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

4

URL *

https://ers.usda.gov/data-products/oil-crops-yearbook

Agency *

The United States Department of Agriculture (USDA)

Event

Title

Oil Crops Yearbook

Crawled by the Internet Archive? *

⦿ Yes

◯ No

Internet Archive URL

https://web.archive.org/web/*/https://ers.usda.gov/data-products/oil-crops-yearbook

Description

Purpose or significance of data

Oilseed, oilmeal, and fats and oils supply and use statistics. Includes oilseed acreage, yield, and production estimates and farm and wholesale price series. Prior years are available at the Mann Library.

- ☐ 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.
- ☐ 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

All data appears to be in easily downloadable formats: PDF, ZIP, XLSX

File formats

PDF, ZIP, XLS

Estimated size in MB

approx 5MB

Related URLs

https://ers.usda.gov/data-products/oil-crops-yearbook

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

◉ Yes

○ No

Harvest method used

jquery-url-extraction to get the list of about 50 links, and then wget-loop to download them.

Notes from Harvest

I believe these links were also acquired by the Internet Archive today but not totally sure.

Also, it looks like there is an included zip file 52529_89002.zip that contains a second copy of all the excel sheets.

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

Harvest URI was missing. Changed name of resource in json file.

Notes from Describe