

# OpenRefine + Wikidata = Magic

A tutorial

01

# What is OpenRefine?

And why do I need it in my life?

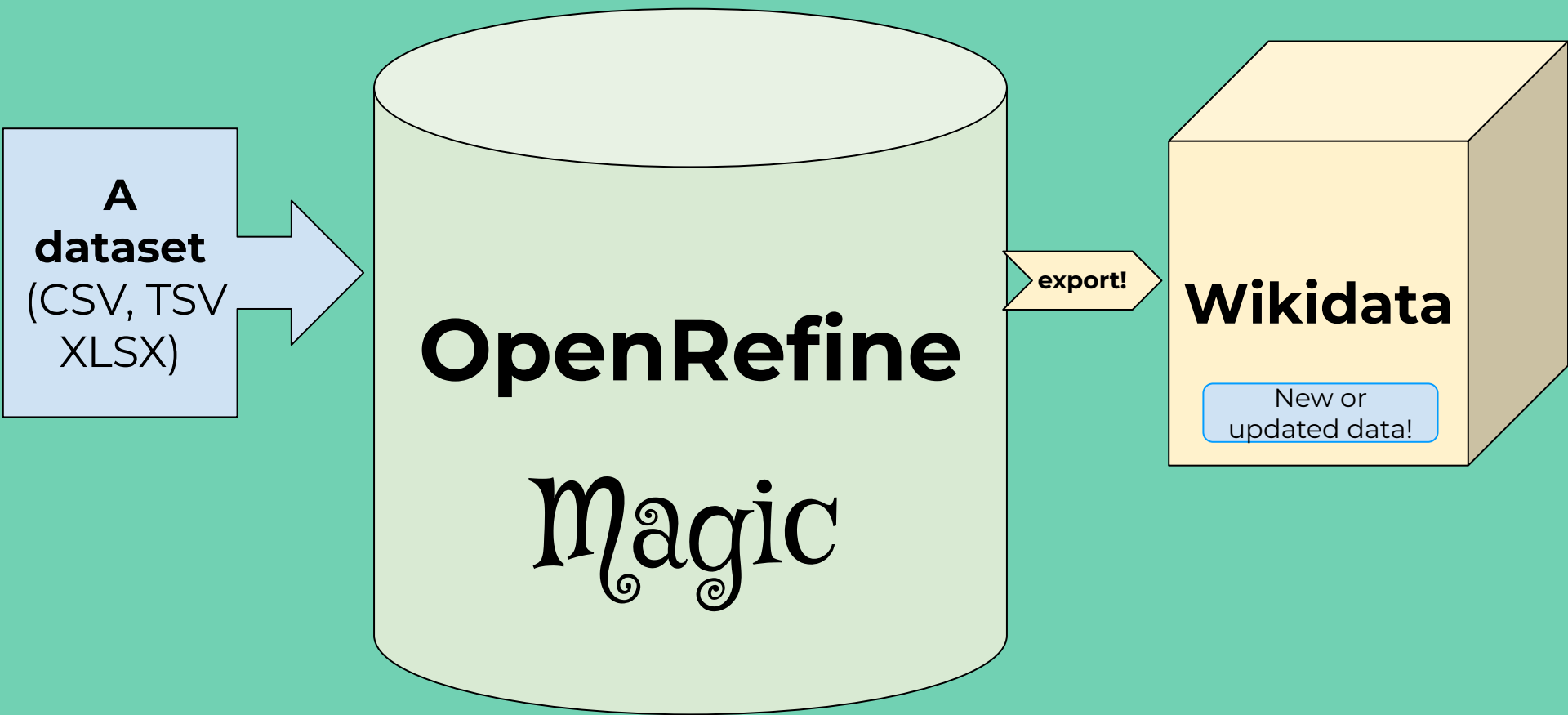
- [OpenRefine](#) is a powerful downloadable tool for working with (even messy!) **tabular data**
- It can be used to **explore** and **transform** data
- It can **match** ("reconcile") data with data from **Wikidata**
- It can **update Wikidata** with **new** or existing items
- It can **upload files to Wikimedia Commons**, and update their attributes

# 02

# Our overall plan

*"First we take Manhattan; then we take Berlin!"*

# Overall plan



# Plan zoom-in

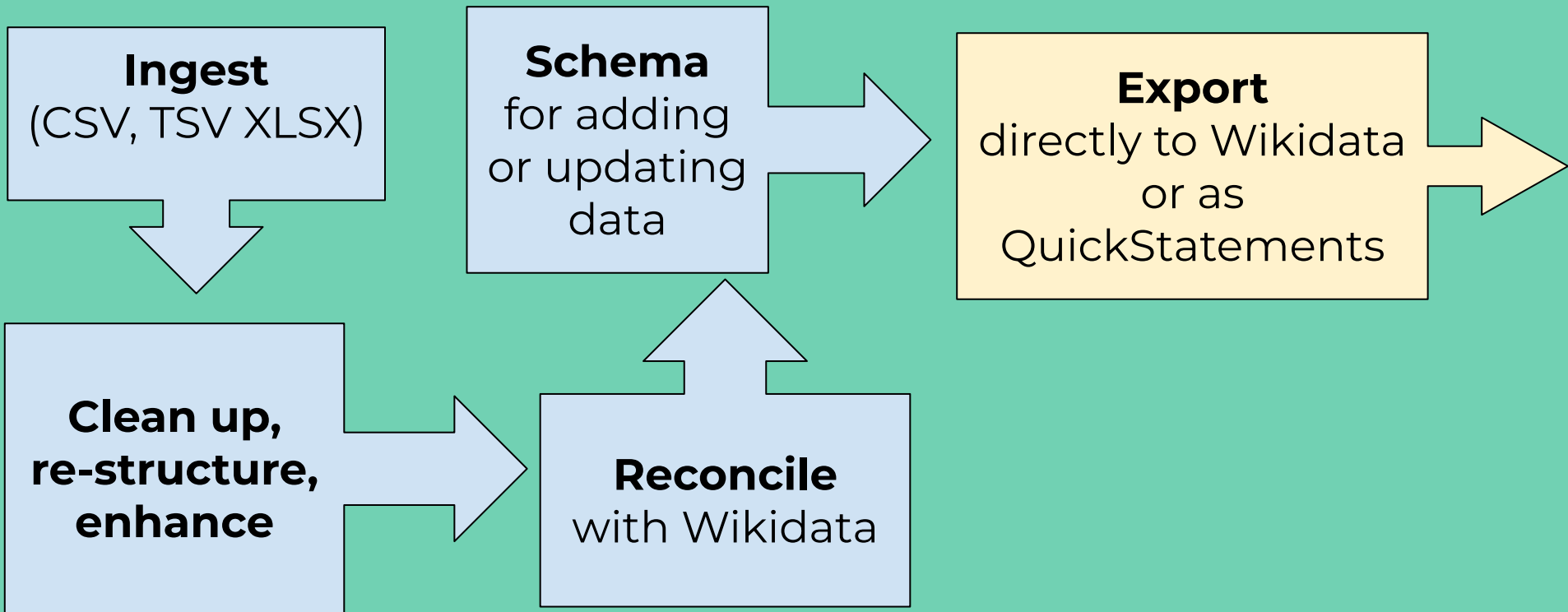
**Ingest**  
(CSV, TSV XLSX)

**Schema**  
for adding  
or updating  
data

**Export**  
directly to Wikidata  
or as  
QuickStatements

**Clean up,  
re-structure,  
enhance**

**Reconcile**  
with Wikidata



# 03

# Ingesting data

**Ingesting** is just a fancy word for "getting something into something"  
Can ingest from CSV, TSV, XLSX, URL, clipboard...

# 04 Exploring and cleaning up data

Get data into the best shape you can *before* feeding it into Wikidata!



# 04a Rows, records, columns...

And columns from columns!

# 04b Facets, filters, and thanks for all the fish!

Your data may surprise you!  
Filter by value/range, sort; save permalinks;

# 04c Editing and transforming

Common transforms, directly editing, splitting

# 04d Undo / Redo

Be confident you can always go back!

# 04e Getting fancy with GREL

One of the most powerful tools you can learn is **regular expressions!**

# 05 Reconciling with Wikidata

# Reconciling?

- Reconciling is just a fancy word for **matching**, or **aligning**
- A reconciled column is one where the values in the cells have been **matched** (or not...) to an external source (Ex: **Wikidata items** (Qnnn))
- You can **help** OpenRefine when it gets it wrong.
- Reconciling is necessary for identifying **which** Wikidata item to update, and for feeding a correct value into Wikidata **when the value is itself a Wikidata item** (and not a string or number)

# Reconciling tips

- It is worth **investigating** why reconciliation is harder than you expected. Often it is a case of a missing label or alias.
- You can **specify the instance-of** for the column you are reconciling
- You can **use values from other columns** to dramatically help the reconciliation.



# 06 Scheming to update Wikidata

Creating the update schema without tears.

# The Schema

- A **Wikibase schema** is a **template** for OpenRefine, telling it which **items** to update, and what **statements** to change or add there, with what **references**.
- It is **very important** to get our schema right *before* updating a large set of data. Best to experiment with one or two items first.
- OpenRefine won't create **exact duplicates**, so it is safe to re-run an upload, and only what hasn't been uploaded yet should get updated.

# 07 Exporting to Wikidata or QuickStatements

If you love your data, set it free!

# 08 Resources

# Resources

- The [Fine Manual](#) is actually really clear and concise!
- [OpenRefine information page](#) on Wikidata
  - [A video tutorial](#) by the main developer of OpenRefine (linked from information page above)
- [OpenRefine information page](#) on Commons
  - [A WikiLearn course on batch-uploading files to Commons, or batch-updating file metadata](#). (not demonstrated today) (also in [FR](#), [ES](#), [IT](#), [PT](#)!)

Which brings me to the final, ultimate secret...

# 09 The `~={secret}=~` to getting help

# The ~={secret}=~

- The ~={secret}=~ to getting the help you need is...
  - ...**asking** for help
  - ...but **doing it right!**
- Asking for help the right way:
  - In the right place (+ [don't ask to ask; just ask!](#))
  - With enough **context, links**, specifics
  - Responding to follow-ups and suggestions

Please please please don't be shy!

# Remember:

With great power comes great responsibility;  
use your power for good and not for evil.

Keep in touch!

Asaf Bartov  
**[asaf@wikimedia.org](mailto:asaf@wikimedia.org)**