archivers.space

Information gathered during Data Rescue Events through the archivers.space app. See http://www.ppehlab.org/datarescueworkflow#retiredworkflow for more information.

Priority *
Number should be 1-10

6

URL *

https://quickstats.nass.usda.gov/

Agency *

USDA

Event

Title

Agricultural Chemical Use Program - Quick Stats

Crawled by the Internet Archive? *

⦿ Yes

◯ No

Internet Archive URL

https://web.archive.org/web/*/https://quickstats.nass.usda.gov/

Description

Purpose or significance of data

NASS publishes U.S., State, County and other geographic level agricultural statistics for many commodities and data series. Quick Stats is a web based application that gives you the ability to query the agricultural census and survey statistics based on the simple filtering rules of What, Where, and When.

- [ ] Do not harvest. All data is small, unstructured, and on a page crawlable by the Internet Archive.

- [ ] Page contains dynamic content (e.g., links loaded by JavaScript).

- [ ] Page contains interactive visualizations.

- [ ] Data is accessible in structured file(s) that can be directly downloaded.

- [x] Data is accessible over FTP.

- [ ] Data is accessible using a documented public API.

- [ ] Data is only accessible using search queries in a web form.

Recommended approach to harvesting data

can just grab the .gzips with save or wget from ftp://ftp.nass.usda.gov/quickstats/

File formats

Estimated size in MB

Related URLs

https://quickstats.nass.usda.gov/

Where you able to capture all of the data at this URL?

○ Yes

○ No

Harvest method used

download the .gz from the developer api download page.

Notes from Harvest

Not sure if all data collected, a few hundred mb of .gz data was downloaded. Not sure if that equals the 34,123,266 of current records.

User certified that to the best of their knowledge this is a well-checked bag that will survive out of context of the site it was harvested from.

◉ Yes

Notes from Bagging

Bagged with Data Conservancy Packaging tool; validated before S3 upload and after S3 download.

Notes from Describe

This dataset (same UUID) was already in ckan. I've included the ckan URL above. - Alison Clemens 3/4/2017