

HTRC Data Capsule Environment

The HTRC Data Capsule environment provides individual, secure computing environments to analyze content in the HathiTrust Digital Library. Researchers can create virtual machines (called Capsules) to which they can import and then analyze HathiTrust text data. Researchers can only perform computational analysis within the secure Data Capsule environment and then export the results of their analysis. Volume text may not be exported outside the HTRC Data Capsule, and data products leaving a Capsule must undergo results review prior to release to ensure they meet the HTRC's policy for [non-consumptive data exports](#).

[Use a Capsule](#) [Read the guide](#) [Follow a tutorial](#)

Capsule specifications

What's in a Capsule?

Out-of-the-box, Capsules are Ubuntu virtual machines with increased security settings. Researchers have the option to set certain parameters for their Capsule when they create it. Capsules come pre-loaded with standard data analysis programs and software. While Capsules come with standard tools pre-installed, ranging from [Anaconda](#) and [R](#) to [Voyant Tools](#), and can be configured with sample public domain data already loaded for testing, any other data or tools the researcher plans to use will need to be brought into the Capsule by the researcher. A Capsule is an almost blank slate that can be customized for each researcher's needs!

Kinds of Capsules

There are two kinds of capsules: Demo Capsules and Research Capsules. Researchers can request for their Research Capsules to have full-corpus access, and approval is limited to those from [HathiTrust member institutions](#).

Using a capsule

Creating a Capsule

Capsules operate from the [HTRC Analytics](#) website, which requires an HTRC account to log-in.

[Create an HTRC Analytics account](#) [Follow a tutorial](#)

You'll use the site to create and administer your Capsule.

[Create a Capsule](#) [Follow a tutorial](#)

Research in a Capsule

In [HTRC Analytics](#), you'll have the option work with your Capsule either via a remote desktop viewer (to see your capsule's desktop) or a terminal viewer (to interact with your capsule via a command line interface).

Capsules are intended for researchers who want access to HathiTrust text data in flexible, individually-driven environment. Capsules can be shared between up to 5 collaborators. Researchers looking for a point-and-click option should explore [HTRC Algorithms](#).

We offer several step-by-step guides for using a Capsule (see links "Read the Guide" and "Follow a Tutorial" located at the top of this page).

Development details

- [For Developers: Developer API](#)
- The Data Capsules threat model and other details are described in report of the environment's prototyping project: <http://hdl.handle.net/2022/19277>

Read more

The HTRC Data Capsule system was prototyped through funding from the Alfred P. Sloan Foundation (2011-2015). The final report is available here: [Final report](#).

Extension of the HTRC Data Capsule project to larger compute resources and better integration with the HTRC worksets was recently funded by a grant from the Andrew T. Mellon Foundation (2016-2018).



Quick links

[Data Capsules specs and usage guide](#)

[Non-consumptive Use Research Policy](#)

[HTRC Data Capsules Terms of Use](#)

[HTRC Algorithms documentation](#)

[Quick Video Demos](#) (out-of-date)

Kevin Borders, Eric Vander Weele, Billy Lau, and Atul Prakash, Protecting Confidential Data on Personal Computers with Storage Capsules. Proceedings of the 18th USENIX Security Symposium, Aug. 2009.

Zeng, J., Ruan, G., Crowell, A., Prakash, A., & Plale, B. (2014, June). Cloud Computing Data Capsules for Non-Consumptive Use of Texts. In Proceedings of the 5th ACM workshop on Scientific cloud computing (pp. 9-16). ACM.

Plale, Beth; Prakash, Atul; McDonald, Robert (2015). The Data Capsule for Non-Consumptive Research: Final Report. Available from <http://hdl.handle.net/2022/19277>