



Kwalk: A Simple Program to Crosswalk Metadata for Repository Uploads

Item Type	Lightning Talk
Authors	Vallee, Kirsten
DOI	10.13028/6tg3-kd24
Publisher	eScholarship@UMassChan
Rights	Copyright © 2023 Vallee. This is an open-access document distributed under the terms of the Creative Commons Attribution 4.0 License (CC BY 4.0). The use, distribution or reproduction in other forums is permitted, provided the original author(s) are credited.
Download date	01/12/2023 15:40:36
Item License	https://creativecommons.org/licenses/by/4.0/
Link to Item	http://hdl.handle.net/20.500.14038/52789



kwalk

a simple program to crosswalk metadata

Kirsten Vallee

December 1, 2023

Scenario

You need to upload 1,000 items to your institutional repository from a source like Lens.org or PLOS journals.

You obtain informal metadata for the items on your own or from another person either from creating the spreadsheet from scratch, exporting the data, or web scraping each individual record.

You might need to

rename

- Rename all the fields from the invented field names to TIND/your repository platform's field names

add

- Add some fields that are missing

exclude

- Exclude some fields you don't want

combine

- Combine several fields into one field

modify

- Modify the values of date formats or author names in a programmatic way

generate

- Generate syntactically correct upload URLs from a simple filename field

What can you do with kwalk?



Write a simple crosswalk that you can apply to each batch of metadata as you receive it



Have multiple crosswalks for multiple projects as you work on them in an intermixed fashion



Apply special functions to modify date formats, combine literal and field name text, generate uniform upload URLs, and more

What do you need to kwalk?

```
*examplecrosswalk - Notepad
File Edit Format View Help
%245_a %title
%269_a %pubdate
%7001_a %author
```

1

A	B	C
title	pubdate	author
Pride and Prejudice	1813	Austen, Jane
Brave New World	1932	Huxley, Aldous
Moby-Dick	1851	Melville, Hermar

2

1. A **crosswalk** in .txt format (i.e. lit-kwalk.txt)
2. A **database** in .csv format (i.e. lit-data.csv)

*This example is a simple one. There are special functions in the program that can modify date formats, combine literal and field name text, generate uniform upload URLs, and more.

How to kwalk

01

Open a program
like Windows
PowerShell

02

Navigate to the
directory where
your .txt and .csv
files are located

03

```
kwalk-tind -c  
crosswalk  
litkwalk.txt  
litdata.csv >  
litupload.csv
```

Outcome

245__a	269__a	7001__a
Pride and Prejudice	1813	Austen, Jane
Brave New World	1932	Huxley, Aldous
Moby-Dick	1851	Melville, Herman

Thank you

Kirsten Vallee

University of Chicago
vallee@uchicago.edu