

Introduction to Open Data: Tools and Techniques for Data Sharing and Collaboration

OpenRefine: The data cleaning and Transformation tool that will make you data sing!

**WIKIMANIA
SINGAPORE**

Joris Darlington Quarshie



Open Data: Tools and Techniques for Data Sharing and Collaboration

OpenRefine: The data cleaning and transformation tool that will make your data sing!

User: Joris Darlington Quarshie

Community Lead & Data Scientist

African Wikimedia Technical Community

Email: joris.quarshie@gmail.com

WIKIMANIA
SINGAPORE



Introduction

**WIKIMANIA
SINGAPORE**



What is OpenRefine?

OpenRefine is a free and open-source data cleaning and transformation tool. It can be used to clean messy data, merge datasets and generate new insights

Why OpenRefine?

- OpenRefine is a powerful tool for cleaning messy data, merging datasets and generating new insights.
- It is easy to use and it is available for free.
- It is a community-driven project, which means that there is a large and active community of users who are willing to help each other.

Who uses OpenRefine

- Data Scientist
- Researchers
- Journalists
- Students
- Data Enthusiasts
- Librarians & Information Specialists
- Governments and Public Sector workers



**WIKIMANIA
SINGAPORE**

Downloading OpenRefine

Downloading OpenRefine

- You can download openrefine from the openrefine website:
<https://openrefine.org/>
- OpenRefine is available for Windows, macOS and Linux.

**WIKIMANIA
SINGAPORE**





**WIKIMANIA
SINGAPORE**

Installing

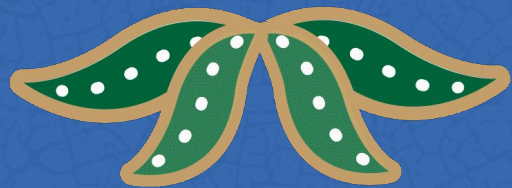
OpenRefine

Installing OpenRefine

- Once you have downloaded OpenRefine, you can install it by following the instructions on the openrefine website:
<https://openrefine.org/>
- Openrefine is a relatively easy tool to install

**WIKIMANIA
SINGAPORE**





**WIKIMANIA
SINGAPORE**

Running OpenRefine

Running OpenRefine

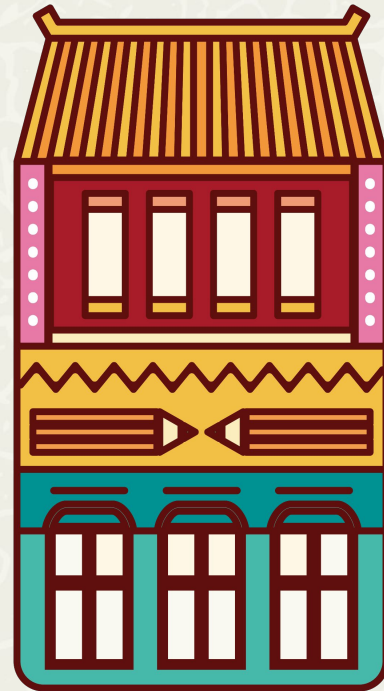
- Once you have installed the openrefine you can run it by opening the openrefine application.
- Openrefine will open a new window.



Starting a Project

OpenRefine

**WIKIMANIA
SINGAPORE**



Starting a Project

- To start a new project in openrefine, click on the “create project” button.
- You will be prompted to select a data source.
- You can select a data source from your computer or url.

**WIKIMANIA
SINGAPORE**



Exploring Data

**WIKIMANIA
SINGAPORE**



Exploring Data

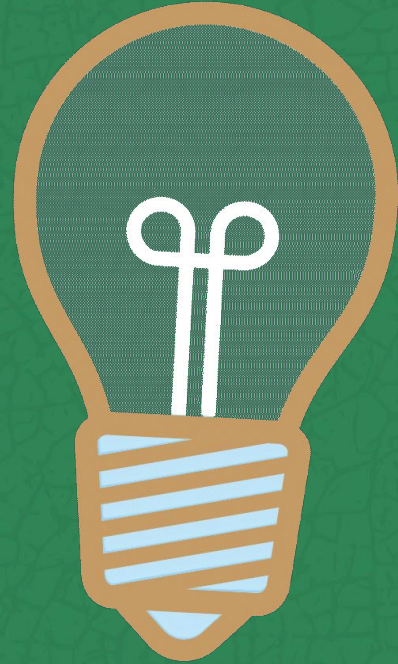
- Once you have imported a data source into openrefine, you can explore the data by using the “explore data” tab.
- The explore tab provides a number of tools for exploring your data.



Transforming Data

OpenRefine

**WIKIMANIA
SINGAPORE**



Transforming Data

- You can download OpenRefine from the OpenRefine website.
<https://openrefine.org/>
- OpenRefine is available on Windows, macOS and Linux

**WIKIMANIA
SINGAPORE**



Reconciling

OpenRefine

**WIKIMANIA
SINGAPORE**



Reconciling

- If you have multiple data sources that contain the same data, you can reconcile them by using the “Reconcile” tab.
- The reconcile tab provides a number of tools for merging and aligning your data sources.



Video-walk through

OpenRefine

WIKIMANIA
SINGAPORE



Wikibase, Wikidata & Wikimedia Commons

OpenRefine

**WIKIMANIA
SINGAPORE**



Wikibase, Wikidata & Wikimedia Commons

- OpenRefine can be used to connect to wikibase, wikidata & wikimedia commons.
- This allows you to use the data in your openrefine project to enrich and improve other wikimedia projects.

**WIKIMANIA
SINGAPORE**



Expressions

OpenRefine

**WIKIMANIA
SINGAPORE**



Expressions

- OpenRefine allows you to use expressions to calculate new values from your data.
- Expressions can be used to calculate simple values such as, the sum of a column of numbers, or they can be used to calculate more complex values, such as a the average of a column of numbers with specific conditions.





**WIKIMANIA
SINGAPORE**

Exporting Data

OpenRefine

Exploring Data

- Once you have finished transforming your data, you can export it from OpenRefine.
- You can export your data in variety of formats, including csv files, XML files and Json files.

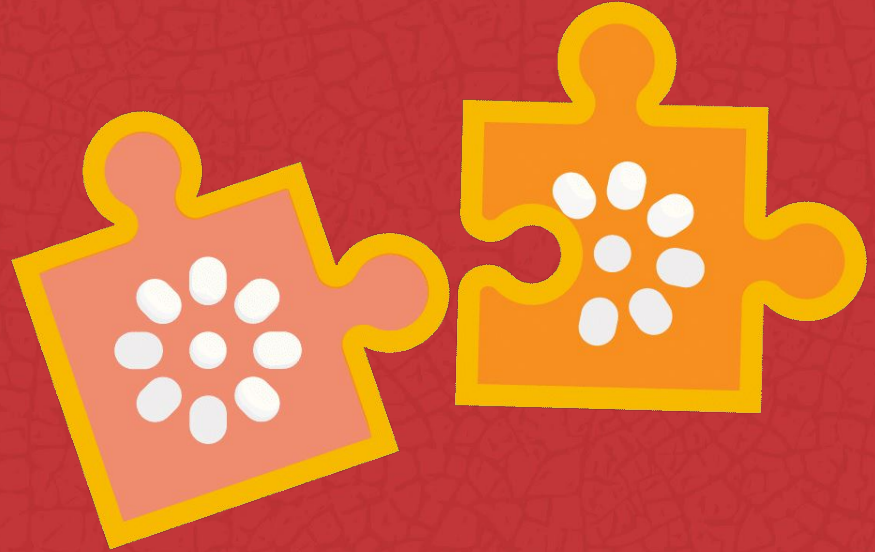
**WIKIMANIA
SINGAPORE**



Troubleshooting

OpenRefine

**WIKIMANIA
SINGAPORE**





**WIKIMANIA
SINGAPORE**

If you encounters any problems while using OpenRefine, you can refer to the OpenRefine documentation:

<https://openrefine.org/docs/> or the OpenRefine forum:

<https://forum.openrefine.org/> or telegram:

<https://t.me/+fRTPbCDKpKwyY2Jk>

Github:

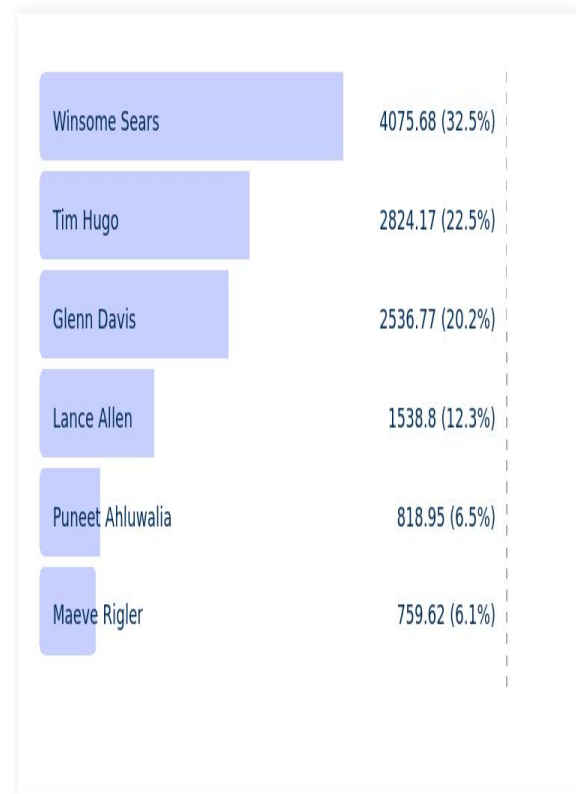
<https://github.com/OpenRefine/OpenRefine/wiki/External-Resources>

OpenRefine

Visualize Data

- Once you have cleaned and transformed your data, you can visualize it using wedodatascience.com:
<https://wedodatascience.com/>
- wedodatascience.com provides a number of tools for visualizing data including charts, graphs and maps

Round 1

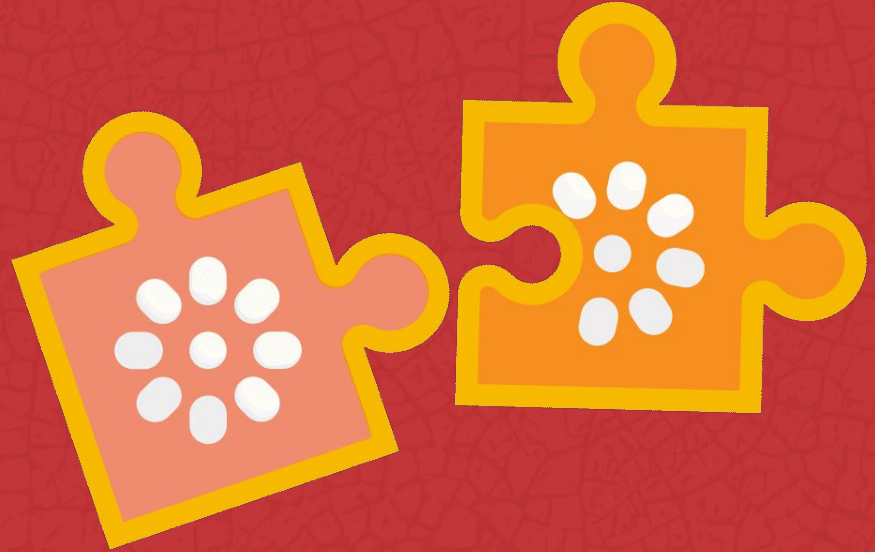


Ranked-Choice Voting results for
Virginia GOP Convention, Lieutenant Governor Nominee

For more tools:
[wikidata:tools](#)

OpenRefine

**WIKIMANIA
SINGAPORE**





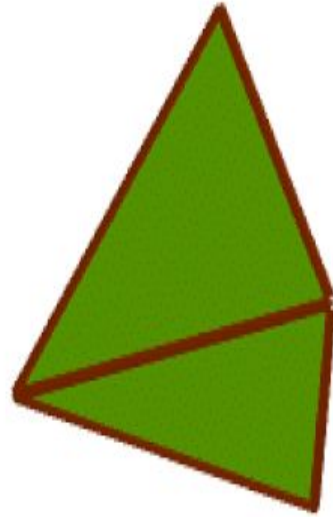
**WIKIMANIA
SINGAPORE**

Conclusion

OpenRefine

Conclusion

- OpenRefine is a powerful tool for cleaning messy data, merging datasets and generating new insights.
- It is easy to use and it is available for free.
- If you are working with data, i encourage you to check out Openrefine.



Questions ?

OpenRefine

**WIKIMANIA
SINGAPORE**





**WIKIMANIA
SINGAPORE**

