

## Page Vault Export Readme

This document describes the contents of your export/ZIP file. The ZIP contains the following:

1. Folder(s) containing captures. Your content will consist of the following:
  - folder(s) with screenshot captures
  - folder(s) with video captures
  - folders(s) with document captures (downloaded in their original format)
2. load-file.dat to import data into eDiscovery tools, e.g. Relativity
3. Page Vault Capture Summary.csv to view data in Excel, Google Sheets, etc.
4. Page Vault Capture Summary.html to view data in a web browser and navigate the document archive.

The .dat, .csv and .html all contain the below fields:

Field Name	Description
control_number	An identifier generated by Page Vault.
begattach	An identifier generated by Page Vault to establish a parent-child link between captures created in various formats stemming from the same original_capture_url
file_path	A relative file path to the capture. Used by the load-file.dat to determine the file to be loaded into a eDiscover tool, and by the .html file to make captures "clickable" so they may be opened directly when viewing that file in a web browser.
capture_url	The URL that the capture was created from.
original_capture_url	The URL that the capture_url was derived from. A common use case for this is social media captures, where the capture_url revealed videos to be downloaded from another URL. In that case, the capture_url would contain the URL of the downloaded video, and the original_capture_url would contain the URL of the social media post that contained the video.
filename	The name of the file when saved. This is most often a derived name based on the site being captured and the type of capture, but may be overridden by the user who created the capture for Page Vault Browser users.
document_title	The title of the document, if any, as reported by the site from which the capture was created.
capture_time	The exact date and time, in UTC, that the capture was recorded.
hash	A SHA256 hash of the capture file.
file_size	The size in bytes of the capture file on disk.
unique_id	An alternate identifier generated by Page Vault.
file_download_type	The MIME type of the captured file.
screenshot_server_ip	For "snapshot" item_type, the IP address of the server through which the capture was made.

screenshot_load_time	The date and time, in UTC, that a webpage was loaded before a capture was created.
screenshot_page_length	For "snapshot" item_type, the number of PDF pages in the captured file.
file_uploaded_at	For "file_capture" item_type, some sites report a date and/or time that the file was uploaded.
file_description	For "file_capture" item_type, some sites report a description of the content.
post_datetime	For social media captures, an attempt is made to parse the datetime associated with a social media post, if the URL captured pertains to a single post. The dates and times are estimates only.
parent_est_post_datetime	For social media captures, in cases where there is a parent-child relationship between captures, this field provides the est_post_datetime of the parent content.
post_author	For social media posts (post_datetime), the author of the post.
post_reaction_count	For social media posts (post_datetime), a count of reactions to the post.
post_comment_count	For social media posts (post_datetime), a count of comments to the post.