Using Wikidata to Capture and Share Information about People in Paleontology

Introduction to efficient/bulk editing in wikidata
30 March 2022

Erica Krimmel
What to do with stuff like this?

Agent Summary for Joseph Grinnell (person)

- Joseph (first name) (search)
- Grinnell (last name) (search)
- Grinnell J. (aka) (search)
- Grinnell, Joseph (aka) (search)
- J. Grinnell (aka) (search)

Remarks about Joseph Grinnell

Museum of Vertebrate Zoology’s founding director in 1908.

Relationships From Joseph Grinnell

- Spouse of Hilda W. Grinnell
- Employee of Museum of Vertebrate Zoology, University of California, Berkeley, 1908
- Grandparent of June R. Grinnell

Relationships To Joseph Grinnell

- From Willard F. Grinnell - child of Hilda W. Grinnell

I was one of those kids who never grew out of my "Dinosaur phase." As children we often show, and then lose, interest in numerous subjects. Dinosaurs are a particularly common interest for young people. For me, I never lost my interest.

As I grew up, I began to show more interest in how the dinosaurs may have behaved. This led me to spend thousands of hours studying the behavior of modern animals so as to understand how prehistoric creatures may have behaved.

1908 - My senior year in high school found me dreaming of becoming a paleontologist. One of my teachers, whom I respected deeply, told me that I should focus on a "real" job. I was devastated. I began to wonder if paleontology was a real job. Perhaps the people I read about were just doing this as a hobby? Young and impressionable I decided to change my career path, I started a career in real and quickly advanced into upper level management positions. I became a director in a major retail chain and oversaw the operations of 14 stores with nearly 200 million dollars in annual sales. But through it all, my love of paleontology continued. I spent my summer vacations traveling throughout the United States, excavating fossils and expanding an extensive collection of fossils and replicas.

Fast forward to 1987. A knock on my front door would change every aspect of my life. I was working in my office when I heard a knock on my door. A young couple was having car trouble and asked me to give them a jump start. The female happened to be a science teacher in an inner city elementary school. She looked in my house and was amazed at my collection. She told me that it was very hard to find people who would come in and speak to her students. She asked me if I could gather some items and come to her school to speak about what I do. It sounded like fun, so I agreed.
What to do with stuff like this...?
Options for getting data into Wikidata

Edit directly in the Wikidata user interface
https://www.wikidata.org

- Get a sense for how Wikidata organizes information
- Easy access to full search results
- Low learning curve

Use QuickStatements
https://www.wikidata.org/wiki/Help:QuickStatements

- Upload a spreadsheet
- QS will check your edits for duplicate statements
- Audit tracking for every edit action

Use Wikidata plugin with OpenRefine
https://www.wikidata.org/wiki/Wikidata:Tools/OpenRefine

- Plugin comes with OpenRefine by default in version 3.0 or later
- One visual interface for editing and pushing changes

Directly via bots scripted in Python, e.g.:

PyWikiBot
https://pypi.org/project/pywikibot/

Wikidataintegrator
https://github.com/SuLab/WikidataIntegrator
Options for getting data into Wikidata

- **Edit directly in the Wikidata user interface**
  
  - [https://www.wikidata.org/](https://www.wikidata.org/)
  
  - Get a sense for how Wikidata organizes information
  
  - Easy access to full search results
  
  - Low learning curve

- **Use QuickStatements**
  
  
  - Upload a spreadsheet
  
  - QS will check your edits for duplicate statements
  
  - Audit tracking for every edit action

- **Use Wikidata plugin with OpenRefine**
  
  
  - Plugin comes with OpenRefine by default in version 3.0 or later
  
  - One visual interface for editing and pushing changes

- **Directly via bots scripted in Python, e.g.:**
  
  - PyWikiBot
    
    - [https://pypi.org/project/pywikibot/](https://pypi.org/project/pywikibot/)
  
  - Wikidataintegrator
    
    - [https://github.com/SuLab/WikidataIntegrator](https://github.com/SuLab/WikidataIntegrator)
Get started with bulk edits in Wikidata

You will need...

1. An autoconfirmed wikidata user account.
   a. This status will automatically be applied when your account is > 4 days old and has >50 edits

2. Familiarity with the kind of editing you want to do.
   a. What wikidata entities (e.g. institutions) have relationships with the entities (e.g. persons) in your data?
   b. What wikidata properties are relevant to your data (e.g. employer - P108)?
Get started with OpenRefine

You can…

2. Learn more about OpenRefine via the user manual at https://docs.openrefine.org/manual.
3. See several OpenRefine demos in our January 2021 Happy Hour.
4. Connect your wikidata account to your OpenRefine instance.
Visualize how the data are structured

<table>
<thead>
<tr>
<th>Charles Schuchert</th>
<th>paleontologist</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q2623977</td>
<td>Q1662561</td>
</tr>
</tbody>
</table>
Understand the process between wikidata and OpenRefine

1. Import your data into OpenRefine
2. Fetch additional data from wikidata into OpenRefine
3. Reconcile your data to entities in wikidata
4. Edit and augment data in OpenRefine
5. Push data from OpenRefine back to wikidata
Import your data into OpenRefine

Tabular data, e.g. a CSV file or Excel spreadsheet
Reconcile your data to entities in wikidata

You may want to…

★ restrict the reconciliation to a wikidata class

★ use additional column(s) in your dataset to refine the reconciliation by acting as a tiebreaker

★ use external identifiers

[column name] → Reconcile → Start reconciling
Reconcile your data to entities in wikidata

You may want to…

★ Hover over reconciliation options to see more details
★ Do your own search in wikidata
★ Follow leads to confirm the identity of an entity
Fetch additional data from wikidata into OpenRefine

Add columns in OpenRefine based on what else wikidata knows about your entities.

Multiple statements for a given property will be grouped as records in OpenRefine

[column name] → Edit column → Add columns from reconciled values
Edit and augment data in OpenRefine

Wrangle data from your own sources into the columns that you want to send to Wikidata.

[column name] → Edit column → Add column based on this column

Your original data

Your data connected to a Wikidata entity by assigning the appropriate Q number
Edit and augment data in OpenRefine

Reconcile the values in your data to entities in wikidata.

[column name] → Reconcile → Use values as identifiers

<table>
<thead>
<tr>
<th>All</th>
<th>family</th>
<th>employer</th>
<th>given</th>
<th>affiliation</th>
<th>P108</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.</td>
<td>Alice Mossie Brues</td>
<td>Harvard University</td>
<td>Alice Mossie</td>
<td>University of Colorado at Boulder</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Choose new match</td>
<td>Choose new match</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>University of Colorado Boulder</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Choose new match</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>University of Oklahoma</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Choose new match</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>Robert Kenneth Godfrey</td>
<td>Robert Kenneth</td>
<td>Florida State University</td>
<td>Q861548</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Choose new match</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>Roger W. Portell</td>
<td>Roger</td>
<td>Florida Museum of Natural History</td>
<td>Q3074272</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Choose new match</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Edit and augment data in OpenRefine

Translate your tabular data (i.e., columns & rows) into a wikidata “schema.”

You are currently working against Wikidata. You should have your data reconciled to reconciliation services linked to Wikidata (such as https://wikidata.reconci.link/en/api) first. The schema below specifies how your tabular data will be transformed into Wikidata edits. You can drag and drop the column names below in most input boxes: for each row, edits will be generated with the values in these columns.

<table>
<thead>
<tr>
<th>family</th>
<th>employer</th>
<th>given</th>
<th>affiliation</th>
<th>P108</th>
</tr>
</thead>
</table>

**Terms**
- no labels, descriptions or aliases added

**Statements**

- employer

- 0 references
Edit and augment data in OpenRefine

You can export any schema you create, and import an existing schema for use with a new dataset.
Push data from OpenRefine back to Wikidata

Address errors flagged by OpenRefine.
Push data from OpenRefine back to Wikidata

Use the preview window to check that your edits appear as you expect them to. Adjust your schema as necessary.
Push data from OpenRefine back to Wikidata

Wikidata → Upload edits to Wikibase
Push data from OpenRefine back to wikidata

What happens when my data overlaps with what is already in wikidata?

★ If you push a **redundant statement**, i.e. the statement you are trying to make has already been made, nothing will happen in wikidata.
★ If you push a **conflicting statement**, i.e. the statement you are trying to make is different than an existing statement for the same property, yours will be added as a second statement.
★ If you include a **new reference** for a statement that already exists, it will be added to the existing statement.
Push data from OpenRefine back to Wikidata

What about creating new items in Wikidata?

★★ OpenRefine can easily push data to Wikidata that creates new items.
★★ Best practice is to treat edits to existing items separately from the creation of new items.

○ As always in OpenRefine, only the rows covered by your facets will be considered when uploading the edits to Wikidata. I.e., if a cell is reconciled to a new item but is excluded by the facet, no new item will be created for it when you push data.
OpenRefine  Live demo
Additional resources

4. Cradle, an additional data entry tool designed to streamline editing for certain types of records, e.g. scientific collector: https://cradle.toolforge.org/#/subject/scientific_collector