HATHI TRUST CAPABILITIES
FROM A RESEARCHER'S PERSPECTIVE
TRUST DATA API  [DEPRECIATED SLIDE]

• Combining This API with…
  • htid2books
  • Distant reader toolbox
  • Data visualization
HATHI USED IN TEXT MINING

• Underwood Bamman and Lee (2018) Used Hathi API data to chart the distinctness of depictions of men and women in literature over the last several hundred years.

HTRC TOOLS

Can be combined with Bookworm  https://bookworm.htrc.illinois.edu/develop/

Data capsules can be created by UNI affiliated persons

Extracted features (downloadable “via rsync”)

Algorithms
LOGGING IN TO HRTC

You should be able to sign in to https://analytics.hathitrust.org

You will need your UNI email and second factor auth.
HRTC DESKTOP (UBUNTU LINUX OPERATING SYSTEM) FOR CREATING AND ACCESSING YOUR DATA CAPSULE.
“WORKSETS”

• A dataset composed of HathiTrust volumes (a dataset without the text itself)

• A workset can be cited and shared with other researchers

• One way to analyse these is with the Algorithms in HTRC tools.
NON CONSUMPTIVE RESEARCH

• Within your “data capsule” (which is a remote virtual machine hosted by the University of Indiana that you can log into) there is data about works that UNI does not necessarily own.

• 1 Use case: GA tech students visualized early 20th century magazines with Voyant.

Policy available https://www.hathitrust.org/htrc_ncup