

# 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

10

URL \*

<https://huduser.gov/portal/datasets/cinch.html>

Agency \*

United States Department of Housing and Urban Development

Event

Title

Components of Inventory Change (CINCH) reports

Crawled by the Internet Archive? \*

☒ Yes

☐ No

Internet Archive URL

[https://web.archive.org/web/\\*/https://huduser.gov/portal/datasets/cinch.html](https://web.archive.org/web/*/https://huduser.gov/portal/datasets/cinch.html)

Description

Purpose or significance of data

The (CINCH) report measures changes in the characteristics of the housing stock of the United States.

- ☐ 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

File formats

PDF SAS ZIP

Estimated size in MB

Related URLs

[huduser.gov/portal/datasets/ci](http://huduser.gov/portal/datasets/ci)

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

☒ Yes

☐ No

Harvest method used

custom python code to pull out urls organized by year and data / related publications then wget to download into sub directories

### Notes from Harvest

Data has been uploaded to here: <https://s3.amazonaws.com/drp-upload/remote/256AA805-CE92-4641-A560-E1582DA1D03F.zip>

The upload form didn't appear to work, but the data did make it to S3 at the link above.

run python tools/get\_urls.py > tools/download.sh

then sh tools/download.sh

Note from bagging: Harvest URL didn't save

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

Extracted without bagging

### Notes from Describe

Extracted and added to ckan without bagging or describing

This form was created outside of your domain.

Google Forms