

QOMPLX:CYBER

Archiving for Investigators

Part 2 - ArkScrape Tool

Nate Johnson

Brief introduction...

Nate Johnson

AKA **Caprico**

OSINT Specialist
QOMPLX Intelligence Unit

 @C4pr1c0



Archiving Tool for the Investigator

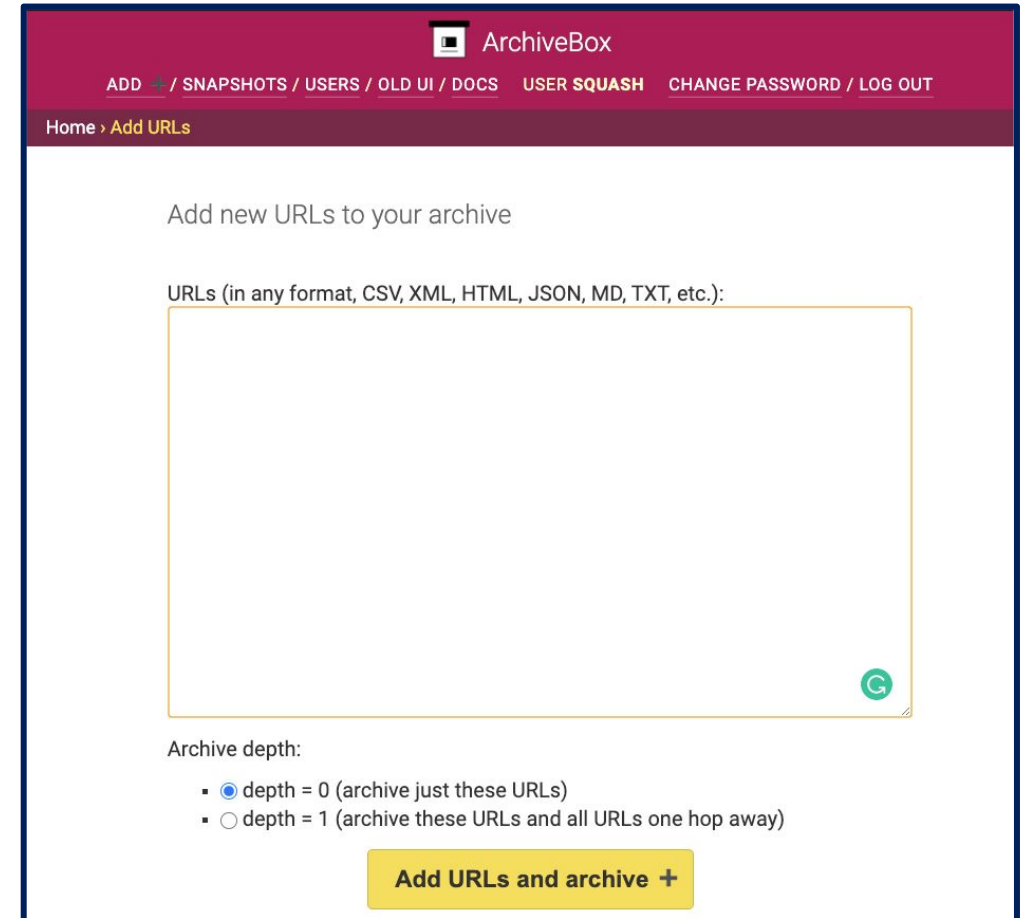
Brief introduction...

- Continuation of Archiving Data
- Shodan
- Archive Case Management tool

Archiving Tool for the Investigator

Archiving websites (ArchiveBox)

- Keith talked through using this
- Use a url to archive it
 - etc.
- How do we organize this?



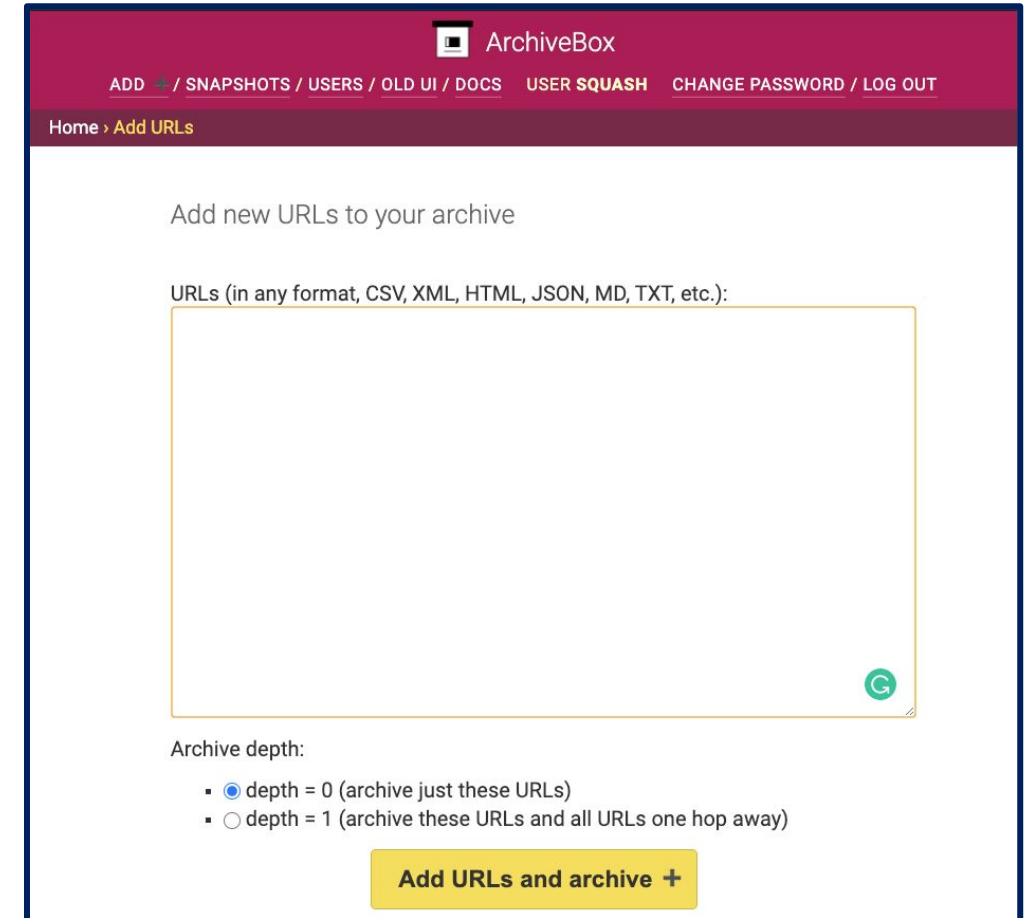
The screenshot shows the ArchiveBox web interface. At the top is a dark red header with the ArchiveBox logo and navigation links: ADD +, SNAPSHOTS, USERS, OLD UI, DOCS, USER SQUASH, CHANGE PASSWORD, and LOG OUT. Below the header is a breadcrumb trail: Home > Add URLs. The main content area has the heading 'Add new URLs to your archive' and a text input field labeled 'URLs (in any format, CSV, XML, HTML, JSON, MD, TXT, etc.):'. Below the input field is a section for 'Archive depth:' with two radio button options: 'depth = 0 (archive just these URLs)' (which is selected) and 'depth = 1 (archive these URLs and all URLs one hop away)'. At the bottom right is a yellow button labeled 'Add URLs and archive +'. A small green circular icon with a white 'G' is visible in the bottom right corner of the input field.

Archiving Tool for the Investigator

Archiving websites (ArchiveBox)

- If you have a single server there is the possibility of cross contamination of data
 - URL 1 has nothing to do with Investigation 2

How do you fix that?

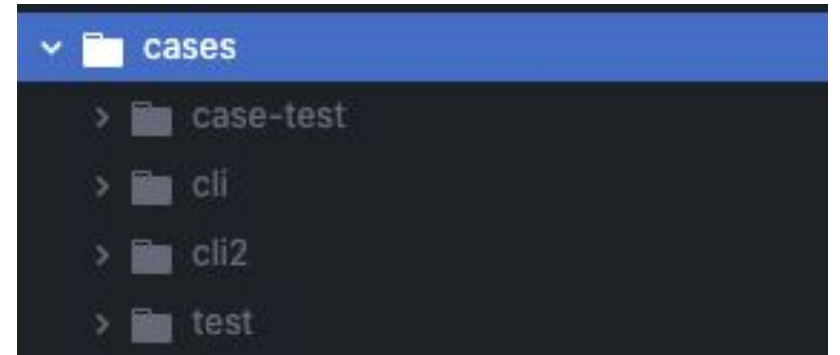


The screenshot shows the ArchiveBox web interface. At the top is a dark red header with the ArchiveBox logo and navigation links: ADD, SNAPSHOTS, USERS, OLD UI, DOCS, USER SQUASH, CHANGE PASSWORD, and LOG OUT. Below the header is a breadcrumb trail: Home > Add URLs. The main content area has the heading 'Add new URLs to your archive'. Below this is a text input field with the placeholder 'URLs (in any format, CSV, XML, HTML, JSON, MD, TXT, etc.):'. To the right of the input field is a green circular icon with a white 'G'. Below the input field is a section titled 'Archive depth:' with two radio button options: 'depth = 0 (archive just these URLs)' (which is selected) and 'depth = 1 (archive these URLs and all URLs one hop away)'. At the bottom right is a yellow button with the text 'Add URLs and archive +'.

Archiving Tool for the Investigator

Archiving websites (ArchiveBox)

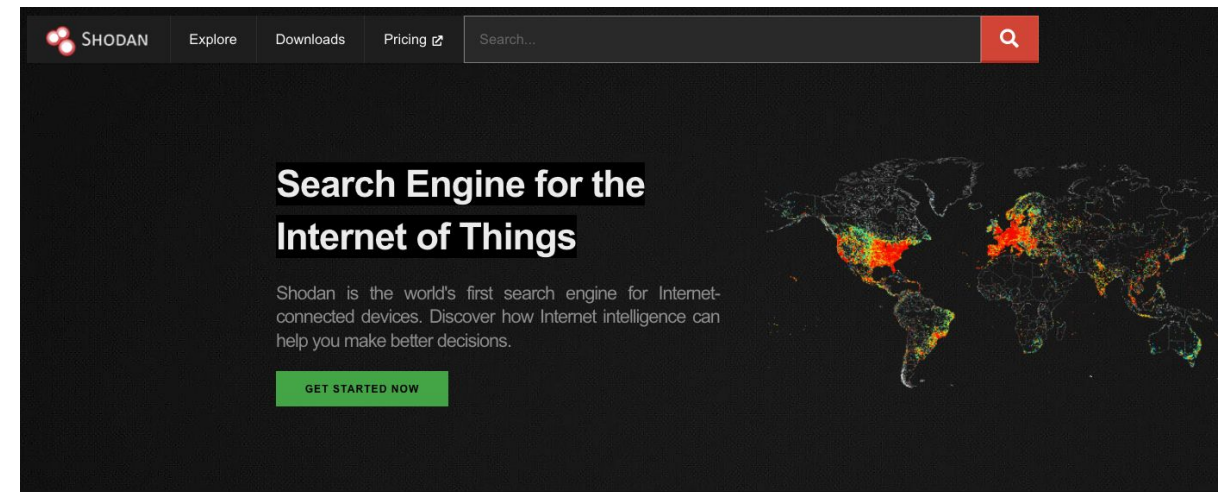
- Scripting (Python 3)
- Separate Archive Databases Into different Cases
- Let's you run databases per operations



Archiving Tool for the Investigator

Archiving Infrastructure (shodan)

- Internet Scanning Tool
 - Price
 - Account (\$)
 - Subscription (\$-\$\$\$\$)
- Archive
 - Location Data
 - How long the site has been up.
 - Who is hosting it.



Archiving Tool for the Investigator

Using our tool

Usage

```
usage: arkscrape.py [-h] [--init] [--shodan SHODAN] [--server SERVER] [--domain DOMAIN]
                  [--scrape SCRAPE]
```

Archive websites, Shodan scan results, and webscraper per a domain

optional arguments:

-h, --help	show this help message and exit
--init	First run setup
--shodan SHODAN	Shodan check: --shodan <archive folder>
--server SERVER	Archive box server start: --server <archive folder>
--domain DOMAIN	Target domain
--scrape SCRAPE	Scrape Urls from a list: --scrape <archive folder>

Archiving Tool for the Investigator

Using our tool

```
$ python3 arkscrape.py --init
Name for Archive Folder: talk-test
=== ARCHIVE SETUP ===
[i] [2021-06-16 19:13:04] ArchiveBox v0.6.2: archivebox init --setup
    > /Users/nathaniel.johnson/arkscrape/cases/talk-test

[+] Initializing a new ArchiveBox v0.6.2 collection...
-----

[+] Building archive folder structure...
    + ./archive, ./sources, ./logs...
    + ./ArchiveBox.conf...

[+] Building main SQL index and running initial migrations...
    Operations to perform:
        Apply all migrations: admin, auth, contenttypes, core, sessions
    Running migrations:
    Applying contenttypes.0001_initial... OK
    Applying auth.0001_initial... OK
```

Archiving Tool for the Investigator

Using our tool

```
$ python3 arkscrape.py --server talk-test
=== Starting Server ===
[i] [2021-06-16 19:37:10] ArchiveBox v0.6.2: archivebox server
> /Users/nathaniel.johnson/arkscrape/cases/talk-test

[+] Starting ArchiveBox webserver...
> Logging errors to ./logs/errors.log
Performing system checks...

System check identified no issues (0 silenced).
June 16, 2021 - 19:37:11
Django version 3.1.8, using settings 'core.settings'
Starting development server at http://127.0.0.1:8000/
Quit the server with CONTROL-C.
```

Archiving Tool for the Investigator

Using our tool

```
$ python3 arkscrape.py --shodan talk-test --domain qomplx.com
Search query:                domain qomplx.com
Total number of results:      0
Query credits left:          9920
Output file:                  2021_6_16-export.json.gz
[#####] 100%
Saved 0 results into file 2021_6_16-export.json.gz
```

Archiving Tool for the Investigator

Using our tool

```
$ python3 arkscrape.py --shodan talk-test --domain qomplx.com
Search query:                domain qomplx.com
Total number of results:      0
Query credits left:           9920
Output file:                  2021_6_16-export.json.gz
[#####] 100%
Saved 0 results into file 2021_6_16-export.json.gz
```

Archiving Tool for the Investigator

Using our tool

```
$ shodan parse --fields ip_str,location.longitude,location.latitude cases/talk-test/shodan/2021_6_18-export.json.gz
121.178.254.145 126.39181 34.81282
216.107.134.6 -97.51643 35.46756
91.201.107.99 30.5238 50.45466
92.118.108.175 28.8575 47.00556
145.239.62.168 2.12807 50.98651
77.223.131.231 27.13838 38.41273
52.33.125.148 -119.70058 45.83986
172.105.59.29 72.88261 19.07283
45.166.51.71 -34.88111 -8.05389
106.14.226.189 114.0683 22.54554
113.43.133.106 139.69171 35.6895
78.46.162.158 7.77746 48.21884
185.225.1.45 -3.70256 40.4165
```

Archiving Tool for the Investigator

Conclusion

- You can find the tool within the leaflet for this presentation
- Feel free to contact me on twitter if you have any questions or suggestions.