



COLLECTION ANALYSIS FOR \$5 A MONTH. **CHEAP**

https://rayvoelker.github.io/oh-iug-2021/collection_analysis_for_cheap

Ray Voelker

ray.voelker@gmail.com | ray.voelker@cincinnati library.org



TALK OVERVIEW

1. Collection Analysis Platform Overview
2. Collection Snapshot Creation: Overview of the workflow and automation involved in extracting the collection metadata from the ILS (Sierra)
3. The Analysis!

Collection Analysis Platform Overview

What is Datasette?

- Datasette is described by its creator, Simon Willison, as "a tool for exploring and publishing data".
- “Datasette is aimed at **journalists, museum curators, archivists, local governments** and anyone else who has data that they wish to share with the world.”

SIMON'S DATASETTE

PRESENTATIONS / DEMOS

- “Introduction to Datasette and sqlite-utils”
<https://www.youtube.com/watch?v=7kDFBnXaw-c>
- “Simon Willison - Instant serverless APIs, powered by SQLite - PyCon 2019”
<https://www.youtube.com/watch?v=pTr1uLQTJNE>
- “Keynote: Simon Willison - Datasette and Dogsheep: Liberating your personal data”
<https://www.youtube.com/watch?v=UFn82w-97kl>

Things I ❤️ about Datasette

- Written in Python / Available via pip
 - Easy and intuitive to use
 - Well-documented
 - Useful, Large, and growing plugin library
 - Built-in API... Using SQL!
 - Open source + supportive developer
 - Flexible deployment with scale-to-zero-cost / free hosting options in mind
- ... or, **CHEAP** hosting options

Cincinnati & Hamilton County Public Library: Collection Analysis

<https://ilsweb.cincinnati-library.org/collection-analysis/>

DATASETTE DEPLOYMENT / HOSTING OPTIONS

- There are lots of ways to deploy Datasette
(Datasette even has a built in command `datasette publish` that can package and deploy to services like Heroku or Google Cloud)

DATASETTE DEPLOYMENT / HOSTING OPTIONS

<https://docs.datasette.io/en/stable/deploying.html#deployment-fundamentals>

- Running Datasette on a full Linux server has a lot of advantages, but does have a steep learning curve
- Luckily there are lots of help-pages, tutorials, and guides available!

DIGITALOCEAN DROPLET (VIRTUAL PRIVATE SERVER)

- **CHEAP**
\$5 per month for a 1 CPU / 1 GB memory / 25 GB Disk Linux Server (plenty for Datasette)
- DigitalOcean has a great community forum with many great guides available for various System Administration tasks

Create DigitalOcean Droplet for Datasette Step by Step

[https://ils-underground.github.io
/python_datasette_vps.html](https://ils-underground.github.io/python_datasette_vps.html)

2. Collection Snapshot Creation

Extracting and Building the Data

Scripts For Extracting Data From ILS (Sierra)

<https://github.com/cincinnati-library/collection-analysis>

- This was originally done to support the task of automating data extraction from the ILS for the CollectionHQ Service
- We don't have to use CollectionHQ--Datasette can act as our data-analysis tool

3. The Analysis!

Looking at Data in Aggregate

- What item locations are defined?
- ... and how many items in each of those locations?

- Datasette Plugin: **Vega** -- A Datasette plugin that provides tools for generating charts using Vega

Named parameters: Datasette provides a nice interface when a named parameter is placed within a SQL query


```
1 select
2   bib.*
3 from
4   bib
5 where
6   indexed_subjects like '%'
7   || :subject
8   || '%'
```


CANNED QUERIES

- Create something that looks more like a traditional REST API

```
# Datasette URI construction
scheme = "https://"
host = "ilsweb.cincinnatilibrary.org"
# path contains the path of the Datasette instance,
# and the name of the database
path = "/collection-analysis/current_collection"
canned_query_path = "/item_lookup_by_barcode"
response_format = '' # '' or '.csv' or '.json'
query_parameter = "?barcode="
query_parameter_value = "A000036107985"

full_request_uri = scheme + host \
    + path + canned_query_path + response_format \
    + query_parameter + query_parameter_value
```


Thanks, and please, let me know if you come up with any cool queries,  Datasette as much as I do, or want to use this at your library!

Ray Voelker

ray.voelker@gmail.com

ray.voelker@cincinnati.library.org