

Troubleshooting and FAQs

When something goes wrong, this page is here to help you identify and hopefully remedy the problem. Please contact HTRC support at htrc-help@hathitrust.org with any questions or bugs or peculiarities you would like to report.

- [HTRC tools and services](#)
 - [Q: What are the HTRC tools and services?](#)
 - [Q: Who can use HTRC?](#)
 - [Q: What is the login timeout for HTRC Analytics?](#)
 - [Q: What are worksets and what do I do with them?](#)
 - [Q: How do I create a workset?](#)
 - [Q: Can I analyze non-HathiTrust data alongside HathiTrust data?](#)
 - [Q: What is the HTRC Data Capsules environment and what can it be used for?](#)
 - [Q: Do I have computational access to the HathiTrust Digital Library's copyrighted content in Data Capsule?](#)
 - [Q: HTRC Analytics showed an error message when I tried to create a Data Capsule. What went wrong?](#)
 - [Q: I have some Python scripts that I want to use in my analysis within the HTRC Data Capsule. How should I start?](#)
 - [Q: Can I import the workset that I have used in HTRC Analytics into the HTRC Data Capsule?](#)
 - [Q: Can you tell me exactly how much data I am allowed to export from my capsule?](#)
 - [Q: How do I use the HTRC Data API?](#)
 - [Q: What is the difference between the HTRC Data API and HathiTrust Data API?](#)
- [What happened to...?](#)
 - [Q: What happened to the Workset Builder?](#)
 - [Q: What happened to the HTRC Solr Proxy API?](#)
 - [Q: What happened to the HTRC Sandbox?](#)
- [User Accounts and Sign-in](#)
 - [Q: Why isn't my institution listed on HTRC's sign-in dropdown menu?](#)
- [HTRC Code and Infrastructure](#)
 - [Q: Can I see the code used to make HTRC tools and services operate?](#)
 - [Q: Where can I learn more about HTRC Data Capsules development project?](#)
 - [Q: To whom can I direct technical questions?](#)
- [Get in touch!](#)
 - [Q: How do I report issues or give feedback?](#)
 - [Q: How do I ask questions or start discussions with other users?](#)

HTRC tools and services

Q: What are the HTRC tools and services?

A: HTRC has created a suite of tools that allow researchers to perform text analysis on content in the HathiTrust Digital Library. Most of these tools are available via the [HTRC Analytics](#) website. They are intended to meet the various needs of HTRC researchers.

- [HTRC Algorithms](#): a set of tools for assembling collections of digitized text and performing text analysis on them.
- [HTRC Extracted Features](#): an openly-available dataset of metadata and derived data from the HathiTrust corpus.
- [HTRC Data Capsule](#): a secure computing environment for performing researcher-driven text analysis on HathiTrust content.
- [HathiTrust+Bookworm](#): a tool for visualizing and analyzing word usage trends in the HathiTrust Digital Library.

Q: Who can use HTRC?

A: Most of HTRC's services require an account on [HTRC Analytics](#) to use. Scholars from non-profit institutions of higher education or other research institutions are eligible for an account, and users don't need to be affiliated with a HathiTrust member institution in order to qualify. Some services within [HTRC Analytics](#) are further restricted: Access to an HTRC Data Capsule with computational access to items in copyright is available ONLY to member-affiliated researchers who complete a Capsule request form. Others require no account to use, such as the [HTRC Extracted Features](#) or [HathiTrust+Bookworm](#).

Q: What is the login timeout for HTRC Analytics?

A: The current login timeout is 1 hour. However, your submitted job won't be affected by this logout time. It will still run even if you logout or if the system logs you out.

Q: What are worksets and what do I do with them?

A: Worksets are sub-collections of HathiTrust volumes created by researchers. You can run HTRC algorithms against worksets in order to analyze them or download their Extracted Features. Worksets can be cited, and researchers can choose to make their worksets public or private. [Learn more about worksets](#).

Q. How do I create a workset?

A: You create a workset by uploading a list of HathiTrust volume IDs to [HTRC Analytics](#) or by importing a publicly-viewable [collection from HathiTrust](#). [You can read more in the tutorial](#).

Q: Can I analyze non-HathiTrust data alongside HathiTrust data?

A: Within the [HTRC Analytics](#) platform, only in the [HTRC Data Capsule](#) environment. [HTRC Algorithms](#) function only on "worksets," which are user-created collections of content from the HathiTrust Digital Library. You can import outside data to your Capsule when it is in maintenance mode, though, and work with it within that system. You can also make use of [HTRC Extracted Features](#) alongside if you prefer to work on your local desktop only.

Q. What is the HTRC Data Capsules environment and what can it be used for?

A. The [HTRC Data Capsule](#) environment provides a secure computing environment to access content in the HathiTrust Digital Library. Users are provisioned virtual machines called capsules to which they can import and then analyze HathiTrust volumes. Users can only perform computational analysis within the secure Data Capsule environment and then export the results of their analysis. Users cannot export volume content outside the HTRC Data Capsule.

Q: Do I have computational access to the HathiTrust Digital Library's copyrighted content in Data Capsule?

A: Computational access to items in copyright is available ONLY to HathiTrust member-affiliated researchers. Existing Data Capsule users from member institutions or new Data Capsule requesters from member institutions have the exclusive option to select "Full Corpus Access," which includes copyrighted items.

Q: HTRC Analytics showed an error message when I tried to create a Data Capsule. What went wrong?

A: Most likely you have reached the maximum amount of space allowed per user in the capsules system. Please delete one of your capsules, or contact HTRC support to solve the issue: htrc-help@hathitrust.org

Q: I have some Python scripts that I want to use in my analysis within the HTRC Data Capsule. How should I start?

- A:
- First store your Python scripts somewhere on Internet.
 - Start your capsule from within the Analytics interface, and make sure your machine is in maintenance mode.
 - Enter your capsule via Terminal viewer or Remote Desktop viewer.
 - Download the Python scripts from the Internet onto your capsule.
 - Switch to secure mode.
 - If you know the volume IDs that you are interested, you can go ahead to fetch content of these volumes by using this sample Python script in [Fetching Volume OCR Content in HTRC Data Capsule \(Secure Mode\)](#).
 - Run your Python scripts against the content.
 - If you don't have the volume IDs of your interest, you can search for volumes in the [HathiTrust Digital Library](#). You can search by subject, topic, author, year, etc., and identify the volumes of interest and save your chosen volumes as a collection in HathiTrust. From there, you can either use the [HTRC Workset Toolkit](#) to load volumes from the collection in your Capsule, or download the collection's metadata to retrieve the volume IDs for the volumes you have selected.
 - Once you have the volume IDs ready, you can go ahead to fetch the volume content in Data Capsule secure mode and perform analysis using your Python scripts as mentioned above.

Q: Can I import the workset that I have used in HTRC Analytics into the HTRC Data Capsule?

A: Currently, there are two ways to do this, depending on whether you have first created a collection in HathiTrust:

1) Download the workset from HTRC Analytic in order to export a list of the volume IDs for that workset, and then use the HTRC Workset Toolkit in the Data Capsule to access the content in those volumes. It is not presently possible to export a workset from HTRC Analytics directly into the HTRC Data Capsule, but we expect to integrate this functionality into future versions.

2) Load volumes from a HathiTrust Digital Library collection into a Capsule using the HTRC Workset Toolkit using the collection's URL. Directions are available here: <https://htrc.github.io/HTRC-WorksetToolkit/cli.html>.

Keep in mind which volumes will be available to you within your Capsule, depending on the kind of Capsule you are using and whether it has access to the full corpus or only "full view"/public domain volumes.

Q: Can you tell me exactly how much data I am allowed to export from my capsule?

A: The standard for non-consumptive export depends on the scope and scale of the data analyzed. The general rule-of-thumb is whether the export would create a substitute for human-reading the original text. (The full [Non-Consumptive Use Research Policy](#) is also available for your reference.) If you would like someone to pre-review a sample file that would represent the kinds of data you would like to export from a capsule before you begin your work, please contact htrc-help@hathitrust.org.

Q: How do I use the HTRC Data API?

A: Check out our [user's guide](#) for more information about using the HTRC Data API in the HTRC Data Capsule.

Q: What is the difference between the HTRC Data API and HathiTrust Data API?

A: This table outlines the differences between the HTRC Data API and HathiTrust Data API

	HTRC Data API	HathiTrust Data API
purpose	to serve high-performance large-scale algorithms and programs	to provide public users some volume retrieval capabilities
throttling enforcement	no	yes
security	JWT	OAuth
bulk retrieval of volumes	yes	no
metadata available	METS	METS, MARC

What happened to...?

Q: What happened to the Workset Builder?

A: As HTRC upgrades its services and builds a new Workset Builder, the retired Workset Builder has been taken offline. The new system of creating a collection in the HathiTrust Digital Library better aligns workset-building with the HathiTrust and offers improved search and selection.

Q: What happened to the HTRC Solr Proxy API?

A: As the HTRC moves to update and improve its search and workset-building services, the Solr Proxy API has been retired. For now, you can search for HathiTrust volumes via the HathiTrust Digital Library interface. Look for improved functionality in the near future, and please reach out with your workset-building scenarios that require additional search functionality.

Q: What happened to the HTRC Sandbox?

A: The HTRC Sandbox, which was a space for testing and experimentation in the early days of the project, has been rolled into our production services available here:

- [HTRC Analytics](#): a set of tools for assembling collections of digitized text and performing text analysis on them.
- [HTRC Data Capsule](#): for use of the production-level HTRC Data API

User Accounts and Sign-in

Q: Why isn't my institution listed on HTRC's sign-in dropdown menu?

coming soon...

HTRC Code and Infrastructure

Q: Can I see the code used to make HTRC tools and services operate?

A: Yes. All of the HTRC services code modules are open source and are available from GitHub: <https://github.com/htrc>.

Q: Where can I learn more about HTRC Data Capsules development project?

A: More information can be found in the public version of the final report of the project as well: <http://hdl.handle.net/2022/19277>

Q: To whom can I direct technical questions?

A: Please email HTRC support: htrc-help@hathitrust.org.

Get in touch!

Q: How do I report issues or give feedback?

A: We welcome your feedback! You can send an email to HTRC Support at htrc-help@hathitrust.org. We track support requests in using JIRA, and you can log-in to see your requests and our responses here: <https://jira.htrc.illinois.edu/servicedesk/customer>.

Q: How do I ask questions or start discussions with other users?

A: Please join the HTRC User Group mailing list.

- Subscribe here: <https://list.indiana.edu/sympa/info/htrc-usergroup-l>
- For questions that you want to discuss with us privately, please write to htrc-help@hathitrust.org, a list subscribed by HTRC internal staff only.
- All users are subscribed to a listserv called HTRC-Announce when they create an HTRC Analytics account. Only approved senders can send mail through this list.