, ABERYSTWYTH, Ceredigion, SY23 3AS, United Kingdom	University of Wales	Llyfrgell Thomas Parry Library	Llanbadarn Fawr		Aberystwyth	Ceredigion	United Kingdom	SY23 3AS
University of Aberdeen, Queen Mother Library, Meston Walk, ABERDEEN, AB24 3UE, United Kingdom	University of Aberdeen	Queen Mother Library	Meston Walk		Aberdeen		United Kingdom	AB24 3UE
University of Birmingham, Barnes Library, Medical School, Edgbaston, BIRMINGHAM, West Midlands, B15 2TT, United Kingdom	University of Birmingham	Barnes Library	Medical School	Edgbaston	Birmingham	West Midlands	United Kingdom	B15 2TT
University of Warwick, Library, Gibbett Hill Road, COVENTRY, CV4 7AL, United Kingdom	University of Warwick	Library	Gibbett Hill Road		Coventry		United Kingdom	CV4 7AL

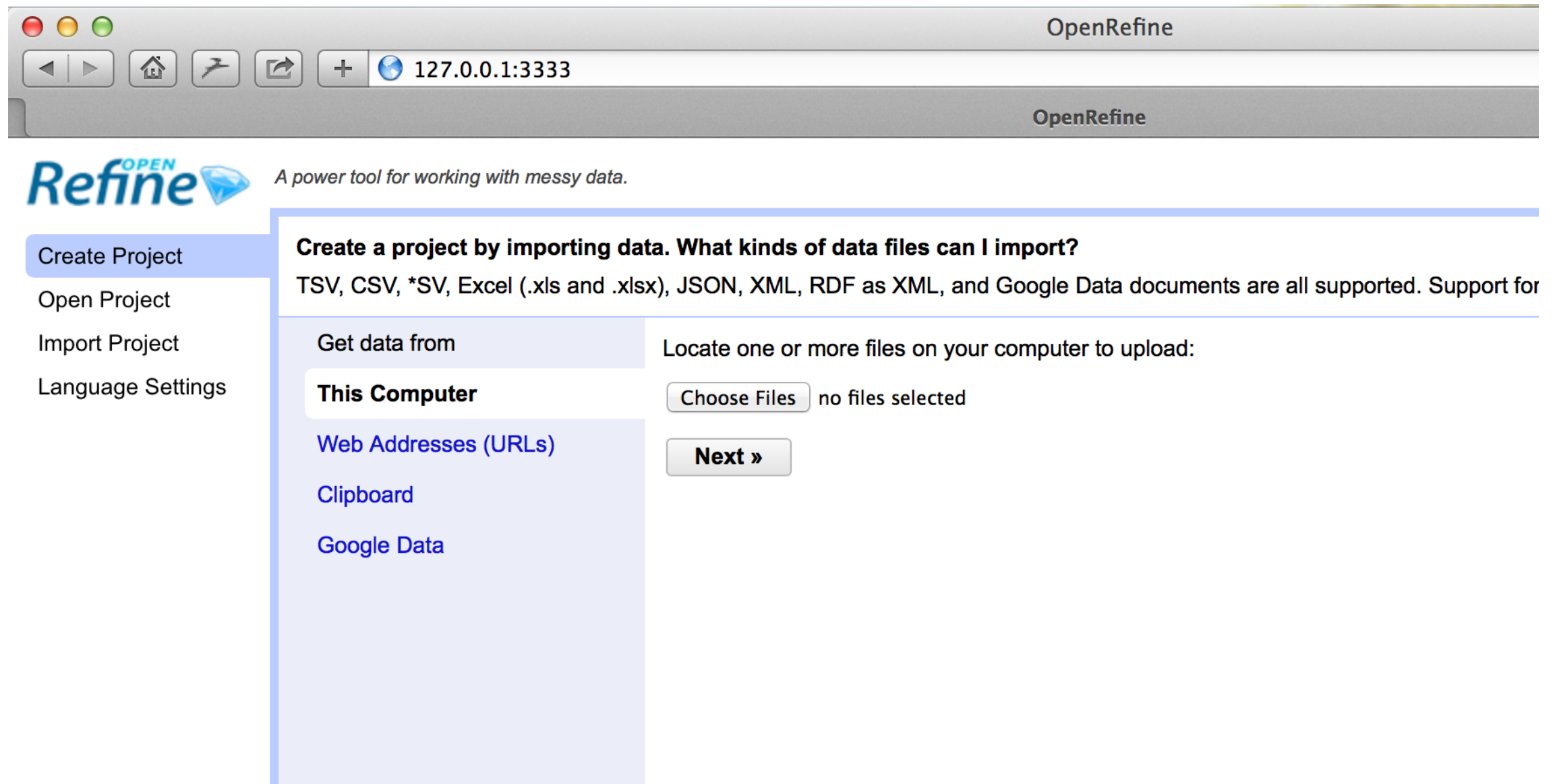
For example...

Data you have	Date of Birth from VIAF (Virtual International Authority File)	Date of Death from VIAF (Virtual International Authority File)
Braddon, M. E. (Mary Elizabeth)	1835	1915
Rossetti, William Michael	1829	1919
Prest, Thomas Peckett	1810	1879

Getting help

- The OpenRefine Wiki <https://github.com/OpenRefine/OpenRefine/wiki>
- The 'Free your metadata' site <http://freeyourmetadata.org/> and book <http://book.freeyourmetadata.org>
- The OpenRefine mailing list and forum <http://groups.google.com/d/forum/openrefine>

Start using OpenRefine



The screenshot shows a web browser window titled "OpenRefine". The address bar displays "127.0.0.1:3333". The page header features the "Refine" logo with "OPEN" in small blue letters above it, followed by the tagline "A power tool for working with messy data.".

On the left side, there is a vertical navigation menu with the following items: "Create Project" (highlighted with a blue background), "Open Project", "Import Project", and "Language Settings".

The main content area is titled "Create a project by importing data. What kinds of data files can I import?". Below this title, it states: "TSV, CSV, *SV, Excel (.xls and .xlsx), JSON, XML, RDF as XML, and Google Data documents are all supported. Support for".

Below the text, there is a section titled "Get data from" which contains a list of options: "This Computer" (highlighted with a light blue background), "Web Addresses (URLs)", "Clipboard", and "Google Data".

To the right of the "Get data from" section, there is a text prompt: "Locate one or more files on your computer to upload:". Below this prompt, there is a button labeled "Choose Files" followed by the text "no files selected". At the bottom of this section, there is a button labeled "Next »".

Demonstrations and follow-along

Download file from: <http://bit.ly/lc-wk4-data>

Using these slides

These slides were developed by Owen Stephens
(owen@ostephens.com) on
behalf of the British Library.

Unless otherwise stated, all images, audio or video content are
separate works with their own licence, and should not be
assumed to be CC-BY in their own right

This work is licensed under a Creative Commons Attribution 4.0
International License <http://creativecommons.org/licenses/by/4.0/>.

It is suggested when crediting this work, you include the phrase
“Developed by Owen Stephens on behalf of the British Library”